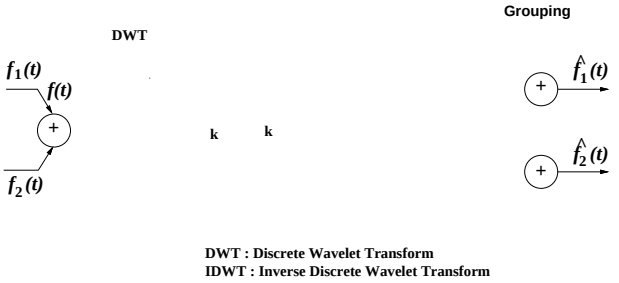
A Method of Signal Extraction from Noisy Signal

Masashi UNOKI and Masato AKAGI
unoki@jaist.ac.jp akagi@jaist.ac.jp
School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Ishikawa 923-12, Japan

Abstract

This paper presents a method of extracting the desired signal from a noise-added signal as a model of acoustic source segregation. Using physical constraints related to the four regularities proposed by Bregman, the proposed method can solve the problem of segregating two acoustic sources. Two simulations were carried out using the following signals: (a) a noise-added AM complex tone and (b) a noisy synthetic vowel. It was shown that the proposed method can extract the desired AM complex tone from noise-added AM complex tone in which signal and noise exist in the same frequency region. The SD was reduced an average of about 20 dB. It was also shown that the proposed method can extract a speech signal from noisy speech.



1 Introduction

Extraction of the desired signal from noisy signal is a important problem not only in robust speech recognition systems but also in various signal processing systems. The aim of this work is to solve the problem by constructing an auditory segregation model based on auditory scene analysis (ASA).

Bregman[1] reported that the human auditory system uses four psychoacoustically heuristic regularities: (i) common onset and offset; (ii) gradualness of change; (iii) harmonicity; and (iv) changes taken in an acoustic event, related to acoustic events for solving the problem of ASA. Typical models of auditory segregation based on ASA are Brown and Cooke's model[2] and Nakatani *et al.*'s model[3]. All these models use regularities (i) and (iii), and an amplitude (or power) spectrum as the acoustic feature. Thus they can not extract the desired signal from a noisy signal completely when the signal and noise exist in the same frequency region. And if background noise increases, it seems that these models can not extract the desired signal with high precision.

In contrast, we have discussed the need for using not only the amplitude spectrum but also the phase spectrum for completely extracting the desired signal from a noisy signal when both signals exist in the same frequency region[4, 5]. In this paper, we present a method for extracting the desired signal from a noisy signal by using physical constraints relat-

Figure 1: Auditory segregation model.

ed to regularities (i) – (iv), as an auditory segregation model. In particular, we consider the problem of extracting the desired signal from the following signals: (a) a noise-added AM complex tone and (b) a noisy synthetic vowel.

2 Auditory segregation model

The auditory segregation model shown in Fig. 1 consists of three parts: (a) auditory filterbank, (b) separation, and (c) grouping. The auditory filterbank is constructed using a gammatone filter as an "analyzing wavelet." The separation block uses physical constraints related to heuristic regularities (ii) and (iv). The grouping block uses physical constraints related to heuristic regularities (i) and (iii), and signal reconstruction in the grouping block is done with the inverse wavelet transform.

2.1 Formulation of the problem of segregating two acoustic sources

In this paper, we define the problem of segregating two acoustic sources as "the segregation of the mixed signal into original signal components, where the mixed signal is composed of two signals generated by any two acoustic sources." We formulate it as follows:

Firstly, we can observe only the signal $f(t)$:

$$f(t) = f_1(t) + f_2(t), \tag{1}$$

where $f_1(t)$ is the desired signal and $f_2(t)$ is a noise. The observed signal $f(t)$ is decomposed into its frequency components by an auditory filterbank. Secondly, outputs of the k-th channel, which correspond

to $f_1(t)$ and $f_2(t)$, are assumed to be

$$A_k(t)\sin(\omega_k t + \theta_{1k}(t)) \tag{2}$$

and

$$B_k(t)\sin(\omega_k t + \theta_{2k}(t)), \tag{3}$$

respectively. Since the output of the k-th channel $X_k(t)$ is represented by

$$X_k(t) = S_k(t)\sin(\omega_k t + \phi_k(t)), \tag{4}$$

where

$$S_k(t) = \sqrt{A_k^2(t) + 2A_k(t)B_k(t)\cos\theta_k(t) + B_k^2(t)} \tag{5}$$

and

$$\phi_k(t) = \tan^{-1}\left(\frac{A_k(t)\sin\theta_{1k}(t) + B_k(t)\sin\theta_{2k}(t)}{A_k(t)\cos\theta_{1k}(t) + B_k(t)\cos\theta_{2k}(t)}\right), \tag{6}$$

then the amplitude envelopes of the two signals $A_k(t)$ and $B_k(t)$ can be determined by

$$A_k(t) = \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)} \tag{7}$$

and

$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)}, \tag{8}$$

respectively, where $\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ and $\theta_k(t) \neq n\pi, n \in \mathbf{Z}$. Thus, if the four parameters, $S_k(t)$, $\phi_k(t)$, $\theta_{1k}(t)$, and $\theta_{2k}(t)$ are calculated, $A_k(t)$ and $B_k(t)$ can be calculated by the above equations. Finally, $f_1(t)$ and $f_2(t)$ can be reconstructed by grouping constraints. $\hat{f}_1(t)$ and $\hat{f}_2(t)$ are reconstructed $f_1(t)$ and $f_2(t)$, respectively.

In this paper, we assume $\theta_{1k}(t) = 0$ and $\theta_k(t) = \theta_{2k}(t)$. Additionally, we consider the problem of segregating two acoustic sources in which the localized $f_1(t)$ is added to $f_2(t)$.

2.2 Calculation of the four parameters

The amplitude envelope $S_k(t)$ and phase $\phi_k(t)$ of $X_k(t)$ are determined by using the amplitude and the phase spectra defined by the complex wavelet transform. Since we assume $\theta_{1k}(t) = 0$, $\theta_k(t) = \theta_{2k}(t)$, we must know the input phase $\theta_k(t)$. The input-phase $\theta_k(t)$ is derived by applying three physical constraints related to regularities (ii) and (iv) as shown below[5].

1. **Gradualness of change**

   This constraint is $dA_k(t)/dt = C_{k,R}(t)$, where $C_{k,R}(t)$ is an Rth-order differentiable polynomial. By putting $dA_k(t)/dt = C_{k,R}(t)$ into equation (7), and solving the resulting linear differential equation, we obtain

   $$\theta_k(t) = \arctan\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C_k(t)}\right), \tag{9}$$

   where unknown function $C_k(t)$ is $-\int C_{k,R}(t)dt + C_{k,0}$. In order to determine $C_k(t)$, we estimate $C_k(t)$ using the Kalman filter.

2. **Smoothness**

   This constraint, the smoothness for $A_k(t)$, is a function of the estimated $C_k(t)$. By considering the relationship between $A_k(t)$ and $C_k(t)$ from Eqs. (7) and (9), we can interpret the smoothness for $A_k(t)$ in order to determine the smoothest $C_k(t)$. Therefore, by calculating the candidates of $C_k(t)$ interpolated using the spline function within the estimated error, and then by calculating a correct solution from the candidates of $C_k(t)$, the smoothest $A_k(t)$ can be determined uniquely.

3. **Changes taken in an acoustic event**

   This constraint is

   $$\frac{A_k(t)}{\|A_k(t)\|} \approx \frac{A_\ell(t)}{\|A_\ell(t)\|}, k \neq \ell. \tag{10}$$

   With this constraint, $\theta_k(t)$ is determined when the correlation between $A_k(t)$ and $A_\ell(t)$ becomes maximum at any $C_k(t)$ within the estimated error-region.

2.3 Grouping constraints

The aim of the grouping constraints is to extract the desired signal from the noise-added signal using regularities (i) and (iii) proposed by Bregman. Therefore, the grouping block takes a solution for the problem of segregating two acoustic sources and applies to $X_k(t)$, in which two acoustic signals exist in the same time region. In other words, it applies the solution to $X_k(t)$, if either of the two physical constraints as shown below are satisfied[5].

1. **Harmonicity**

   When, the channel number k corresponds to an integer multiple of the fundamental frequency F_0, $n \cdot F_0, n = 1, 2, \cdots, N_{F_0}$. This constraint means that "when a body vibrates with a repetitive period, its vibrations give rise to an acoustic pattern in which the frequency components are multiples of a common fundamental."

2. **Common onset and offset**

   When onset and offset of $X_k(t)$ for $f_1(t)$ match those of $X_k(t)$, corresponding to the fundamental frequency F_0. This constraint means that "unrelated sounds seldom start or stop at exactly the same time."

3 Simulations and Results

We carried out two simulations on segregating two-acoustic sources using noise-added signal $f(t)$ to show that the proposed method can extract the desired signal $f_1(t)$ from it. The two simulations are composed as follows:

1. Extracting an AM complex tone from a noise-added AM complex tone.

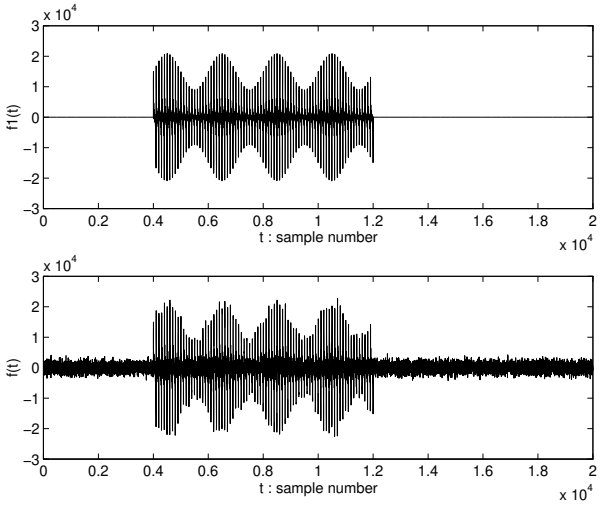


Figure 2: $f_1(t)$ and $f(t)$ (SNR=10 dB).

2. Extracting a speech signal from a noisy speech.

We use two types of measures to evaluate the performance of the segregation using the proposed method.

One is the power ratio in terms of the amplitude envelope $A_k(t)$, i.e., the likely SNR. The aim of using this measure is to evaluate the segregation in terms of the amplitude envelope where signal and noise exist in the same frequency region. This measure is called "Precision," and is defined by

$$\text{Precision}(k) := 10\log_{10}\frac{\int_0^T A_k^2(t)dt}{\int_0^T (A_k(t)-\hat{A}_k(t))^2 dt}, \tag{11}$$

where $A_k(t)$ is the amplitude envelope of original signal $f_1(t)$ and $\hat{A}_k(t)$ is the amplitude envelope of the segregated signal $\hat{f}_1(t)$.

The other is spectrum distortion (SD). The aim of using this measure is to evaluate the extraction of a desired signal $\hat{f}_1(t)$ from noise-added signal $f(t)$. This measure is defined by

$$\text{SD} := \sqrt{\frac{1}{W}\sum_{\omega}^{W}\left(20\log_{10}\frac{\tilde{F}_1(\omega)}{\tilde{\hat{F}}_1(\omega)}\right)^2}, \tag{12}$$

where $\tilde{F}_1(\omega)$ and $\tilde{\hat{F}}_1(\omega)$ are the amplitude spectra of $f_1(t)$ and $\hat{f}_1(t)$, respectively. Moreover, the frame length is 51.2 ms, the frame shift is 25.6 ms, W is analyzable bandwidth of filterbank (about 6 kHz), and the window function is Hamming.

The reduced SD of $f_1(t)$ is the SD difference between $f(t)$ and $\hat{f}_1(t)$.

3.1 Simulation 1

This simulation assumes that $f_1(t)$ is an AM complex tone as shown in Fig. 2, where $F_0 = 200$ Hz, $N = 10$,

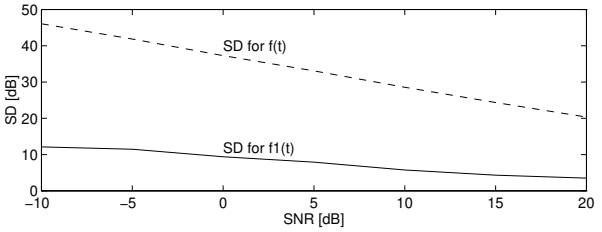


Figure 3: SDs of $\hat{f}_1(t)$ and $f(t)$.

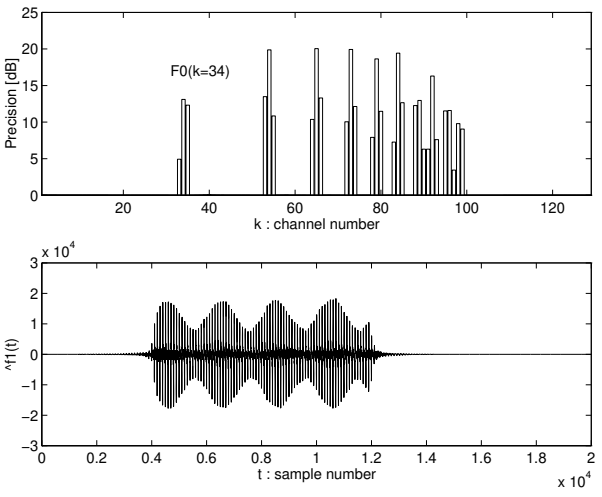


Figure 4: Precision property for $\hat{f}_1(t)$.

and envelope of $f_1(t)$ is sinusoidal (10 Hz), and $f_2(t)$ is a bandpassed random noise, where the bandwidth of $f_2(t)$ is about 6 kHz. Seven types of $f(t)$ are used as simulation stimuli, where the SNRs of $f(t)$ are from -10 to 20 dB in 5-dB steps. The mixed signal for SNR= 10 dB is plotted in Fig. 2.

The simulations were carried out using the seven mixed signals. The average SDs of $f_1(t)$ and $f(t)$ are shown in Fig. 3. As a result, it is possible to reduce the SD by about 20 dB as noise reduction by using the proposed method. For example, when the SNR of $f(t)$ is 10dB, the proposed method can segregate $A_k(t)$ with high precision as shown in Fig. 4, and it can extract the $\hat{f}_1(t)$ shown in Fig. 4 from the $f(t)$ shown in Fig. 2. The signal is reconstructed by considering $\theta_{1k}(t) = \phi_k(t)$, because phase information can not be determined by the assumption $\theta_{1k}(t) = 0$. The proposed model can extract the amplitude information of AM complex tone from a noise-added signal $f(t)$ with a high precision in which signal and noise exist in the same frequency region. Moreover, the proposed model can also extract the desired AM complex tone from mixed AM complex tones with different fundamental frequencies.

3.2 Simulation 2

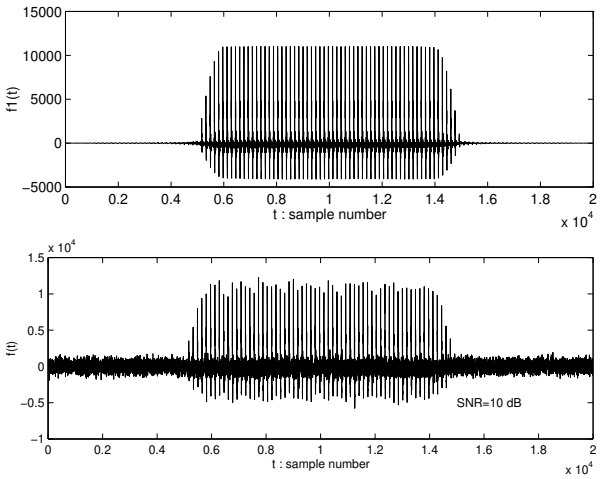


Figure 5: $f_1(t)$ and $f(t)$ (SNR=10dB).

This simulation assumes that $f_1(t)$ is a synthetic vowel, as shown in Fig. 5, where $F_0 = 125$ Hz, $N_{F_0} = 40$, and it is the vowel /a/ synthesized by LMA, and that $f_2(t)$ is a bandpassed random noise with a bandwidth of about 6 kHz. Three types of $f(t)$ are used as simulation stimuli, where the SNRs of $f(t)$ are from 0 to 20 dB in 10-dB steps. The mixed signal for SNR= 10 dB is plotted in Fig. 5.

The simulations were carried out using the three mixed signals. The average SDs of $f_1(t)$ and $f(t)$ are shown in Fig. 6. Hence, it is possible to reduce the SD by about 15 dB as noise reduction by using the proposed method. For example, when the SNR of $f(t)$ is 10 dB, the proposed method can segregate $A_k(t)$ with high precision, as shown in Fig. 7, and it can extract the $\hat{f}_1(t)$ shown in Fig. 7 from the $f(t)$ shown in Fig. 5. Therefore, the proposed model can also extract the amplitude information of speech $f_1(t)$ from a noisy speech $f(t)$ with high precision when speech and noise exist in the same frequency region. Hence, this method can be used to extract a speech signal from noisy speech.

4 Conclusion

In this paper, we proposed a method of signal extraction from noisy signal using physical constraints related to the four regularities proposed by Bregman, and by solving the problem of segregation two acoustic sources. We carried out two simulations on segregating two-acoustic sources using noise-added signal $f(t)$ to show that the proposed method can extract the desired signal $f_1(t)$ from it. Simulation 1, showed that the proposed method can extract an AM complex tone from a noise-added AM complex tone in which signal and noise exist in the same frequency region, with high precision. In particular, using the proposed method, it is possible to reduce the SD by about 20 dB. Moreover, simulation 2 showed that the proposed method can also extract a speech signal from noisy speech.

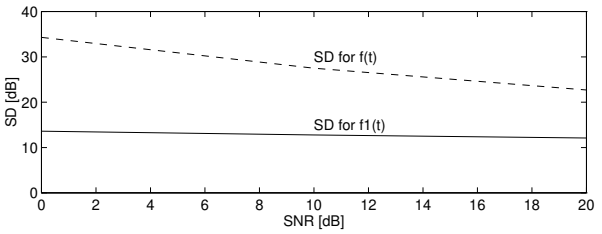


Figure 6: SDs of $\hat{f}_1(t)$ and $f(t)$.

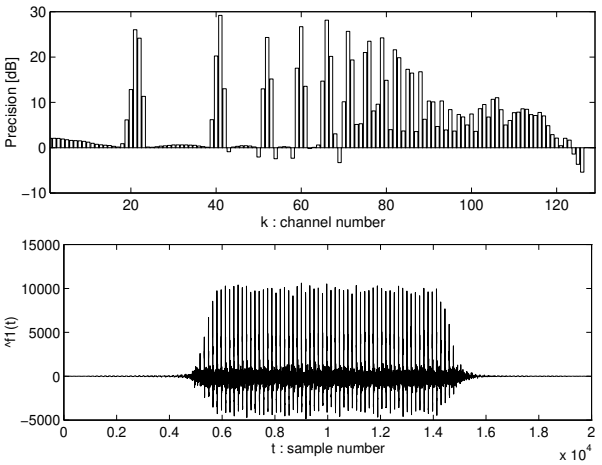


Figure 7: Precision property for $\hat{f}_1(t)$.

References

[1] A. S. Bregman, "Auditory Scene Analysis: hearing in complex environments," in Thinking in Sounds, (Eds. S. McAdams and E. Bigand), pp. 10–36, Oxford University Press, 1993.

[2] Guy J. Brown and Martin Cooke, "Computational auditory scene analysis: Exploiting principles of perceived continuity," Speech Communication, pp. 391-399, North Holland, 13, 1993.

[3] T. Nakatani, H.G. Okuno and T. Kawabata, "Unified Architecture for Auditory Scene Analysis and Spoken Language Processing," ICSLP '94, 24, 3, 1994.

[4] Masashi Unoki and Masato Akagi. "A Method of Signal Extraction from Noise-Added Signal," IEICE,vol.J80-A, no.3, March 1997 (in Japanese).

[5] Masashi Unoki and Masato Akagi. "A Method of Signal Extraction from Noisy Signal based on Auditory Scene Analysis," IJCAI-97 Workshop on CASA'97, Nagoya, Japan, August 1997.