A fundamental study on modeling of auditory segregation based on auditory scene analysis

Masashi UNOKI
School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 1996

Abstract

Key words : auditory scene analysis, two acoustic sources segregation, gammatone filter, wavelet filterbank, co-modulation masking release(CMR)

1 Introduction

Recentry, some computational models based on auditory scene analysis(ASA) have been proposed. The aims of these models are to construct a robust speech recognition system and to model cocktail party effects. This can lead construction of computational theory on audition. As a modeling of auditory segregation, Brown and Cooke[1] proposed a segregation model based on acoustic events, Ellis[2] proposed a segregation model based on psychoacoustic grouping rules, Nakatani et al.[3] proposed stream segregation agents, and Kashino et al.[4] proposed comprehensive music-understanding system. All these segregation models were used amplitude(or power) supectrum as acoustic features. Thus, these models can not segregate mixed signals in same frequency region completely.

This paper presents an extraction method of the signal from noise-added signals as a modeling of acoustic source segregation, to segregate mixed signals in same frequency region completely by using amplitude and phase spectrum on auditory filterbank. The extraction method is based on some constraints of ASA and can solve the problems to segregate two acoustic sources by using three physical clues : phase difference between two acoustic sources $\theta_k(t)$, amplitude and phase of auditory filter outputs, $S_k(t)$ and $\phi_k(t)$. The physical clues are derived from two of

four constraints proposed by Bregman[5]. The simulation in this paper deals with a pure tone mixed with a narrow-band white signal. The results show that the proposed method can extract the pure tone when it is mixed with an amplitude modeulated band-passed noise and can not extract the pure tone mixed with a narrow-band white noise. This result indicates that the proposed method is one of the computational model to be able to explain the CMR.

2 Formulation of a mixed waveform segregation problem

A source waveform segregation problem is formulated as follows:

1. An observed signal is $f(t) = f_1(t) + f_2(t)$, mixed with two signals, $f_1(t)$ and $f_2(t)$.

2. The signal $f(t)$ is passed through an auditory filterbank. This filterbank is constructed by the wavelet transform using the gammatone filters as an analysing wavelet.

3. Outputs of the k-th channel related to two signals, $f_1(t)$ and $f_2(t)$ are

$$f_1(t) \ : \ A_k(t)\sin(\omega_k t) \tag{1}$$
$$f_2(t) \ : \ B_k(t)\sin(\omega_k t + \theta_k(t)). \tag{2}$$

where ω_k is the center-frequency of the k-th channel and $\theta_k(t)$ is the phase defference between two acoustic sources. Then, the output of the k-th channel $X_k(t)$ is represented as follows:

$$\begin{aligned} X_k(t) &= A_k(t)\sin\omega_k t + B_k(t)\sin(\omega_k t + \theta_k(t)) \\ &= S_k(t)\sin(\omega_k t + \phi_k(t)) \end{aligned} \tag{3}$$

where $S_k(t)$ and $\theta_k(t)$ are

$$S_k(t) = \sqrt{A_k^2(t) + 2A_k(t)B_k(t)cos\theta_k(t) + B_k^2(t)} \tag{4}$$

$$\phi_k(t) = \tan^{-1}\left(\frac{B_k(t)\sin(\theta_k(t))}{A_k(t) + B_k(t)\cos\theta_k(t)}\right). \tag{5}$$

4. Since the amplitude $S_k(t)$ and center frequency ω_k of k-th channel output are able to observe, if the phases $\phi_k(t)$ and $\theta_k(t)$ are determined, the amplitudes of two signals, $A_k(t)$ and $B_k(t)$, are estimated as follows:

$$A_k(t) = \frac{S_k(t)\sin(\theta_k(t) - \phi_k(t))}{\sin\theta_k(t)} \tag{6}$$

$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t))}{\sin\theta_k(t)}. \tag{7}$$

5. Finaly, each signal can be reconstructed by grouping all separated signals calculated in each channel through Step 3 to 4.

3 Extraction of three physical clues

Extraction of three physical clues is carried out as follows:

1. The amplitude and phase of auditory filter outputs, $S_k(t)$ and $\phi_k(t)$, are estimated by using the amplitude and phase spectra defined by complex wavelet tramsform.

2. The input-phase $\theta_k(t)$ can be derived by applying following Steps related to three constraints proporsed by Bregman.

   (a) Gradualness of change : A single sound tends to change its properties smoothly and slowly.

   (b) Continuity : Three physical parameters, $A_k(t)$, $B_k(t)$ and $\theta_k(t)$, must be continuous in separation-boundary.

   (c) Many changes that take place in an acoustic event will affect all the components of the resulting sound in the same way and at the same time.

Step i. The constraint(a) is translated as $dA_k(t)/dt = 0$. Then, the input-phase $\theta_k(t)$ is determined as

$$\theta_k(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C_0}\right), \tag{8}$$

where C_0 is unknown parameter.

Step ii. The constraint(b) leads

$$\left|[Q_k(t)]_{R:t\to+T_t} - [Q_k(t)]_{R:t\to-T_r}\right| \le \Delta Q, \tag{9}$$

where $Q_k(t)$ is $A_k(t)$, $B_k(t)$ or $\theta_k(t)$ and ΔQ is ΔA, ΔB or $\Delta\theta$. Then, the region where all parameters satisfies constraint(b), $C_\alpha \leq C_0 \leq C_\beta$, is pressed.

Step iii. The constraint(c) is regarded as

$$\max_{C_\alpha \leq C_0 \leq C_\beta} \left(\frac{< B_k, \hat{B}_k >}{||B_k|| ||\hat{B}_k||} \right), \tag{10}$$

where $\hat{B}_k(t) = (B_{k-1}(t) + B_{k+1}(t))/2$. Then, C_0 is determined when the correlation between $B_k(t)$ and adjacent filters becomes maximum at any C_0 within $[C_\alpha, C_\beta]$. This constraint comes from knowledge about CMR.

4 Experiments and Results

Experiments are carried out as follows:

1. A separating section is determined by onset and offset detected from $dS_k(t)/dt$ and $d\phi_k(t)/dt$.
2. This section is divided into small segments, M/f_0 where f_0 is a center frequency on auditory filterbank.
3. By applying equations(6) to (10), segregation of two waveforms is done.

This paper supposes that $f_1(t)$ is pure tone and $f_2(t)$ is amplitude modulated band-passed noise or narrow-band white noise. Additionally, parameters for calculation are set that the length of the small segment is $3/f_0$, $\Delta B = 50$, $\Delta\theta = \pi/20$ and ΔA is amplitude difference between those of pre- and post segments. When the signal $f_2(t)$ is an amplitude modulated band-passed noise, the mixed signal $f_C(t)$ is supposed the CMR situation. In this case, this model can extract the pure tone $f_1(t)$ from the mixed signal $f_C(t)$. The SNR on $f_1(t)$ is 15.82[dB] and the SNR on $f_2(t)$ is 11.76[dB]. On the other hand, when the $f_2(t)$ is a narrow-band white noise, the mixed signal $f_M(t)$ is supporsed masking situation. In this case, this model can not extract the pure tone $f_1(t)$ from the mixed signal $f_M(t)$. The SNR on $f_1(t)$ is 0.13[dB] and the SNR on $f_2(t)$ is 6.47[dB].

To simulate CMR, band widths of the auditory filterbank are tuned as the same as the human auditory system. In this case, the more number of

adjacent channels are used, the more completely the segregation method can extract the pure tone in the mixed signal $f_C(t)$. However, the case of the mixed signal $f_M(t)$ shows worse results.

5 Conclusion

The new model for auditory segregation was prororsed. The results show that the proposed method can extract the pure tone when it is mixed with an amplitude modeulated band-passed noise and can not extract the pure tone mixed with a narrow-band white noise. This result indicates that the proposed method is one of the computational model to be able to explain the CMR.

References

[1] Guy J.Brown and Martin Cooke：“Comutational auditory scene analysis,” Computer Speech and Language , pp.297-336, 8(1994).

[2] D.P.W.Ellis: “A Computer Implementation of Psycoacoustic Grouping Rules,” Proc. 12th Int. Conf. on Pattern Recognition(1994).

[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] Kunio Kashino,“Toward computational auditory scene analysis – A first step –,” The Jurnal of the Acoustical Society on Japan, vol.50 No.12, pp.1023-1028(1994).

[5] 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).