Title	振幅変調マスキングの知見基づいた音源分離に関する研究
Author(s)	三浦, 慎
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1262
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

Study on sound source segregation based on knowledge of Amplitude modulation masking

Shin MIURA

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

February 15, 1999

Keywords: amplitude modulation masking, modulation-frequency map, auditory scene analysis, double vowels, sound source segregation

1 Introduction

Most naturally occurring sounds show complex amplitude envelope patterns reflecting characteristics of its sources. Namely, this is an amplitude modulation (AM). The analysis of AM plays important roles in detection, separation and identification of sound sources.

It is well established that the initial stage of auditory processing can be modeled as a bank of bandpass filters (auditory filters). On the other hand, discussions continue concerning the mechanism used to analyze AM patterns of output signals from auditory filters. In majority of cases, this process is modeled as an envelope detector. Namely, this is a lowpass filter.

An alternative hypothesis is that AM is analyzed by an array of channels tuned to different modulation frequencies[1]. There is amplitude modulation masking as well known psychoacoustic experimental results [1][2][3]. In typical experiments, the detection threshold for target modulated components was obtained. It has been shown that modulation masking becomes greatest when the modulation frequencies of target and masker signals are identical, becomes smaller as the modulation frequency difference between the two signals increases.

This paper proposes modulation-frequency map to determine modulation frequency based on this concept. This concept is shown in figure 1 (left panel). For example, when signal component and noise component overlap in the frequency domain, detecting difference of amplitude modulation can be detected, signal component and noise component can be segregated (figure 1 right panel).

Furthermore, this paper proposes a method for segregating concurrent vowels based on the modulation-carrier frequency detection map and the auditory scene analysis [4].

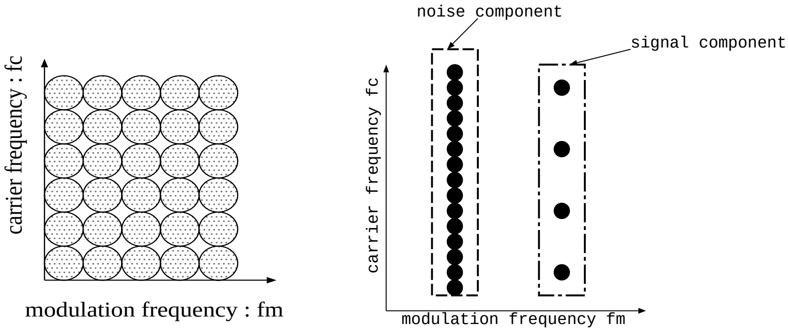

Figure 1: modulation-carrier frequency map(left), example of detection(right)

2 Detection of modulation-carrier frequency map

Figure 2 shows a model of modulation-carrier frequency map. The modulation-carrier frequency map is detected as follows:

- 1. Input signal passes Filter bank1 to separate into each carrier frequency
- 2. instantaneous amplitude envelope is obtained from output of Filter bank1
- 3. Down sampling
- 4.
  - 4-1. DC component is obtained using LPF
  - 4-2. Passed to Filter bank2 AM component is obtained passing through Filter bank2

Filter bank1 is a bank of BPFs to filter carrier frequency . Filter bank2 is a bank of BPFs to filter modulation frequency .

DC component includes Low frequency component and DC because LPF is used . Both Filterbank1 and Filterbank2 are constructed by wavelet filterbank. The basis of wavelet filterbank is a Gabor function.

Experimental signals are shown in figure 3. The left panel signals are the mixed AM sound and masker (AM sound and Band noise) . The right panel signals are the mixed vowel (a male vowel /a/ and a female vowel /i/).

The detected results are shown in figure 4. The result of segregating the mixed AM sound and masker signals show that if the difference of modulation frequencies can be detected , the AM sound and the masker can be segregated (figure 4:left panel).

The result of the mixed vowel signal shows that the fundamental frequency of each vowel can be obtained on the high modulation frequency region in the modulation-carrier frequency map. (figure 4:right panel).

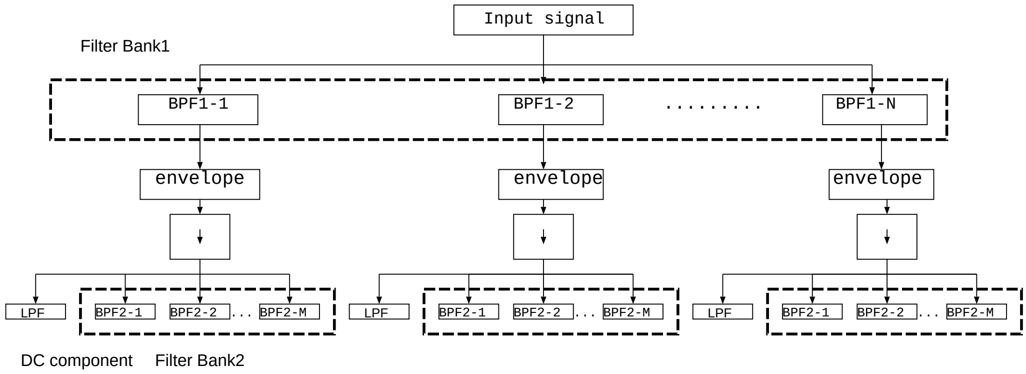

Figure 2: modulation-frequency map model

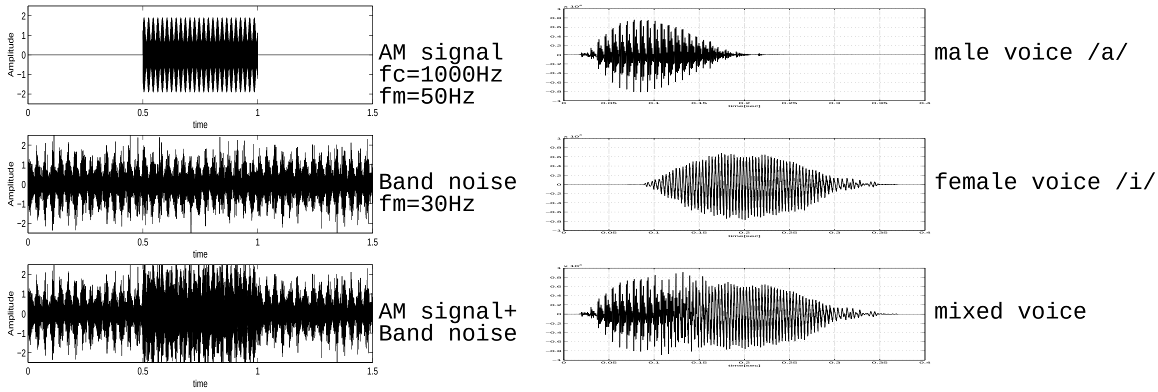

Figure 3: experimental signals

3 Double vowels segregation

The proposed method uses two reguralities of the auditory scene analysis and the modulation-carrier map to segregate mixed sounds. The two reguralities are "Unrelated sounds seldom start or stop at exactly the same time" and "When a body with a repeative perid, its vibration gives rise to an acoustic pattern in which the frequency components are multiples of a common fundamental". The segregated results are shown as follows :

- Before segregation
  - male vowel /a/: 37.3[dB]
  - female vowel /i/: 30.9[dB]
- After segregation
  - male vowel /a/: 18.9[dB]

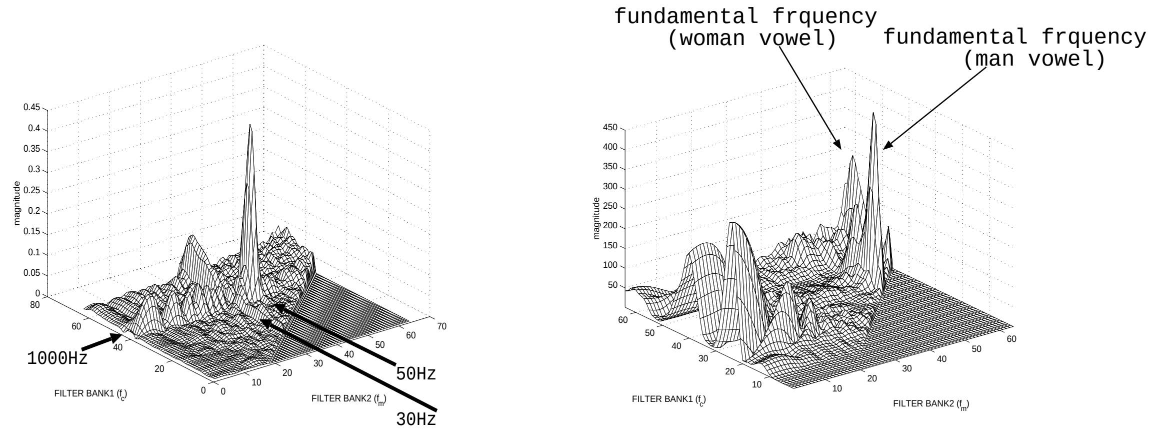

Figure 4: results of modulation-carrier frequency map

– female vowel /i/: 18.3[dB]

The SD value is calculated as follows :

$$SD = 10 \log \frac{1}{\omega} \int_0^{\omega} (|S(\omega)| - |S(\hat{\omega})|)^2 d\omega$$

$S(\omega)$: original signal spectrum
$S(\hat{\omega})$: resolved signal spectrum

4 Conclusion

This paper a model of the proposed modulation-carrier frequency detection map based on knowledge of amplitude modulation masking. The results indicate that AM sound and masker detection are one of the computational model to be able to explain.

Additionary, the result of segregating the mixed AM sound and masker signals show that if the difference of modulation frequencies can be detected, the AM sound and the masker can be segregated. The segregated results showed that double vowels can be segregated by using the modulation-carrier map and regularities of auditory scene analysis.

References

[1] M. Kasino, "A modulation filterbank as a basis for auditory spectro-temporal analysis," ICONIP (1998)

[2] T. Houtgast, "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am 85(4), April 1989

[3] Sid P. Bacon and D. Wesely Grantham, "Modulation masking: Effects of modulation frequency depth, and phase", J. Acoust. Soc. Am 85(6), June 1989

[4] Albert S Bregman , "Auditory scene analysis: hearing in complex environments, in Thinking sounds", pp. 10-36, Oxford Universiyt Press, (1993)