

Title	振幅変調マスク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	<a href="http://hdl.handle.net/10119/1262">http://hdl.handle.net/10119/1262</a>
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

# 振幅変調マスキングの知見に基づいた音源分離に関する研究

三浦 慎

北陸先端科学技術大学院大学 情報科学研究科

1999年2月15日

キーワード： 振幅変調マスキング，変調 搬送周波数マップ，聴覚の情景解析，二重母音，音源分離.

## 1 はじめに

我々が日々耳にする音のほとんどは複雑に時間変化する振幅包絡，すなわち振幅変調 (AM) を伴っている．この振幅変調は音色にも大きな影響を与え音源の識別や分離においても重要な情報となっていることが知られている．

また，聴覚処理における初期の処理過程が BPF のフィルタバンクとしてモデル化できることはよく知られているが，聴覚フィルタ出力の AM 処理のメカニズムに関してまだ定説と呼べるものは存在しない．

多くの場合においては，この AM 処理のプロセスは包絡検出器すなわち LPF としてモデル化されている [1] ．

このような単一の LPF の代わりに，それぞれ異なった変調に同調する BPF 群の存在を仮定している振幅変調マスキングの知見が知られている [2] [3] ．この知見は，純音と帯域雑音を振幅変調したとき，その振幅変調周波数が近いとマスキング量が増加し，逆に変調周波数が離れるとマスキング量が減少する現象である．

本論文では振幅変調マスキングの知見を工学的に図 1 左側のように捕らえ直すことで時間的に変化する音の振幅変調周波数と搬送周波数を検出する振幅 搬送周波数検出マップを構築する．ここで図 1 右側は，振幅変調マスキングを工学的に考えた例である．例えば図のように信号成分と雑音成分の搬送周波数成分 ( $f_c$ ) に重なりがあっても変調周波数 ( $f_m$ ) の違いを検出することによって図のように信号，雑音の成分グルーピングすることができる．

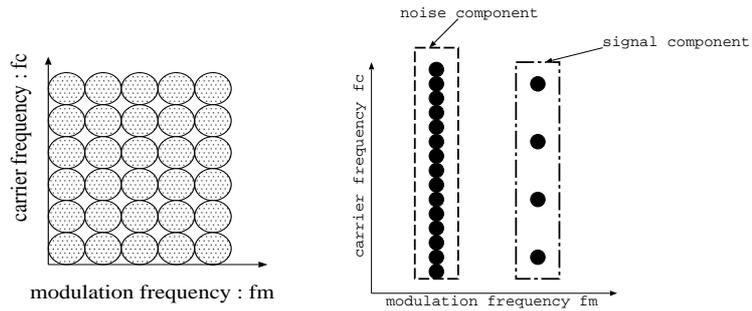


図 1: 変調 搬送周波数マップ (右) とその例 (左)

さらに, 検出された変調 搬送周波数マップと聴覚の情景解析を使って二重母音の分離を試みる.

## 2 変調 搬送周波数マップの構築とシミュレーション

図 2 に構築した変調 搬送周波数マップ検出モデルを示す. 変調 搬送周波数マップは次のように検出される

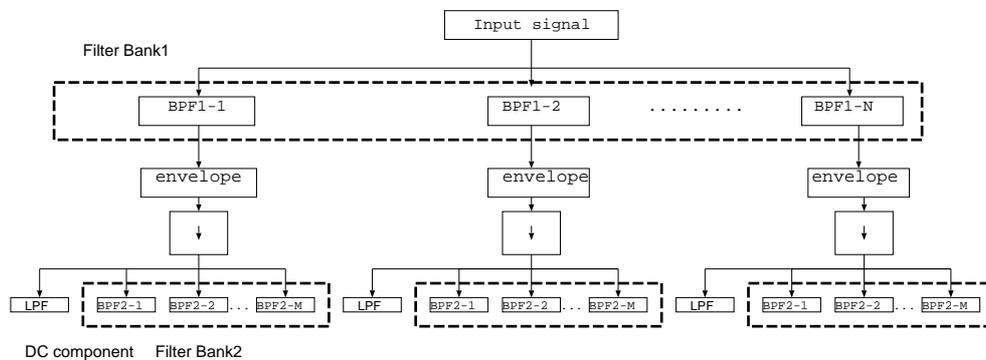


図 2: 変調 搬送周波数マップ検出モデル

- 1. 入力信号を各搬送周波数  $f_c$  に分けるために FILTER BANK1 に通す.
- 2. 各搬送周波数に対する振幅包絡 (envelope) を求めるために FILTER BANK1 の出力より瞬時振幅を求める.
- 3. 求めた瞬時振幅をダウンサンプリングする
- - 4-1. 各搬送周波数成分より LPF を用いて DC 成分を求める

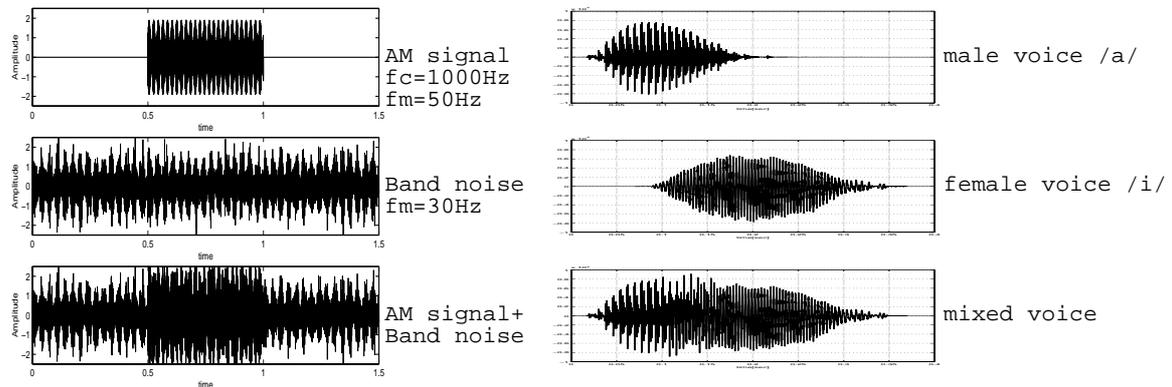


図 3: 解析信号 (左: AM 音とマスキング, 右; 二重母音)

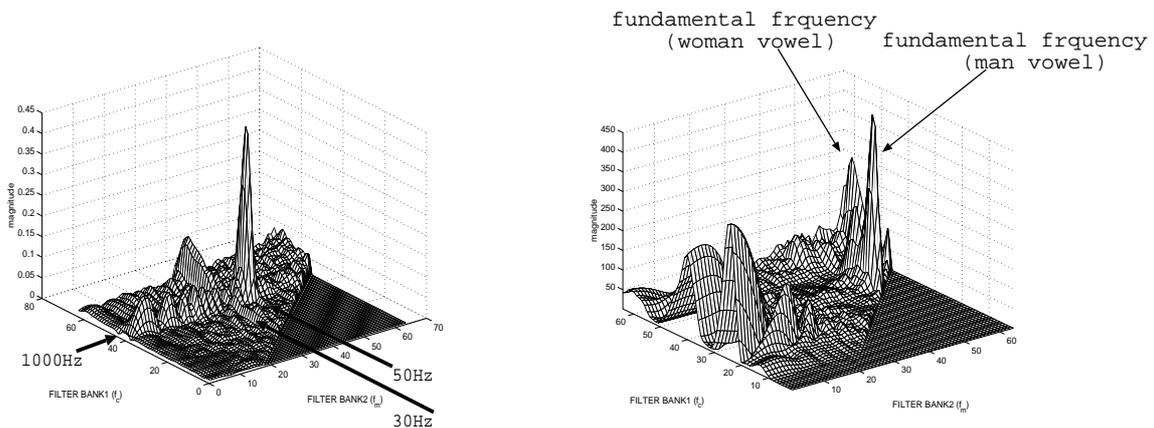


図 4: 変調 搬送周波数マップの検出結果 (時間平均)

- 4-2. 各搬送周波数成分を FILTER BANK2 に通して各搬送周波数成分の変調周波数成分  $f_m$  を求める .

ここで FILTER BANK1 は、搬送周波数を分けるためにある帯域幅を持つ BPF 群である . 同様に FILTER BANK2 は変調周波数を分けるためにある帯域幅を持つ BPF 群である .

ここでの DC 成分とは LPF を使用しているため、実際に求まっているのは DC 成分と非常に低い周波数成分になっている . ここでの BPF 群は Gabor 関数を用いて Wavelet 分析合成系で構築している .

次に構築したモデルを用いて解析を行った . 解析に用いた信号を図 3 に示す . 図左側は信号として搬送周波数 1kHz の純音に変調深さ 0.5 で 50Hz の変調をかけ , マスキングとして広帯域雑音 (1Hz ~ 4800Hz ) に変調深さ 0.5 で 30Hz で変調をかけたものである . 図右側は男性母音 /a/ と女性母音 /i/ を混合した信号である .

図 4 に検出した変調 搬送周波数マップの検出結果を示す。

AM 音とマスキングの検出結果では図 1：右側のように検出され搬送周波数に重なりがあっても変調周波数が違うことによって 2 つの AM 音とマスキングの分離・抽出が行える（図 4：左側）。また，二重母音の解析結果は，高い変調周波数の位置にそれぞれの音声の基本周波数が表れている（図 4：右側）。

### 3 二重母音の分離・抽出

図 4 の検出結果を基に，二重母音の分離・抽出を行った。ここでの分離・抽出は変調搬送周波数マップで検出された基本周波数成分と時間変化を用いて Bregman によって提案された聴覚の情景解析 [4] の 2 つの規則”調波関係に関する規則”と”共通の立ち上がり/立ち下がりに関する規則”を実装することで行った。以下に二重母音を分離・抽出した結果を示す。

- 分離前

- 男性母音/a/ 平均 SD 値：37.3 [dB]
- 女性母音/i/ 平均 SD 値：30.9 [dB]

- 分離後

- 男性母音/a/ 平均 SD 値：18.9 [dB]
- 女性母音/i/ 平均 SD 値：18.3 [dB]

- 改善量

- 男性母音/a/ 平均 SD 値：18.4 [dB]
- 女性母音/i/ 平均 SD 値：12.6 [dB]

ここで結果は平均 SD 値で示している。SD 値の計算は下式を用いて行った。

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

$S(\omega)$  : 原信号のスペクトル

$S(\hat{\omega})$  : 解析信号のスペクトル

### 4 まとめ

本論文では振幅変調マスキングの知見に基づいて，これを工学的に捕らえ直すことで搬送 変調周波数マップを構築した。構築した変調 搬送周波数マップ検出モデルで解析し

た AM 音とマスクの検出結果が、振幅変調マスクの工学的な説明としても解釈できることが確認された。

さらに、この変調搬送周波数マップで検出された時間変化及び基本周波数成分と聴覚の情景解析の2つの規則を用いることで二重母音の分離ができた。

## 参考文献

- [1] M. Kasino, "A modulation filterbank as a basis for auditory spectro-temporal analysis", ICONIP (1998)
- [2] T. Houtgast Frequency selective time amplitude-modulation detection, *J. Acoust. Soc. Am.* 85(4), April 1989
- [3] S. P. Bacon and D. Wesely Grant ham Modulation masking: Effect 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 University Press, (1993)