共変調マスキング解除の計算モデルの高性能化*

◎ 鵜木　祐史　　赤木　正人 (北陸先端科学技術大学院大学)

1. はじめに

1984 年、Hall らは聴覚フィルタ間の比較によって変動する雑音にマスクされた正弦波信号の検出が容易になるという可能性を示した。この現象は、共変調マスキング解除 (Co-modulation Masking Release : CMR) [1] として知られており、多くの聴覚心理実験によって同様の結果が得られている。しかし、CMR が起こる条件が知られているのにもかかわらず、この条件を利用した工学的モデルはほとんどない。

一方、著者らは Bregman によって提唱された聴覚の情景解析 (Auditory Scene Analysis) に基づく音源分離モデルの研究を行なっている[3]。この成果の１つとして、人間の聴覚特性に合わせて分析合成系を設計し、著者らによって提案されている２波形分離問題の解法を適用することで CMR の計算モデルが得られている[4]。本稿では、この CMR の計算モデルについて再考し、モデルの高性能化を計る。

2. 共変調マスキング解除の計算モデル

著者らによって提案された共変調マスキング解除の計算モデル[4]は、図１のように Gammatone filter を基底関数とした wavelet 分析合成系で構成される。これは、解析周波数範囲が 60 ~ 6000Hz、フィルタ数が $K = 128$ であり、中心周波数 $f_0 = 600$Hz のとき、Gammatone filter の等価矩形帯域幅 (ERB) をちょうど幅１とした定 Q フィルタバンクとして設計されている[4]。また、図１の各パラメータは、図２に示すような２波形分離問題の定式化に従っている。ここで、振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ、複素 wavelet 変換で定義された振幅スペクトルと位相スペクトルから導出でき、入力位相 $\theta_k(t)$ は、図３に示すような３つの物理的制約条件を適用することで得られる[4]。特に、規則 (i),(iii) は、Bregman が提唱した発見的規則[2]を物理的制約条件に捉え直したものであり、規則 (iii) は CMR のよい説明になっている[2]。

このモデルでは、$f_1(t)$ を純音 (正弦波信号)、$f_2(t)$ をその純音の周波数を中心周波数とする２種類のマスカー (振幅変調された (AM) 帯域雑音 or ランダム帯域雑音) とし、$f_2(t)$ が存在している状態で $f_1(t)$ が加算される状況を想定している。また、CMR の実験条件として Hall らは帯域雑音の帯域幅を関数とした正弦波信号の閾値を測った[1]が、このモデルでは

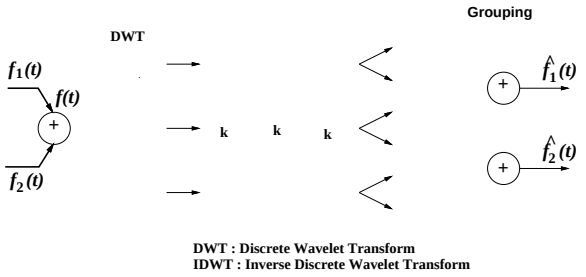

図 1 . wavelet 分析合成系

1. 各音響信号の周波数成分

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

2. k番目の分析フィルタの出力

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

3. 振幅包絡 $S_k(t)$

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

4. 出力位相 $\theta_k(t)$

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

5. ２波形の振幅包絡 $A_k(t)$ と $B_k(t)$

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

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

図 2 . ２波形分離問題の定式化

帯域雑音の帯域幅を固定にして、式 (10) の隣接する聴覚フィルタ数 L を関数とした純音の分離精度 (SN 比) を測ることでこれと等価と見なした。

このとき、式 (10) の L を増やしていけば、マスカーが AM 帯域雑音のとき純音 $\hat{f}_1(t)$ の検出が容易になるが、ランダム帯域雑音のときほとんど検出されないという特性が得られた[4]。この L–SN 比特性が CMR の特性と類似していることから、このモデルを CMR の計算モデルと解釈した。

しかし、この計算モデルでは、Hall らによって行なわれた CMR の実験を模擬しているものの分析合成系の関係で純音の中心周波数が異なっていた。また、式 (10) の雑音の振幅包絡 $\hat{B}(t)$ を求めるとき、Gammatone filter による群遅延による影響や雑音のパワーが平坦でない場合を考慮していなかった。

---

* A Study on refined Computational model of Co-modulation Masking Release

By Masashi Unoki and Masato Akagi (JAIST)

(i) 漸近的変化 (**Bregman**)

$$\frac{dA_k(t)}{dt} = C_{k,R}(t)$$

$$\Downarrow$$

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

$$C_k(t) = -\int C_{k,R}(t)dt + C_{k,0}$$

但し、微小区間で未定関数を定数とする ($C_k(t) = C_{k,0}$)

(ii) 連続性 (時間的近接)

$$\begin{aligned} |A_k(T_r+0) - A_k(T_r-0)| &\le \Delta A \\ |B_k(T_r+0) - B_k(T_r-0)| &\le \Delta B \\ |\theta_k(T_r+0) - \theta_k(T_r-0)| &\le \Delta\theta \\ \Rightarrow \quad C_{k,\alpha} &\le C_{k,0} \le C_{k,\beta} \end{aligned} \quad (9)$$

(iii) 1つの音響事象に生じる変化 (**Bregman**)

$$B_k(t) \approx B_{k\pm\ell}(t), \quad \ell = 1,2,\cdots,L$$

$$\hat{\hat{B}}_k(t) = \frac{1}{2L}\sum_{\ell=-L,\ell\neq 0}^{L}\hat{B}_{k+\ell}(t) \quad (10)$$

$$\Rightarrow \max_{C_{k,\alpha}\le C_{k,0}\le C_{k,\beta}} \frac{<\hat{B}_k, \hat{\hat{B}}_k>}{||\hat{B}_k||||\hat{\hat{B}}_k||} \quad (11)$$

$\Rightarrow$ $C_{k,0}$より入力位相 $\theta_k(t)$ を決定できる (式 (8))

図 3．入力位相 $\theta_k(t)$ の決定

3. モデルの改良点

本稿では、上記の点から、(a) wavelet 分析合成系の中心周波数 (1kHz) と解析周波数範囲 (0.1 ~ 10kHz) の変更、(b) Gammatone filter の群遅延の補正 (alinement 処理)、(c) 振幅包絡 $\hat{\hat{B}}_k(t)$ の正規化を考慮し、モデルの高性能化を計ることにした。特に、(a) では中心周波数 $f_0 = 1$kHz のときに ERB が幅 1 となるように wavelet 分析合成系を再設計し、(c) では式 (10) を次式のように変更した。

$$\hat{\hat{B}}_k(t) = \frac{1}{2L}\sum_{\ell=-L,\ell\neq 0}^{L}\frac{\hat{B}_{k+\ell}(t)}{\|\hat{B}_{k+\ell}(t)\|} \quad (12)$$

4. 2 波形分離のシミュレーション

Hall らによる実験と等価な条件を考慮して、実験データは、周波数を 1kHz、呈示時間を 400msec、振幅を一定とした正弦波信号 $f_1(t)$ と 2 種類のマスカー (振幅変調された帯域雑音 $f_{21}(t)$ とランダム帯域雑音 $f_{22}(t)$) を用意した。ここで、帯域雑音の中心周波数を 1kHz、帯域幅を 1kHz、$f_{21}(t)$ の変調速度を 50Hz とした。このとき、純音と帯域雑音の SN 比は -6.61dB であった。刺激は、開始時刻を変化させた純音 $f_1(t)$ を 10 個、乱数の種を変えて作成した 2 種類のマスカーをそれぞれ 5 個とし、合計 50 個の混合信号を用意した。また、2 波形分離に必要な各パラメータ[4]は、微小区間を$\Delta t = 3/f_0$、$S_k(t)$ の最大値を $S_{\max}$ としたときに$\Delta B = 0.01S_{\max}$、$\Delta\theta = \pi/20$、

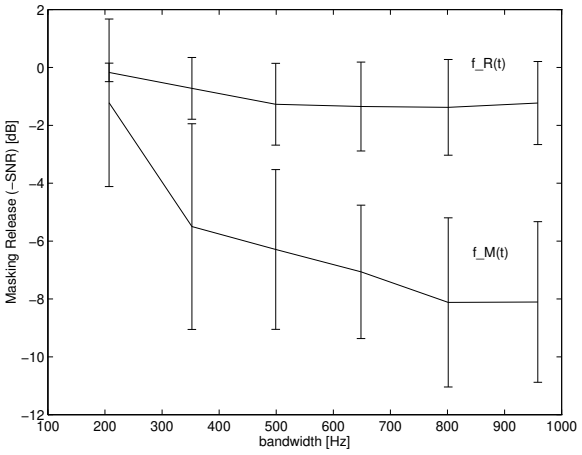

図 4．帯域幅とマスキング解除の関係

$\Delta A = |A_k(T_r - \Delta t) - A_k(T_r - 2\Delta t)|$ とし、隣接する聴覚フィルタ数を $L = 1, 3, \cdots, 11$ とした。

上記の条件に従い CMR を想定した 2 波形分離の結果を図 4 に示す。この図の縦軸は抽出された純音の SN 比の向上量を下向きに表すことで、SN 比向上量をマスキング解除量と読み替えている。また、横軸は隣接する聴覚フィルタ数 L に対応した帯域幅を表し、曲線は SN 比の平均を、縦棒は標準偏差を表している。この結果、混合信号 $f_M(t) = f_1(t) + f_{21}(t)$ の場合、隣接する聴覚フィルタの参照数 L を増加させると、抽出された純音$\hat{f}_1(t)$ の SN 比が向上する傾向がみられたが、混合信号 $f_R(t) = f_1(t) + f_{22}(t)$ の場合、参照数 L を増加させても、純音はほとんど検出されず、SN 比はあまり変わらなかった。この結果から、文献 [3] の結果と同様に本モデルが CMR の工学的モデルと解釈できる。尚、本シミュレーションでは、共変調マスキング解除量が最大約 8dB であった。

5. まとめ

著者らによって提案された CMR の計算モデルについて再考した。特に、Gammatone フィルタの群遅延、Hall らによる CMR の実験と等価な実験条件を考慮して高性能化を計った。この結果、計算モデルにより Hall らの実験結果と同様の結果が得られた。

謝辞　本研究の一部は文部省科学研究費補助金 (課題番号 07308026) によって行なわれたものである。

参考文献

[1] Brian C.J.Moore (大串健吾監訳):“聴覚心理学概論”, 第 3 章, 誠信書房 (1994).

[2] 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).

[3] 鵜木, 赤木, “帯域雑音中の AM 調波複合音の一抽出法,” 音響学春季講論, 1997.

[4] 鵜木, 赤木, “共変調マスキング解除の計算モデルに関する一考察,” 信学技報, SP-96-37,July 1996.