共変調マスキング解除の計算モデルの提案

鵜木　祐史　　赤木　正人

北陸先端科学技術大学院大学　情報科学研究科

〒 923-1292　石川県能美郡辰口町旭台 1-1

E-mail : unoki@jaist.ac.jp　akagi@jaist.ac.jp

あらまし　本論文では、共変調マスキング解除 (CMR) の計算モデルを提案する。このモデルは、著者らが提案した二波形分離モデル (モデル A) とマスキングのパワースペクトルモデル (モデル B) の二つのモデル、及び、二つのモデルの処理結果を選択する選択処理部で構成される。モデル A では聴覚フィルタ群の出力を、モデル B では単一の聴覚フィルタの出力を利用して、マスクされた信号から純音を分離抽出する。選択処理部では、この二つの分離抽出された純音からマスキングしきい値の低い方の純音を選択する。本モデルに対し、Hall らによる CMR の実験を想定したシミュレーションを行った結果、分離抽出された純音のマスキングしきい値の変化は、Hall らが示した CMR の結果と類似する傾向が示された。また、このときの共変調マスキング解除量は最大約 8 dB であった。

キーワード　聴覚の情景解析、共変調マスキング、二波形分離、パワースペクトルモデル

A Computational Model of Co-modulation Masking Release

Masashi Unoki　and　Masato Akagi

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

1-1 Asahidai, Tatsunokuchi, Nomigun, Ishikawa 923-1292 Japan
E-mail : unoki@jaist.ac.jp　akagi@jaist.ac.jp

Abstract　This paper proposes a computational model of co-modulation masking release (CMR). It consists of two models, our auditory segregation model (model A) and the power spectrum model of masking (model B), and a selection process that selects one of their results. Model A extracts a sinusoidal signal using the outputs of multiple auditory filters and model B extracts a sinusoidal signal using the output of a single auditory filter. The selection process selects the sinusoidal signal with the lowest signal threshold from the two extracted signals. For the proposed model, simulations similar to Hall *et al.*'s demonstrations were carried out. As a result, the signal threshold of the pure tone extracted using the proposed model shows the similar properties to Hall *et al.*'s demonstrations. The maximum amount of CMR in the proposed model is about 8 dB.

Key words　Auditory Scene Analysis, CMR, segregation, power spectrum model

1　はじめに

聴覚系の周波数選択性の研究において、マスキングの現象を説明するモデルとして、マスキングのパワースペクトルモデル [1] が広く受け入れられている。このモデルでは、聴取者が、背景雑音中で特定の中心周波数をもつ正弦波信号を検知しようとするとき、信号周波数付近で中心周波数をもち、信号対雑音比の最も高くなる単一の聴覚フィルタの出力を利用するものと仮定している。また、刺激は長時間パワースペクトルとして表現されており、信号のマスキングしきい値は、聴覚フィルタを通過する雑音の量によって決定されるものと仮定している。この一連の仮定により、パワースペクトルモデルは同時マスキングといった多くの現象をよく説明できるが、成分音間の相対位相やマスカーの短時間変動を無視しているため、説明できないマスキング現象もいくつか存在した。

Hall (1984) らは聴覚フィルタ間の比較によって、振幅包絡が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 [2]。このような検知能力の向上が生じるための決定的な条件は、異なる周波数帯域間で振幅包絡の変動が一致しているか、あるいは相関があるということであった。Hall らはこの異なる周波数帯域間の振幅包絡の一致を “共変調” と呼び、これによる検知能力の向上、つまりマスキングの解除を共変調マスキング解除 (Co-modulation Masking Release: CMR) と呼んだ。この現象については、多くの聴覚心理実験 [3, 4, 5, 6] が行われており、同様の結果が得られている。しかし、CMR が起こるための条件が知られているにもかかわらず、この条件を利用した計算モデルはほとんど報告されていない。

一方、著者らは、Bregman によって提案された聴覚の情景解析 [7, 8] に基づく二波形分離モデルの研究に取り組んでおり、帯域雑音に埋もれた信号を分離・抽出する方法を提案した [9]。これは、同一周波数領域において信号と雑音を完全に分離するために振幅スペクトルと位相スペクトルを考慮し、Bregman のいう四つの発見的規則 [8] :

(i) 共通の立上がり / 立下がり
(ii) 漸近的変化
(iii) 調波性
(iv) 一つの音響事象に生じる変化

のうち、規則 (ii) と規則 (iv) を物理的制約条件にとらえ直すことで、二波形分離問題を解いている。

本論文では、同時マスキング現象の説明に利用されてきたマスキングのパワースペクトルモデル

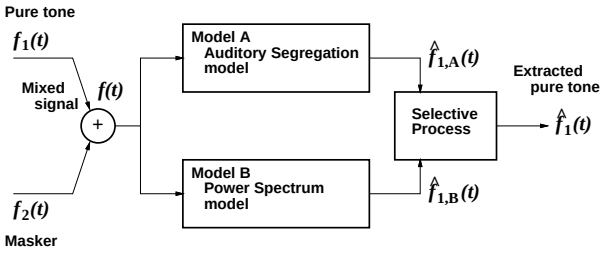


図 1: CMR の計算モデル

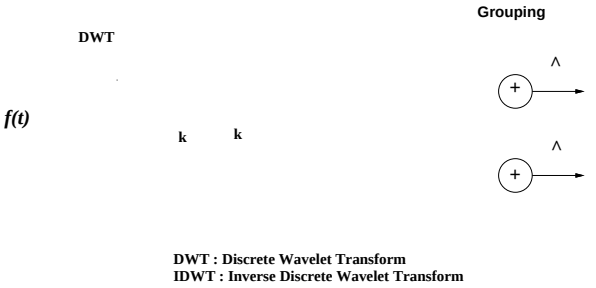
と、著者らが提案する二波形分離モデルの二つのモデルを利用し、これに二つのモデルの結果を選択する処理を付加することで、CMR の計算モデルを提案する。

2　CMR の計算モデル

CMR の計算モデルを図 1に示す。本モデルは、2 つのモデル (A, B) と選択処理で構成される。また、$f_1(t)$ 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二種類のマスカー (ランダム帯域雑音と振幅変調されたランダム帯域雑音) 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 (同時マスキング) を想定している。本モデルはこの混合信号 $f(t)$ だけを受音でき、二つのモデル (A, B) を用いて、純音 $f_1(t)$ を分離抽出する。モデル A は、二波形分離モデル [9] であり、モデル B はマスキングのパワースペクトルモデル [1] である。CMR の計算モデルでは、これら二つのモデルが並行に動作し、それぞれ、マスクされた信号から純音を分離抽出する。ここで、モデル A によって分離抽出された純音を $\hat{f}_{1,\mathrm{A}}(t)$、モデル B によって分離抽出された純音を $\hat{f}_{1,\mathrm{B}}(t)$ とする。最後に、選択処理部は、二つのモデルにより分離抽出された $\hat{f}_{1,\mathrm{A}}(t)$ と $\hat{f}_{1,\mathrm{B}}(t)$ のうちマスキングしきい値の低い方を選択し、これを計算モデルで分離抽出された純音 $\hat{f}_1(t)$ とする。この計算モデルの根本的な考えは、Hall らの実験結果 [2] において、マスカー帯域幅が単一の聴覚フィルタの帯域幅 (1 ERB) を越えるか越えないかでマスキングの傾向が異なっていたことに由来する。言い換えると、Hall らの実験結果において、マスカー帯域幅が 1 ERB を越えないときに、単一の聴覚フィルタの出力を手がかり (モデル B) として説明でき、マスカー帯域幅が 1 ERB を越えるときに、聴覚フィルタ群の出力を手がかり (モデル A) として説明できるというものである。

3　モデル A: 二波形分離モデル

二波形分離モデルを図 2に示す。これは、(a) 聴覚フィルタ群、(b) 波形分離部、(c) グルーピング部の三部で構成される。聴覚フィルタ群は、gammatone フィルタバンクで構成される。波形分離部は、Bregman によって提唱された発見的規則 (ii) と (iv) に関係した物理的制約条件を利用し、各物理パラメータを求める。グルーピング部は、分離された各物理パラメータをそれぞれ合成し、wavelet 逆変換を用いて信号を再構成する。



3.1　聴覚フィルタ群

聴覚フィルタ群は、gammatone フィルタのインパルス応答の実部と虚部が Hilbert 変換で結ばれるような関数を基底関数 $\psi(t)$ とした wavelet 変換を用いて構成される [9, 10]。

$$\psi(t) = At^{N-1}e^{j2\pi f_0 t - 2\pi b_f t} \tag{1}$$

但し、$\mathrm{ERB}(f_0) = 24.7(4.37f_0/1000 + 1)$ である。また、N と b_f はガンマ分布関数のパラメータであり、それぞれ $N = 4$、$b_f = 1.019\mathrm{ERB}(f_0)$ とした。このフィルタ群は、中心周波数 $f_0 = 1$ kHz、通過帯域が 100 Hz ~ 10 kHz、フィルタ数 $K = 128$ の定 Q フィルタバンクであり、中心周波数が 1 kHz のとき、聴覚フィルタの帯域幅が 1 ERB になるように設計された。実際の聴覚フィルタ群は定 Q の特性を持たないが、中心周波数が 800 Hz より高域においておおよそ定 Q であると見なせることと、信号の分析合成を考慮していることから、本研究では、定 Q の gammatone フィルタバンクを利用する。

3.2　分離部とグルーピング部

モデル A は次に示す二波形分離問題の定式化 [9] に従っている。

はじめに、ある二つの音響信号 $f_1(t)$ と $f_2(t)$ が $f(t) = f_1(t) + f_2(t)$ に加算され、混合信号 $f(t)$ のみが観測できるものとする。観測された信号 $f(t)$ は聴覚フィルタ群により周波数分解される。このとき、$f_1(t)$ と $f_2(t)$ に対する k 番目の分析フィルタの出力は、それぞれ

$$f_1(t) : A_k(t)\sin(\omega_k t + \theta_{1k}(t)) \tag{2}$$
$$f_2(t) : B_k(t)\sin(\omega_k t + \theta_{2k}(t)) \tag{3}$$

となる。但し、ω_k は分析フィルタの中心角周波数、$\theta_{1k}(t)$ は $f_1(t)$ のもつ入力位相、$\theta_{2k}(t)$ は $f_2(t)$ の

図 2: モデル A: 二波形分離モデル

もつ入力位相である。このとき、k 番目の分析フィルタの出力 $X_k(t)$ は、式 (2) と式 (3) の和であり、

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

と表される。ここで、$S_k(t)$ は振幅包絡、$\phi_k(t)$ は出力位相である。従って、波形分離部において、四つの物理パラメータ ($S_k(t)$, $\phi_k(t)$, $\theta_{1k}(t)$, $\theta_{2k}(t)$) を求めることができれば、二波形の振幅包絡 $A_k(t)$ と $B_k(t)$ を

$$A_k(t) = S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))/\sin\theta_k(t) \tag{5}$$
$$B_k(t) = S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))/\sin\theta_k(t) \tag{6}$$

で一意に求めることができる [9]。但し、$\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ であり、$\theta_k(t) \neq n\pi, n \in \mathbf{Z}$ である。最後に、グルーピング部で式 (2) と式 (3) より、それぞれの周波数成分を合成することで、$f_1(t)$ と $f_2(t)$ を再構成する。但し、再構成された $f_1(t)$ と $f_2(t)$ をそれぞれ、$\hat{f}_{1,\mathrm{A}}(t)$、$\hat{f}_{2,\mathrm{A}}(t)$ とする。

本論文では、$f_1(t)$ を純音、$f_2(t)$ をマスカーとし、$f_1(t)$ の信号周波数とフィルタの中心周波数が一致しているものと考えるため、入力位相 $\theta_{1k}(t) = 0$、$\theta_k(t) = \theta_{2k}(t)$ として二波形分離問題を考える。

3.3 四つのパラメータの計算方法

振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ 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求めることができる [9]。ここで、$\theta_{1k}(t) = 0$ の仮定から、残る入力位相 $\theta_{2k}(t)$ を求めなければならない。そこで、Bregman によって提唱された発見的規則 (ii) と (iv) から導かれた三つの物理的制約条件を用いることで、$\theta_{2k}(t)$ を求める [10]。

1. 漸近的変化 (なめらかさ)

   発見的規則 (ii) は、“一つの音の振幅包絡はゆっくりと滑らかに変化する”[8] ということ

を意味する。この “ゆっくりと” という規則をとらえ直した物理的制約条件は、$dA_k(t)/dt = C_{k,R}(t)$ である [10]。ここで、$C_{k,R}(t)$ は R 回微分可能な R 次多項式である。この制約を式 (5) に適用して得られた線形微分方程式を解くことで、入力位相 $\theta_{2k}(t)$ の一般解：

$$\theta_{2k}(t) = \arctan\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C_k(t)}\right) \tag{7}$$

を得る。但し、$C_k(t) = -\int C_{k,R}(t)dt + C_{k,0}$ である。ここで、微小区間 Δt において $C_k(t) = C_{k,0}$、すなわち $dA_k(t)/dt = 0$ であるものとする。

2. 漸近的変化 (なめらかに)

“なめらかに” をとらえ直した物理的制約条件は、分離を行った微小区間 ($T_r - \Delta t \le t < T_r$) と分離を行う微小区間 ($T_r \le t < T_r + \Delta t$) の境界 T_r において、

$$\begin{aligned} |A_k(T_r+0) - A_k(T_r-0)| &\le \Delta A \\ |B_k(T_r+0) - B_k(T_r-0)| &\le \Delta B \\ |\theta_{2k}(T_r+0) - \theta_{2k}(T_r-0)| &\le \Delta\theta \end{aligned} \tag{8}$$

を満たすことである [10]。上式の関係から、$C_{k,0}$ を決定するために、この物理的制約条件を $C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}$ に制限することと解釈する。但し、$C_{k,\alpha}$ と $C_{k,\beta}$ はこの境界における $C_{k,0}$ の上限と下限である。

3. 一つの音響事象に生じる変化

発見的規則 (iv) は “一つの音響事象に生じる変化は、その音を構成する各成分に同じような影響を与える”[8] ということを意味する。これをとらえ直した物理的制約条件は、

$$\frac{B_k(t)}{\|B_k(t)\|} \approx \frac{B_{k\pm\ell}(t)}{\|B_{k\pm\ell}(t)\|}, \qquad \ell = 1, 2, \cdots, L \tag{9}$$

である [10]。この制約を、マスカーの振幅包絡間の相関を尺度とした

$$C_{k,0} = \arg \max_{C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}} \frac{\langle \hat{B}_k, \hat{\hat{B}}_k \rangle}{||\hat{B}_k|| ||\hat{\hat{B}}_k||} \tag{10}$$

を用いて最適な $C_{k,0}$ を選ぶことと解釈する。ここで、$B_k(t)$ は、式 (6) と式 (7) から $C_{k,0}$ の関数であり、ある $C_{k,0}$ によって決定された振幅包絡を $\hat{B}_k(t)$ とする。また、$\hat{\hat{B}}_k(t)$ は次のように隣接する聴覚フィルタ ($k \pm \ell$) の出力から得られるマスカーの振幅包絡であり、

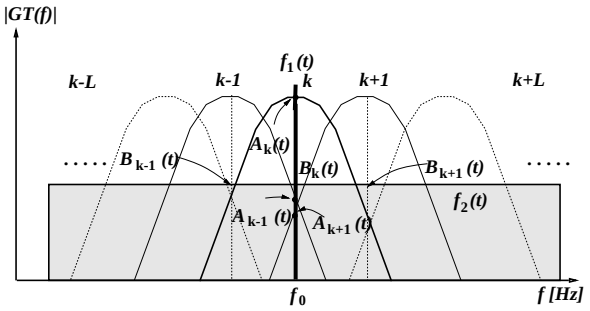


図 3: 隣接する聴覚フィルタ出力における純音 $f_1(t)$ とマスカー $f_2(t)$ の特性

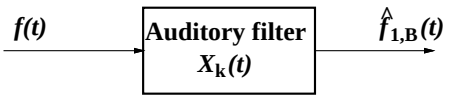


図 4: モデル B: パワースペクトルモデル

ある $C_{k,0}$ によって決定された純音の振幅包絡 $A_k(t)$ と図 3の振幅特性の関係を利用して決定される。

$$\hat{\hat{B}}_k(t) = \frac{1}{2L} \sum_{\ell=-L,\ell\neq 0}^{L} \frac{\hat{B}_{k+\ell}(t)}{\|\hat{B}_{k+\ell}(t)\|} \tag{11}$$

上記の計算過程を要約すると、制約条件 1 より、$\theta_{2k}(t)$ の一般解が導出され、それを一意に決定するために、制約条件 2 より未定係数の候補を求め、制約条件 3 より最適な未定係数を決定することで、一意な $\theta_{2k}(t)$ が決定される。

4 モデル B: パワースペクトルモデル

Patterson と Moore によって提案されたマスキングの “パワースペクトルモデル” [1] では、聴取者が背景雑音から正弦波信号を検知するときに、信号周波数近傍の中心周波数をもち、信号対雑音比が最も高い単一の聴覚フィルタの出力を利用するものと仮定した。従って、雑音の成分のうち、このフィルタを通過する成分だけが信号のマスキングに影響を及ぼすものと考えることができる。また、信号のしきい値は聴覚フィルタを通過する雑音の量によって決定される。

本論文では、このパワースペクトルモデルを図 4に示すモデルで構成する。ここで、$X_k(t)$ はモデル A で構成された聴覚フィルタ群のうちの一つの聴覚フィルタの出力である。また、聴覚フィル

タは、中心周波数が 1 kHz、帯域幅が 1 ERB の gammatone フィルタで構成される。このモデルでは、混合信号 $f(t)$ に対し、単一聴覚フィルタ (帯域幅が 1 ERB の帯域通過フィルタ) を通過した出力を、分離抽出された純音 $\hat{f}_{1,\mathrm{B}}(t)$ とする。

ここで、純音の分離抽出に関するモデル A とモデル B の決定的な違いは、図 3の単一フィルタ出力 (例えば、中心周波数が f_0 のとき) において、その出力をそのまま利用して純音を分離抽出するか、あるいは隣接するフィルタ出力も利用して純音を分離抽出するかということである。

5 シミュレーション

5.1 共変調マスキング解除 (CMR)

Hall らは、CMR の実験の一つで、1 kHz、400 msec の正弦波信号のしきい値をスペクトルレベルを一定に保った雑音マスカーの帯域幅の関数として測定した [2, 4]。また、マスカーの中心周波数は 1 kHz であり、次のような二種類のマスカーが用いられた。

- ランダム帯域雑音：振幅は不規則にかつ異なる周波数領域において独立に変動する。
- 振幅変調されたランダム帯域雑音：ランダム帯域雑音であるが、ランダム帯域雑音の振幅を不規則なゆっくりとした速度で変調 (50 Hz の低域通過フィルタリング) したものである。振幅変動は異なる周波数領域において等しい。

この二種類のマスカーを用いて正弦波信号の検知能力を測定したところ、図 5に示す結果が得られた。図中の点 R はランダム帯域雑音の場合の信号のしきい値を表し、点 M は振幅変調された雑音の場合の信号のしきい値を表している。この結果、帯域雑音の帯域幅が、この中心周波数での聴覚フィルタの帯域幅 (約 130 Hz) を越えない場合、いずれの帯域雑音についてもマスキング量が増加している。一方、この帯域幅を越える場合、ランダム帯域雑音ではマスキング量が変動しないのに対し、振幅変調されたランダム帯域雑音の場合、マスカー帯域幅の増加に従って、マスキング量が減少している。この結果から、Hall らは、異なる聴覚フィルタ間の比較によって、聴取者は信号検出能力を高めることができることを示し、この現象を CMR と呼んだ。この実験では、共変調マスキング解除量は最大約 10 dB であった。

本論文では、Hall らの実験と等価な条件を考慮し、本モデルが CMR の特性を模擬することを検証するため、次のような計算機シミュレーションを行う。

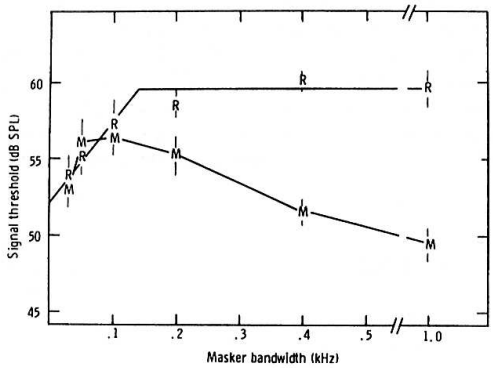


図 5: CMR の実験結果 (Hall et al., 1984)

5.2 モデル A のシミュレーション

5.2.1 刺激とシミュレーション条件

実験データは、Hall らの実験と等価な条件を考慮するため、サンプリング周波数 20 kHz、周波数を 1 kHz、呈示時間を 400 msec、振幅を一定とした正弦波信号 $f_1(t)$ と $f_1(t)$ の周波数を中心周波数とした二種類のマスカー $f_2(t)$(ランダム帯域雑音と振幅変調されたランダム帯域雑音) を用意した。ここで、$f_{21}(t)$ はランダム帯域雑音であり、ある乱数の種を設定することで生成される白色雑音を基に、これを帯域制限することで得らる。また、$f_{22}(t)$ は振幅変調されたランダム帯域雑音であり、変調速度が 50 Hz、変調度が 100 %で $f_{21}(t)$ を振幅変調 (50 Hz の低域通過フィルタリング) したランダム帯域雑音である。このとき、$f_2(t)$ のパワーが $\sqrt{f_{21}(t)^2/f_{22}(t)^2} = 1$ となるように調整され、$f_1(t)$ と $f_2(t)$ の SNR (signal to noise ratio) は -6.61 dB であった。これらの実験データを図 6 (左側) に示す。ここで、各混合信号は $f_R(t) = f_1(t) + f_{21}(t)$, $f_M(t) = f_1(t) + f_{22}(t)$ であり、それぞれ Hall らの実験データで用いられた点 R、点 M の刺激に対応する。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 (5 種類) を変化させて作成した二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。このときの混合信号の一例を図 6 (右側) に示す。ここで、いずれの混合信号においても、純音 $f_1(t)$ は視覚的にマスカーに埋もれていたが、聴覚的には $f_M(t)$ で純音を容易に検知でき、$f_R(t)$ で純音を検知することが困難であった。

次に、Hall らの実験と等価なシミュレーション

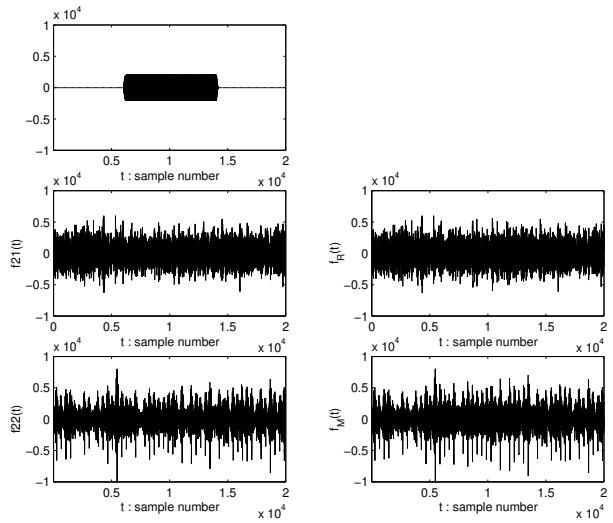


図 6: 刺激 :(左上) 原信号 : 純音 $f_1(t)$, (左中) ランダム帯域雑音 $f_{21}(t)$, (左下) 振幅変調されたランダム帯域雑音 $f_{22}(t)$, (右上) 混合信号 $f_R(t)$, (右下) 混合信号 $f_M(t)$

条件を考える際、CMR で利用する手がかりの幅を制御するために聴覚フィルタ間の帯域幅を知る必要があるが、この実験において人間がどの程度の幅の聴覚フィルタ間の手がかりを利用したか分からない。そのため、本研究では、CMR を起こすために与えた手がかりの帯域幅 (マスカー帯域幅) と手がかりを扱える帯域幅 (聴覚フィルタ間の帯域幅) を等価と考える。従って、ここでは、Hall らによるマスカー帯域幅の関数としてマスキングしきい値を測定した方法を、マスカー帯域幅をあらかじめ広めに固定 (1 kHz) にしておき、隣接する聴覚フィルタの参照数 L (利用する聴覚フィルタ間の全帯域幅に対応) の関数としてしきい値を測定することと見なす。また、しきい値は分離抽出された $\hat{f}_{1,A}(t)$ の SN 比 (分離精度) と見なし、マスキングからの解除量をちょうど SN 比の改善量に対応させる。このとき、入力位相 $\theta_{2k}(t)$ は、式 (7) と (11) から、隣接する聴覚フィルタの参照数 L の関数として求められた $\hat{\hat{B}}_k(t)$ によって一意に決定される。ここで、参照数は $L = 1, 3, 5, 7, 9, 11$ とし、これに対応する帯域幅はそれぞれ 207, 352, 499, 648, 801, 958 Hz である。

最後に、本論文では、二波形分離アルゴリズムで用いる各パラメータの値を $\Delta t = 3/f_0$、$\Delta A = |A_k(T_r - \Delta t) - A_k(T_r - 2\Delta t)|$, $\Delta B = 0.01 S_{\max}$, $\Delta\theta = \pi/20$ とした。但し、$S_{\max}$ は $S_k(t)$ の最大値である。

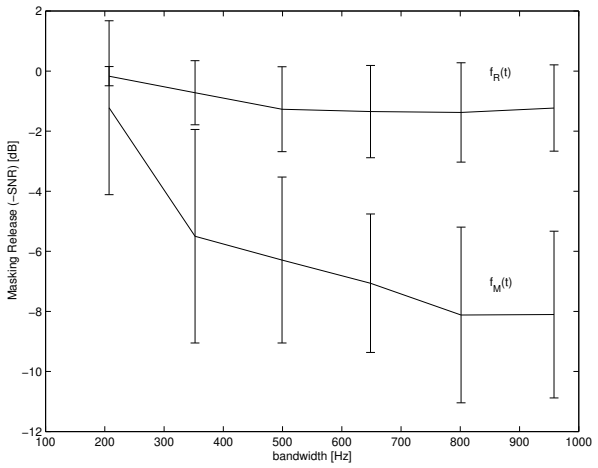


図 7: 隣接する聴覚フィルタの参照数 L に対応した帯域幅と $\hat{f}_{1,\mathrm{A}}(t)$ の分離精度 (SNR) の関係

5.2.2 結果と考察

シミュレーション条件に従い、各混合信号についてシミュレーションを行った。このときの結果を図 7に示す。この図の縦軸は分離抽出された純音の SN 比の向上量を下向きに表し、横軸は、隣接する聴覚フィルタの参照数 L に対応した帯域幅を表している。また、図中の実線と縦棒はそれぞれ 50 個の混合信号に対して分離抽出された正弦波信号 $\hat{f}_{1,A}(t)$ の SN 比の平均値と標準偏差を表している。この結果、混合信号 $f_M(t)$ 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、分離抽出された純音 $\hat{f}_{1,A}(t)$ の SN 比が向上する傾向が見られた。しかし、混合信号 $f_R(t)$ の場合、隣接する聴覚フィルタの参照数 L を増加させても、純音はほとんど抽出されず、SN 比はほとんど変わらなかった。従って、この結果は、マスカーの振幅成分が異なる周波数領域において同じ振幅変調パターンをもつとき、すなわち、マスカーの振幅包絡間の相関が高いとき、純音 $f_1(t)$ をより分離抽出しやすくなるという結果を示している。故に、この結果から、モデル A は、複数の聴覚フィルタ出力を利用し、マスカーの振幅包絡間の相関を手がかりにマスキング解除のメカニズムを模擬しているといえる。

5.3 モデル B のシミュレーション

5.3.1 刺激とシミュレーション条件

実験データは、モデル A で利用したものと同様開始時刻を変化させた 10 個の純音 $f_1(t)$ を利用するが、二種類のマスカーについては、乱数の種

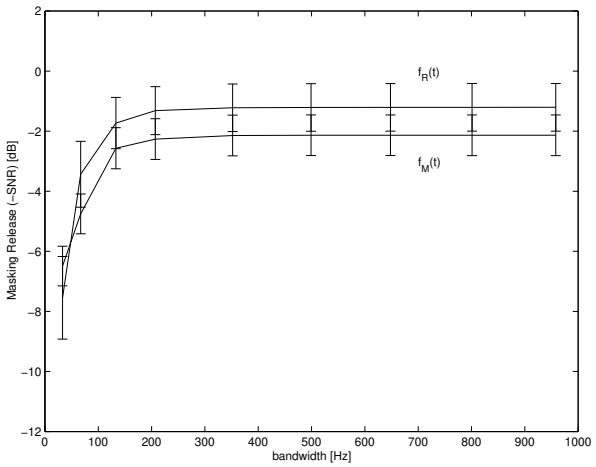


図 8: マスカー帯域幅と $\hat{f}_{1,\mathrm{B}}(t)$ の分離精度 (SNR) の関係

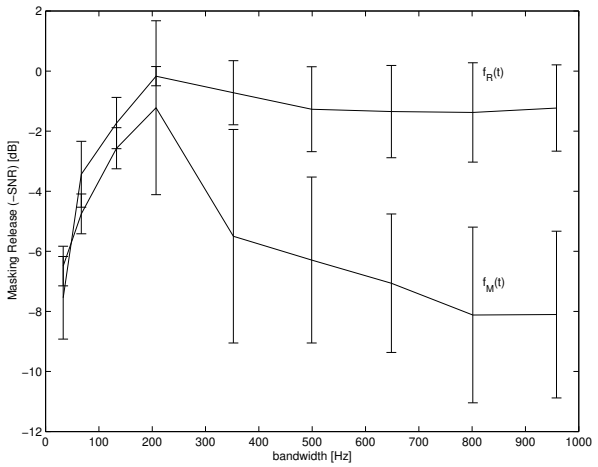


図 9: マスカー帯域幅と純音 $\hat{f}_1(t)$ の分離精度 (SNR) の関係

(5 種類) と帯域幅 (9 種類) を変化させて作成した 45 個とし、合計 450 個の混合信号を用意した。ここでマスカー帯域幅は、1/4 ERB, 1/2 ERB, 1 ERB に対応した 33, 67, 133 Hz の他、モデル A での $L = 1, 3, \cdots, 11$ に対応した 207, 352, 499, 648, 801, 958 Hz である。

モデル B では、Hall らの条件と同様、マスカー帯域幅の関数としてマスキングしきい値を測定する。また、モデル A の条件と同様に $\hat{f}_{1,B}(t)$ の SN 比をしきい値と見なす。

5.3.2 結果と考察

シミュレーション条件に従い、各刺激についてシミュレーションを行った。このときの結果を図 8に示す。この図の縦軸は分離抽出された純音の SN 比の向上量を下向きに表したものであり、横軸はマスカー帯域幅を表している。また、図中の実線と縦棒は、それぞれ SN 比の平均と標準偏差を表している。この結果、マスカーの種類に関係なく、マスカー帯域幅の増加とともにマスキングしきい値が変化していることがわかる。マスカー帯域幅が単一の聴覚フィルタの帯域幅に相当する 1 ERB を越えない場合、マスカー帯域幅の関数としてしきい値は増加しているが、マスカー帯域幅が 1 ERB を越える場合、しきい値はそれ以上増加せず一定になっている。

5.4 CMR の計算モデルの特性

二つのモデルについて、シミュレーションを行った結果、モデル A ではマスカー帯域幅が 1 ERB を越えたとき、マスカーの振幅包絡間の変動の一致 / 不一致による共変調マスキング解除 / マスキングの現象を模擬していることがわかる。また、モデル B ではマスカー帯域幅が 1 ERB を越えるまでマスキングしきい値が増加し、1 ERB を越えた後しきい値がそれ以上増加せず一定になるというマスキング現象を模擬していることがわかる。選択処理では、これら二つのモデルの処理結果のうち、マスキングしきい値の低いもの、言い換えると分離抽出された正弦波信号の SN 比の高いものを選択することで、図 7と図 8の結果から、図 9に示すような結果を得る。この特性は、図 5に示した Hall らの実験結果と類似した結果を示す。従って、本モデルは共変調マスキング解除の計算モデルと解釈できる。特に、共変調マスキング解除量は、Hall らの結果では最大約 10 dB であったのに対し、本モデルでは最大約 8 dB であった。

6 まとめ

本論文では、共変調マスキング解除の計算モデルを提案した。このモデルは、二波形分離モデル (モデル A) とマスキングのパワースペクトルモデル (モデル B) の二つのモデルと、この二つのモデルの結果を選択する処理で構成される。マスカーから純音を検出するメカニズムについて、モデル A では、複数の聴覚フィルタの出力を手がかりにしているのに対し、モデル B では、単一の聴覚フィルタからの出力を手がかりにしている。Hall らによる共変調マスキング解除の実験を想定したシミュレーションを二つのモデルについてそれぞれ行った。モデル A では、マスカーの種類によってマスキングしきい値に変動があった。これは、

ランダム帯域雑音の場合、マスカー帯域幅の増加に関係なくしきい値は変動しなかったものの、AM ランダム帯域雑音の場合、マスカー帯域幅の増加とともにマスキング解除が起こるという結果が得られた。モデル B ではマスカーの種類に関係なく、マスカー帯域幅の増加とともにマスキングしきい値が増加した。このしきい値は、マスカー帯域幅が 1 ERB を越えるまで増加したが、1 ERB を越えてからはそれ以上増加せず一定であった。この結果に対し、選択処理は二つのモデルの結果から分離抽出した純音のマスキングしきい値の低いものを選択することで、Hall らが示した CMR の結果と同様の傾向を示す特性が得られた。このとき共変調マスキング解除量は最大約 8 dB であった。

以上の結果から、本モデルが CMR の計算モデルと解釈できる。また、CMR の手がかりとして、規則 (iv) が有効であることも確認できる。

謝辞

本研究の一部は、科学技術振興事業団 (CREST) 及び日本学術振興会特別研究員研究奨励金の援助を受けて行なわれた。

参考文献

[1] Patterson, R. D. and Moore, B. C. J. Auditory filters and excitation patterns as representations of frequency resolution. In Frequency Selectivity in Hearing (ed. B. C. J. Moore), Academic Press, London and New York, 1986.

[2] Hall, J. W., Haggard, M. P. and Fernandes, M. A. "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am., 76, 50–56, 1984.

[3] Hall, J. W. and Grose, J. H. "Comodulation masking release: Evidence for multiple cues," J. Acoust. Soc. Am. 84, pp. 1669–1675, 1988.

[4] Brain C.J. Moore. An Introduction to the Psychology of Hearing, 4th ed., Academic Press, San Diego, 1997.

[5] Brain C.J. Moore. "Comodulation Masking release and Modulation Discrimination Interface," in The Auditory Processing of Speech, from Sound to Words, (Edited by M. E. H. Schouten), pp. 167–183, Mouton de Gruyter, NewYork, 1992.

[6] Willen A. C. van den Brink, Tammo Houtgast, and Guido F. Smoorenburg. (1992). "Effectiveness of Comodulation Masking Release," in The Auditory Processing of Speech, from Sound to Words, (Eds. M. E. H. Schouten), pp. 167–183, Mouton de Gruyter, NewYork.

[7] A. S. Bregman. Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, Cambridge, Mass, 1990.

聴覚フィルタ群により $f(t)$ を K 個の周波数成分に分解する (alignment 処理により群遅延を補正) ;
for $k := 1$ **to** K **do**
  入力位相を $\theta_{1k}(t) = 0$, $\theta_k(t) = \theta_{2k}(t)$ とする;
  $S_k(t)$ と $\phi_k(t)$ を求める;
  $dS_k(t)/dt$ と $d\phi_k(t)/dt$ から立上がり $T_{k,\mathrm{on}}$ と立下がり $T_{k,\mathrm{off}}$ を求める;
  分離区間を $T_{k,\mathrm{on}} \le t \le T_{k,\mathrm{off}}$ とする;
  分離区間を I 個の微小区間 $\Delta t = M/f_0$ に分割する;
  for $i := 1$ **to** I **do**
    $C_{k,0}$ の制限範囲 $C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}$ を求める;
    for $C_{k,0} := C_{k,\alpha}$ **to** $C_{k,\beta}$ **do**
      $C_{k,0}$ に対する入力位相 $\hat{\theta}_k(t)$ を求める;
      振幅包絡 $\hat{A}_k(t)$ と $\hat{B}_k(t)$ を求める;
      隣接する聴覚フィルタ出力 (図 3)($\ell = 1, 2, \cdots, L$) において以下の処理をする;
      (a) 図 3の振幅特性から $\hat{A}_{k\pm\ell}(t)$ を求める;
      (b) $S_{k\pm\ell}(t)$ と $\phi_{k\pm\ell}(t)$ を求める;
      (c) 式 (5) から、$\hat{A}_{k\pm\ell}(t)$, $S_{k\pm\ell}(t)$, $\phi_{k\pm\ell}(t)$ を用いて $\hat{\theta}_{k\pm\ell}(t)$ を求める;
      (d) 式 (6) から $\hat{B}_{k\pm\ell}(t)$ を求める;
      (e) 式 (11) から $\hat{\hat{B}}_k(t)$ を求める;
      (f) マスカーの振幅包絡間の相関値を求める
    end
    式 (10) から、$C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}$ において相関が最大になる未定係数 $C_{k,0}$ を求める;
    式 (7) から、$\theta_k(t)$ を求める;
    式 (5) から $A_k(t)$ を求める;
    式 (6) から $B_k(t)$ を求める;
  end
  式 (2) と (3) から各周波数成分を求める;
end
$\hat{f}_{1,\mathrm{A}}(t)$ と $\hat{f}_{2,\mathrm{A}}(t)$ を再構成する;

図 10: 二波形分離アルゴリズム

[8] 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, New York, 1993.

[9] 鵜木, 赤木. "雑音が付加された波形からの信号波形の一抽出法," IEICE, Vol. J80-A, No. 3, pp. 444–453, March, 1997.

[10] 鵜木, 赤木. "共変調マスキング解除の計算モデルの高性能化," 音響学会春季講論, 3-8-2, March, 1997.

付録

モデル A の処理過程 (アルゴリズム) を図 10に示す。聴覚フィルタ群の処理において、式 (1) のインパルス応答から推測できるように、各スケールに対して群遅延が生じる。この遅延は、実際の基底膜応答としてよく模擬されているが、物理的制約条件 3 において、振幅包絡間の相関を取る際に問題となる。そのため、ここでは alignment 処理として、式 (1) の振幅包絡のピークを各スケール毎にそろえることで群遅延を補正している。