# 共変調マスキング解除の計算モデルに関する一考察

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

〒923-12　石川県能美郡辰口町旭台 1-1

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

本論文では、音源分離のモデル化の試みの１つとして共変調マスキング解除 (CMR) を想定した２波形分離問題を取り上げ、CMR の工学的モデル化を行なう。まず、２波形分離問題の定式化を行ない、聴覚フィルタ群として Gammatone filter を基底関数とした wavelet 分析合成系を設計する。次に、２波形分離に必要な３つの物理量（振幅包絡、出力位相、入力位相）の計算方法を明らかにする。ここで、入力位相は Bregman が提唱した (1) １つの音響事象に生じる変化と (2) 漸近的変化に関する発見的規則を物理的制約条件として用いることで求められる。CMR のシミュレーションとして、Hall らによる CMR の実験を想定した２波形分離問題を解いたところ、本モデルが CMR の工学的モデルと解釈できる結果が得らた。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。

共変調マスキング解除、聴覚の情景解析、Gammatone filter、wavelet 分析合成系

# A Study on Computational model of Co-modulation Masking Release

Masashi Unoki　and　Masato Akagi

Japan Advanced Institute of Science and Technology, Hokuriku

1-1 Asahidai, Tatsunokuchi, Nomigun, Ishikawa 923-12 Japan

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

This paper presents a computational model of co-modulation masking release(CMR) as a model of acoustic source segregation. Firstly, problems of two acoustic sources segregation are formulated, and a wavelet filterbank, whose analysing wavelet is a gammatone filter, is designed as an auditory filterbank. Next, a derivation method of three physical clues(amplitude envelope, phase, and phase difference between two acoustic sources of the auditory filter outputs) is introduced. In particular, input phase is derived from two of the four constraints proposed by Bregman. The experiment simulating CMR in this paper deals with the same experimental stimuli as used by Hall *et al.*. The results indicate that the proposed model is one of the computational models representing the CMR. Maximum amount of co-modulatin masking release is about 8 dB.

co-modulation masking release(CMR), auditory scene analysis, gammatone filter, wavelet filterbank

# 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る。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 (co-modulation)” 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 [1]。この現象については、多くの聴覚心理実験が行なわれており、同様の結果が得られている。しかし、CMR が起こるという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の報告はない。

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) [2, 3] に基づく音源分離モデルの研究を行なっており、狭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 [4, 5]。これは、同一周波数領域においても信号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 4 つの発見的規則 ：

(1) 共通の立ち上がり/立ち下がりに関する規則

(2) 漸近的変化に関する規則

(3) 調波関係に関する規則

(4) 1 つの音響事象に生じる変化に関する規則

のうち規則 (2) と規則 (4) を物理的制約条件に捉え直すことで、 2 波形分離問題を解いている [4, 5]。また、この方法を用いた分離例として、CMR と類似した結果も得られている [5]。

そこで、本論文ではこの類似した結果を積極的に検証するために、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本モデルが CMR の工学的モデルになることを示す。

# 2　2 波形分離問題の定式化

本論文では、“ある 2 つの独立な音源で生じた音響信号が加算された状況から、それぞれの音響信号に分離・再合成すること” を 2 波形分離問題と定義する。この 2 波形分離問題は以下のように定式化される。

はじめに、ある 2 つの音響信号 $f_1(t)$ と $f_2(t)$ が

$$f(t) = f_1(t) + f_2(t) \tag{1}$$

に加算され、信号 $f(t)$ のみが受聴できるものとする。これは、図 1に示す $K$個の分析 (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

$$f_1(t) \quad : \ A_k(t)\sin(\omega_k t) \tag{2}$$

$$f_2(t) \quad : \ B_k(t)\sin(\omega_k t + \theta_k(t)) \tag{3}$$

に周波数分解される。但し、$\omega_k$は分析フィルタの中心角周波数、$\theta_k(t)$ は $f_2(t)$ のもつ $f_1(t)$ に対応した入力位相である。このとき、$k$番目の分析フィルタの出力 $X_k(t)$ は、

$$\begin{aligned} X_k(t) &= A_k(t)\sin\omega_k t + B_k(t)\sin(\omega_k t + \theta_k(t)) \\ &= S_k(t)\sin(\omega_k t + \phi_k(t)) \end{aligned} \tag{4}$$

と表される。ここで、振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ

$$S_k(t) = \sqrt{A_k^2(t) + 2A_k(t)B_k(t)\cos\theta_k(t) + B_k^2(t)} \tag{5}$$

$$\phi_k(t) = \arctan\left(\frac{B_k(t)\sin\theta_k(t)}{A_k(t) + B_k(t)\cos\theta_k(t)}\right) \tag{6}$$

である。このとき、角周波数 $\omega_k$は任意に設定可能であり、かつ振幅包絡 $S_k(t)$ は観測可能であることから、出力位相 $\phi_k(t)$ と入力位相 $\theta_k(t)$ がわかれば 2 波形の振幅包絡 $A_k(t)$ と $B_k(t)$ を

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

$$B_k(t) = \frac{S_k(t)\sin\phi_k(t)}{\sin\theta_k(t)} \tag{8}$$

のように解析的に解くことができる。最後に、すべての分析フィルタについて、振幅包絡 $A_k(t)$ と $B_k(t)$ に分離し、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 $f_2(t)$ を再構成できる。但し、再構成された信号をそれぞれ$\hat{f}_1(t)$、$\hat{f}_2(t)$ とする。

本論文では、図 1の分析合成系を wavelet 変換によって実現する。また、CMR のシミュレーション

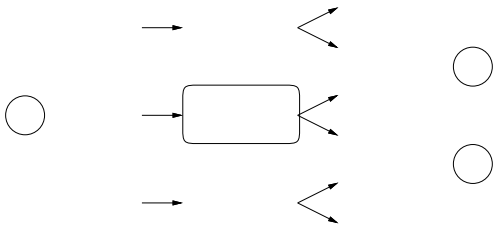

図 1: wavelet 分析合成系

を試みるため、$f_1(t)$ を純音、$f_2(t)$ をその純音の周波数を中心周波数とする狭帯域雑音とし、$f_2(t)$ が存在している状態で $f_1(t)$ が加算されるといった状況を想定することで 2 波形分離問題を議論する。

## 3 Gammatone filter を基底関数とした wavelet 分析合成系

本研究では、聴覚系を模擬した聴覚フィルタ群を構築するために、基底膜の特性をより良く模擬した聴覚フィルタとして知られている Gammatone filter を基底関数とすることで、分析合成系を実現している [4, 5]。これは、基本 wavelet を

$$\psi(t) = At^{n-1}e^{j2\pi f_0 t - 2\pi b_f t} \tag{9}$$

と定義し、wavelet 変換対 ：

$$\begin{aligned} \tilde{f}(a,b) &= \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \\ &= |\tilde{f}(a,b)|e^{j\arg(\tilde{f}(a,b))} \end{aligned} \tag{10}$$

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(a,b)\psi\left(\frac{t-b}{a}\right)\frac{dadb}{a^2} \tag{11}$$

を用いて分析合成系を構築している。ここで、$\hat{\psi}(\omega)$ を $\psi(t)$ の Fourier 変換としたとき、$C_\psi$ は許容条件 ：

$$C_\psi := \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega < \infty \tag{12}$$

を満たす正数であり、$\psi$ がこの条件を満たす限り再構成式 (11) は必ず存在する。また、Gammatone filter はこの許容条件を近似的に満たすため、基本 wavelet として十分利用できる [4]。

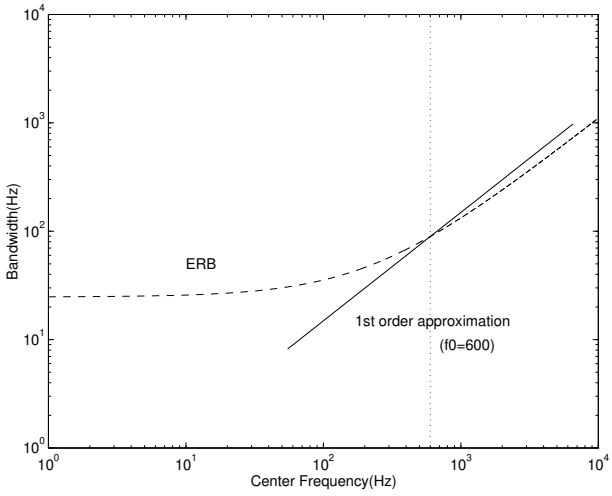


図 2: 中心周波数と ERB の関係 [6]

本論文では、中心周波数 $f_0 = 600$[Hz]、通過帯域が 60 ~ 6000[Hz]、フィルタ数 (チャネル数) が $K = 128$ の wavelet 分析合成系を設計した。便宜上、ここでは構築方法を連続 wavelet 変換で表現しているが、計算機への実装には、サンプリング周波数 $f_s = 20$[kHz]、スケールパラメータ $a = \alpha^i, -\frac{K}{2} \le i \le \frac{K}{2}, \alpha = 10^{2/K}$、シフトパラメータ $b = j/f_s$ として離散 wavelet 変換 $(i, j \in \mathbf{Z})$ を用いている [4]。また、図 2の実線のように中心周波数 600[Hz] のとき、Gammatone filter の等価矩形帯域幅 (Equivalent Rectangular Bandwidth) をちょうど幅 1 として定 Q フィルタの分析合成系を設計する。このときの wavelet 分析合成系の周波数特性を図 3に示す。この図から、分析合成系の周波数特性は解析周波数帯域を完全に被覆していることがわかる。

## 4 物理パラメータの導出方法

### 4.1 振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ の計算方法

式 (5) の振幅包絡 $S_k(t)$ と式 (6) の出力位相 $\phi_k(t)$ は次の補題 1 で得られる。

**補題 1** 振幅包絡 $S_k(t)$ は、振幅項 $|\tilde{f}(a,b)|$ から

$$S_k(t) = |\tilde{f}(\alpha^{k-\frac{K}{2}}, t)| \tag{13}$$

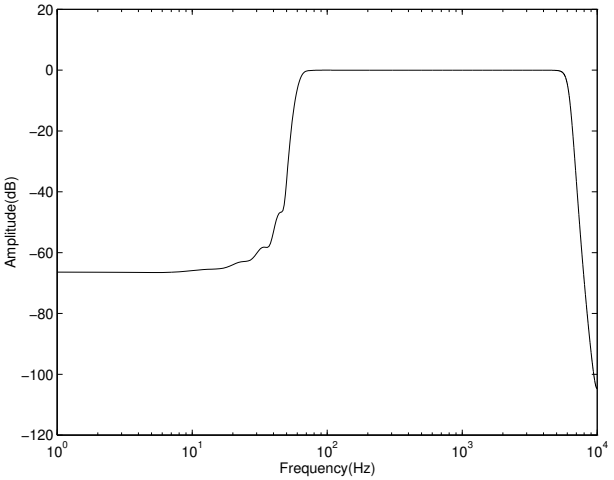


図 3: wavelet 分析合成系の周波数特性 $(K = 128)$

で求めることができる。また、出力位相 $\phi_k(t)$ は、位相項 $\arg(\tilde{f}(a,b))$ から

$$\phi_k(t) = \int \left( \frac{d}{dt} \arg \left( \tilde{f}(\alpha^{k-\frac{K}{2}}, t) \right) - \omega_k \right) dt \quad (14)$$

で求めることができる [4]。 ■

## 4.2 入力位相 $\theta_k(t)$ の計算方法

入力位相 $\theta_k(t)$ は、3 つの物理的制約条件 :(i) 漸近的変化に関する規則、(ii) 連続性 (時間的近接) の規則、(iii) 1 つの音響事象に生じる変化に関する規則を用いることで得られる。特に、規則 (i) と (iii) は Bregman によって提唱された発見的規則 (2) と (4) である。

はじめに、規則 (i) を適用する。これは、“1 つの音の振幅包絡はゆっくりと滑らかに変化する” という規則 [3] であり、これを使用するため次のような物理的制約条件に捉え直す。

制約条件 **1** (漸近的変化) 振幅包絡 $A_k(t)$ の時間変化は $L$ 回微分可能な $L$ 次多項式で近似される :

$$\frac{dA_k(t)}{dt} = C_{k,L}(t) \quad (15)$$

■

この制約を式 (7) に適用することで、1 階線形微分方程式 :

$$y'(t) + \frac{P'(t)}{P(t)} y(t) = \frac{Q'(t) - C_{k,L}(t)}{P(t)} \quad (16)$$

を得る。但し、$P(t) = S_k(t) \sin \phi_k(t)$、$Q(t) = S_k(t) \cos \phi_k(t)$、$y(t) = \cot \theta_k(t)$ である。入力位相 $\theta_k(t)$ は、この微分方程式 (16) を解くことで得られる。

補題 **2** 入力位相 $\theta_k(t)$ は、微分方程式 (16) の一般解

$$\theta_k(t) = \arctan \left( \frac{S_k(t) \sin \phi_k(t)}{S_k(t) \cos \phi_k(t) + C_k(t)} \right) \quad (17)$$

として得られる。但し、$C_k(t)$ は各項に未定係数をもつような未定関数であり、$C_k(t) = -\int C_{k,L}(t)dt + C_{k,0}$ と表される。 ■

ここで未定関数 $C_k(t)$ を決定できれば、一意に $\theta_k(t)$ を求めることができる。最適化問題として $C_k(t)$ の各係数 $C_{k,\ell}, \ell = 0, 1, \cdots, L$ を推定することは可能であるが、計算量を抑えるために未定関数を定数項 $(C_{k,L}(t) = C_{k,0})$ として取り扱うことにする。このとき、制約条件 (15) は $dA_k(t)/dt = 0$ となり、微小区間 $\Delta t$ で振幅包絡は変化しないことを意味する。

次に、この微小区間 $\Delta t$ における分離を行なうために規則 (ii) を適用する。これは、“分離を行なう前後の境界において、各物理パラメータが連続性を保持しなければならない” という規則であり、これを使用するため次のような物理的制約条件に捉え直す。

制約条件 **2** (連続性) 各物理パラメータ $Q_k(t)(= A_k(t), B_k(t), \theta_k(t))$ は、分離を行なう前後の境界 $(t = T_r)$ においてある幅 $\Delta Q(= \Delta A, \Delta B, \Delta\theta)$ 以内で接合されていなければならない :

$$|Q_k(T_r + 0) - Q_k(T_r - 0)| \leq \Delta Q \quad (18)$$

■

ここで、式 (7)、式 (8)、式 (17) から振幅包絡 $A_k(t)$ と $B_k(t)$ および入力位相 $\theta_k(t)$ が未定係数 $C_{k,0}$の関数となっていることに着目すれば、この制約条件は、ある境界 $T_r$における連続性を保持した形で未定係数 $C_{k,0}$の取り得る範囲を

$$C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta} \quad (19)$$

に限定することと解釈できる。但し、$C_{k,\alpha}$と $C_{k,\beta}$ は、この境界における未定係数$C_{k,0}$の上限と下限である。

最後に規則 (iii) を適用する。これは、“１つの音響事象に生じる多くの変化は、その音を構成する各成分に同じような影響を与える” という規則 [3] であり、これを使用するため次のような物理的制約条件に捉え直す。

制約条件 **3** (１つの音響事象に生じる変化) 振 幅包絡 $B_k(t)$ は $2L$ 個の隣接する聴覚フィルタから得られる $B_{k\pm L}(t)$ に強い相関がなければならない：

$$B_k(t) \approx B_{k\pm\ell}(t), \qquad \ell = 1, 2, \cdots, L \tag{20}$$

■

ここで、式 (8) と式 (17) から振幅包絡 $B_k(t)$ は未定係数 $C_{k,0}$の関数であることに着目し、ある未定係数 $C_{k,0}$により決定された振幅包絡を$\hat{B}_k(t)$ とおく。また、振幅包絡間の相関を

$$\mathrm{Corr}(\hat{B}_k, \hat{\hat{B}}_k) = \frac{\langle \hat{B}_k, \hat{\hat{B}}_k \rangle}{||\hat{B}_k|| ||\hat{\hat{B}}_k||} \tag{21}$$

と定義する。但し、$L$ 個の隣接する聴覚フィルタから得られる振幅包絡$\hat{\hat{B}}_k(t)$ は

$$\hat{\hat{B}}_k(t) = \frac{1}{2L} \sum_{\ell=-L, \ell\neq 0}^{L} \hat{B}_{k+\ell}(t) \tag{22}$$

であり、相関を求める時間領域 (積分範囲) はある過去の時刻 $T_B$ から分離を行なう微小区間までとする。また、振幅包絡$\hat{B}_{k\pm\ell}(t)$ はある未定係数 $C_{k,0}$によって得られた$\hat{A}_k(t)$ と$\hat{B}_k(t)$ を用いて図 4の周波数特性から求めることができる。このとき、制約条件３は、帯域雑音の振幅包絡間の相関を最大とする未定係数 $C_{k,0}$を最適値として選ぶことと解釈できる。これは、

$$\max_{C_{k,\alpha} \le C_{k,0} \le C_{k,\beta}} \mathrm{Corr}(\hat{B}_k, \hat{\hat{B}}_k) \tag{23}$$

なる未定係数 $C_{k,0}$を見つけ出すことで実現できる。そうすれば、式 (17) によって一意に入力位相 $\theta_k(t)$ を決定できる。

### 4.3 2 波形分離アルゴリズム

前節で得られた補題から次の定理が得られる。

定理 **1** (２波形分離アルゴリズム) ２波形分離問題において、物理的制約条件１～３を適用すれば、式 (17) と式 (23) によって入力位相 $\theta_k(t)$ を求めることができる。従って、振幅包絡 $S_k(t)$、出力位相 $\phi_k(t)$ とあわせて、式 (1) ~ (8) から２波形分離問題を解くことができる。

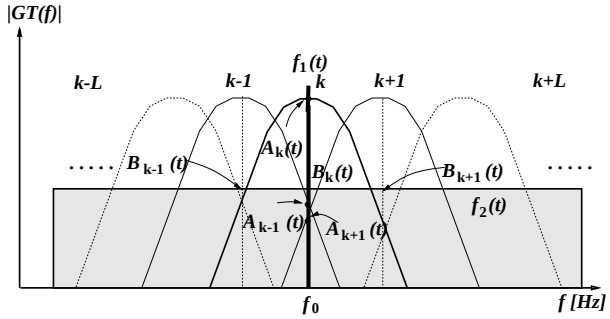


図 4: 隣接する聴覚フィルタにおける純音 $f_1(t)$ と狭帯域雑音 $f_2(t)$ の通過特性。

(証明)　補題１と２より明らか。　■

定理１を用いて２波形分離問題を解く場合、２つの音響信号が同時に存在する区間を知る必要がある。第 2 節で述べたように、本論文では $f_2(t)$ が存在している状態で $f_1(t)$ が加算されるような状況を想定しているのだから、この区間は $f_1(t)$ の立ち上がりと立ち下がりにより知ることができる。そのため、ここでは振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ の時間変化に着目することで次のように $f_1(t)$ の立ち上がりと立ち下がりを求める。

1. 立ち上がり時刻 $T_{\mathrm{on}}$ :$dS_k(t)/dt$ の極大点近傍にある $|d\phi_k(t)/dt|$ の局所的な極大点
2. 立ち下がり時刻 $T_{\mathrm{off}}$ :$dS_k(t)/dt$ の極小点近傍にある $|d\phi_k(t)/dt|$ の局所的な極大点

ここで、$t_{\mathrm{on}} \le t \le t_{\mathrm{off}}$が分離すべき区間となる。

次に、この定理１に基づく２波形分離アルゴリズムを図 5に示す。これは、混合信号 $f(t)$ を wavelet 分析系により周波数分解した後に、すべての分析フィルタ $X_k(t), k = 1, 2, \cdots, K$ に対して適用される。最後に、分離されたすべての振幅包絡 $A_k(t)$ と $B_k(t)$ に対し、式 (2) と式 (3) から各信号の周波数成分を合成し、wavelet 合成系により$\hat{f}_1(t)$ と$\hat{f}_2(t)$ を再合成する。

## 5 共変調マスキング解除を想定した２波形分離シミュレーション

### 5.1 共変調マスキング解除 (CMR)

Hall らは共変調マスキング解除の実験のうちの１つで、1[kHz]、400[ms] の正弦波信号の閾値をス

**for** $k := 1$ **to** $K$ **do**
　補題 1 より，振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ を求める;
　$dS_k(t)/dt$ と $d\phi_k(t)/dt$ から分離抽出する信号音の立ち上がり時刻 $T_{\rm on}$ と立ち下がり時刻 $T_{\rm off}$ を求める;
　分離区間を $T_{\rm on} \leq t \leq T_{\rm off}$ とする;
　分離区間を $N$ 個の微小区間 $\Delta t = M/f_0$ に分割する;
　**for** $n := 1$ **to** $N$ **do**
　　未定係数の制限範囲 $C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta}$ を求める;
　　**for** $C_{k,0} := C_{k,\alpha}$ **to** $C_{k,\beta}$ **do**
　　補題 2 より，$C_{k,0}$ に対する入力位相 $\hat{\theta}_k(t)$ を求め、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 (図 4) において，次の処理をする $(\ell = 1, 2, \cdots, L)$;
　　(1) 周波数特性から振幅包絡 $\hat{A}_{k\pm\ell}(t)$ を求める;
　　(2) 補題 1 より，$S_{k\pm\ell}(t)$ と $\phi_{k\pm\ell}(t)$ を求める;
　　(3) 式 (7) から，$\hat{A}_{k\pm\ell}(t)$，$S_{k\pm\ell}(t)$，$\phi_{k\pm\ell}(t)$ を用いて入力位相 $\hat{\theta}_{k\pm\ell}(t)$ を求める;
　　(4) 式 (8) から，振幅包絡 $\hat{B}_{k\pm\ell}(t)$ を求める;
　　(5) 式 (22) から、$\hat{\hat{B}}_k(t)$ を求める;
　　(6) 式 (21) から，雑音の振幅包絡間の相関値 $\mathrm{Corr}(\hat{B}_k(t), \hat{\hat{B}}_k(t))$ を求める;
　　**end**
　　式 (23) から，$C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta}$ において相関が最大になる未定係数 $C_{k,0}$ を求める;
　　式 (17) から入力位相 $\theta_k(t)$ を求める;
　　式 (7) と式 (8) から振幅包絡 $A_k(t)$ と $B_k(t)$ を求める;
　**end**
**end**

図 5: 2 波形分離アルゴリズム

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 [1]。また、マスカーの中心周波数は 1[kHz] であり、次のような 2 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で独立に変動する。
- 振幅変調された帯域雑音 :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) したものであり、振幅変動は異なる周波数領域においても等しい。

この 2 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 6 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閾値を表し、点 M は振幅変調された雑音の場合の閾値を表している。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

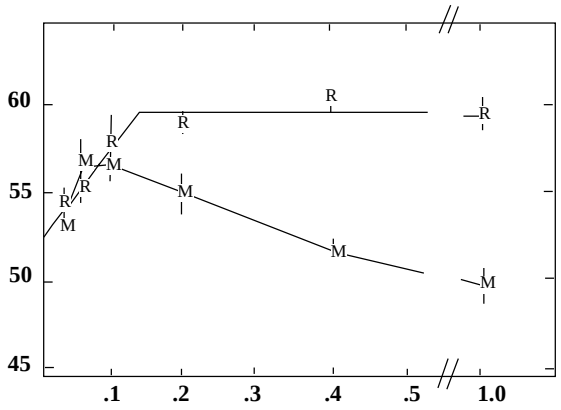


図 6: 共変調マスキング解除 (Hall ら,1984)。文献 [1] より転載。

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は変動しないのに対し、振幅変調された帯域雑音の場合、雑音の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、被験者は信号検知能力を高めることができることを示し、この現象を共変調マスキング解除 (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 [1]。

## 5.2 実験データ

ここでは、上記で述べた CMR を想定して 2 波形分離のシミュレーションを行なう。分析合成系のフィルタ数や解析周波数帯域の関係上、Hall らと全く同じ実験データを作成できないが、中心周波数、変調速度を除くすべてのパラメータをできる限り一致させた。

実験データは、$f_1(t)$ として純音、$f_2(t)$ として振幅変調された狭帯域雑音 $f_{21}(t)$ とランダム狭帯域雑音 $f_{22}(t)$ の 2 種類の信号であり、それぞれ次のように与えられる。

$$f_1(t) = \begin{cases} 1200\sin(2\pi f_0 t), \\ \qquad 0.3 + T_n \leq t \leq 0.7 + T_n \\ 0, \quad \mathrm{otherwise} \end{cases} \quad (24)$$

$$f_{21}(t) = \sum_{f=f_0-500}^{f_0+500} B(t)\sin(2\pi f t + R(f)),$$

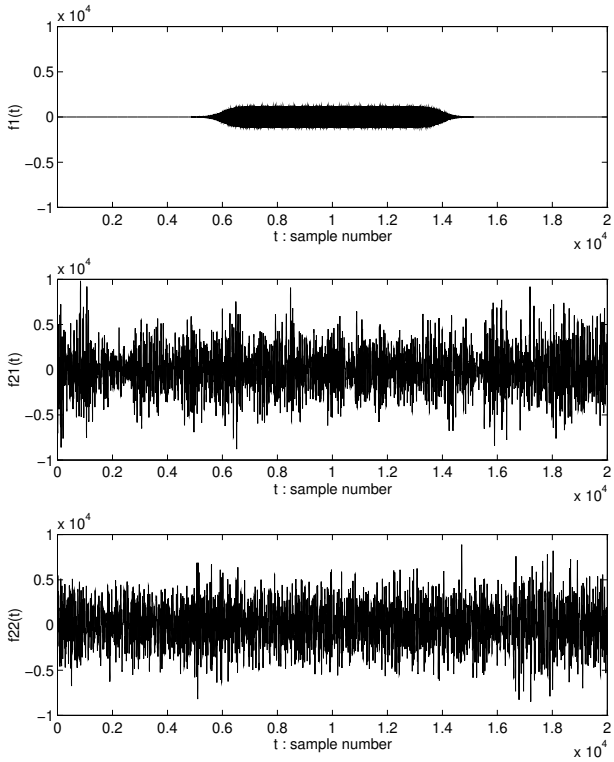


図 7: 音響信号 ：$f_1(t), n=0$、$f_{21}(t)$、$f_{22}(t)$

$$0 \le t \le 1.0 \tag{25}$$

$$f_{22}(t) \quad = \quad \sum_{f=f_0-500}^{f_0+500} C(f,t)\sin(2\pi f t + R(f)), \qquad 0 \le t \le 1.0 \tag{26}$$

但し、$f_0 = 600[\mathrm{Hz}]$、$T_n = 0.0125n, n = 0, 1, \cdots, 9$、$R(f)$ は $[-\pi, \pi]$ の一様乱数である。ここで、$B(t)$ と $C(f,t)$ は白色雑音に $30[\mathrm{Hz}]$ の低域通過フィルタをかけたものであり、過変調を起こさないようにあるバイアス値を足している。特に、$C(f,t)$ は周波数 $f$ 間で振幅が独立に変動するものとし、$\sqrt{f_{21}(t)^2/f_{22}(t)^2} = 1$ となるようにパワーを調整している。このとき、純音と帯域雑音の SN 比は $-8.51[\mathrm{dB}]$ である。図 7 に、これらの信号を wavelet 分析合成系で再構成したものを示す。

次に、$f_2(t) = f_{21}(t)$ とした混合信号 $f_M(t) = f_1(t) + f_{21}(t)$ と $f_2(t) = f_{22}(t)$ とした混合信号 $f_R(t) = f_1(t) + f_{22}(t)$ を用意する。これらは、それぞれ図 6 の点 M と点 R に対応した混合信号である。この 2 種類の混合信号を図 8 に示す。

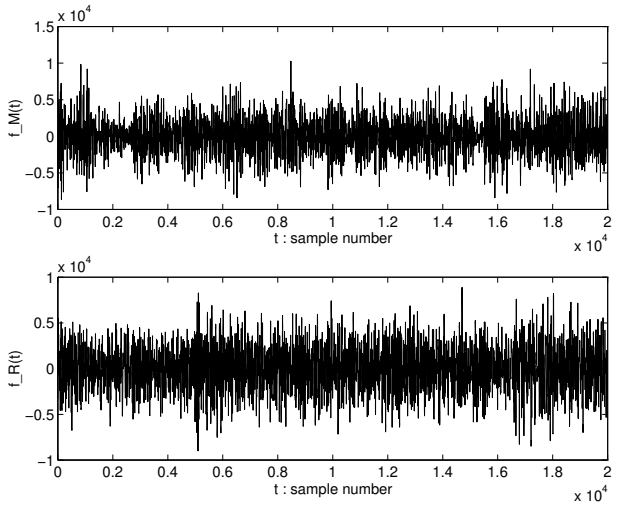


図 8: 混合信号 ：$f_M(t)$ と $f_R(t)$

### 5.3 シミュレーション条件

本シミュレーションでは、$f_1(t)$ の開始時刻を式 (24) の $n$ の範囲 ($n = 0, 1, \cdots, 9$) で変化させ、計 10 個の混合波形を用いる。また、微小区間を $\Delta t = 3/f_0$、$T_B$ を $100\Delta t$ だけ過去の時刻、$\Delta A$ を分離前の $A_k(t)$ の差分値、$\Delta B = 50.0$, $\Delta\theta = \pi/20$ として 2 波形分離のシミュレーションを行なう。

次に、Hall らの実験 [1] では、帯域雑音の帯域幅を可変にしていたが、ここでは隣接する聴覚フィルタ数 $L$ を可変とすることで手がかりとなる帯域幅を変化させる。そのため、式 (22) のように振幅包絡 $\hat{B}_k(t)$ を隣接する聴覚フィルタ数 $L$ の関数として求めることで、式 (17) と式 (23) から入力位相 $\theta_k(t)$ を一意に求める。ここで、隣接するフィルタ数は $L = 1, 2, 3, 4, 5, 6, 7, 8$ であり、これに対応する帯域幅は、それぞれ $134, 178, 220, 264, 308, 352, 297, 441[\mathrm{Hz}]$ となる。

### 5.4 分離結果と考察

上記の条件に従い CMR を想定した 2 波形分離の結果を図 9 に示す。ここでは、マスキング解除量の 1 つの評価として SN 比を取り入れている。そのため、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、この SN 比向上量をマスキング解除量と読み替えることにする。また、横軸は $L$ 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 $f_M(t)$ の場合、隣接する聴覚フィルタの

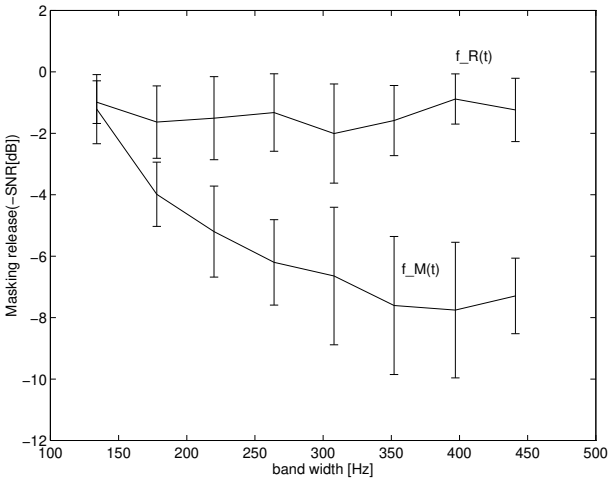


図 9: 帯域幅とマスキング解除の関係

参照数を増加させると、純音$\hat{f}_1(t)$ が検出しやすくなり、SN 比が向上する傾向がみられた。これに対し、混合信号 $f_R(t)$ の場合、参照数を増加させても $\hat{f}_1(t)$ はほとんど検出されず、SN 比はあまり変わらなかった。この実験結果は、振幅変調された狭帯域雑音の場合、帯域幅を広げることでマスキング解除が起こり、ランダム狭帯域雑音の場合、帯域幅を広げてもマスキング量が変らないという結果を模擬していることがわかる。従って、この結果から本モデルは CMR の工学的モデルと解釈できる。尚、この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 8[dB] であった。

本モデルでは、少なくとも 1 つの隣接する聴覚フィルタからの手がかりを利用して分離を行なうため、中心周波数が 600[Hz] のとき 134[Hz] が最小周波数の測定限界となる。そのため、Hall らが示したように臨界帯域幅以内において 2 つの混合信号のマスキング量が増加するという現象を確認できないが、それ以上においては同様の結果が得られている。また、本モデルが CMR の工学的モデルになった主な要因は、Bregman によって提唱された発見的規則 (4) を物理的制約条件に捉え直したことにある。この発見的規則は CMR の現象も十分に説明できるものとして知られていたが、工学的にも有効な規則であることが明らかになった。

## 6 おわりに

本研究では、Bregman によって提唱された 4 つの発見的規則を物理的制約条件に捉え直すことで、音源分離モデルの構築を試みている。本論文では、この試みの 1 つとして、共変調マスキング解除を想定した 2 波形分離問題を取り上げ、この工学的モデル化を行なった。まず、2 波形分離問題を定式化し、wavelet 分析合成系を人間の聴覚特性に合わせて設計した。その後で、著者らによって提案された 2 波形分離問題の解法を適用し、Hall らによって行なわれた CMR の実験を想定した 2 波形分離シミュレーションを行なった。この結果、振幅変調された狭帯域雑音の場合、帯域幅を広げることでマスキング解除が起こり、ランダム狭帯域雑音の場合、帯域幅を広げてもマスキング量が変わらないという特性が得られ、本モデルが CMR の工学的モデルになることを示した。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。また、この結果から、発見的規則 (4) が CMR を工学的に説明する有効な規則であることも確認できた。

今後の課題は、残りの発見的規則 (1) と (3) を同様に物理的制約条件に捉え直し、一般的な音源分離問題の解法を求めることである。

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: The Perceptual Organization of Sound,” MIT Press, 1990.

[3] 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.

[4] 鵜木 祐史, 赤木 正人, “雑音が付加された波形からの信号波形の抽出法,” 音響学聴覚研資, H-95-79, Nov. 1995.

[5] 鵜木 祐史, 赤木 正人, “帯域雑音に埋もれた信号波形の一抽出法,” 音響学春季講論, 3-3-15, 1996.

[6] 赤木 正人, “聴覚フィルタとそのモデル,” 信学誌 vol. 77 no.9 pp.948–956 Sep. 1994.