雑音が付加された波形からの信号波形の抽出法

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 15

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

あらまし　　本論文では、音源分離のモデル化の試みとして 2 波形分離問題を取り上げ、雑音が付加された波形から原信号と雑音を分離、抽出する方法を提案する。この方法は、聴覚の情景解析に基づくものであり、wavelet 分析合成系の各フィルタ出力から得られる振幅包絡と出力位相、入力信号間の位相を用いることで 2 波形分離を可能にする。この 3 つの物理的手がかりは、Bregman が示した共変調マスキング解除と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導出される。この方法を用いた分離例として、帯域雑音中に純音が混入した 2 波形分離問題の解法を示す。特に、振幅変調された雑音が混合された場合、分離が容易になり、ランダム帯域雑音の場合、分離が困難になるという、共変調マスキング解除の工学的な説明として解釈できる結果が得られた。

キーワード　　共変調マスキング解除 (CMR)、2 波形分離、Gammatone filter、wavelet 分析合成系、聴覚の情景解析

An Extraction Method of the Signal from Noise-Added Signals

Masashi Unoki　and　Masato Akagi

Japan Advanced Institute of Science and Technology, Hokuriku

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

Abstract　　This paper presents an extraction method of the signal from noise-added signals as a modeling of acoustic source separation. The extraction method is based on auditory scene analysis and can solve a problem of two acoustic source segregation by using three physical clues : phase difference between two acoustic sources $\theta_i(t)$ and amplitude and phase of auditory filter outputs, $S_i(t)$ and $\phi_i(t)$. The physical clues are derived from two of four constraints proposed by Bregman. The simulation in the paper deals with a band-passed noise signal. The results show that the proposed method can extract the pure tone if it is mixed with an amplitude modulated band-passed noise and can not extract the pure tone mixed with a random band-passed noise. This result indicates that the proposed method is one of the model to be able to explain the CMR computationally.

Key words　　co-modulation masking release(CMR), two acoustic source segregation, Gammatone filter, wavelet filterbanks, auditory scene analysis

1 はじめに

近年、聴覚は能動的な環境把握システムの一環として考えられるようになり、聴覚の情景解析 (Auditory Scene Analysis)[1,2] に基づく音源分離の研究が盛んに行なわれるようになった。音源分離問題を解くことができれば、雑音中の特定の音を聴きとるという点でカクテルパーティー効果のモデル化も可能になり、機械による音声認識などにも応用できる。また、聴覚の計算理論の構築[3] に向けて新たな視点を提供してくれる。

計算機モデルの研究として、Sheffield 大学の Brown と Cooke ら[4,5] による音響イベントに着目した分凝モデル、MIT の Ellis[6] による心理音響学的グルーピング規則を取り入れた分凝モデル、NTT の中谷ら[7] による聴覚の情景解析をマルチエージェントシステムによって実現した分凝の実装例がある。また、東大の柏野ら[8] は、２つの周波数成分の分離知覚に関して、スペクトログラム上の複数の特徴と分離知覚の生じる割合との定量的関係をモデル化している。しかし、いずれの計算機モデルもスペクトログラム上での分離を考えているため、２つの信号が同じ周波数領域の成分を含むような場合、完全に分離できているとは言い難い。

我々は、同一周波数領域において完全に分離するためには、スペクトルの他に位相も考慮しなければならないという立場に立ち、基本的な音源分離問題である２波形分離問題に取り組んでいる。本論文では、この２波形分離問題の解法の１つとして、雑音が付加された波形からスペクトルと位相の両方を考慮して信号波形を抽出する方法を提案する。

本方法のモデルは、図１のように構成される。図１の聴覚フィルタ群は基底膜の特性を模擬している Gammatone filter とこれを基底とする wavelet 分析系で構成され、分離規則は、Bregman が提唱した４つの発見的規則[2] のうちの２つ (共変調マスキング解除と漸近的変化に関する規則) を考慮した物理的制約条件として記述されている。

2 ２波形分離問題の定式化

ある２つの音響信号波形 $f_1(t)$ と $f_2(t)$ が波形 $f(t) = f_1(t) + f_2(t)$ に合成された状況を想定する。これらは、図１に示した N 個の聴覚フィルタで構成される wavelet 分析系により周波数分解される。ここで、$f_1(t)$ と $f_2(t)$ がそれぞれ k 番目の分析フィルタにより

$$f_1(t) \quad : A_k(t)\sin(\omega_k t) \tag{1}$$

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

に周波数分解されたものとする。但し、ω_k は周波数、$\theta_k(t)$ 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、分析フィルタの出力 $X_k(t)$ は、

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

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

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

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

となる。従って、振幅包絡 $S_k(t)$ と周波数 ω_k が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波形の振幅情報 $A_k(t), B_k(t)$ を

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

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

のように解析的に解くことができる。この処理過程は、図１の分離部で行なわる。最後に、すべての分析フィルタ $(X_k(t), 1 \le k \le N)$ について、振幅包絡 $A_k(t)$ と $B_k(t)$ に分離し、それぞれを同じ信号成分のものとして wavelet 合成系により合成 (グルーピング) することで、$f_1(t)$ と $f_2(t)$ を再構成できる。

本研究では、$f_1(t)$ を純音、$f_2(t)$ を純音の周波数を中心周波数とする帯域雑音として２波形分離問題を議論する。

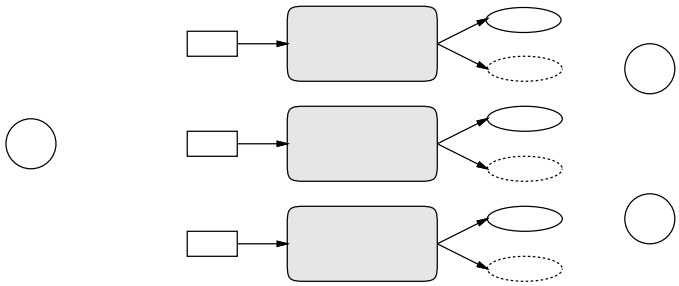

図 1: Gammatone filter を基底関数とした wavelet 分析合成系の構成

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

3.1 Gammatone filter の特性

Patterson[9] によって設計された聴覚フィルタは、Gammatone filter と呼ばれ、基底膜の特性を模擬したものとして知られている。Gammatone filter は

$$gt(t) = At^{n-1}e^{-2\pi b_f t}\cos(2\pi f_0 t + \phi), \quad (t \geq 0) \tag{8}$$

のインパルス応答を持ち、ガンマ分布を表す振幅項 $At^{n-1}e^{-2\pi b_f t}$ と周波数 f_0 で表される。また、Gammatone filter の周波数特性は、$f_0 >> b_f$ のとき、

$$GT(f) \approx \left[1 + \frac{j(f-f_0)}{b_f}\right]^{-n}, \quad 0 < f < \infty \tag{9}$$

となる。$GT(f)$ は、$gt(t)$ の Fourier 変換を周波数 f の関数で表したものであり、中心周波数を f_0 とする帯域通過フィルタの形態を示している[11]。図 2 に Gammatone filter の時間領域および周波数領域の特性を示す。

3.2 Gammatone filter を基底関数とした wavelet 変換

wavelet 変換で用いられる基底関数は、直交基底と擬似直交基底とに分けられるが、後者の場合、逆変換を定義するために基底 wavelet が許容条件を満たさなければならない[10]。

$$C_\psi := \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{10}$$

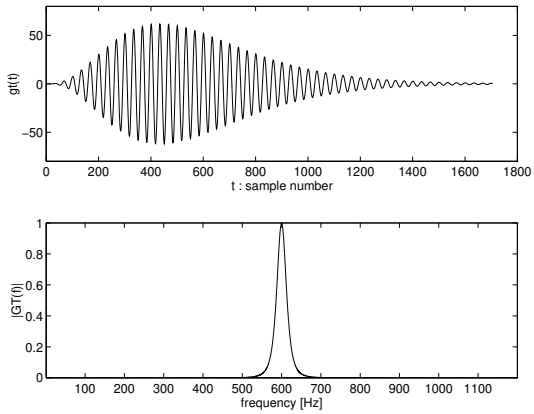

図 2: Gammatone filter のインパルス応答と周波数特性 (中心周波数 600[Hz]、$n = 4, b_f = 22.9945$)

但し、$\hat{\psi}(\omega)$ は $\psi(t)$ の Fourier 変換であり、基底関数の時間平均は 0 ($\int \psi(t)dt = 0$) でなければならない。

本研究では、擬似直交基底にある基本 wavelet を Gammatone filter とする ($\psi(t) = gt(t)$)。この基本 wavelet は、図 2 に示すように $GT(0) = 0$ と式 (10) の許容条件を近似的に満たしている。このとき、連続 wavelet 変換とその逆変換は、

$$\hat{f}(a,b) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a,b}(t)}dt \tag{11}$$

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{f}(a,b)\psi_{a,b}(t)\frac{dadb}{a^2} \tag{12}$$

と定義できる。但し、a はスケールパラメータ、b はシフトパラメータであり、$\overline{\psi_{a,b}}$ は $\psi_{a,b}$ の複素共役である。また、wavelet は、基本 wavelet を a 倍

のスケール変換と b だけシフトしたものであり、

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{13}$$

と表される $(a, b \in \mathbf{R})$。便宜上、連続 wavelet 変換で議論しているが、計算機への実装には、スケールパラメータ $a = \alpha^i$、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。

次に wavelet 分析合成系において位相を定義するために、実部と虚部が Hilbert 変換で結ばれるような基底関数として Gammatone filter を再定義する。

$$\Psi(t) = At^{n-1}e^{j2\pi f_0 t - 2\pi b_f t} \tag{14}$$

このとき複素 wavelet 変換は、

$$\tilde{\mathbf{f}}(a,b) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} f(t)\overline{\Psi\left(\frac{t-b}{a}\right)}dt \tag{15}$$

と定義され、振幅項 $|\tilde{\mathbf{f}}(a,b)|$ と位相項 $\arg(\tilde{\mathbf{f}}(a,b))$

$$\tilde{\mathbf{f}}(a,b) = |\tilde{\mathbf{f}}(a,b)|e^{j\arg(\tilde{\mathbf{f}}(a,b))} \tag{16}$$

で表すことができる。

3.3 wavelet 分析合成系

本論文で作成したものは、中心周波数 $f_0 = 600$[Hz]、通過帯域を 60 ~ 6000[Hz]、フィルタ数 (チャネル数) を $N = 128$ とした。これより、スケールパラメータ $a = \alpha^i$ は、i が $-\frac{N}{2} \le i \le \frac{N}{2}$, α が $10^{2/N}$ となる[11]。また、シフトパラメータ $b = j/f_s$ は、サンプリング周波数 $f_s = 20$[kHz] で離散化される。ここで作成した wavelet 分析合成系の周波数特性は、図 3 のように完全に通過帯域を被覆している。

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 (4) における振幅包絡 $S_k(t)$ は、複素 wavelet 変換で定義された振幅スペクトル $|\tilde{\mathbf{f}}(a,b)|$ から

$$S_k(t) = |\tilde{\mathbf{f}}(a,b)| \tag{17}$$

で求めることができる。また、位相スペクトル $\arg(\tilde{\mathbf{f}}(a,b))$ の時間変化は、周波数 ω_k をバイアス

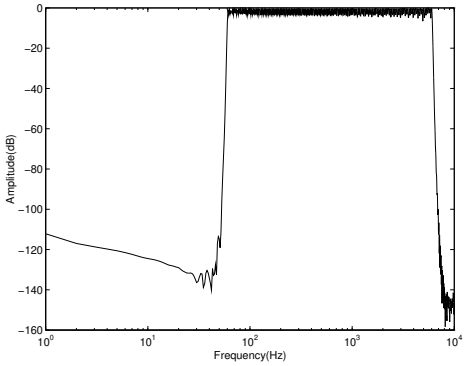

図 3: wavelet 分析合成系の周波数特性 $(N = 128)$

とした出力位相の時間変化を表す。従って、式 (5) における出力位相 $\phi_k(t)$ は、複素 wavelet 変換で定義された位相スペクトル $\arg(\tilde{\mathbf{f}}(a,b))$ から

$$\phi_k(t) = \int \left(\frac{d}{db}\arg\left(\tilde{\mathbf{f}}(a,b)\right) - \omega_k\right) db \tag{18}$$

で求めることができる。

4.2 入力位相 $\theta_k(t)$ の計算方法

Bregman が提唱した発見的規則には、“単一音は急に変化しないし、ゆっくりと滑らかに変化する” という漸近的変化の規則がある[2]。そこで、これを物理的制約条件に捉え直し、近似的に “微小区間において、$A_k(t)$ は変化しない” ものとすれば、

$$\frac{dA_k(t)}{dt} = 0 \tag{19}$$

と表すことができる。この制約条件を式 (6) に適用することにより、1 階線形微分方程式 ：

$$y'(t) + \frac{P'(t)}{P(t)}y(t) = \frac{Q'(t)}{P(t)} \tag{20}$$

が導びかれる。但し、$y(t) = \cot\theta_k(t)$、$P(t) = S_k(t)\sin\phi_k(t)$、$Q(t) = S_k(t)\cos\phi_k(t)$ である。この微分方程式の一般解 $\theta_k(t)$ は

$$\theta_k(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C}\right) \tag{21}$$

となる。尚、Cは未定係数である。ここで未定係数 Cを決定できれば、一意に $\theta_k(t)$ を求めることができるが、明確な決定法が明らかになっていない。そこで、CMR の知見と連続性の概念を物理的制約条件として考慮することで、ある境界条件で決定される未定係数 Cの最適値を求める。

4.3 2波形分離方法

Bregmanは、CMRの現象[12]を発見的規則の1つである“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”という点で説明している[2]。この知見は、信号が同じ音源から出ている場合、分析フィルタ間の振幅包絡に強い相関があるということを意味している。そこで、この規則を帯域雑音に適用し、物理的制約条件に捉え直すと、“着目している振幅包絡$B_k(t)$は、図4に示すような隣接する聴覚フィルタから得られる$B_{k-1}(t)$と$B_{k+1}(t)$に強い相関がなければならない”となり、

$$B_k(t) \approx B_{k-1}(t) \quad B_k(t) \approx B_{k+1}(t) \qquad (22)$$

と表すことができる。この制約から帯域雑音の振幅間の相関が最も強くなるときの未定係数Cが最適なものと解釈できる。

次に、未定係数Cは何らかの境界条件で決定されるものであり、この境界においてそれぞれ分離された信号は連続性を保持しなければならない。連続性に関する発見的規則は、漸近的変化[2]やold-plus-newの規則[1]で述べられているが、物理的制約として、“雑音の振幅包絡$B_k(t)$は分離境界の接合部$(t = T_r)$で、ある幅ΔBで接合されていなければならない”とすると、

$$\left| [B_k(t)]_{R:t\to+T_t} - [B_k(t)]_{R:t\to-T_r} \right| \le \Delta B \quad (23)$$

と表すことができる。但し、Rはそれぞれ境界を表し、振幅包絡$[B_k(t)]_{R:t\to+T_t}$は、境界Rにおける$t = Tr$のときの値を示す。この制約から、境界Rにおける連続性を許容する幅ΔBに関して、未定係数Cに制限範囲$\alpha(t) \le C \le \beta(t)$を設けることができる。但し、$\alpha(t)$と$\beta(t)$は、この境界における未定係数の上限と下限である。従って、これらの物理的制約を考慮することで、接合境界における連続性を保持した形で、最適な未定係数Cを決定できる。

次に、上記の制約を考慮した2波形分離法を説明するが、これには

1. 入力位相$\theta_k(t), k = i-1, i, i+1$はある区間であまり変化しない。
2. 出力位相$\phi_k(t), k = i-1, i, i+1$は測定可能である。
3. 振幅包絡$A_{i-1}(t)$、$A_{i+1}(t)$は聴覚フィルタの周波数特性から求めることができる。

という仮定を設けている。また、微小区間M/f_0において、制約条件$\frac{dA(t)}{dt} = 0$が成り立つものとしている。

全解析区間内における分離区間は、立ち上がり(onset)と立ち下がり(offset)によって決定され、微小区間$1/f_0$の整数倍Mの間隔M/f_0で解析的に解くことができる。ここで、立ち上がりと立ち下がりは、振幅包絡$S_i(t)$と位相$\phi_i(t)$の時間微分から検出可能である。また、微小区間の整数倍Mの区間M/f_0は、隣接する聴覚フィルタ間での振幅包絡の変化を十分観測できる区間である。この区間M/f_0の反復的な解析による2波形分離法を図5に示す。この分離法では、分離区間内で区間M/f_0が重複しないように繰り返し解かれる。尚、図5において、振幅間の相関をとるとき、境界M/f_0とそれより過去の境界区間すべてを利用しており、記号として境界M/f_0における振幅包絡を$[B_k(t)]_{M/f_0}$、それより過去を含めたものを$[B_k(t)]_F$と表している。

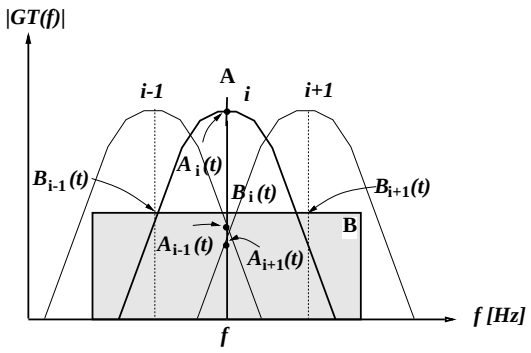

図 4: 隣接する聴覚フィルタの特性

5 2波形分離のシミュレーション

5.1 実験データ

上記で説明した分離方法を用いて2波形分離が可能であることを示す。まず、実験データとして、部分的な定常音$f_1(t)$と振幅変調された帯域雑音$f_2(t)$、ランダム帯域雑音$f_3(t)$の3つの信号を用意する。この実験データを図6に示す。

$$f_1(t) = \begin{cases} A(t)\sin(2\pi f_0 t) & 0.3 \le t \le 0.7 \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(t) = 1000 \sum_{f=f_0-500}^{f_0+500} B(t)\sin(2\pi f t + R(f))$$

1. 未定係数Cの制限範囲 $\alpha(t) \leq C \leq \beta(t)$ を決める。
2. 振幅包絡 $[S_i(t)]_{M/f_0}$、出力位相 $[\phi_i(t)]_{M/f_0}$ を求める。
3. $\alpha(t) \leq C \leq \beta(t)$ の範囲内で以下の処理をする。
   (a) 2波形の振幅包絡 $[\hat{A}_i(t)]_{M/f_0}$、$[\hat{B}_i(t)]_{M/f_0}$ を求める。
   (b) 隣接する聴覚フィルタにおいて
      i. 振幅包絡 $[\hat{A}_{i-1}(t)]_{M/f_0}$,$[\hat{A}_{i+1}(t)]_{M/f_0}$
      ii. 振幅包絡 $[S_{i-1}(t)]_{M/f_0}$,$[S_{i+1}(t)]_{M/f_0}$
      iii. 出力位相 $[\phi_{i-1}(t)]_{M/f_0}$,$[\phi_{i+1}(t)]_{M/f_0}$
      iv. 振幅包絡 $[\hat{B}_{i-1}(t)]_{M/f_0}$,$[\hat{B}_{i+1}(t)]_{M/f_0}$
      を求める。
   (c) 振幅包絡 $[\hat{\hat{B}}_i(t)]_{M/f_0}$ を推測し、$[\hat{\hat{B}}_i(t)]_F$ を求める。

$$[\hat{\hat{B}}_i(t)]_{M/f_0} = \frac{[\hat{B}_{i-1}(t)]_{M/f_0} + [\hat{B}_{i+1}(t)]_{M/f_0}}{2}$$

   (d) $[\hat{B}_i(t)]_F$ と $[\hat{\hat{B}}_i(t)]_F$ の相関値を求める。

$$Corr([\hat{B}_i(t)]_F, [\hat{\hat{B}}_i(t)]_F) = \frac{([\hat{B}_i(t)]_F, [\hat{\hat{B}}_i(t)]_F)}{||[\hat{B}_i(t)]_F|| \cdot ||[\hat{\hat{B}}_i(t)]_F||}$$

4. $Corr([\hat{B}_i(t)]_F, [\hat{\hat{B}}_i(t)]_F)$ が最大になるような未定係数 C を求める。

$$\max_{C, \alpha(t) \leq C \leq \beta(t)} Corr([\hat{B}_i(t)]_F, [\hat{\hat{B}}_i(t)]_F)$$

5. 式 (18) から $[\theta_i(t)]_{M/f_0}$ を求める。
6. $[A_i(t)]_{M/f_0}$ と $[B_i(t)]_{M/f_0}$ を求める。
7. $[A_i(t)]_F$ と $[B_i(t)]_F$ を求める。

図 5: 2波形分離方法

$$f_3(t) = A_C \sum_{f=f_0-500}^{f_0+500} C(f,t) \sin(2\pi f t + R(f))$$

$$\begin{aligned} A(t) &= 1200 \\ B(t) &= 1.5 * W_m - W(t) \\ C(f,t) &= 1.5 * W_m - W(t - \Delta T \cdot R(f)) \end{aligned}$$

但し、$f_0 = 600[\mathrm{Hz}]$、$W_m = \max(W(t))$ であり、$W(t)$ 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る。また、A_Cは$\sqrt{f_2(t)^2/f_3(t)^2} = 1$ となる定数、ΔTは任意の定数、$R(f)$ は $[-\pi, \pi]$ の一様乱数である。

このとき、混合波形 $f_C(t) = f_1(t) + f_2(t)$ は CMR の場合を想定し、混合波形 $f_M(t) = f_1(t) + f_3(t)$ はマスキングの場合を想定している。これらの混合波形を図 7 に示す。

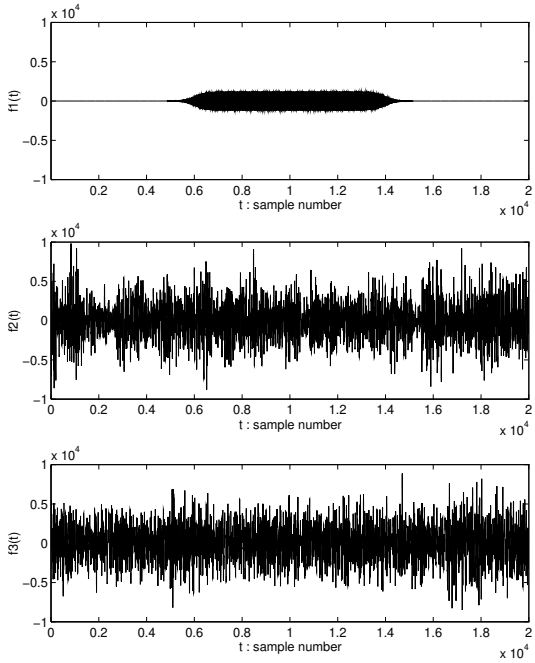

図 6: 音響信号 :$f_1(t)$、$f_2(t)$、$f_3(t)$

5.2 分離結果

混合信号 $f_C(t)$ の分離結果を示す。混合信号 $f_C(t)$ は、wavelet 分析系により周波数分解され、分析フィルタの出力における振幅包絡 $S_i(t)$ と出力位相$\phi_i(t)$ が求められる。次に、振幅包絡 $S_i(t)$ から $\frac{dS_i(t)}{dt}$の極大点と極小点を求め、その近傍にある $\frac{d\phi_i(t)}{dt}$の局所的な極大点 (あるいは極小点) を図 8 のように求める。このときの局所的な極大点を立ち上がり (onset) と立ち下がり (offset) とし、分離区間 (onset – offset) を決定する。入力位相

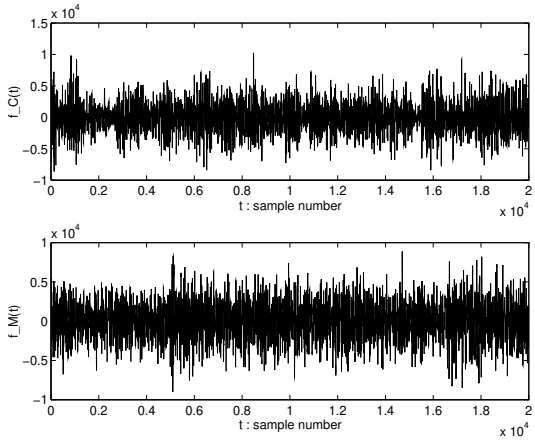

図 7: 混合信号 :$f_C(t)$ と $f_M(t)$

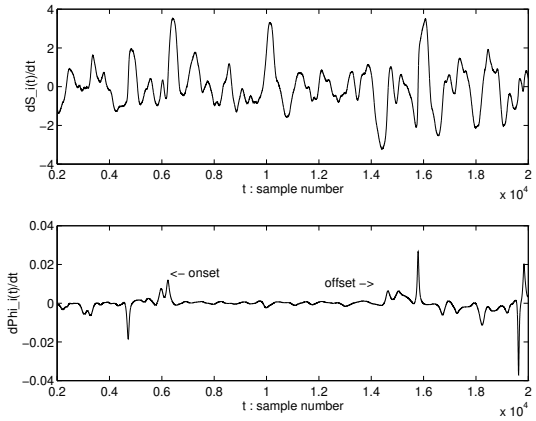

図 8: 振幅包絡と位相の時間微分 :$\frac{dS_i(t)}{dt}$ と$\frac{d\phi_i(t)}{dt}$

$\theta_i(t)$ は、図 5 の分離方法 ($M = 3$、区間 $3/f_0$、$\Delta B = 50.0$) に従って未定係数 Cの最適値を求めることで、一意に求まる。これらの物理的手がかりを求めることで、式 (6),(7) より 2 波形の振幅包絡 $A_i(t)$ と $B_i(t)$ が計算される。中心周波数 f_0の分析フィルタにおける振幅包絡の分離結果を図 9 に示す。最後に、すべてのチャネルにわたって分離を行ない、それぞれの分離結果を wavelet 合成系により合成 (グルーピング) することで、$\hat{f}_1(t)$ と$\hat{f}_2(t)$ を再構成できる。この結果を図 10 に示す。尚、$\hat{f}_1(t)$ については、周波数 f_0が既知であるため、それを中心周波数とする分析フィルタについてだけ分離を行ない、再構成している。

同様に、混合波形 $f_M(t)$ の場合、振幅包絡 $S_i(t)$ と出力位相$\phi_i(t)$ を求め、入力位相$\theta_i(t)$ が一意に求まることで、2 波形の振幅包絡 $A_i(t)$ と $B_i(t)$ が図 11 のように分離され、図 12 のように再構成される。

これらの再構成された結果をみると、振幅変調された帯域雑音の方が、ランダム帯域雑音よりも容易に純音 $f_1(t)$ を抽出できることがわかる。また、再構成された信号波形について、時間領域における SN 比 :

$$SNR = 10\log_{10}\frac{\int_0^T f_i^2(t)dt}{\int_0^T (f_i(t) - \hat{f}_i(t))^2 dt}, \quad i = 1, 2, 3 \tag{24}$$

を用いて定量的な評価を行なったところ $f_C(t)$ のとき、$\hat{f}_1(t)$ が 30.99[dB]、$\hat{f}_2(t)$ が 23.13[dB]、$f_M(t)$ のとき、$\hat{f}_1(t)$ が 3.18[dB]、$\hat{f}_3(t)$ が 16.03[dB] であった。この結果から、混合信号 $f_C(t)$ の方が混合信号 $f_M(t)$ よりも正確に純音を抽出でき、更に分離した雑音もより正確に分離し、再構成できることがわかる。

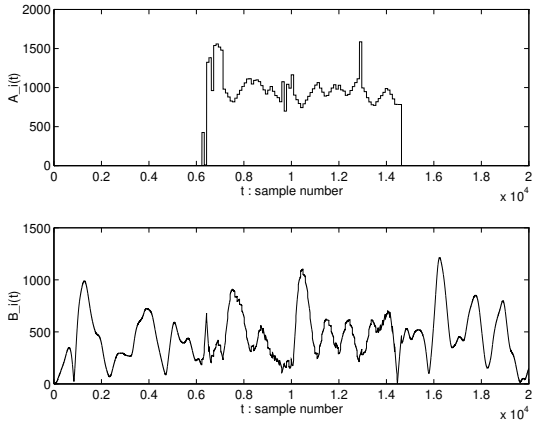

図 9: 分離された振幅情報 :$A_i(t)$ と $B_i(t)$

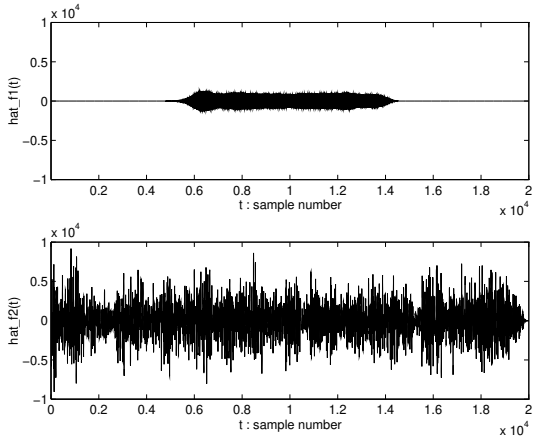

図 10: 再構成された信号 :$\hat{f}_1(t)$ と$\hat{f}_2(t)$

5.3 考察

混合波形が $f_C(t)$、$f_M(t)$ のいずれの場合も、原信号 $f_1(t)$ は、雑音を分離した残りものとして再構成されている。しかし、混合波形 $f_M(t)$ の場合、$f_1(t)$ は純音とかなり異なる信号として抽出されているのに対し、混合波形 $f_C(t)$ の場合、純音にかなり類似した信号として抽出されている。この結果は、雑音の振幅包絡 $B_i(t)$ と隣接する聴覚フィルタから得られる雑音の振幅包絡 $B_{i+1}(t)$,$B_{i-1}(t)$ に強い相関があるため、分離された残りものが純音として抽出されたものと解釈できる。また、CMR のモデル化とも解釈できる。この分離法にはまだ改良する余地があり、純音の振幅包絡 $A_i(t)$ にも連続性の制約を設けることで、純音 $f_1(t)$ をより正確に抽出できるものと考えられる。

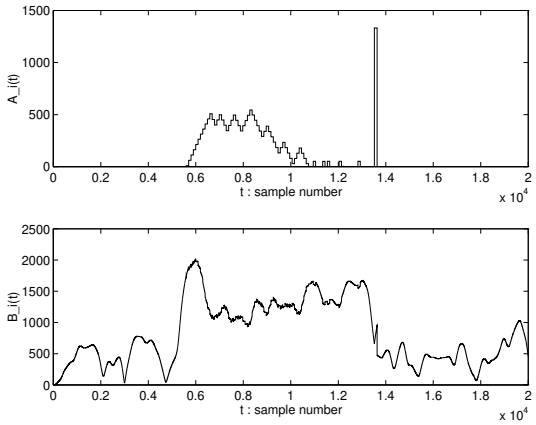

図 11: 分離された振幅情報 :$A_i(t)$ と $B_i(t)$

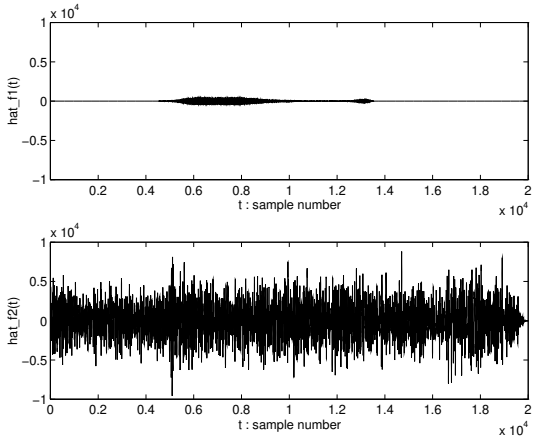

図 12: 再構成された信号 :$\hat{f}_1(t)$ と$\hat{f}_3(t)$

6 まとめ

雑音が付加された信号を wavelet 分析合成系で周波数分解し、振幅包絡と位相の変化及び、入力位相の変化に着目することで、原信号を抽出する方法を提案した。本方法では、振幅変調された帯域雑音が純音に付加された場合、純音の抽出が容易になり、ランダム帯域雑音が付加された場合、困難になることを示した。また、帯域雑音についてもかなり高い精度で分離ができることを示した。この結果は、CMR の工学的な説明として解釈でき、この枠組での CMR のモデル化となっている。

本論文では、Bregman のいう４つの発見的規則[1]のうち２つを物理的制約条件として取り入れたものであるが、残り２つ (調波関係、共通の立ち上がりと立ち下がり) についても同様に物理的制約条件として取り入れることで、雑音が付加された複合音や音声からの信号波形の抽出も可能と考えられる。

最後に、本研究で提案した抽出法において、音を分凝するためにどんな物理的手がかりを必要とし、どんな規則で解けばいいのかという必要十分条件を明らかにできれば、聴覚の計算理論の構築[3]に向けて一歩前進できるものと思われる。

謝辞

wavelet 分析合成系の構築およびその議論をして頂いた野田宗裕氏 (現スズキ (株)) に感謝する。

参考文献

[1] A. S. Bregman : “Auditory Scene Analysis: The Perceptual Organization of Sound,” MIT Press(1990).

[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] 河原 : “聴覚の計算理論の構築に向けて,” 日本音響学会聴覚研究会資料,H-94-63(1994).

[4] M. P. Cooke : “Modelling Auditory Processing and Organization,” Ph D Thesis, University of Sheffield (1991).

[5] G. J. Brown : “Computational Auditory Scene Analysis : A Representational Approach, ” Ph D Thesis, University of Sheffield(1992).

[6] D.P.W.Ellis: “A Computer Implementation of Psycoacoustic Grouping Rules,” Proc. 12th Int. Conf. on Pattern Recognition(1994).

[7] T.Nakatani, H.G.Okuno and T.Kawabata : “Unified Architecture for Auditory Scene Analysis and Spoken Language Processing,” ICSLP’94, 24,3(1994).

[8] 柏野、田中 :“２つの周波数成分の分離知覚に関する工学的モデル,” 信学論 (A), J77-A, 5, pp. 731-740(1994).

[9] Roy D.Patterson and John Holdsworth :“A Functional Model of Neural Activity Patterns and Auditory Images,” Advances in speech, Hearing and Language Processing, Vol3. JAI Press, London(1991).

[10] Chui C.K :“An Introduction to Wavelets”, Academic Press(1992).

[11] 野田 :“人間の聴覚特性を考慮した波形分離法に関する基礎的研究,” JAIST 修士論文 (1995).

[12] 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).