

Title	雑音が付加された波形からの信号波形の一抽出法
Author(s)	鷓木, 祐史; 赤木, 正人; UNOKI, Masashi; AKAGI, Masato
Citation	電子情報通信学会論文誌 A, J80-A(3): 444-453
Issue Date	1997-03
Type	Journal Article
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/3985">http://hdl.handle.net/10119/3985</a>
Rights	Copyright (C) 1997 IEICE ( 許諾番号 : 07RB0181 ) <a href="http://www.ieice.org/jpn/trans_online/">http://www.ieice.org/jpn/trans_online/</a>
Description	

雑音が付加された波形からの信号波形の一抽出法

鵜木 祐史<sup>†</sup> 赤木 正人<sup>†</sup>

A Method of Signal Extraction from Noise-Added Signal

Masashi UNOKI<sup>†</sup> and Masato AKAGI<sup>†</sup>

あらまし 本論文では、音源分離のモデル化の試みとして二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 (Auditory Scene Analysis) 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二波形分離を可能にする。この三つの物理的手がかりのうち、振幅包絡と出力位相は直接求められ、入力位相は Bregman が示した一つの音響事象に生じる変化と漸近的变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。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二波形分離問題の解法を示す。特に、本モデルを人間の聴覚特性に合わせてパラメータ設定すれば、振幅変調された帯域雑音と純音が混合されたとき純音の分離が容易になり、ランダム帯域雑音のとき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モデル化が可能になる。

キーワード 聴覚の情景解析, 二波形分離, 共変調マスキング解除, Gammatone filter, wavelet 分析合成系

1. ま え が き

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: ASA) [1],[2] に基づく音源分離の研究が盛んに行われるようになった。この側面から、音源分離やカクテルパーティ効果をモデル化できれば、聴取者にとって必要な音だけを選択し、他の音を除外するということが可能になり、雑音に強い音声認識などに応用できる [4]。また、Marr [5] による視覚の計算理論の構築のアナロジーから聴覚の計算理論を構築することは現段階でかなりの時間が必要であると考えられるが、ASA は聴覚の計算理論を構築するための一方向を提供してくれることから、この側面からのモデル化は聴覚の計算理論の構築 [6]~[8] に向けても新たな視点を提供できるものと考えられる。

Bregman は、音を通じて環境を把握する情景解析問題を解くために、聴覚が利用している制約条件のいくつかを音響事象に関係する四つの発見的規則：

- (1) 共通の立上り/立下りに関する規則
- (2) 漸近的变化に関する規則

(3) 調波関係に関する規則

(4) 一つの音響事象に生じる変化に関する規則

としてまとめている [2],[3]。これらの規則に基づく計算機モデルの研究として、Sheffield 大学の Brown と Cooke ら [9]~[11] による音響事象に着目した分凝モデル、MIT の Ellis [12] 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 [13],[14] 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モデルがある。また、東大の柏野ら [15],[16] は、二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。これらの計算機モデルはいずれも Bregman の言う規則 (1) と (3) を利用したものである。また、振幅あるいはパワースペクトルを用いた分離を考えているため、二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。

一方、筆者らは、同一周波数領域において信号を完全に分離するためには振幅スペクトルのほかに位相も考慮しなければならないという立場に立ち [17]、Bregman の言う規則 (2) と (4) を利用して基本的な音源分離問題である二波形分離問題に取り組んでいる [18],[24]。

本論文では、この二波形分離問題の解法の一つと

<sup>†</sup> 北陸先端科学技術大学院大学情報科学研究科, 石川県 Japan Advanced Institute of Science and Technology, Ishikawa-ken, 923-12 Japan

して、雑音が付加された波形から wavelet 変換を用いて振幅スペクトルと位相スペクトルを計算し、これを用いて信号波形を抽出する方法を提案する。また、本モデルを人間の聴覚特性に合わせてパラメータ設定した場合、共変調マスキング解除 (co-modulation masking release: CMR) [19] の工学的モデルになることを示す。

以下、2. で本方法のモデルの概要を説明し、二波形分離問題の定式化を行う。3. で wavelet 分析合成系の設計方法と諸特性を示す。4. で定式化された物理パラメータの導出方法とこれを用いた二波形分離アルゴリズムを説明する。5. で本手法の有効性を示すために、二波形分離のシミュレーションを行う。6. で分析合成系を人間の聴覚特性に合わせたパラメータ設定を行うことにより、CMR の工学的モデル化が可能であることを示す。最後に、7. で本論文のまとめを述べる。

## 2. 二波形分離問題の定式化

本論文では、“ある二つの独立な音源で生じた音響信号が加算された信号から、それぞれの音響信号に分離すること”を二波形分離問題と定義する。この二波形分離問題は以下のように定式化される。

はじめに、ある二つの音響信号  $f_1(t)$  と  $f_2(t)$  が

$$f(t) = f_1(t) + f_2(t) \quad (1)$$

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

$$f_1(t) : A_k(t) \sin(\omega_k t + \theta_{1k}(t)) \quad (2)$$

$$f_2(t) : B_k(t) \sin(\omega_k t + \theta_{2k}(t)) \quad (3)$$

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、 $\theta_{1k}(t)$  は  $f_1(t)$  のもつ入力位相、 $\theta_{2k}(t)$  は  $f_2(t)$  の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、

$$X_k(t) = S_k(t) \sin(\omega_k t + \phi_k(t)) \quad (4)$$

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

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

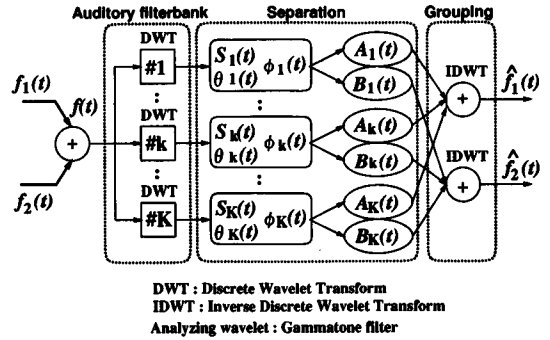


図 1 wavelet 分析合成系  
Fig.1 Wavelet analysis-system.

$$\phi_k(t) = \arctan\left(\frac{A_k(t) \sin \theta_{1k}(t) + B_k(t) \sin \theta_{2k}(t)}{A_k(t) \cos \theta_{1k}(t) + B_k(t) \cos \theta_{2k}(t)}\right) \quad (6)$$

である。但し、 $\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$  であり、 $\theta_k(t) \neq n\pi, n \in \mathbf{Z}$  とする。このとき、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  $\theta_{1k}(t)$ 、 $\theta_{2k}(t)$  がわかれば二波形の振幅包絡  $A_k(t)$  と  $B_k(t)$  を

$$A_k(t) = \frac{S_k(t) \sin(\theta_{2k}(t) - \phi_k(t))}{\sin \theta_k(t)} \quad (7)$$

$$B_k(t) = \frac{S_k(t) \sin(\phi_k(t) - \theta_{1k}(t))}{\sin \theta_k(t)} \quad (8)$$

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 変換によって実現し、入力位相を  $\theta_{1k}(t) = 0$ 、 $\theta_k(t) = \theta_{2k}(t)$ 、 $f_1(t)$  を振幅変化音 (AM 音:  $A_k(t)$  が一定ならば純音) に限定する。また、分析フィルタの中心角周波数  $\omega_k$  を  $f_1(t)$  の中心角周波数に一致させ、 $f_2(t)$  を  $f_1(t)$  の周波数を中心周波数とする帯域雑音とし、 $f_2(t)$  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した場合の二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

#### 3.1 wavelet 変換の諸定義

はじめに, wavelet 分析合成系を設計するために必要な範囲で wavelet 変換の性質を以下にまとめる.

関数  $f(t)$  の wavelet 変換  $\tilde{f}(a, b)$  は,

$$\tilde{f}(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \quad (9)$$

なる積分変換で定義される. 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,  $\overline{\psi}$  は  $\psi$  の複素共役である. 積分核は関数  $\psi(t)$  を  $a$  倍のスケール変換と  $b$  だけシフトしたものとなっている. この関数  $\psi(t)$  の選択には数学的に大きな自由度をもつが, 一般に許容条件 (admissibility condition):

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

を満たす正数  $D_\psi$  が存在するような 2 乗可積分関数とする. 但し,  $\hat{\psi}(\omega)$  は  $\psi(t)$  の Fourier 変換である. このとき,  $\psi$  を基本 wavelet と言い, 次のような逆変換 (再構成) が存在する [20].

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

更に,  $\psi$  が絶対可積分ならば, 許容条件は  $\hat{\psi}(0) = 0$  と同値である.

また, 基本 wavelet が複素数値をとるとき, wavelet 変換を振幅項  $|\tilde{f}(a, b)|$  と位相項  $\arg(\tilde{f}(a, b))$  で表すこともできる [21].

$$\tilde{f}(a, b) = |\tilde{f}(a, b)| e^{j \arg(\tilde{f}(a, b))} \quad (12)$$

本論文では, 聴覚系を模擬した聴覚フィルタ群を構築するために, 次のような基底膜の特性を模擬した Gammatone filter を基本 wavelet として用いる.

#### 3.2 Gammatone filter の特性

Gammatone filter は, Patterson [22] によって設計された聴覚フィルタであり, 基底膜の特性をより良く模擬したものとして知られている. これは

$$gt(t) = At^{N-1} e^{-2\pi b_f t} \cos(2\pi f_0 t), \quad t \geq 0 \quad (13)$$

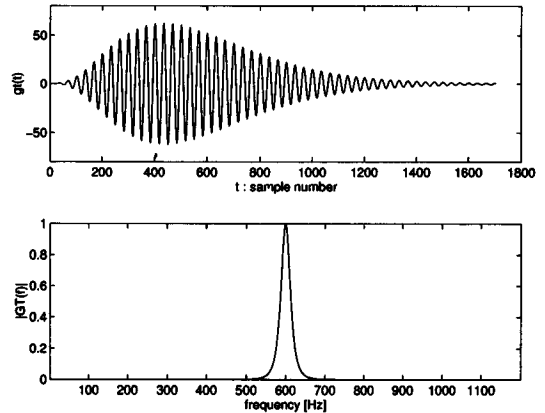


図2 Gammatone filter のインパルス応答と周波数特性 (中心周波数: 600 Hz,  $N = 4$ ,  $b_f = 22.9945$ )

Fig.2 Impulse response and amplitude of gammatone filter ( $f_0 = 600$  Hz,  $N = 4$ ,  $b_f = 22.9945$ ).

のインパルス応答をもち, ガンマ分布を表す振幅項  $At^{N-1} e^{-2\pi b_f t}$  と周波数  $f_0$  の余弦波で表される. また, この周波数特性は,  $f_0 \gg b_f$  のとき, 近似的に

$$GT(f) \approx \left[ 1 + \frac{j(f - f_0)}{b_f} \right]^{-N}, \quad 0 < f < \infty \quad (14)$$

と表せる.  $GT(f)$  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  $f_0$  とする帯域フィルタの形態を示している [22]. 図 2 に Gammatone filter の時間領域および周波数領域の特性を示す. この図からもわかるように  $GT(0) \approx 0$  となっていることから, Gammatone filter は許容条件を近似的に満たすことができるので, 基本 wavelet として十分利用できることがわかる.

#### 3.3 wavelet 分析合成系

本論文では, wavelet 変換を式 (12) のように表現するために, 実部と虚部が Hilbert 変換で結ばれるような関数として Gammatone filter を再定義する. 次に, これを基本 wavelet :

$$\psi(t) = At^{N-1} e^{j2\pi f_0 t - 2\pi b_f t} \quad (15)$$

とすることで, 中心周波数  $f_0 = 600$  Hz, 通過帯域が 60 ~ 6000 Hz, フィルタ数が  $K = 128$  の wavelet 分析合成系を設計した. 便宜上, ここでは連続 wavelet 変換で議論しているが, 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

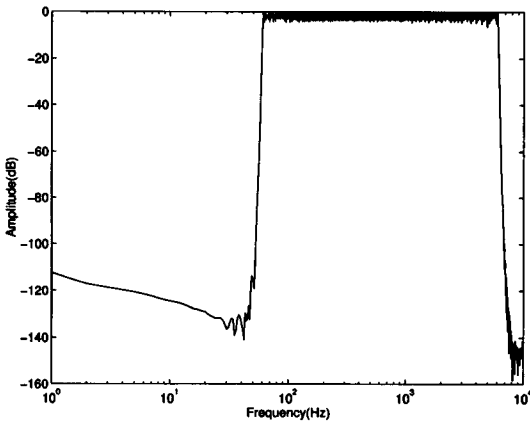


図3 wavelet 分析合成系の周波数特性 ( $K = 128$ )  
Fig.3 Frequency characteristics of wavelet filterbank.

$a = \alpha^p, -\frac{K}{2} \leq p \leq \frac{K}{2}, \alpha = 10^{2/K}$ , シフトパラメータ  $b = q/f_s$  として離散 wavelet 変換 ( $p, q \in \mathbf{Z}$ ) を用いている [18]. 図 3 に wavelet 分析合成系の周波数特性を示す。ここで、各フィルタの方形帯域幅は重複せず、図 3 のように完全に通過帯域を被覆している [18].

#### 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

$$S_k(t) = |\tilde{f}(\alpha^{k-\frac{K}{2}}, t)| \quad (16)$$

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

$$\phi_k(t) = \int \left( \frac{d}{dt} \arg \left( \tilde{f}(\alpha^{k-\frac{K}{2}}, t) \right) - \omega_k \right) dt \quad (17)$$

で求めることができる。

(証明) 付録 1 参照。 □

##### 4.2 入力位相 $\theta_k(t)$ の計算方法

入力位相  $\theta_k(t)$  は、三つの物理的制約条件: (i) 漸近的变化に関する規則, (ii) 連続性 (時間的接続) の規則, (iii) 一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、“一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 [2] 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とらえ直す。

[制約条件 1] (漸近的变化) 振幅包絡  $A_k(t)$  の時間変化は  $R$  回微分可能な  $R$  次多項式  $C_{k,R}(t)$  で近似される:

$$\frac{dA_k(t)}{dt} = C_{k,R}(t) \quad (18)$$

□

この制約を式 (7) に適用することで、

$$y'(t) + \frac{P'(t)}{P(t)}y(t) = \frac{Q'(t) - C_{k,R}(t)}{P(t)} \quad (19)$$

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。 $\theta_k(t)$  は、この微分方程式 (19) 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 (19) の一般解

$$\theta_k(t) = \arctan \left( \frac{S_k(t) \sin \phi_k(t)}{S_k(t) \cos \phi_k(t) + C_k(t)} \right) \quad (20)$$

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、 $C_k(t) = -\int C_{k,R}(t)dt + C_{k,0}$  と表される。 □

ここで  $C_k(t)$  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  $C_{k,r}, r = 0, 1, \dots, R$  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 ( $C_k(t) = C_{k,0}$ ) として取り扱うことにする。このとき、制約条件 (18) は  $dA_k(t)/dt = 0$  となり、微小区間  $\Delta t$  で  $A_k(t)$  は変化しないことを意味する。

次に、この微小区間  $\Delta t$  において分離を行うために規則 (ii) を適用する。これは、“分離を行った微小区間 ( $T_r - \Delta t \leq t < T_r$ ) と分離を行う微小区間 ( $T_r \leq t < T_r + \Delta t$ ) の境界  $T_r$  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とらえ直す。

[制約条件 2] (連続性) 時間領域  $T_r \leq t < T_r + \Delta t$  において分離を行うとき、 $A_k(t), B_k(t), \theta_k(t)$  は、分離境界 ( $t = T_r$ ) の前後において、ある幅  $\Delta A, \Delta B, \Delta \theta$  以内で接合されていなければならない:

$$\begin{cases} |A_k(T_r + 0) - A_k(T_r - 0)| \leq \Delta A \\ |B_k(T_r + 0) - B_k(T_r - 0)| \leq \Delta B \\ |\theta_k(T_r + 0) - \theta_k(T_r - 0)| \leq \Delta \theta \end{cases} \quad (21)$$

□

ここで、式 (7)、式 (8)、式 (20) から  $A_k(t)$  と  $B_k(t)$  および  $\theta_k(t)$  が未定係数  $C_{k,0}$  の関数となっているこ

```

入力位相  $\theta_{1k}(t) = 0, \theta_k(t) = \theta_{2k}(t);$ 
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  $d\phi_k(t)/dt$  から分離抽出の対象となる信号
  の立上り時刻  $T_{on}$  と立下り時刻  $T_{off}$  を求める;
  分離区間を  $T_{on} \leq t \leq T_{off}$  とする;
  この分離区間を  $I$  個の微小区間  $\Delta t = M/f_0$  に分割する;
  for  $i := 1$  to  $I$  do
    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;
      (1) 周波数特性から振幅包絡  $\hat{A}_{k\pm 1}(t)$  を求める;
      (2) 補題 1 より,  $S_{k\pm 1}(t)$  と  $\phi_{k\pm 1}(t)$  を求める;
      (3) 式 (7) から,  $\hat{A}_{k\pm 1}(t), S_{k\pm 1}(t), \phi_{k\pm 1}(t)$ 
      を用いて入力位相  $\hat{\theta}_{k\pm 1}(t)$  を求める;
      (4) 式 (8) から, 振幅包絡  $\hat{B}_{k\pm 1}(t)$  を求める;
      (5)  $\hat{B}_k(t) = (\hat{B}_{k-1}(t) + \hat{B}_{k+1}(t))/2;$ 
      (6) 式 (24) から, 雑音の振幅包絡間の相関値
       $\text{Corr}(\hat{B}_k(t), \hat{B}_k(t))$  を求める;
    end
    式 (25) 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 (20) 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
  式 (2) と式 (3) から各周波数成分を求める;
end
 $\hat{f}_1(t)$  と  $\hat{f}_2(t)$  に再構成する;

```

図 5 二波形分離アルゴリズム  
Fig. 5 Signal segregation algorithm.

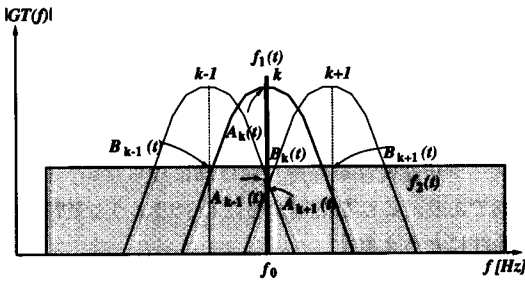


図 4 隣接する聴覚フィルタにおける純音  $f_1(t)$  と帯域雑音  $f_2(t)$  の通過特性  
Fig. 4 Characteristic of adjacent auditory filters.

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  $C_{k,0}$  の取り得る範囲を

$$C_{k,\alpha} \leq C_{k,0} \leq C_{k,\beta} \quad (22)$$

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, “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” という規則 [2] 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とらえ直す。[制約条件 3] (一つの音響事象に生じる変化) 振幅包絡  $B_k(t)$  は隣接する聴覚フィルタから得られる振幅包絡  $B_{k\pm 1}(t)$  に強い相関がなければならない:

$$B_k(t) \approx B_{k\pm 1}(t) \quad (23)$$

□

ここで, 式 (8) と式 (20) から  $B_k(t)$  は  $C_{k,0}$  の関数であることに着目し, ある  $C_{k,0}$  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

$$\text{Corr}(\hat{B}_k, \hat{B}_k) = \frac{\langle \hat{B}_k, \hat{B}_k \rangle}{\|\hat{B}_k\| \|\hat{B}_k\|} \quad (24)$$

と定義する。但し,  $\hat{B}_k(t) = (\hat{B}_{k+1}(t) + \hat{B}_{k-1}(t))/2$  であり, ある過去の時刻を  $T_B$ , 分離境界を  $T_r$  としたとき

相関を求める時間領域（積分範囲）は  $T_B \leq t < T_r + \Delta t$  とする。また、 $\hat{B}_{k\pm 1}(t)$  はある  $C_{k,0}$  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 は式 (24) を最大とする  $C_{k,0}$  を最適値として選ぶことと解釈できる。これは、

$$C_{k,\alpha \leq C_{k,0} \leq C_{k,\beta}} \text{Corr}(\hat{B}_k, \hat{B}_k) \quad (25)$$

なる未定係数  $C_{k,0}$  を見つけ出すことで実現でき、式 (20) によって一意に  $\theta_k(t)$  を決定できる。

### 4.3 二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] (二波形分離アルゴリズム) 入力位相を  $\theta_{1k}(t) = 0$ ,  $\theta_k(t) = \theta_{2k}(t)$  とした二波形分離問題において、制約条件 1~3 を適用すれば、式 (20) と式 (25) 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) から二波形分離問題を解くことができる。

(証明) 補題 1 と 2 より明らか。 □

この定理 1 に基づく二波形分離アルゴリズムを図 5 に示す。ここで、定理 1 を用いて二波形分離問題を解く場合、二つの音響信号が同時に存在する区間を知る必要がある。4.2 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  $f_1(t)$  の立上りと立下りにより知ることができる。そのため、ここでは  $S_k(t)$  と  $\phi_k(t)$  の時間変化に着目し、次のように  $f_1(t)$  の立上りと立下りを求める。

- (1) 立上り時刻  $T_{\text{on}} : dS_k(t)/dt$  の極大点近傍 ( $\pm 0.25 s$ ) にある  $|d\phi_k(t)/dt|$  の極大点
  - (2) 立下り時刻  $T_{\text{off}} : dS_k(t)/dt$  の極小点近傍 ( $\pm 0.25 s$ ) にある  $|d\phi_k(t)/dt|$  の極大点
- ここで、 $t_{\text{on}} \leq t \leq t_{\text{off}}$  が分離すべき区間となる。

## 5. 二波形分離のシミュレーション

ここでは、前章で述べた分離方法を用いて二波形分離のシミュレーションを行う。いずれの混合信号に対しても、微小区間を  $\Delta t = 3/f_0$ ,  $T_B$  を  $100\Delta t$  だけ過去の時刻とした。また、分離区間 ( $T_{\text{on}} \leq t \leq T_{\text{off}}$ ) において、 $S_k(t)$  の最大値を  $S_{\text{max}}$  としたとき、 $\Delta B = 0.027S_{\text{max}}$ ,  $\Delta\theta = \pi/20$  と一定値にしたが、 $\Delta A$  は  $C_k(t) = C_{k,0}$ ,  $T_r \leq t < T_r + \Delta t$  の影響により一定値にすることが困難であるため、式 (21) に基づき  $\Delta A = |A_k(T_r - \Delta t) - A_k(T_r - 2\Delta t)|$  とした。

### 5.1 実験データ

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

$$f_1(t) = F_{\text{BP}}(g_1) \quad (26)$$

$$g_1(t) = \begin{cases} 1200 \sin(2\pi f_0 t), & 0.3 + T_m \leq t \leq 0.7 + T_m \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

$$f_{21}(t) = \sum_{f=f_0-500}^{f_0+500} E_M(t) \sin(2\pi ft + R(f)), \quad 0 \leq t \leq 1.0 \quad (28)$$

$$f_{22}(t) = \sum_{f=f_0-500}^{f_0+500} E_R(f, t) \sin(2\pi ft + R(f)), \quad 0 \leq t \leq 1.0 \quad (29)$$

但し、 $F_{\text{BP}}(\cdot)$  は中心周波数が  $f_0$  で帯域幅が 23 Hz の帯域フィルタ (図 2 参照) を表し、 $f_0 = 600 \text{ Hz}$ ,  $T_m = 0.0125 m$ ,  $m = 0, 1, \dots, 9$ ,  $R(f)$  は  $[-\pi, \pi]$  の一様乱数である。ここで、 $E_M(t)$  と  $E_R(f, t)$  は白色

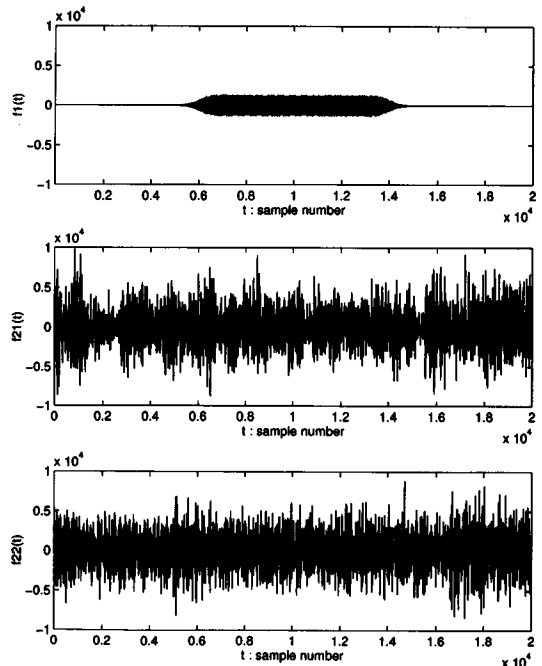


図 6 音響信号： $f_1(t)$ ,  $m = 0$ ,  $f_{21}(t)$ ,  $f_{22}(t)$   
Fig.6 Acoustic signals :  $f_1(t)$ ,  $m = 0$ ,  $f_{21}(t)$ , and  $f_{22}(t)$ .

雑音に 30 Hz の低域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、 $E_R(f, t)$  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、帯域雑音の帯域幅は 1 kHz であり、純音と帯域雑音の SN 比は  $-8.5$  dB である。図 6 にこれらの信号を示す。

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。人間がこれらの混合信号を聞くと、 $f_R(t)$  ではマスキングが起こり純音を聞き取れないが、 $f_M(t)$  では共変調マスキング

解除が起こり純音を聞き取ることができる。図 7 にこれらの混合信号を示す。

本シミュレーションでは、 $f_1(t)$  の開始時刻を式 (27) の  $m$  の範囲 ( $m = 0, 1, \dots, 9$ ) で変化させ、計 10 個の混合波形を用いて二波形分離を行う。

### 5.2 分離結果

はじめに、10 個の混合信号  $f_M(t)$  に対する二波形分離を行った。 $f_M(t)$  は wavelet 分析系により周波数分解され、図 5 の二波形分離アルゴリズムが適用される。このとき、 $S_k(t)$  と  $\phi_k(t)$  は図 8 のように求められ、 $f_1(t)$  の立上り時刻  $T_{on}$  と立下り時刻  $T_{off}$  は図 9

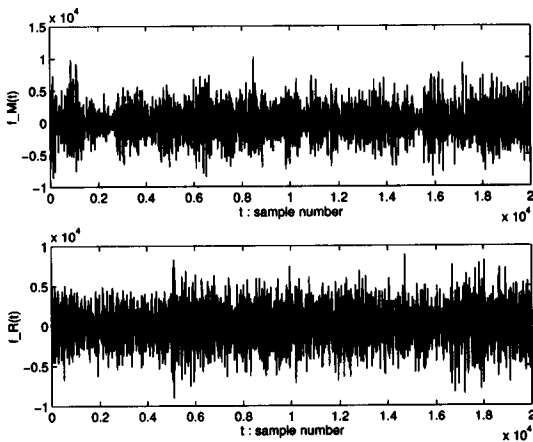


図7 混合信号： $f_M(t)$  と  $f_R(t)$   
Fig.7 Noise-added signals :  $f_M(t)$  and  $f_R(t)$ .

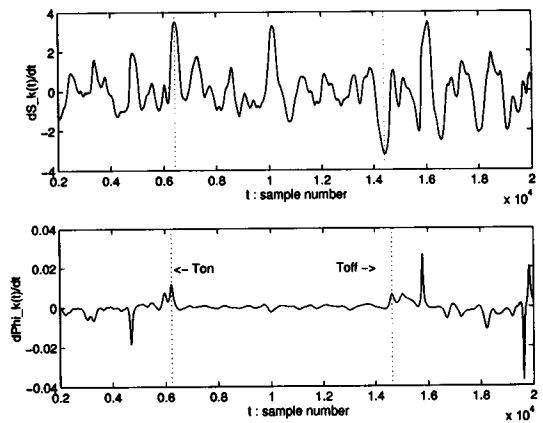


図9 振幅包絡  $S_k(t)$  と出力位相  $\phi_k(t)$  の時間微分 (混合信号  $f_M(t)$  の場合) :  $dS_k(t)/dt$  と  $d\phi_k(t)/dt$   
Fig.9  $dS_k(t)/dt$  and  $d\phi_k(t)/dt$  (Case of noise-added signal  $f_M(t)$ ).

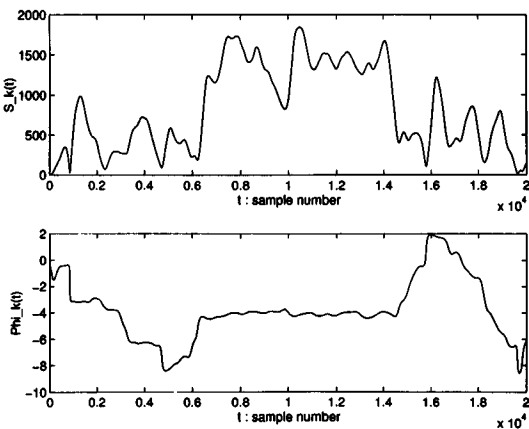


図8 振幅包絡  $S_k(t)$  と出力位相  $\phi_k(t)$  (混合信号  $f_M(t)$  の場合)  
Fig.8 Amplitude envelope  $S_k(t)$  and output phase  $\phi_k(t)$  (Case of noise-added signal  $f_M(t)$ ).

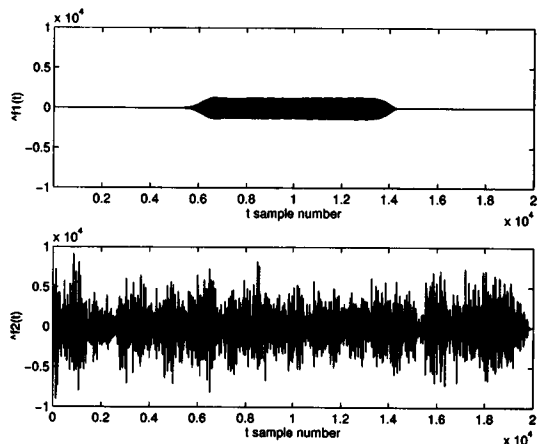


図10 分離結果 (混合信号  $f_M(t)$  の場合) :  $f_1(t)$  と  $f_2(t)$   
Fig.10 Extracted result (Case of noise-added signal  $f_M(t)$ ):  $f_1(t)$  and  $f_2(t)$ .



のように求められる。ここで、図 6 の純音 ( $m = 0$ ) に対する分離結果を図 10 に示す。この図から、純音の分離抽出が可能になっていることがわかる。

同様に、10 個の混合信号  $f_R(t)$  に対しても二波形分離を行ったところ、 $\hat{f}_1(t)$  はほとんど抽出されず、 $\hat{f}_2(t)$  は  $f(t)$  とほぼ同じ波形になった。

次に、10 個の混合信号に対する分離・再構成された信号波形について、時間領域における SN 比：

$$SNR = 10 \log_{10} \frac{\int_0^T f_i^2(t) dt}{\int_0^T (f_i(t) - \hat{f}_i(t))^2 dt}, \quad i = 1, 2 \quad (30)$$

の平均値を求めたところ、 $f_M(t)$  の分離では、 $\hat{f}_1(t)$  が 12.9 dB (標準偏差 2.58)、 $\hat{f}_2(t)$  が 10.1 dB (標準偏差 0.20)、 $f_R(t)$  のとき、 $\hat{f}_1(t)$  が 1.6 dB (標準偏差 1.58)、 $\hat{f}_2(t)$  が 8.7 dB (標準偏差 0.48) であった。例えば、中心周波数が 600 Hz、帯域幅が 23 Hz の帯域フィルタ (BPF) を用いて  $f(t)$  の雑音を抑制し、これを  $\hat{f}_1(t)$  とするような分離方法でも、 $\hat{f}_1(t)$  の分離精度は 8.1 dB しかない。この結果から、本方法は BPF を用いた分離方法と比較しても分離精度が優れていることがわかる。また、 $f_M(t)$  の方が  $f_R(t)$  よりも正確に純音を抽出できると同時に分離した雑音もより正確に分離できることもわかる。工学的応用としては、すべての信号波形を分離抽出できた方がよいと思われるが、本モデルは聴覚の情景解析の枠組で考えられるような信号 ( $f_M(t)$ ) の分離抽出を実現している。特に、上記の結果は人間の聴覚現象で知られている CMR と類似した特性である。そこで、次章ではこのことを積極的に検証するために、分析系のパラメータを人間の聴覚特性に合わせて同様の実験を行う。

## 6. 人間の聴覚特性に合わせたパラメータ設定

### 6.1 wavelet 分析合成系の再設計

図 1 の wavelet 分析合成系を人間の聴覚特性に合わせてパラメータ設定する。今まで聴覚フィルタの等価方形帯域幅 (Equivalent Rectangular Bandwidth: ERB) [23] をフィルタ数に応じて変化させていたが ( $K = 128$  のとき約 1/4ERB)、聴覚特性に合わせて中心周波数 600 Hz のときちょうど 1ERB として定 Q フィルタの分析合成系を設計する ( $K = 128$ ) [24].

### 6.2 実験条件と分離結果

実験データは、前章で用いた二つの混合信号 ( $f_M(t)$ )

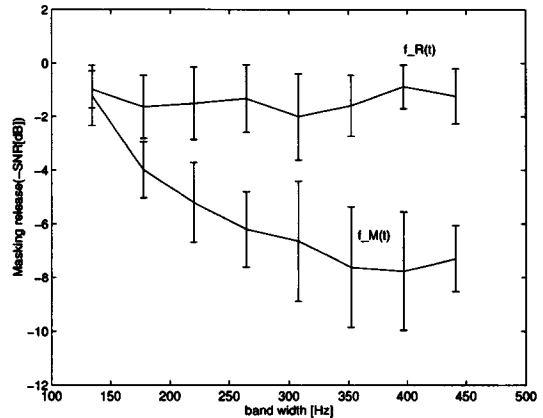


図 11 帯域幅とマスク解除の関係  
Fig. 11 Relationship between bandwidth and masking release.

と  $f_R(t)$  10 個ずつとする。CMR の実験 [19] 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の参照数  $L$  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、制約条件 3 における振幅包絡  $\hat{B}_k(t)$  を

$$\hat{B}_k(t) = \frac{1}{2L} \sum_{\ell=-L, \ell \neq 0}^L \hat{B}_{k+\ell}(t) \quad (31)$$

として、式 (20) と式 (25) から入力位相  $\theta_k(t)$  を一意に求める。

このときの帯域幅と SN 比向上の関係を図 11 に示す。この図の縦軸は SN 比の向上量を下向きに表し、横軸が  $L$  に対応した帯域幅を表している。また、曲線は SN 比の平均を、縦棒は標準偏差を表している。この結果から、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、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。これに対し、混合信号  $f_R(t)$  の場合、参照数  $L$  を増加させてもあまり SN 比が変わらなかった。ここで、抽出された純音の SN 比が向上することはマスク解除されたことに相当し、この実験結果は、振幅変調された帯域雑音の場合、帯域幅を広げることでマスク解除が起こり、ランダム帯域雑音の場合、帯域幅を広げてもマスク量が変わらないという結果を模擬していることがわかる。従って、この結果は、CMR の工学的な説明として解釈できる。

## 7. むすび

雑音が付加された信号を wavelet 分析合成系で周波数分解し、振幅包絡と出力位相の変化および、入力位相の変化に着目することで、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。工学的応用としては、すべての信号波形を分離抽出できた方がよいと思われるが、無条件にそれを実現することは不可能である。これに対し、本モデルは聴覚の情景解析に基づいた条件を設定することで信号の分離抽出を実現している。

また、wavelet 分析合成系を人間の聴覚特性に合わせてパラメータ設定した場合、振幅変調された帯域雑音に埋もれた純音は帯域幅を広げることにより、マスキング解除が起こるのに対し、ランダム帯域雑音に埋もれた純音は帯域幅の広がりに関わらずマスキング解除が起こらないという現象が見られた。この結果は、CMR の工学的モデルと解釈できる。

本論文では、入力位相を求めるために、Bregman によって提唱された四つの発見的規則のうち (2) 漸近的な変化と (4) 一つの音響事象に生じる変化に関する規則を物理的制約条件として取り入れている。今後は、入力位相  $\theta_{1k}(t)$ ,  $\theta_{2k}(t)$  を求めることが課題となるが、これを求めるために残り二つの規則 ((1) 共通の立上りと立下りと (3) 調波関係) 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になるものと考えられる。

**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) によって行われたものである。

## 文 献

- [1] A.S. Bregman, Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, Cambridge, Mass. 1990.
- [2] A.S. Bregman, "Auditory Scene Analysis: hearing in complex environments," in Thinking in Sounds, ed. S. McAdams and E. Bigand, pp.10-36, Oxford University Press, New York, 1993.
- [3] 河原英紀, "音声コミュニケーションにおける聴覚的情景解析," 音響学会秋季講論, 2-7-13, 1994.
- [4] 赤木正人, "カクテルパーティー効果とそのモデル化," 信学誌, vol.78, no.5, pp.450-453, May 1995.
- [5] D. Marr, "Vision," Freeman, 1992 (邦訳, 乾 敏郎, 安藤広志, "ビジョン," 産業図書, 1992.).
- [6] 河原英紀, "聴覚の計算理論の構築に向けて," 音響学会聴覚研資, H-94-63, Nov. 1994.
- [7] 入野俊夫, Roy D. Patterson, "音響事象検出・強調の計算理論," 音響学会聴覚研資, H-94-64, Nov. 1994.
- [8] 入野俊夫, "聴覚末梢系の計算理論," 信学技報, SP95-40, July 1995.
- [9] M.P. Cooke, "Modelling Auditory Processing and Organization," Ph.D. Thesis, University of Sheffield, 1991 (Cambridge University Press, Cambridge, 1993).
- [10] G.J. Brown, "Computational Auditory Scene Analysis: A Representational Approach," Ph.D. Thesis, Dep. Comput. Sci. University of Sheffield, 1992.
- [11] M.P. Cooke and G.J. Brown, "Computational auditory scene analysis: Exploiting principles of perceived continuity," Speech Communication, pp.391-399, North Holland, 13, Dec. 1993.
- [12] D.P.W. Ellis, "A computer implementation of psychoacoustic grouping rules," Proc. 12th Int. Conf. on Pattern Recognition, Oct. 1994.
- [13] 中谷智広, 川端 豪, 奥乃 博, "計算論的アプローチによる音響ストリームの分離," 音響学会聴覚研資, H-93-83, Dec. 1993.
- [14] T. Nakatani, H.G. Okuno, and T. Kawabata, "Unified architecture for auditory scene analysis and spoken language processing," ICSLP'94, vol.24, no.3, Sept. 1994.
- [15] 柏野邦夫, 田中英彦, "二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), vol.J77-A, no.5, pp.731-740, May 1994.
- [16] 柏野邦夫, "計算機による聴覚的情景解析-はじめの一步," 音響誌, vol.50, no.12, pp.1023-1028, Dec. 1994.
- [17] 井川博行, 赤木正人, "位相の変化に着目した 2 波形分離法," 音響学会春季講論, 3-4-15, March 1994.
- [18] 鶴木祐史, 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," 音響学会聴覚研資, H-95-79, Nov. 1995.
- [19] 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.
- [20] C.K. Chui, An Introduction to Wavelets, Academic Press, Boston, MA, 1992.
- [21] 河原英紀, "ウェーブレット解析の聴覚研究への応用," 音響誌, vol.47, no.6, pp.424-429, June 1991.
- [22] Roy D. Patterson and John Holdsworth, A Functional Model of Neural Activity Patterns and Auditory Images, Advances in speech, Hearing and Language Processing, vol.3, JAI Press, London, 1991.
- [23] 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, Sept. 1994.
- [24] 鶴木祐史, 赤木正人, "共変調マスキング解除の計算機モデルに関する一考察," 信学技報, SP96-37, July 1996.

## 付 録

## 1. 補題 1 の証明

式 (12) の wavelet 変換は、式 (4) の分析フィルタの出力  $X_k(t)$  を複素数表現したものである。

$$X_k(t) = S_k(t)e^{j(\omega_k t + \phi_k(t))}$$

$$:= \tilde{f}(a, b), \quad a = \alpha^{k - \frac{K}{2}}, b = t \quad (\text{A}\cdot 1)$$

従って、両辺の絶対値を取ると

$$|X_k(t)| = S_k(t) = |\tilde{f}(\alpha^{k - \frac{K}{2}}, t)| \quad (\text{A}\cdot 2)$$

を得る。

同様に、式 (A・1) と式 (12) の位相項を比較すると

$$\omega_k t + \phi_k(t) = \arg(\tilde{f}(a, b)) \quad (\text{A}\cdot 3)$$

を得る。ここで、位相スペクトル  $\arg(\tilde{f}(a, b))$  は、

$$\arg(\tilde{f}(a, b)) = \tan^{-1} \frac{\text{Im}\{\tilde{f}(a, b)\}}{\text{Re}\{\tilde{f}(a, b)\}} \quad (\text{A}\cdot 4)$$

と表されるため、 $-\pi \leq \arg(\tilde{f}(a, b)) \leq \pi$  の区間ランブ関数となることがわかる。そこで、式 (A・3) の両辺を微分すると、

$$\omega_k + \frac{d\phi_k(t)}{dt} = \frac{\partial}{\partial t} \arg(\tilde{f}(\alpha^{k - \frac{K}{2}}, t))$$

となり、これを整理すると、

$$\frac{d\phi_k(t)}{dt} = \frac{\partial}{\partial t} \arg(\tilde{f}(\alpha^{k - \frac{K}{2}}, t)) - \omega_k$$

を得る。故に、出力位相  $\phi_k(t)$  は

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

となる。 □

(平成 8 年 6 月 11 日受付, 9 月 24 日再受付)



鷗木 祐史 (学生員)

平 6 能開大・情報卒。平 8 北陸先端科学技術大学院大学情報科学研究科博士前期課程了。現在、同大学院博士後期課程在学中。音響情報に対する信号処理および聴覚のモデル化の研究に従事。日本音響学会会員。



赤木 正人 (正員)

昭 54 名工大・工・電子卒。昭 59 東工大大学院博士課程情報工学専攻了。工博。同年電電公社研究所入社。以来、ATR 視聴覚機構研究所、NTT 基礎研究所を経て、現在、北陸先端科学技術大学院大学情報科学研究科助教授。この間、昭 63 米国 MIT 客員研究員。平 5 オランダアムステルダム大学客員研究員。音声信号処理、聴覚のモデル化の研究に従事。昭 62 年度本会論文賞受賞。IEEE, ASA, ESCA, 日本音響学会各会員。