帯域雑音中の AM 調波複合音の一抽出法

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 1-1

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

あらまし　本論文では、帯域雑音中の AM 調波複合音を分離抽出する方法を提案する。著者らはこれまでに、Bregman によって提唱された四つの発見的規則のうちの二つ (漸近的変化、一つの音響事象で生じる変化) を利用し、帯域雑音中に純音が混入した二波形分離問題の解法を提案した。本論文で新たに提案する方法は、著者らによって提案された解法に、残りの二つの発見的規則 (立上り・立下り、調波関係) を加えて発展させたものである。この方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った。この結果、AM 調波複合音と帯域雑音が同一周波数領域に存在しても、正確に AM 調波複合音の分離抽出が可能であることが示された。特に、SD 値で平均約 20dB の改善 (雑音除去) が得られた。

キーワード　聴覚の情景解析、波形分離、カルマンフィルタ、スプライン補間

An Extraction Method of the AM complex tone from Noise-added AM complex tone

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

Abstract　This paper presents an extraction method of AM complex tone from noise-added AM complex tone. We have proposed a method for solving the problem of segregating a sinusoidal signal from noise-added signal, using physical constraints. These constraints are two of the four psychoacoustically heuristic regularities: gradualness of change and changes taken in an acoustic event, proposed by Bregman. The proposed method also uses grouping constraints which are the two of the remain regularities: onset/offset and harmonicity, proposed by Bregman. The simulations were carried out using mixed signals of AM complex tone and a random bandpassed noise. The average of reduced SD is about 20dB. As a result, the proposed method can extract an AM complex tone from noise-added AM complex tone in which signal and noise exist in the same frequency region.

Key words　Auditory Scene Analysis, Segregation, Kalman filter, Spline Interpolation

1　はじめに

近年、聴覚の情景解析 (Auditory Scene Analysis:ASA) という言葉が Bregman の著書 [1] により広く知られるようになった。ASA とは、音を通じて環境を把握することである。我々が日常経験する環境では、話声や雑音、残響など様々な音が混在するが、聴覚はいともたやすく情景解析問題を解いているように思われる。しかし、同じ環境で受音 (片耳あるいは両耳) された音響信号から工学的に情景解析問題を解くことは、音源や環境に関する制約が無い限り一意な解を得ることができない問題である。

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

としてまとめている [2]。

著者らは、これらの発見的規則を物理的制約条件としてとらえ直すことにより、計算論的な聴覚の情景解析の問題を解くことが可能であると考えている。もし、ASA の問題を解くための第一歩として、必要な音だけを選択し、他の音を除外するというような音源分離問題を解くことができれば、カクテルパーティー効果のモデル化やロバストな音声認識システムの実現が期待できる [3]。また、共変調マスキング解除といった様々な聴覚的現象の工学的モデル化も可能になると考えられる。

一方、四つの発見的規則のいくつかを利用した計算論的な音源分離モデルとして、ボトムアップ処理に基づくモデル [4] やトップダウン処理に基づくモデル [5, 6] がある。これらは、いずれも規則 (1) と (3) を利用したものである。また、柏野ら [7] は、二つの周波数成分の分離知覚に関して、規則 (1) と (3) の定量的関係をモデル化している。しかし、これらのモデルでは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。また、分離抽出したい信号を妨害する雑音等のパワーが増大すれば、これらのモデルのもつ分離精度が低下するものと考えられる。

これに対し、著者らは、複数の音源が同時に存在する状況から特定の音源を分離抽出する音源分離問題を解くことを念頭におき、基本的な問題として二波形分離問題に取り組んでいる [10]。これまでに、著者らは、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 [8]、Bregman のいう規則 (2) と (4) を利用した二波形分離問題の解法を提案した [9, 10]。この解法により、帯域雑音に埋もれた純音の分離抽出が可能になった [10] ほか、提案した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の計算モデルになるという結果も得られた [9]。

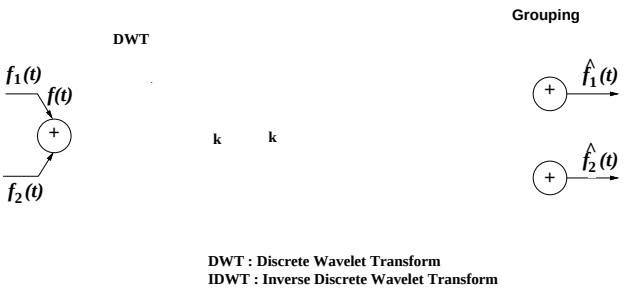

図 1: モデル構成

本論文では、著者らによって提案された解法 [8, 9, 10] に、残りの発見的規則 (1) と (3) を加えて発展させ、帯域雑音が付加された AM 調波複合音から AM 調波複合音を分離抽出する方法を新たに提案する。

2　モデル構成と定式化

本方法のモデルは、図 1 のように (a) 分析フィルタ群、(b) 波形分離、(c) グルーピングの 3 層で構成される。分析フィルタ群は、Gammatone filter を基底関数とした wavelet 分析合成系で構成される。波形分離では、発見的規則 (2) と (4) を考慮した物理的制約条件を利用し、グルーピングでは、発見的規則 (1) と (3) を考慮した物理的制約条件を利用する。また、グルーピングにおける信号の再構成は、wavelet 合成系により行われる。このモデルの波形分離部は、次に示す二波形分離問題の定式化に従っている。

2.1　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 $f(t) = f_1(t) + f_2(t)$ 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

ため、k 番目の分析フィルタの出力 $X_k(t)$ は、

$$X_k(t) = S_k(t)\sin(\omega_k t + \phi_k(t)) \quad (1)$$

と表される。これは、$f_1(t)$ と $f_2(t)$ に対する k 番目の分析フィルタの出力が、それぞれ

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

と仮定できることから、式 (2) と式 (3) の和になっている。ここで、ω_k は分析フィルタの中心角周波数、$\theta_{1k}(t)$ と $\theta_{2k}(t)$ はそれぞれ $f_1(t)$ と $f_2(t)$ のもつ入力位相であり、振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ

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

$$\phi_k(t) = \tan^{-1}\left(\frac{A_k(t)\sin\theta_{1k}(t) + B_k(t)\sin\theta_{2k}(t)}{A_k(t)\cos\theta_{1k}(t) + B_k(t)\cos\theta_{2k}(t)}\right) \quad (5)$$

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、四つの物理パラメータ ($S_k(t)$, $\phi_k(t)$, $\theta_{1k}(t)$, $\theta_{2k}(t)$) がわかれば、二波形の振幅包絡 $A_k(t)$ と $B_k(t)$ を

$$A_k(t) = \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)} \quad (6)$$
$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)} \quad (7)$$

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の出力 ($X_k(t)$,$1 \le k \le K$) に対し、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、これより得られた二波形の振幅包絡 $A_k(t)$ と $B_k(t)$ を式 (2) と式 (3) を用いて合成することで、wavelet 合成系より $f_1(t)$ 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、入力位相を $\theta_{1k}(t) = 0$、$\theta_k(t) = \theta_{2k}(t)$、$f_1(t)$ を振幅変化音に限定している。また、$f_2(t)$ 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、分離抽出したい信号を $f_1(t)$ とした二波形分離問題を議論する。

2.2 wavelet 分析合成系

本研究では、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関数として Gammatone filter を基本 wavelet：

$$\psi(t) = At^{N-1}e^{j2\pi f_0 t - 2\pi b_f t} \quad (8)$$

と定義し、wavelet 変換対：

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

$$f(t) = \frac{1}{D_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(a,b)\psi\left(\frac{t-b}{a}\right)\frac{dadb}{a^2} \quad (10)$$

を用いて分析合成系を構築している [10]。

本論文では、中心周波数 $f_0 = 600$Hz、通過帯域が 60 ~ 6000Hz、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 $a = \alpha^p, -\frac{K}{2} \le p \le \frac{K}{2}, \alpha = 10^{2/K}$、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 [10]。ここで、各フィルタの矩形帯域幅は重複せず、完全に通過帯域を被覆している [10]。

3 物理パラメータの導出方法

3.1 $S_k(t)$ と $\phi_k(t)$ の計算方法

式 (4) の振幅包絡 $S_k(t)$ と式 (5) 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

$$S_k(t) = |\tilde{f}(\alpha^{k-\frac{K}{2}}, t)| \quad (11)$$

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

$$\phi_k(t) = \int\left(\frac{d}{dt}\arg\left(\tilde{f}(\alpha^{k-\frac{K}{2}}, t)\right) - \omega_k\right)dt \quad (12)$$

で求めることができる。

(証明)　文献 [10] 付録参照。 ■

3.2 $\theta_k(t)$ の計算方法

本論文では、入力位相を $\theta_{1k}(t) = 0$ に限定しているため、もう 1 つの入力位相 $\theta_{2k}(t)$ については入力位相差 $\theta_k(t) = \theta_{2k}(t)$ を求めればよいことになる。これは、Bregman によって提唱された発見的規則 (2) と (4) に関係した三つの物理的制約条件を用いることで得られる。

はじめに、規則 (2) を利用する。これは、“一つの音の振幅包絡はゆっくりとなめらかに変化する” という定性的条件 [2] である。この規則を振幅包絡 $A_k(t)$ に適用するため、次のような物理的制約条件にとらえ直す。

物理的制約条件 **1** (漸近的変化) 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：

$$\frac{dA_k(t)}{dt} = C_{k,R}(t) \tag{13}$$

■

この制約条件を式 (6) に適用することで、一階線形微分方程式が得られ、これを解くことで入力位相 $\theta_k(t)$ の一般解が得れる。

補題 **2** 入力位相 $\theta_k(t)$ の一般解は、

$$\theta_k(t) = \arctan\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C_k(t)}\right) \tag{14}$$

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。 ■

$C_k(t)$ を決定できれば、式 (14) から一意に $\theta_k(t)$ を求めることができる。また、未定関数 $C_k(t)$ は、最適化問題として $C_k(t)$ の各係数 $C_{k,r}, r = 0, 1, \cdots, R$ を推定することで求めることができる。今までは、推定の計算量を抑えるために未定関数を定数項として扱ってきたが ($C_k(t) = C_{k,0}$) [10]、より精度の高い分離を行うためにはこの未定関数 $C_k(t)$ を正確に求める必要がある。そこで、本論文では $C_k(t)$ をカルマンフィルタにより推定する。

3.2.1 カルマンフィルタによる $C_k(t)$ の推定

はじめに、未定関数 $C_k(t)$ の推定問題をカルマンフィルタで解くために問題設定を行う。

式 (9) の wavelet 変換は、式 (1) 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。

$$\begin{aligned} X_k(t) &= S_k(t)e^{j(\omega_k t + \phi_k(t))} \\ &:= \tilde{f}(a,b), \quad a = \alpha^{k-\frac{K}{2}}, b = t_m \end{aligned} \tag{15}$$

但し、$t_m = m/f_s, m = 0, 1, \cdots, M$ である。また、式 (2) と式 (3) に着目すれば、これは $f_1(t)$ と $f_2(t)$ を wavlet 変換したものの和に等しいことがわかる。

$$\begin{aligned} \tilde{f}(\alpha^{k-\frac{K}{2}}, t_m) &= \tilde{f}_1(\alpha^{k-\frac{K}{2}}, t_m) \\ &\quad + \tilde{f}_2(\alpha^{k-\frac{K}{2}}, t_m) \end{aligned} \tag{16}$$

$$\tilde{f}_1(\alpha^{k-\frac{K}{2}}, t_m) = A_k(t_m)e^{(j\omega_k t_m + \theta_{1k}(t))} \tag{17}$$

$$\tilde{f}_2(\alpha^{k-\frac{K}{2}}, t_m) = B_k(t_m)e^{(j\omega_k t_m + \theta_{2k}(t))} \tag{18}$$

一方、式 (13) を積分した結果と式 (14) の $C_k(t)$ から、

$$C_k(t) = -A_k(t) \tag{19}$$

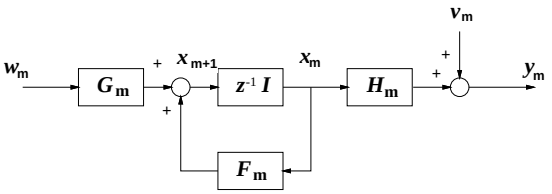

図 2: Kalman filter の基本システム

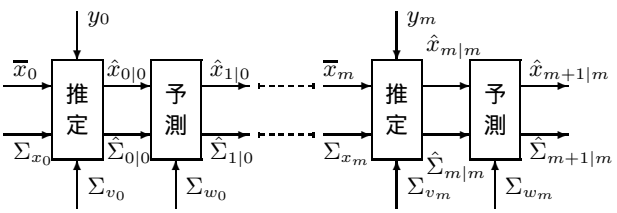

図 3: Kalman フィルタの適用手順

という関係があることがわかる。そこで、未定関数の離散時間における変化が

$$C_k(t_{m+1}) = C_k(t_m)\Delta C_k + w_m \quad (20)$$

$$\Delta C_k = 1 + \frac{C_k(t_m) - C_k(t_{m-1})}{C_k(t_m) \cdot f_s} \quad (21)$$

であるものと考える。つまり、$C_k(t_{m+1})$ は $C_k(t_m)$ の ΔC_k 倍であり、その誤差が平均 0、分散 σ_w の白色雑音 w_m で表されるものとする。

ここで、未定関数の推定問題とは、式 (16) で観測される情報 ($X_k(t)$) から式 (20) の未定関数 $C_k(t)$ を推定することである。

次に、上記の問題にカルマンフィルタを適用するためには、まず式 (16) を状態方程式と観測方程式からなる確率システムで表す必要がある。ここで、観測信号を $\mathbf{y}_m = \tilde{f}(\alpha^{k-\frac{k}{2}}, t_m)$、状態変数を $\mathbf{x}_m = -C_k(t)$、観測雑音を $\mathbf{v}_m = \tilde{f}_2(\alpha^{k-\frac{k}{2}}, t_m)$、システム雑音を $\mathbf{w}_m = w_m$ と考えれば、式 (20) と式 (16) は次のような複素確率システムで表すことができる。

$$\mathbf{x}_{m+1} = \mathbf{F}_m\mathbf{x}_m + \mathbf{G}_m\mathbf{w}_m \quad (\text{状態方程式}) \quad (22)$$

$$\mathbf{y}_m = \mathbf{H}_m\mathbf{x}_m + \mathbf{v}_m \quad (\text{観測方程式}) \quad (23)$$

但し、状態遷移行列 $\mathbf{F}_m = \Delta C_k$、観測行列 $\mathbf{H}_m = e^{j\omega_k t_m}$、駆動行列 $\mathbf{G}_m = -1$ である。これは、基本システムと呼ばれ [11]、図 2のように表される。また、この問題に適用される複素カルマンフィルタは次式で表され、図 3のような手順で適用される [11]。

(a) フィルタ方程式

$$\hat{\mathbf{x}}_{m|m} = \hat{\mathbf{x}}_{m|m-1} + \mathbf{K}_m(\mathbf{y}_m - \mathbf{H}_m\hat{\mathbf{x}}_{m|m-1}) \quad (24)$$

$$\hat{\mathbf{x}}_{m+1|m} = \mathbf{F}_m\hat{\mathbf{x}}_{m|m} \quad (25)$$

(b) カルマン (フィルタ) ゲイン

$$\mathbf{K}_m = \frac{\hat{\mathbf{\Sigma}}_{m|m-1}\mathbf{H}_m^{*T}}{\mathbf{H}_m\hat{\mathbf{\Sigma}}_{m|m-1}\mathbf{H}_m^{*T} + \mathbf{\Sigma}_{v_m}} \tag{26}$$

(c) 誤差の共分散行列方程式

$$\hat{\mathbf{\Sigma}}_{m|m} = \hat{\mathbf{\Sigma}}_{m|m-1} - \mathbf{K}_m\mathbf{H}_m\hat{\mathbf{\Sigma}}_{m|m-1} \tag{27}$$
$$\hat{\mathbf{\Sigma}}_{m+1|m} = \hat{\mathbf{F}}_m\mathbf{\Sigma}_{m|m}\mathbf{F}_m^{*T} + \mathbf{G}_m\mathbf{\Sigma}_{w_m}\mathbf{G}_m^{*T} \tag{28}$$

但し、初期値は、$\hat{\mathbf{x}}_{0|-1} = 0$、$\hat{\mathbf{\Sigma}}_{0|-1} = S_k(t_0)$、$\hat{\mathbf{\Sigma}}_{w_m} = 0.01$、である。また、$\hat{\mathbf{\Sigma}}_{v_m}$ は、$\tilde{f}_2(\alpha^{k-\frac{K}{2}}, t_m)$ の分散であるが、本論文では $f_2(t)$ が存在している状態で $f_1(t)$ が加算される状況を想定していることから、$f_2(t)$ のみが存在している区間での $X_k(t_m)$ の分散から知ることができる。

この方法により、最小分散推定値 $\hat{C}_k(t)$ と推定誤差 $P_k(t)$ はそれぞれ

$$\hat{C}_k(t) = -|\hat{\mathbf{x}}_{m|m}| \tag{29}$$
$$P_k(t) = |\hat{\mathbf{\Sigma}}_{m|m}| \tag{30}$$

で得られる。この推定値を用いることで、一意に入力位相を決定できるが、制約条件 1 では $A_k(t)$ の時間変化に関してだけ制約を与えていたため、残りの $A_k(t)$ の “なめらかさ” を定義する必要がある。

3.2.2 スプライン補間によるなめらかさの定義

ある未定関数 $C_k(t)$ によって得られた $f_1(t)$ の振幅包絡を $\hat{A}_k(t)$ とし、$t_1, t_2, \cdots, t_i$ は開区間 (t_a, t_b) に含まれ、$t_a < t_1 < \cdots < t_i < t_b$ とする。また、各時刻 t_i における振幅包絡値を $\hat{A}_{k,i} := \hat{A}_k(t_i)$ とする。ここで、$A_k(t_i) = \hat{A}_{k,i}, i = 1, 2, \cdots, I$ となる I 個の点を通る最もなめらかな補間関数 $A_k(t)$ を求めることは、閉区間 $[t_a, t_b]$ で定義される $A_k(t)$ が r 階微分可能であり、積分 $\sigma = \int_{t_a}^{t_b}[A_k^{(r)}(t)]^2dt$ が可能な限り小さい補間関数を求めることに等しい [12]。

そこで、$A_k(t)$ の “なめらかさ” を表現するために、規則 (2) を次のような物理的制約条件にとらえ直す。

物理的制約条件 **2** (連続性 (なめらかさ))　区間 $[t_a, t_b]$ で定義され、物理的制約条件 1 を満たす振幅包絡 $A_k(t)$ は、最もなめらかである：

$$\sigma = \int_{t_a}^{t_b}[A_k^{(R+1)}(t)]^2dt \quad \Rightarrow \min \tag{31}$$

■

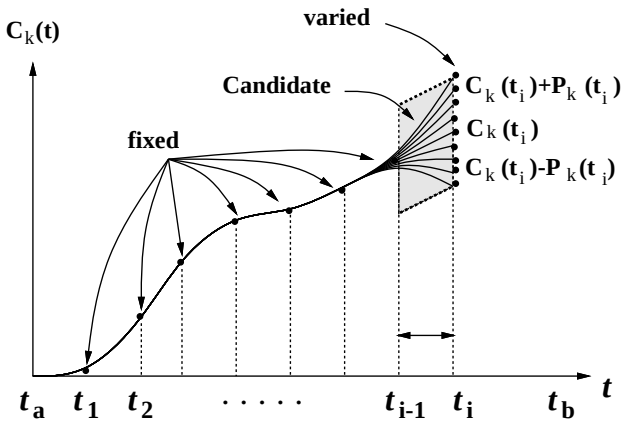

図 4: スプライン補間された $C_k(t)$ 候補

この制約での最良補間関数は、$2R+1$ 次スプライン関数であり、唯一存在する [12]。ここで、式 (6) と式 (14) から $A_k(t)$ が $C_k(t)$ の関数であることに着目すれば、物理的制約条件 2 は、真の $A_k(t)$ の解が存在すると考えられる推定誤差範囲：

$$\hat{C}_k(t) - P_k(t) \leq C_k(t) \leq \hat{C}_k(t) + P_k(t) \qquad (32)$$

においてスプライン補間された $C_k(t)$ を求めることと解釈できる。従って、推定誤差範囲内でスプライン補間された $C_k(t)$ の各候補を求め、その候補から正しい解を 1 つ求めることで、最もなめらかな振幅包絡 $A_k(t)$ の真の解を一意に求めることができる。時刻 t_i における $C_k(t)$ のスプライン補間の方法を図 4に示す。既に得られた時刻 $t_1, \cdots, t_{i-1}$ までの $C_k(t_1), \cdots, C_k(t_{i-1})$ を固定し、$C_k(t_i)$ を $C_k(t_i) \leq |\hat{C}_k(t_i) + P_k(t_i)|$ の範囲で変動させてスプライン補間し、各候補を求める。ここで補間間隔を $\Delta t = 15/f_0$ とした。

3.2.3 相関を手がかりにした $C_k(t)$ の決定

最後に、スプライン補間された $C_k(t)$ の候補を絞り込むために規則 (4) を利用する。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” という定性的条件 [2] であり、この規則を利用するために次のような物理的制約条件にとらえ直す。

物理的制約条件 **3** (一つの音響事象に生じる変化) 振幅包絡 $A_k(t)$ は隣接する聴覚フィルタにおける振幅包絡 $A_\ell(t)$ に強い相関がなければならない：

$$A_k(t) \approx A_\ell(t), \quad k \neq \ell \qquad (33)$$

■

このとき、制約条件 3 は振幅包絡 $A_k(t)$ 間の相関を最大とする $C_k(t)$ を最適値として選ぶことと解釈できる。これは、

$$\max_{\hat{C}_k - P_k \le C_k \le \hat{C}_k + P_k} \frac{< \hat{A}_k, \hat{\hat{A}}_k >}{||\hat{A}_k|| ||\hat{\hat{A}}_k||} \tag{34}$$

なる $C_k(t)$ を見つけ出すことで実現でき、式 (14) によって一意に $\theta_k(t)$ を決定できる。ここで、$\hat{A}_k(t)$ はスプライン補間された $C_k(t)$ により得られた振幅包絡である。$\hat{\hat{A}}_k(t)$ については次節で説明する。

4 二波形分離とグルーピング

ここでは、二つの信号波形が同時に存在する $X_k(t)$ に対し、グルーピングの制約条件を用いることで二波形分離問題の解法を適用する。これは、基本周波数 F_0 の倍音成分が存在する $X_k(t)$ か、または立上りと立下りが同期する $X_k(t)$ に対して二波形分離問題の解法を適用するものであり、次のような二つの物理的制約条件が適用される。

一つ目の規則として、規則 (3) を利用する。これは、“物が繰り返し振動するときには、共通の基本周波数の整数倍の音響的成分が発生する” という定性的条件 [2] である。この規則を利用するために、次のような物理的制約条件にとらえ直す。

物理的制約条件 **4** (調波関係)　基本周波数を F_0、高調波の次数を N_{F_0} としたとき、調波関係にある信号成分が存在する分析フィルタの番号 ℓ は

$$\ell = \frac{K}{2} - \left\lceil \frac{\log(n \cdot F_0/f_0)}{\log \alpha} \right\rceil \pm 1, \quad n = 1, 2, \cdots, N_{F_0} \tag{35}$$

でなければならない。 ■

もう一つの規則として、規則 (1) を利用する。これは、“関連の無い音が一緒に始まったり終ったりすることはない” という定性的条件 [2] である。この規則を利用するために、次のような物理的制約条件にとらえ直す。

物理的制約条件 **5** (立上り・立下りの同期)

ある音源で生じた信号 $f_1(t)$ の立上りを $T_{\rm S}$、立下りを $T_{\rm E}$ としたとき、同じ音源で生じた信号成分であれば、各分析フィルタにおける $f_1(t)$ の立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ は

$$|T_{\rm S} - T_{k,\rm on}| \le 50ms, \quad |T_{\rm E} - T_{k,\rm off}| \le 100ms \tag{36}$$

を満たさなければならない。 ■

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 , $S_k(t)$ と $\phi_k(t)$ を求める;
　立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ を求める;
　分離区間を $T_{k,\rm on} \le t \le T_{k,\rm off}$ とする;
　if 制約条件 4 or 5 を満たす **then**
　　$C_k(t)$ をカルマンフィルタにより推定する;
　　I 個の補間区間 ($\Delta t = 15/f_0$) を求める;
　　for $i = 1$ **to** I **do**
　　$\hat{C}_k(t_i) - P_k(t_i) \le C_k(t_i) \le \hat{C}_k(t_i) + P_k(t_i)$ でスプライン補間された未定関数 $C_k(t)$ の候補を求める (図 4) ;
　　式 (14) から入力位相 $\hat{\theta}_k(t)$ を求める;
　　式 (6) から振幅包絡 $\hat{A}_k(t)$ を求める;
　　式 (37) から振幅包絡 $\hat{\hat{A}}_k(t)$ を求める;
　　式 (34) から $\mathrm{Corr}(\hat{A}_k(t), \hat{\hat{A}}_k(t))$ を求める;
　　end
　　$\mathrm{Corr}(\hat{A}_k(t), \hat{\hat{A}}_k(t))$ が最大になる $C_k(t)$ を求める;
　　式 (14) から $\theta_k(t)$ を求める;
　else
　　$A_k(t) = 0$, $B_k(t) = S_k(t)$, $\theta_k(t) = \phi_k(t)$;
　end
　式 (6) と式 (7) から $A_k(t)$ と $B_k(t)$ を求める;
　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

k 番目の分析フィルタにおける $f_1(t)$ の立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ は、それぞれ

(a) 立上り時刻 $T_{k,\rm on}$: $dS_k(t)/dt$ の極大点近傍 (± 0.25s) にある $|d\phi_k(t)/dt|$ の極大点

(b) 立下り時刻 $T_{k,\rm off}$: $dS_k(t)/dt$ の極小点近傍 (± 0.25s) にある $|d\phi_k(t)/dt|$ の極大点

で求めることができる。また、$T_{\rm S}$ と $T_{\rm E}$ は F_0 の周波数成分が存在する $X_k(t)$ での立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ であらかじめ知ることができる。

ここで、式 (34) における $\hat{\hat{A}}_k(t)$ は、$\hat{A}_\ell(t)$ の加算平均とする。

$$\hat{\hat{A}}_k(t) = \frac{1}{N_{F_0}} \sum_{\ell \in \mathbf{L}} \frac{\hat{A}_\ell(t)}{\|\hat{A}_\ell(t)\|} \tag{37}$$

但し、$\mathbf{L}$ は式 (35) を満たす ℓ の集合である。

以上、四つの発見的規則を利用した二波形分離問題の解法は、図 5に示すアルゴリズムでまとめられる。

5 二波形分離のシミュレーション

ここでは、前節で述べた波形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。

5.1 実験データ

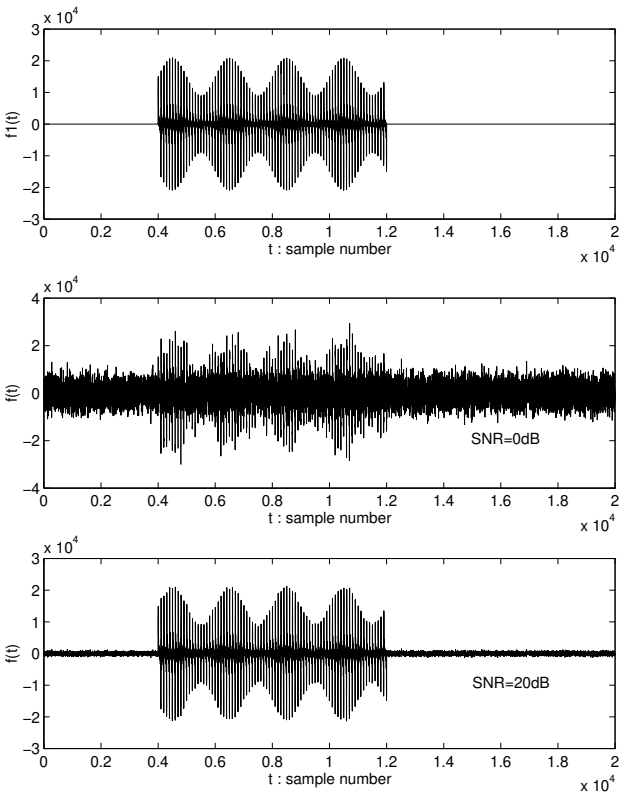

図 6: AM 調波複合音 $f_1(t)$ と混合信号 $f(t)$

実験データは、AM 調波複合音 $f_1(t)$ とこれに SNR= -10dB ~ 20dB (5dB 刻み) のランダム帯域雑音 $f_2(t)$ を加算した合計 7 種類の混合信号 $f(t)$ を用意する。但し、AM 調波複合音は、基本周波数 F_0 を 200Hz 一定、高調波の次数を $N_{F_0} = 10$、振幅包絡を 10Hz の正弦波とし、ランダム帯域雑音の帯域幅は 60 ~ 6000Hz とした。図 6に AM 調波複合音と SNR が 0dB と 20dB の混合信号 $f(t)$ を示す。

5.2 分離精度の評価

ここでは、分離精度の評価尺度を二種類用意した。

一つは、信号と雑音が同一周波数領域において正確に分離できるかを調べるために、振幅包絡 $A_k(t)$ に関するパワー比を用いた。これは一種の SNR であり、この評価尺度を次式で定義される Precision と呼ぶことにする。

$$\text{Precision} := 10\log_{10}\frac{\int_0^T A_k^2(t)dt}{\int_0^T (A_k(t)-\hat{A}_k(t))^2 dt} \quad (38)$$

ここで、$A_k(t)$ は $f_1(t)$ のみを分析フィルタ群に展開したときの振幅包絡であり、$\hat{A}_k(t)$ は本方法によって分離抽出された $\hat{f}_1(t)$ の振幅包絡である。

もう一つは、分離抽出された $\hat{f}_1(t)$ がどの程度、雑音除去されたかを定量的に調べるため、次式で定義される SD (Spectrum Distortion) を用いた。

$$\text{SD} := \sqrt{\frac{1}{N_{\text{SD}}}\sum_{n=1}^{N_{\text{SD}}}\left(20\log_{10}\frac{\tilde{F}_1(\omega_n)}{\tilde{\hat{F}}_1(\omega_n)}\right)^2} \quad (39)$$

ここで、$\tilde{F}_1(\omega_n)$ と $\tilde{\hat{F}}_1(\omega_n)$ は、フレーム長を 51.2ms、フレームシフトを 25.6ms、$N_{\text{SD}} = 256$、窓関数を Hamming とした $f_1(t)$ と $\hat{f}_1(t)$ の振幅スペクトルである。

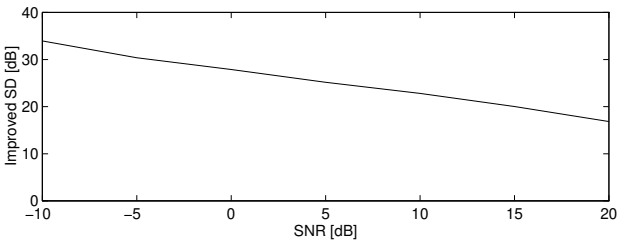

図 7: SD 改善量

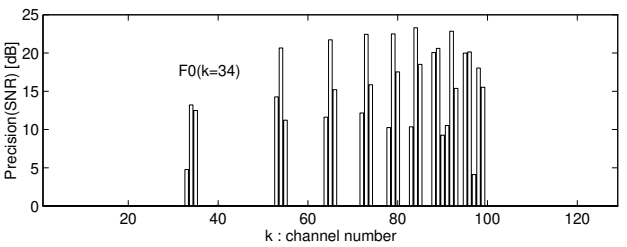

図 8: 分離された $A_k(t)$ の分離精度 (SNR= 20dB)

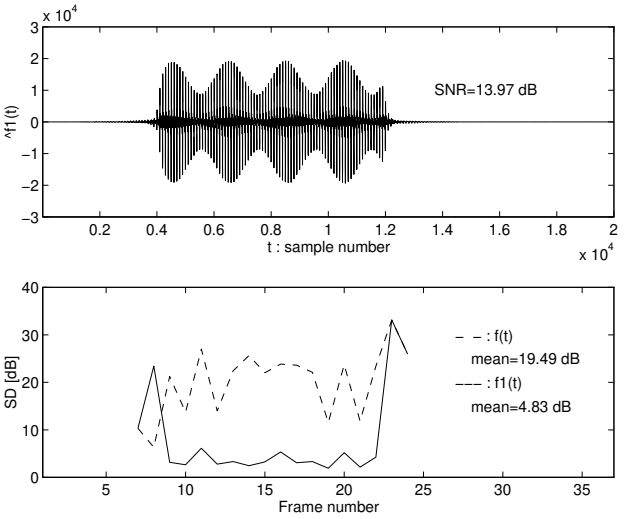

図 9: $\hat{f}_1(t)$ と分離精度 (SNR= 20dB)

5.3 分離結果と考察

各混合信号 $f(t)$ に対して二波形分離のシミュレーションを行った。この結果、$f(t)$ と $\hat{f}_1(t)$ の SD 値の差を改善量として評価すれば、図 7に示すよう

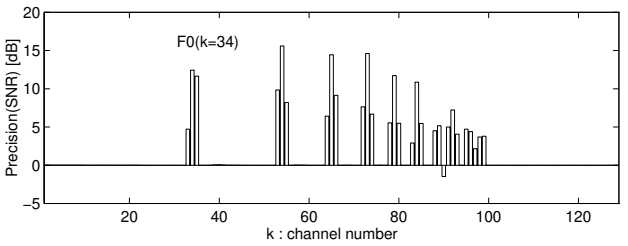

図 10: 分離された $A_k(t)$ の分離精度 (SNR= 0dB)

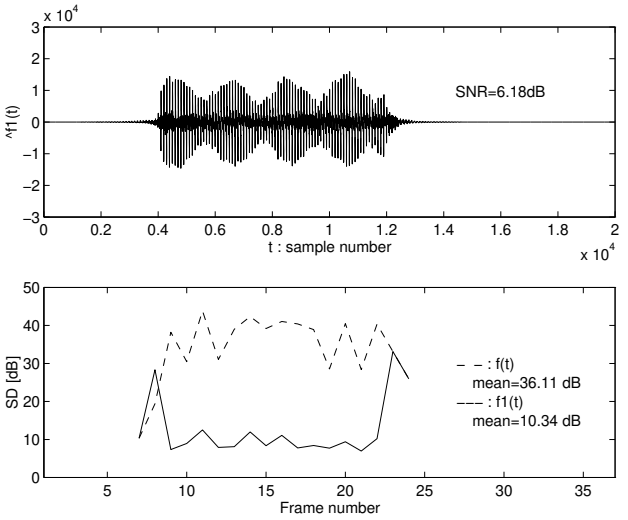

図 11: $\hat{f}_1(t)$ と分離精度 (SNR= 0dB)

に、本方法を利用することで約 20dB の改善 (雑音除去) が得らることがわかる。例えば、図 6に示す SNR= 20dB の混合信号 $f(t)$ に対して、図 8のように各分析フィルタの出力では $A_k(t)$ が高い精度で分離され、図 9に示すような $\hat{f}_1(t)$ に分離抽出される。また、AM 調波複合音が雑音に完全に埋もれてしまうような混合信号 $f(t)$ (図 6に示す SNR= 0dB) に対しても、図 10、図 11に示すようにある程度の精度で $\hat{f}_1(t)$ を分離抽出できる。尚、$\theta_{1k}(t) = 0$ の仮定により、$\hat{f}_1(t)$ の位相情報を完全に復元できないため、ここでは $\theta_{1k}(t) = \phi_k(t)$ として再合成した。

上記の結果より、本方法を用いることで、帯域雑音が付加された AM 調波複合音から AM 調波複合音の振幅情報を正確に分離抽出できることがわかる。

6 おわりに

Bregman によって提唱された四つの発見的規則を利用し、著者らが提案した二波形分離問題の解法を適用することで、帯域雑音が付加された AM 調波複合音を分離抽出する方法を提案した。この方法により、AM 調波複合音と帯域雑音が同一周波数領域に存在しても、正確に AM 調波複合音の分離抽出が可能になった。特に、本方法により SD 値で平均約 20dB の改善 (雑音除去) が得られた。

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、$\theta_k(t)$ を求めるときと同様に物理的制約条件を適用することで推定は可能であると考えられる。この課題がクリアされれば、分離抽出したい AM 調波複合音を波形レベルで完全に復元できるだけでなく、FM 音の分離抽出問題に拡張することができる。更に、雑音中の音声の分離抽出問題というような現実的な問題にも拡張できる。

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] 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, (Eds. S. McAdams and E.Bigand), pp. 10–36, Oxford University Press, New York, 1993.

[3] 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.

[4] 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.

[5] D.P.W.Ellis, “A Computer Implementation of Psycoacoustic Grouping Rules,” Proc. 12th Int. Conf. on Pattern Recognition, Jerusalem, Oct. 1994.

[6] 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.

[7] 柏野 邦夫, 田中 英彦, “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), vol.J77-A, no. 5, pp. 731-740, May 1994.

[8] 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,” 音響学会聴覚研資, H-95-79, Nov. 1995.

[9] 鵜木 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.

[10] 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一抽出法,” 信学論 (A) , vol.J80-A, no.3, March 1997.

[11] 中野 道雄, 西山 清, パソコンで解くカルマンフィルタ, 丸善, 1993.

[12] 桜井 明, スプライン補間入門, 東京電機大学出版局, 1981.