基本周波数の時間変動を考慮した調波複合音の抽出法

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

〒 923-1292　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　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を利用することで、AM 調波複合音を分離抽出する二波形分離問題の解法を提案した。しかし、この解法では、基本周波数が一定であり、その値が既知であるという仮定があった。本論文では、この解法に、河原によって提案された TEMPO による基本周波数の抽出と抽出された基本周波数の時間変化に制約を設けることで、雑音が付加された調波複合音から望みの調波複合音を分離抽出する方法を提案する。この方法の有効性を示すために、(a) 雑音が付加された AM 調波複合音、(b)AM 調波複合音同士、(c) 雑音が付加された合成母音の三つの信号を用いてシミュレーションを行なった。この結果、本モデルが、SD 値で約 15 dB 程度の雑音除去を可能とし、雑音が付加された調波複合音を分離抽出することができることが示された。

キーワード　聴覚の情景解析、二波形分離、グルーピング、基本周波数、TEMPO

An Extraction Method of the complex tone considering temporal variation of the fundamental frequency

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　The authors have proposed an extraction method of the desired AM complex tone from a noisy AM complex tone, using physical constraints which are related to the four regularities proposed by Bregman. However, in the method, it was assumed that the fundamental frequencies are constant and are known. This paper presents an extraction method of the complex tone considering temporal variation of the fundamental frequency. The method adopts TEMPO proposed by Kawahara for extracting fundamental frequencies. Three simulations were performed using the following signals: (a) a noise-added AM complex tone, (b) a mixed AM complex tones, and (c) noisy synthesized vowel. The results show that the proposed method can extract the desired complex tone from a noisy complex tone. Mean of the reduced SD was about 15 dB.

Key words　Auditory Scene Analysis, segregation, grouping, fundamental frequency, TEMPO

1　はじめに

我々が日常経験する環境では、話し声や雑音、残響など様々な音が混在するが、聴覚はこういった環境の中でいともたやすく目的の音を分離抽出している。これは聴覚の情景解析 (Auditory Scene Analysis: ASA) [1] と呼ばれる聴覚の処理能力の一つである。Bregman は、聴覚の情景解析の問題を解くために、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 ((i) 共通の立上り/立下り、(ii) 漸近的変化、(iii) 調波関係、(iv) 一つの音響事象に生じる変化) としてまとめている [2]。

ここで、先ほどと同じ環境で受音 (片耳あるいは両耳) された音響信号から工学的に同様の処理を実現することを考えてみる。これは様々な音が混在する信号からそれぞれの信号を求める一種の逆問題と考えることができる。それぞれの信号が時間的にあるいは周波数的に独立して存在したのであれば受音した信号からの分離は容易になると思われるが、日常の環境では時間的あるいは周波数的に重なって存在するため、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は難しい。

著者らは、Bregman がまとめた発見的規則を物理的制約条件としてとらえ直し、これを利用することで、計算論的に聴覚の情景解析の問題を解くことが可能であると考える。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雑音や残響に対してロバストな音声認識システムの実現が期待できる。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 [3, 4, 5] やトップダウン処理に基づくモデル [6, 7] がある。これらのモデルは、いずれも音響的特徴として振幅 (パワー) スペクトルを利用し、発見的規則 (i) と (iii) を用いて望みの信号成分を抽出している。そのため、これらのモデルでは、二つ (あるいはそれ以上) の信号が時間–周波数平面上に重って存在するとき、抽出された信号成分には他の信号成分が付加されたままであり、望みの信号を完全に分離抽出できているとは言い難い。

これに対し、著者らは、二つの信号成分が時間–周波数平面上で重って存在する場合、これらを完全に分離するためには振幅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をとる。そこで、

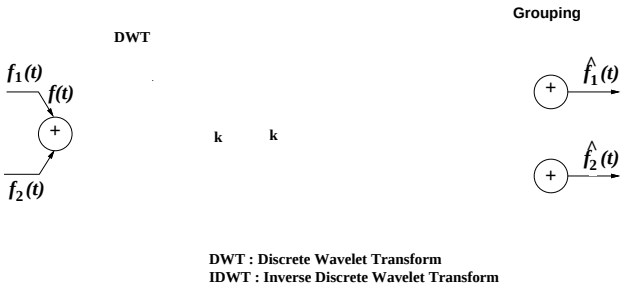

図 1: 二波形分離モデル

複数の音源で生じた信号が時間–周波数平面上に重なって存在する状況から特定の音源で生じた信号を分離抽出する音源分離問題を解くことを目標に、この基本的な問題として二波形分離問題に取り組む。

この研究の第一歩として、発見的規則 (ii) と (iv) を利用した二波形分離モデルを提案することで、雑音が付加された正弦波信号から望みの正弦波信号の分離抽出が可能になった [8]。また、この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]。

次の一歩として、先に提案した二波形分離モデルに、残りの発見的規則 (i) と (iii) を加えて発展させることで、雑音が付加された AM 調波複合音から望みの AM 調波複合音の分離抽出が可能になった [10]。しかし、この方法では、調波複合音の基本周波数が一定であり、かつその値が既知であるという仮定があったため、二波形分離の対象となる信号を実音声などに拡張する場合、この仮定が極めて大きな制約となっていた。

本論文では、TEMPO[11] を利用して基本周波数を抽出し、この基本周波数の時間変化に発見的規則 (ii) を利用した制約を設けることで、これらの仮定を取り除き、雑音が付加された調波複合音から望みの調波複合音を分離抽出する方法を提案する。

2 二波形分離モデル

二波形分離モデルは、図 1 に示すように (a) 分析フィルタ群、(b) 波形分離部、(c) グルーピング部の 3 ブロックで構成され、2.2 節で述べる二波形分離問題の定式化に従う。分析フィルタ群は、Gammatone filter を基底関数とした wavelet 分析系 (2.1 節参照) で構成され、混合信号を周波数成分に分解する。次に、波形分離部では、発見的規則 (ii) と (iv) を考慮した物理的制約条件を利用して、周波数分解された二波形の信号成分を分離する。最後に、グルーピング部では、発見的規則 (i) と (v) を考慮した物理的

制約条件を利用して、分離された信号成分をグループ化し、wavelet 合成系で信号を再構成する。

2.1 wavelet 分析合成系

本研究では、信号の分析合成において位相情報を利用するために、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として、基本 wavelet：

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

を定義し、wavelet 変換対：

$$\begin{aligned}\tilde{f}(a,b) &= \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \\ &= |\tilde{f}(a,b)|e^{j\arg(\tilde{f}(a,b))}\end{aligned} \tag{2}$$

$$f(t) = \frac{1}{D_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(a,b)\psi\left(\frac{t-b}{a}\right)\frac{dadb}{a^2} \tag{3}$$

を用いて分析フィルタ群を構築する [8]。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、$a, b \in \mathbf{R}$、$a \neq 0$ である。

本論文では、中心周波数が $f_0 = 600$ Hz、通過帯域が 60 ~ 6000 Hz、フィルタ数が $K = 128$ の分析フィルタ群を設計した。ここでは、サンプリング周波数 $f_s = 20$ kHz、スケールパラメータ $a = \alpha^p, -\frac{K}{2} \leq p \leq \frac{K}{2}, \alpha = 10^{2/K}$、シフトパラメータ $b = q/f_s$ として離散 wavelet 変換 $(p, q \in \mathbf{Z})$ を用いて計算機上に分析フィルタ群を実装した。このとき、分析フィルタ群は約 1/4 ERB の定 Q フィルタバンクである。また、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ため、alinement 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た。

2.2 二波形分離問題の定式化

本研究では、“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 [8]。

はじめに、信号 $f(t) = f_1(t) + f_2(t)$ のみが受聴できるものとする。観測された信号 $f(t)$ は、K 個の分析フィルタ群により周波数分解される。次に、$f_1(t)$ と $f_2(t)$ に対応する k 番目の分析フィルタの出力は、それぞれ

$$f_1(t) : A_k(t)\sin(\omega_k t + \theta_{1k}(t)) \tag{4}$$

$$f_2(t) : B_k(t)\sin(\omega_k t + \theta_{2k}(t)) \tag{5}$$

と仮定される。但し、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る。また、k 番目の分析フィルタの出力 $X_k(t)$ は、式 (4) と式 (5) の和であり、

$$X_k(t) = S_k(t)\sin(\omega_k t + \phi_k(t)) \tag{6}$$

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

$$S_k(t) = \sqrt{A_k^2(t) + 2A_k(t)B_k(t)\cos\theta_k(t) + B_k^2(t)} \tag{7}$$

$$\phi_k(t) = \tan^{-1}\left(\frac{A_k(t)\sin\theta_{1k}(t) + B_k(t)\sin\theta_{2k}(t)}{A_k(t)\cos\theta_{1k}(t) + B_k(t)\cos\theta_{2k}(t)}\right) \tag{8}$$

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。ここで、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

$$A_k(t) = \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)} \tag{9}$$

$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)} \tag{10}$$

で解析的に解くことができる。最後に、すべての $A_k(t)$ と $B_k(t)$ をそれぞれ式 (4) と式 (5) を用いて合成することで、$f_1(t)$ と $f_2(t)$ を再構成できる。ここで、再構成された信号を $\hat{f}_1(t)$、$\hat{f}_2(t)$ とする。

上記の定式化において、$\theta_{1k}(t)$ と $\theta_{2k}(t)$ の解を一意に決定することは困難であるが、分析フィルタの帯域幅が狭く、フィルタ数 K が十分大きければ、信号の各周波数成分は分析フィルタの中心周波数におおよそ一致する。そこで、本論文では、$f_1(t)$ を望みの信号、$f_2(t)$ を雑音とし、$\theta_{1k}(t) = 0$ と仮定することで、$\theta_{2k}(t)$ を決定する。また、$f_2(t)$ が存在している状態で $f_1(t)$ が加算される問題を想定する。

3 四つの物理パラメータの計算方法

3.1 $S_k(t)$ と $\phi_k(t)$ の計算方法

式 (7) の振幅包絡 $S_k(t)$ と式 (8) の出力位相 $\phi_k(t)$ はそれぞれ式 (2) の振幅項 $|\tilde{f}(a,b)|$ と位相項 $\arg(\tilde{f}(a,b))$ から

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

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

で求めることができる。

物理的制約条件 1（ゆっくりと：漸近的変化）

$$dA_k(t)/dt = C_{k,R}(t) \quad (13)$$

$$\Rightarrow \quad \theta_{2k}(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t)+C_k(t)}\right) \quad (14)$$

物理的制約条件 2（なめからに：漸近的変化）

$$\sigma = \int_{t_a}^{t_b}\left[A_k^{(R+1)}(t)\right]^2 \Rightarrow \min \quad (15)$$

物理的制約条件 3（一つの音響事象に生じる変化）

$$\frac{A_k(t)}{\|A_k(t)\|} \approx \frac{A_\ell(t)}{\|A_\ell(t)\|}, \quad \ell \neq k \quad (16)$$

$$\Rightarrow \max_{\hat{C}_k - P_k \le C_k \le \hat{C}_k + P_k} \frac{< \hat{A}_k, \hat{\hat{A}}_k >}{||\hat{A}_k|| ||\hat{\hat{A}}_k||} \quad (17)$$

図 2: 二波形分離の物理的制約条件

3.2 $\theta_{1k}(t)$ と $\theta_{2k}(t)$ の計算方法

本論文では、入力位相を $\theta_{1k}(t)=0$ とするため、$\theta_k(t)=\theta_{2k}(t)-\theta_{1k}(t)$ の関係から、残る入力位相 $\theta_{2k}(t)$ を求めなければならない。これは、Bregman によって提唱された発見的規則 (ii) と (iv) を考慮した三つの物理的制約条件を用いることで決定できる [10]。はじめに、物理的制約条件 1 では、振幅包絡 $A_k(t)$ に式 (13) の制約を設けることで、$\theta_{2k}(t)$ の一般解 (14) を得る。次に、これを決定するためには未定関数 $C_k(t)$ を知る必要があるため、これを Kalman フィルタを用いて推定する。ここで、$\hat{C}_k(t)$ を最小分散推定値、$P_k(t)$ を推定誤差とする。次に、物理的制約条件 2 では、式 (15) のなめらかさの評価基準を満たす $A_k(t)$ を決めるために、推定誤差 ($\hat{C}_k(t)-P_k(t) \le C_k(t) \le \hat{C}_k(t)+P_k(t)$) 内で Spline 補間された $C_k(t)$ の候補を求める。最後に、物理的制約条件 3 では、分離・抽出したい信号の振幅包絡 $A_k(t)$ 間の相関が最大になる $C_k(t)$ を式 (17) から求めることで、最適な $C_k(t)$ を求め、$\theta_{2k}(t)$ を一意に決定する。

4 二波形分離とグルーピング

ここでは、グルーピングについて説明する。グルーピングの制約の狙いは、Bregman によって提唱された発見的規則 (i) と (iii) を用いて雑音が付加された信号から望みの信号を抽出することである。そのため、グルーピング部では二波形分離問題の解を式 (6) のすべての $X_k(t)$ に適用するのではなく、二つの信号波形が同時に存在する $X_k(t)$ のみに適用し、その結果から二波形の振幅包絡をそれぞれグループ化する。言い換えると、グルーピングに関する二つの物理的制約条件を満たすとき、グルーピング部は二波形分離問題の解を $X_k(t)$ に適用し、二波形の振幅包絡をグループ化した後で信号波形に再構成する。

ここで、発見的規則 (iii) では調波複合音の基本周波数を利用するため、次に基本周波数の推定方法とその時間変化の対応を述べる。

4.1 基本周波数の推定とその時間変動の対応

本論文では、河原によって提案された TEMPO (Time-domain Excitation extraction based on a Minimum Perturbation Operator) [11] を利用して調波複合音の基本周波数を抽出する。この方法は基本波の瞬時周波数を基本周波数と定義して「基本波らしさ」という概念を導入することで、基本周波数の推定とその抽出精度の推定を同時に行なうことを可能にした方法である。また、これは、調波複合音を定 Q フィルタバンクで分析した結果から、基本波のみが含まれているフィルタ出力を選択する方法であり、本モデルの分析系にも実装できる。

一般に、基本周波数は時間的に変化するため、グルーピングの制約を用いて二波形分離を行なうとき、この時間変化に応じた処理を行なわなければならない。そこで、次に基本周波数の時間変化に対する処理を説明する。

はじめに、TEMPO で推定された基本周波数を $F_0(t)$ とおく。次に、発見的規則 (ii) の漸近的変化を再度考える。この規則は、“基本波の周波数変化がゆっくりとなめらかである” と解釈できるし、言い換えると、“基本波の周波数変化が急激には起こらない” と解釈できる。そこで、基本周波数の時間変化に対しても発見的規則 (ii) を利用するため、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 4 (基本周波数の時間変化)　ある微小区間における基本周波数は一定である：

$$\frac{dF_0(t)}{dt} = 0 \quad (18)$$

■

この制約条件における微小区間は、基本周波数 $F_0(t)$ の分散量がある範囲内にあるときの区間と解釈できる。これは、次式を用いることで決定できる。

$$\frac{1}{t_h - t_{h-1}}\int_{t_{h-1}}^{t_h}\left|F_0(t)-\overline{F_0(t)}\right|^2 dt \le \Delta F_0^2 \quad (19)$$

ここで、微小区間は $t_h - t_{h-1}$ であり、$\overline{F_0(t)}$ は $F_0(t)$ の平均値、ΔF_0^2 は分散量の上限である。本論文では、ΔF_0 を 1 Hz とした。

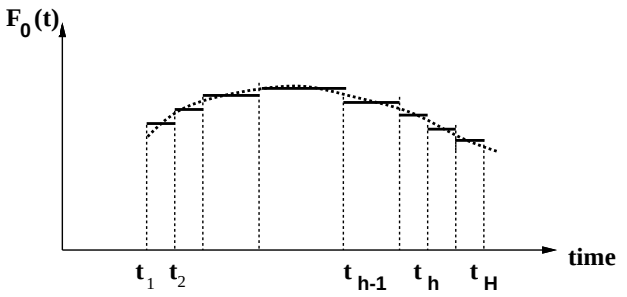

図 3: 基本周波数の時間変動

この物理的制約条件を適用した基本周波数 $F_0(t)$ と微小区間の関係を図 3に示す。図中の破線で示される基本周波数 $F_0(t)$ に対し、式 (19) の制約により、$H-1$ 個の微小区間に分割される。

次に、この基本周波数を利用したグルーピングの制約について説明する。

4.2 グルーピングの制約条件

一番目の規則として、発見的規則 (iii) 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” という規則 [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** (調波関係)　F_0 を式 *(19)* で決定された微小区間における基本周波数、N_{F_0} を高調波の次数とする。このとき、調波関係にある信号成分が分析フィルタ $X_\ell(t)$ に存在すれば、そのチャネル番号 ℓ は

$$\ell = \frac{K}{2} - \left\lceil \frac{\log(n \cdot F_0/f_0)}{\log \alpha} \right\rceil, \quad n = 1, 2, \cdots, N_{F_0} \tag{20}$$

を満たさなければならない。但し、α はスケールパラメータであり、$\lceil \cdot \rceil$ は $\cdot$ を正の無限大方向に最も近い正数値へ丸める記号である。 ■

二番目の規則として、発見的規則 (i) 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” という規則 [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 **6** (立上り・立下りの同期)

信号 $f_1(t)$ を調波複合音とする。また、物理的制約条件 4 を満たす $f_1(t)$ の基本波成分の立上りを $T_{\rm S} = t_{h-1}$、立下りを $T_{\rm E} = t_h$ とする。このとき、ある分析フィルタで得られた音響事象がこの基本波の高調波成分であれば、その分析フィルタで得られた高調波成分の立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ は

$$|T_{\rm S} - T_{k,\rm on}| \le 50{\rm ms}, \quad |T_{\rm E} - T_{k,\rm off}| \le 100{\rm ms} \tag{21}$$

を満たさなければならない。 ■

本論文では、k 番目の分析フィルタの出力 $X_k(t)$ における $f_1(t)$ の高調波成分の立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ を、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

で決定する。

更に、物理的制約条件 3 (式 (17)) における $\hat{\hat{A}}_k(t)$ は、

$$\hat{\hat{A}}_k(t) = \frac{1}{N_{F_0}} \sum_{\ell \in \mathbf{L}} \frac{\hat{A}_\ell(t)}{\|\hat{A}_\ell(t)\|} \tag{22}$$

で決定される。但し、$\mathbf{L}$ は式 (20) を満たす ℓ の集合である。

ここで、これら二つのグルーピングのうち、物理的制約条件 5 は主に $f_1(t)$ の高調波成分の分離抽出を制約するもの、物理的制約条件 6 は $f_1(t)$ の非高調波成分の分離抽出を制約するものに対応する。

以上、六つの物理的制約条件を利用した二波形分離問題の解法は、図 4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

本モデルが、雑音が付加された信号 $f(t)$ から望みの信号 $f_1(t)$ を分離抽出できることを示すために、三種類の混合信号 $f(t)$ を用いて二波形分離のシミュレーションを行う。三種類のシミュレーションは、次のものである。

1. 雑音が付加された AM 調波複合音の分離
2. AM 調波複合音同士の分離
3. 雑音が付加された合成母音の分離

但し、1. と 2. では基本周波数が一定の場合、3. では基本周波数が時間的に変動する場合のシミュレーションである。

本論文では、次式で定義されるスペクトル歪み (SD: Spectrum Distortion) を用いて二波形分離の精度を評価する。

$${\rm SD} := \sqrt{\frac{1}{W} \sum_{\omega}^{W} \left(20 \log_{10} \frac{\tilde{F}_1(\omega)}{\tilde{\hat{F}}_1(\omega)} \right)^2} \tag{23}$$

wavelet 変換を用いて $f(t)$ を式 (6) の成分に分解する;
TEMPO を用いて基本周波数 $F_0(t)$ を抽出する;
式 (19) から微小区間の数 H を求める;
for $k := 1$ **to** K **do**
  $\theta_{1k}(t) = 0$;
  補題 1 から $S_k(t)$ と $\phi_k(t)$ を求める;
  for $h := 2$ **to** H **do**
    $T_{\rm S} = t_{h-1}$, $T_{\rm E} = t_h$;
    分離区間を $t_{h-1} \le t \le t_h$ とする;
    立上り時刻 $T_{k,\rm on}$ と立下り時刻 $T_{k,\rm off}$ を求める;
    if 制約条件 5 または 6 を満たす **then**
      Kalman filter を用いて $C_k(t)$ を推定する;
      補間区間を決定し、I を補間点数とする;
      for $i = 1$ **to** I **do**
        $\hat{C}_k(t_i) - P_k(t_i) \le C_k(t_i) \le \hat{C}_k(t_i) + P_k(t_i)$ 内で Spline 補間された $C_k(t)$ の候補を求める;
        式 (14) から $\hat{\theta}_{2k}(t)$ を求める;
        式 (9) から $\hat{A}_k(t)$ を求める;
        式 (22) から $\hat{\tilde{A}}_k(t)$ を求める;
        式 (17) から $\mathrm{Corr}(\hat{A}_k(t), \hat{\tilde{A}}_k(t))$ を求める;
      end
      推定誤差内で $\mathrm{Corr}(\hat{A}_k(t), \hat{\tilde{A}}_k(t))$ が最大になるときの $C_k(t)$ を求める;
      式 (14) から $\theta_{2k}(t)$ を求める;
    else
      $A_k(t) = 0$, $B_k(t) = S_k(t)$, $\theta_{2k}(t) = \phi_k(t)$;
    end
    式 (9) と式 (10) から $A_k(t)$ と $B_k(t)$ を求める;
  end
  式 (4) と式 (5) から $f_1(t)$ と $f_2(t)$ の成分を求める;
end
式 (9) と式 (10) から wavelet 逆変換を用いて $\hat{f}_1(t)$ と $\hat{f}_2(t)$ を再構成する;

図 4: 二波形分離アルゴリズム

但し、$\tilde{F}_1(\omega)$ と $\hat{\tilde{F}}_1(\omega)$ は、それぞれ $f_1(t)$ と $\hat{f}_1(t)$ の振幅スペクトルである。また、フレーム長は 51.2 ms、フレームシフトは 25.6 ms、w は分析合成系の周波数帯域 (約 6 kHz)、窓関数は Hamming 窓である。ここで、$f_1(t)$ の雑音除去特性は、$f(t)$ と $\hat{f}_1(t)$ の SD の差 (改善量) として評価される。

5.1 シミュレーション 1

このシミュレーションでは、$f_1(t)$ を図 5に示される AM 調波複合音、$f_2(t)$ を帯域制限されたピンク雑音とする。但し、$f_1(t)$ の基本周波数は 200 Hz 一定、高調波の次数は $N_{F_0} = 10$、振幅包絡の変動は 10 Hz の正弦波信号である。また、$f_2(t)$ の帯域幅は、6 kHz である。混合信号 $f(t)$ の SNR が 0 dB から 20 dB まで 5 dB 刻みに変化させた 5 種類の $f(t)$ をシミュレーションデータとして利用する。

例えば、図 6に示すように $f(t)$ の SNR が 10 dB のとき、本モデルは高い精度で $A_k(t)$ を分離でき、図 7に示すように、混合信号 $f(t)$ から $\hat{f}_1(t)$ を分離抽出できる。また、5 つのシミュレーションに対し、$\hat{f}_1(t)$ と $f(t)$ の SD 値の平均は図 8に示す結果となった。この結果、本モデルを用いることで、SD 値で約 15 dB 程度雑音を除去できることがわかる。従って、本モデルは雑音が付加された AM 調波複合音 $f(t)$ から高い精度で AM 調波複合音 $f_1(t)$ を分離抽出できることがわかる。

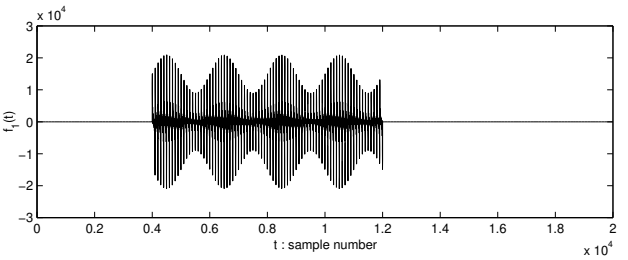

図 5: AM 調波複合音 $f_1(t)$

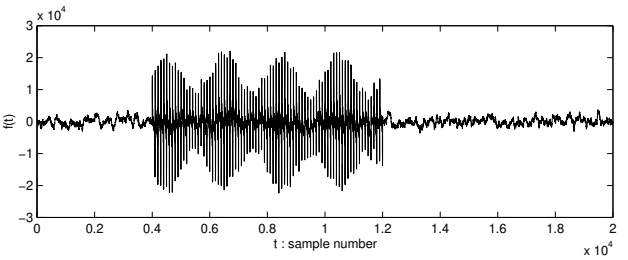

図 6: 混合信号 $f(t)$ (SNR= 10 dB)

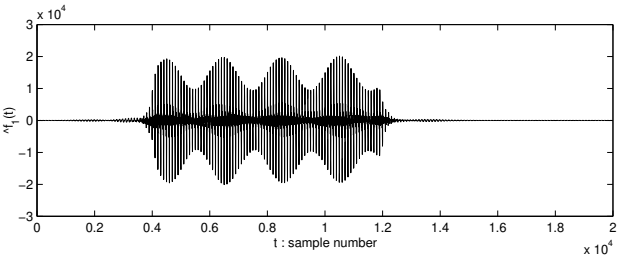

図 7: 分離抽出された信号 $\hat{f}_1(t)$ (SNR= 10 dB)

5.2 シミュレーション 2

このシミュレーションでは、$f_1(t)$ を図 5と同じ AM 調波複合音、$f_2(t)$ をもう一つの AM 調波複合音とする。但し、$f_2(t)$ の基本周波数は $F_0 = 300$ Hz、高調波の次数は $N_{F_0} = 10$、振幅包絡の変動は

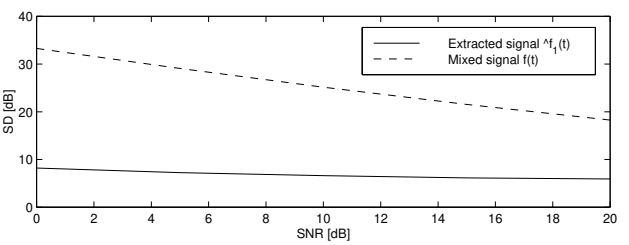

図 8: $\hat{f}_1(t)$ の SD 特性

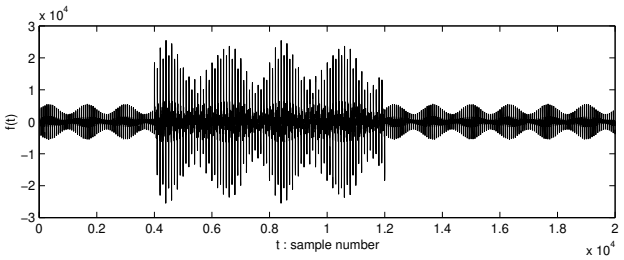

図 9: 混合信号 $f(t)$ (SNR= 10 dB)

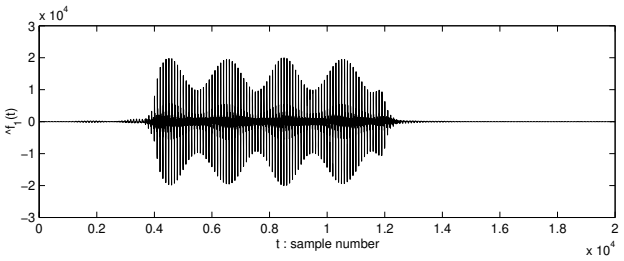

図 10: 分離抽出された信号 $\hat{f}_1(t)$ (SNR= 10 dB)

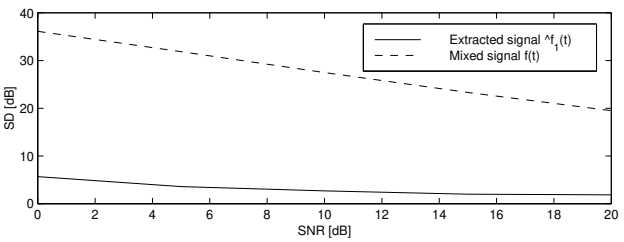

図 11: $\hat{f}_1(t)$ の SD 特性

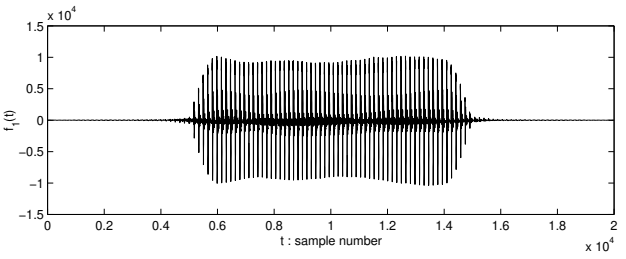

図 12: LMA 合成母音 $f_1(t)$

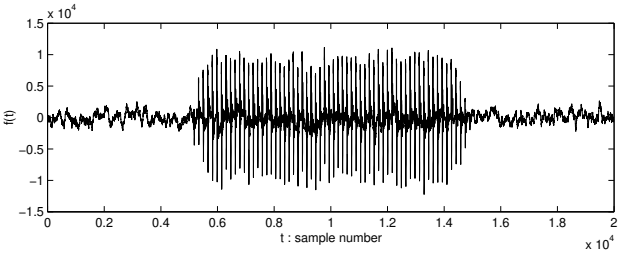

図 13: 混合信号 $f(t)$ (SNR=10 dB)

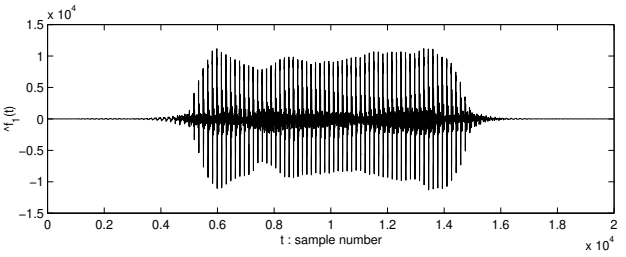

図 14: 分離抽出された母音 $\hat{f}_1(t)$ (SNR=10 dB)

15 Hz の正弦波信号である。従って、$f_1(t)$ と $f_2(t)$ の共通の高調波成分は 600 Hz の整数倍となり、例えば $f_1(t)$ の 3 次高調波と $f_2(t)$ の 2 次高調波が同じ周波数領域に存在する。混合信号 $f(t)$ の SNR が 0 dB から 20 dB まで 5 dB 刻みに変化させた 5 種類の $f(t)$ をシミュレーションデータとして利用する。

例えば、図 9に示すように $f(t)$ の SNR が 10 dB のとき、本モデルは高い精度で $A_k(t)$ を分離でき、図 10に示すように、混合信号 $f(t)$ から $\hat{f}_1(t)$ を分離抽出できる。また、5 つのシミュレーションに対し、$\hat{f}_1(t)$ と $f(t)$ の SD 値の平均は図 11に示す結果となった。この結果、本モデルを用いることで、SD 値で約 20 dB と雑音を除去できることがわかる。従って、先のシミュレーション結果も含め、本モデルは二つの信号が時間–周波数平面上に重なって存在しても、高い精度で混合信号から望みの AM 調波複合音を分離抽出できることがわかる。

5.3 シミュレーション 3

このシミュレーションでは、$f_1(t)$ を図 12に示す LMA 合成母音、$f_2(t)$ をシミュレーション 1 で用いたピンク雑音とする。但し、$f_1(t)$ の基本周波数は平均で 125 Hz、ダイナミックレンジが 5 Hz (123 ~ 128 Hz) であり、LMA で合成された母音/a/とした。このとき、高調波の次数は $N_{F_0} = 40$ とみなした。混合信号 $f(t)$ の SNR が 0 dB から 20 dB まで 5 dB 刻みに変化させた 5 種類の $f(t)$ をシミュレーションデータとして利用する。

例えば、図 13に示すように $f(t)$ の SNR が 10 dB のとき、本モデルは高い精度で $A_k(t)$ を分離でき、図 14に示すように、混合信号 $f(t)$ から $\hat{f}_1(t)$ を分離抽出できる。また、5 つのシミュレーションに対し、$\hat{f}_1(t)$ と $f(t)$ の SD 値の平均は図 16に示す結果となった。この結果、本モデルを用いることで、SD 値で約 15 dB と雑音を除去できることがわかる。このとき、$f_1(t)$ と $\hat{f}_1(t)$、および $f(t)$ における振幅スペクトルの比較 (式 (23) と同じ条件で計算) を行なったところ、図 15のようになり、明らかに雑音成分を除去できていることがわかる。従って、本モデルは合成母音と雑音が同一周波数領域に存在しても、高い精度で雑音中から合成母音を分離抽出できることがわかる。

以上、三つのシミュレーションを行なった結果、

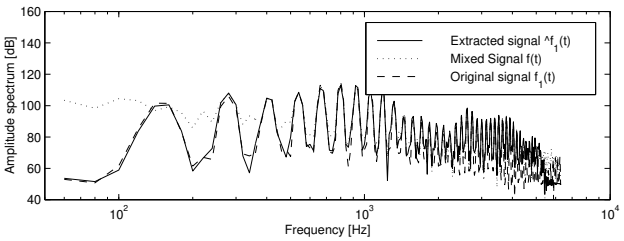

図 15: 振幅スペクトルの比較 (SNR=10 dB)

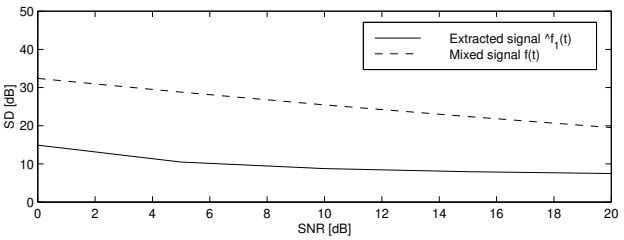

図 16: $\hat{f}_1(t)$ の SD 特性

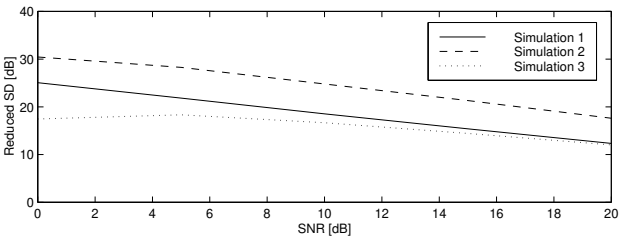

図 17: 雑音除去特性

本モデルは図 17に示すような雑音除去特性をもつことがわかる。この特性図から、全体的に混合信号の SNR が低い方ときに雑音除去の特性が高くなる傾向にあることがわかるが、少なくとも SD 値で約 15 dB 程度の雑音除去が可能であるといえる。

6 おわりに

本論文では、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、雑音が付加された調波複合音を分離抽出する方法を提案した。ここでは、本方法が混合信号から望みの信号を分離抽出できることを示すために、三種類の混合信号：

1. 雑音が付加された AM 調波複合音の分離抽出
2. AM 調波複合音同士の分離抽出
3. 雑音が付加された合成母音の分離抽出

に対する二波形分離のシミュレーションを行なった。シミュレーション 1 と 2 の結果、本方法により、AM 調波複合音と帯域雑音 (あるいは他の AM 調波複合音) が同一周波数領域に存在しても、正確に AM 調波複合音を分離抽出することが可能になった。特に、本方法により SD 値で平均約 15 dB の改善 (雑音除去) が得られた。更に、シミュレーション 3 の結果、本方法は雑音が付加された音声から望みの音声を分離抽出することもできた。

今後は、入力位相 $\theta_{1k}(t) = 0$ の仮定を取り除いたときの二つの入力位相 ($\theta_{1k}(t)$ と $\theta_{2k}(t)$) の決定方法の考案と本方法の実音声への適用が課題である。

謝辞　本研究の一部は、科学技術振興事業団 (CREST) の援助を受けて行なわれた。

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

[4] M. P. Cooke. “Modelling Auditory Processing and Organization,” Ph. D. Thesis, University of Sheffield, 1991 (Cambridge University Press, Cambridge, 1993).

[5] M. Abe, S. Ando. “Computational Auditory Scene Analysis Based on Loudness/Pitch/Timbre Decomposition,” Working Notes on the CASA Workshop, IJCAI-97, pp. 47–54, 1997.

[6] D. P. W. Ellis. “A Computer Implementation of Psychoacoustic 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] 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, pp. 444–453, March 1997.

[9] 鵜木 祐史, 赤木 正人, “共変調マスキング解除の計算モデルに関する一考察,” 信学技報, SP96-37, July 1996.

[10] M. Unoki and M. Akagi. “A Method of Signal Extraction from Noisy Signal,” In Proc. EuroSpeech’97, vol. 5, pp. 2583-2586, RHODOS-GREECE, Sept. 1997.

[11] H. Kawahara, “STRAIGHT - TEMPO : A Universal Tool to Manipulate Linguistic and Para-Linguistic Speech Information,” In Proc. SMC-97, Oct. 12-15, Orland, Florida, USA.