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)

1. はじめに

Bregman は、Auditory Scene Analysis (ASA) で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４つの発見的規則 ((1) 立ち上がり・立ち下がり、(2) 漸近的変化、(3) 調波関係、(4) １つの音響事象に生じる変化) としてまとめている[1]。この ASA に基づく音源分離モデルとして、著者らは、２つの信号が同一周波数領域にあるような場合でも完全に分離するために振幅スペクトルの他に位相も考慮し、規則 (2) と (4) を利用した２波形分離問題の解法を提案した[2]。本稿では、更に規則 (1) と (3) を利用し、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・抽出する方法を提案する。

2. モデル構成と２波形分離問題

本方法のモデルは、図１のように Gammatone filter を基底関数とした wavelet 分析合成系 (中心周波数 $f_0 = 600$Hz、チャネル数 $K = 128$、解析周波数範囲 60 ~ 6000Hz) で構成される[2]。

本研究では、“ある２つの独立な音源で生じた音響信号 $f_1(t)$ と $f_2(t)$ が加算された信号 $f(t) = f_1(t) + f_2(t)$ から、それぞれの音響信号 ($\hat{f}_1(t)$ と $\hat{f}_2(t)$) に分離すること” を２波形分離問題と定義しており、本モデルは次のような定式化[2]に従っている。

混合信号 $f(t)$ は K 個の分析フィルタ群により周波数分解されるため、$f_1(t)$ と $f_2(t)$ の周波数成分

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

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

の和として、分析フィルタの出力 $X_k(t)$ は表される。

$$X_k(t) = S_k(t)\sin(\omega_k t + \phi_k(t)) \qquad (3)$$

このとき、２波形の振幅包絡 $A_k(t)$ と $B_k(t)$ は

$$A_k(t) \quad = \quad \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)} \qquad (4)$$

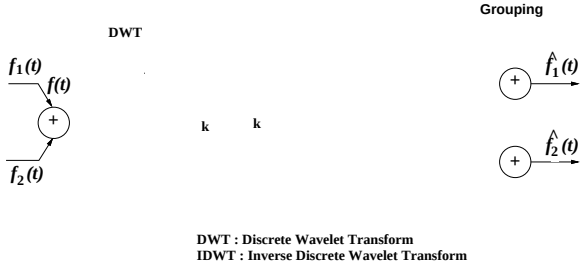

図 1 . wavelet 分析合成系

$$B_k(t) \quad = \quad \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)} \qquad (5)$$

のように解析的に解くことができる。但し、$\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ である。従って、本モデルでの２波形分離とは、すべての分析フィルタの出力 $(X_k(t), 1 \le k \le K)$ に対し、４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２波形の振幅包絡 $A_k(t)$ と $B_k(t)$ を式 (1) と式 (2) より合成して $f_1(t)$ と $f_2(t)$ を再構成することである[2, 3]。

本稿では、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た２波形分離問題の解法の１つ[2]を適用する。

2.1 物理パラメータの計算方法

振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ複素 wavelet 変換で定義された振幅スペクトルと位相スペクトルから導出できる[2]。また、入力位相は、$\theta_{1k}(t) = 0$、$\theta_k(t) = \theta_{2k}(t)$ であり、$\theta_k(t)$ は図２に示す３つの制約条件を用いることで得られる[3]。

制約条件 (i) では、$A_k(t)$ を式 (6) で制約することで入力位相 $\theta_k(t)$ を得るが、未定関数 $C_k(t)$ を知る必要がある。そこで、$C_k(t)$ を Kalman filter で推定し (推定値 $\hat{C}_k(t)$ と推定誤差 $P_k(t)$)、制約条件 (ii) で、$C_k(t)$ が変動できる範囲を式 (8) の推定誤差範囲に制限する。制約条件 (iii) 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 (10) から求めることで、入力位相 $\theta_k(t)$ を一意に求める。ここで、$\hat{A}_k(t)$ は式 (8) の誤差範囲でスプライン補間された $C_k(t)$ から得られたものであり、$\hat{\hat{A}}_k(t)$ は式 (11) を満たす $X_k(t)$ において $\hat{C}_k(t)$ より得られた $\hat{A}_k(t)$ の (正規化した) 平均である。

2.2 ２波形分離とグルーピング

ここでは、２波形の周波数成分が同時に存在する $X_k(t)$ に２波形分離問題の解法を適用するために、図３に示すグルーピングの制約条件を用いる。これは、基本周波数 F_0 の倍音成分が存在する $X_k(t)$ か (制約条件 (iv))、または立ち上がりと立ち下がりが同期する (制約条件 (v)) $X_k(t)$ に対して２波形分離問題の解法を適用するものである。ここで、F_0 は一定であり、既知であるものとする。また、$T_{k,\mathrm{on}}$ と $T_{k,\mathrm{off}}$ は $dS_k(t)/dt$ と $d\phi_k(t)/dt$ から検出可能な $X_k(t)$ での立ち上がりと立ち下がりであり、T_S と T_E は F_0 の周波数成分が存在する $X_k(t)$ での $T_{k,\mathrm{on}}$ と $T_{k,\mathrm{off}}$ である。

---

* An Extraction Method of the AM Complex tone from Bandpassed-Noise added AM Complex tone
By Masashi Unoki and Masato Akagi (JAIST)

(i) 漸近的変化 (**Bregman**)

$$dA_k(t)/dt = C_{k,R}(t) \quad (6)$$

$$\Rightarrow \quad \theta_k(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t)+C_k(t)}\right) \quad (7)$$

$$C_k(t) = -\int C_{k,R}(t)dt + C_{k,0}$$

(ii) 連続性

$$\hat{C}_k(t) - P_k(t) \le C_k(t) \le \hat{C}_k(t) + P_k(t) \quad (8)$$

(iii) １つの音響事象に生じる変化 (**Bregman**)

$$A_k(t) \approx A_\ell(t), \quad \ell \ne k \quad (9)$$

$$\Rightarrow \max_{\hat{C}_k - P_k \le C_k \le \hat{C}_k + P_k} \frac{< \hat{A}_k, \hat{\hat{A}}_k >}{||\hat{A}_k||||\hat{\hat{A}}_k||} \quad (10)$$

図 **2** ．２波形分離の制約条件 (入力位相 $\theta_k(t)$ の決定)

(iv) 調波関係 (**Bregman**)

$$k = \frac{K}{2} - \left\lceil \frac{\log(n\cdot F_0/f_0)}{\log\alpha} \right\rceil \pm 1, \qquad n = 1,2,\cdots,N \quad (11)$$

(v) 共通の立ち上がり・立ち下がり (**Bregman**)

$$|T_{\rm S} - T_{k,{\rm on}}| \le 50{\rm msec}, \quad |T_{\rm E} - T_{k,{\rm off}}| \le 100{\rm msec} \quad (12)$$

図 **3** ．グルーピングの制約条件

ここで、図２と図３の物理的制約条件 (i), (iii), (iv), (v) 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したものである[3]。

3. ２波形分離のシミュレーション

実験データとして、AM 調波複合音 $f_1(t)$ (図４) とこれに SNR=0, 10, 20dB のランダム帯域雑音 $f_2(t)$ を加算した３種類の混合信号 $f(t)$ を用意する。但し、AM 調波複合音は、基本周波数を 200Hz 一定、高調波の次数を $N = 10$、振幅包絡を 10Hz の正弦波とし、ランダム雑音の帯域幅は解析周波数範囲とした。

分離精度の評価については、$f_1(t)$ の信号区間でフレーム (フレーム長 51.2msec、フレームシフト 25.6msec、Hamming 窓) ごとに求めた SD (Spectrum Distortion) の平均値とし、$f(t)$ と $\hat{f}_1(t)$ の SD 値の差を改善量として評価した。

３種類の混合信号 $f(t)$ に対して２波形分離のシミュレーションを行なった。この結果、SNR が 0dB,10dB,20dB の混合信号に対して、それぞれ SD 値で 25.8dB, 20.6dB, 14.7dB の改善 (雑音除去) が得られた。例えば、SNR=10dB の混合信号 $f(t)$ (図４) に対して、図５のような分離結果が得られた。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 $\hat{f}_1(t)$ を $\theta_{1k}(t) = \phi_k(t)$ として再構成した。上記の結果より、本手法を用いることで、帯域雑音中の 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。

4. まとめ

Bregman によって提唱された４つの発見的規則を利用し、著者らが提案した２波形分離問題の解法を

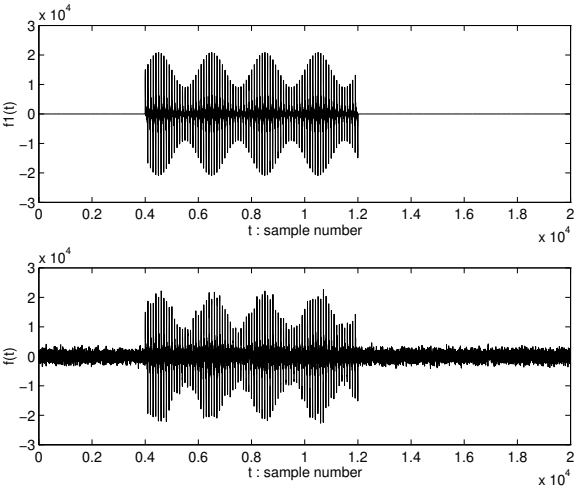

図 **4** ．信号 $f_1(t)$ と混合信号 $f(t)$ (**SNR=10dB**)

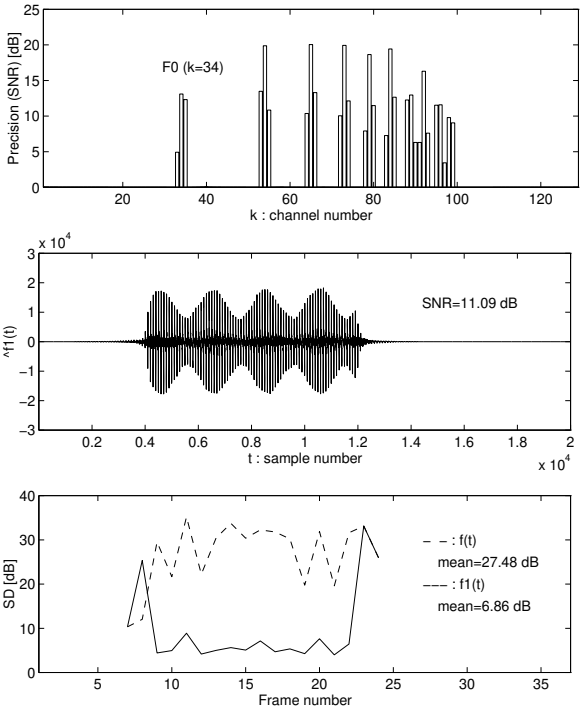

図 **5** ．分離抽出された信号 $\hat{f}_1(t)$ と分離精度の評価

適用することで、帯域雑音中の AM 調波複合を分離・抽出する方法を提案した。この方法により、信号と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。

今後は、入力位相 $\theta_{1k}(t)$ の推定法を考案し、FM 音の分離・抽出問題に拡張していく予定である。

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] 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).

[2] 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) (採録決定).

[3] 鵜木 祐史、赤木 正人 :“帯域雑音中からの AM 調波複合音の一抽出法,” 音響学音声研資,1997.