

F-001

## GHA と聴覚の情景分析を用いた音源分離

Sound source separation based on GHA and Auditory Scene Analysis

松居 俊孝†  
Toshitaka Matsui甲藤 二郎†  
Jirou Katto

## 1. まえがき

近年、様々な音が含まれている音の中から目的の音を取得する音源分離について様々な研究がされている[1]~[3]。本手法では単一の観測音、つまりモノラルの音源とし、複数の楽器音で構成された混合音からの音源分離を試みる。具体的には一般調和解析(GHA: Generalized Harmonic Analysis)とくし型フィルタを併用し、混合音を構成する支配的な基本周波数の組み合わせを取り出すことによって音源分離を実現する。

## 2. 従来手法

一般調和解析を用いた音源分離手法としては[1][2]に挙げた手法が知られている。この手法では周波数成分ごとの時間変化の相関の高いもの同士を同一音源とすることによって、モノラル音源からの音源分離を試みている。

しかし、この従来手法ではパワーの大きい周波数と小さい周波数との間で時間変化の幅が大きく異なってしまう、分離が困難な場合が頻繁に発生する。これに対して本手法では、時間変化によらない処理の改善を組み入れる事により、分離精度の改善を目指す。

## 3. 一般調和解析

一般調和解析は原信号から残差のエネルギーが最小となる純音成分を逐次計算し正弦波を取り出していく解析手法である。以下に解析方法を示す。

観測区間長  $L(t)$  で観測された連続信号  $x_0(t)$  があるとする。Fourier 係数

$$S(f) = \frac{2}{nT} \int_0^{nT} x_0(t) \sin(2\pi ft) dt \quad (3.1)$$

$$C(f) = \frac{2}{nT} \int_0^{nT} x_0(t) \cos(2\pi ft) dt \quad (3.2)$$

をそれぞれ計算する。ここで  $f$  (Hz) は任意の周波数、 $T=1/f$  (s) は周期、 $n$  は整数、 $nT \leq L$  である。上記より求めた係数を元に観測区間における残差

$$\varepsilon(t, f) = x_0(t) - S(f) \sin(2\pi ft) - C(f) \cos(2\pi ft) \quad (3.3)$$

のエネルギー

$$E(f) = \int_0^L \varepsilon(t, f)^2 dt \quad (3.4)$$

を周波数ごとに計算し、エネルギーを最小とする周波数と Fourier 係数を求める。原信号から求めた周波数と Fourier

係数を用い周波数成分を取り除く。この過程を繰り返す。

この周波数解析の特徴として FFT や wavelet に比べて時間窓にとらわれずに高い周波数解像度で解析を行うことができる。これにより、きめ細やかな周波数解析結果が期待される。

## 4. くし型フィルタ

本稿で使用するくし型フィルタを以下に定義する。

$$\text{Comb}(f, n) = \begin{cases} 2, & f = nf_1 + \alpha, \text{ if } (1 \leq n \leq 3) \\ 1, & f = nf_1 + \alpha, \text{ if } (4 \leq n \leq N) \\ 0, & (\text{otherwise}) \end{cases} \quad (4.1)$$

$N$ : 倍音の数

ただし、

$$\begin{cases} \alpha = 0, & \text{if } (n=1) \\ -3 \leq \alpha \leq 3, & (\text{otherwise}) \end{cases} \quad (4.2)$$

とし、次式に従い、基本周波数を決定する。

$$\hat{f} = \arg \max_{1 \leq L} \sum_f \text{Comb}(f, n) f(t) \quad (4.3)$$

$f(t)$  は時刻  $t$  における周波数分布

このくし型フィルタではまず仮基本周波数  $f_1$  を決めてやり、その仮基本周波数から導いた  $N$  倍音までに対して式(4.1)のようなフィルタを設定し、このフィルタを通過したパワーの総和を計算する。パワーが高い順から  $L$  個の  $f_1$  に対してこの計算を行い、その中で通過したパワーの総和が最も大きい仮基本周波数  $f_1$  を基本周波数と決め、その基本周波数と通過した倍音の周波数を選び出す。ただし、倍音の周波数が必ずしも整数倍とは限らないので式(4.2)のようなずれを設定し補正する。

このくし型フィルタよりある音の基本周波数とそれより求められる倍音による構造が推測できる。

## 5. 本手法の流れ

以上の一般調和解析とくし型フィルタを併用し、本手法では以下の手順に従って音源分離を試みる。

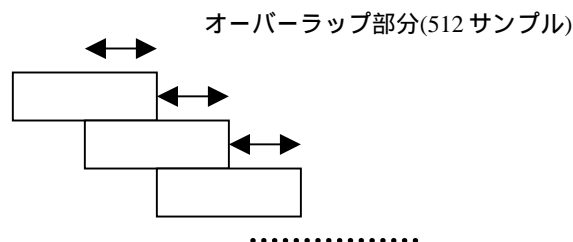


図 5.1 入力音分割イメージ

まず、いくつかの音源(本論文では異なる楽器音とする)からなる入力音を 1 フレーム 1024 サンプルで前フレーム

† 早稲田大学理工学研究所

と 512 サンプルの重なりをもって分割を行い、それぞれのフレームに対して一般調和解析を行う(図 5.1)。

次に、一般調和解析から得られたフレームごとの周波数分布に対してくし型フィルタを適用する。この際、一般調和解析においてパワーの大きい周波数から順に仮基本周波数としてくし型フィルタに通し、さらに通過したパワーの総和の順にこの仮基本周波数群を並べ替える。

また、くし型フィルタから求めた仮基本周波数群からよりふさわしい基本周波数を求めるために、補正の処理を行う。基本的には、くし型フィルタから求められた仮基本周波数群とその倍音を用いて算出した波形と前フレームの重なり部分において相関係数を計算する。この中で最も相関係数の高くなるものを基本周波数とする。その求められた基本周波数とそれより求まる倍音から算出される波形を分離結果として出力する。

分離した音のフレーム間接続に関しては、まず現フレームと前フレームにおいて重なり部分同士の相関係数を算出する。そして、相関係数の高いフレームの組み合わせを同一音源と認識して時間的に接続する。

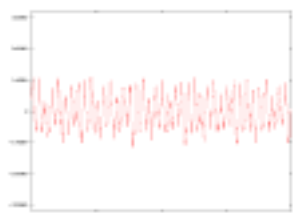


図 5.2 混合音

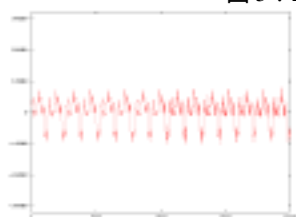


図 5.3 本手法結果 1

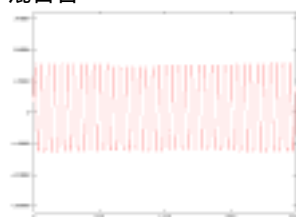


図 5.4 本手法結果 2

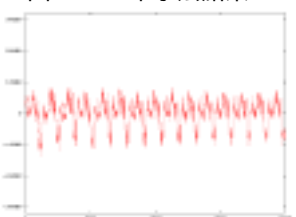


図 5.5 ギター原音

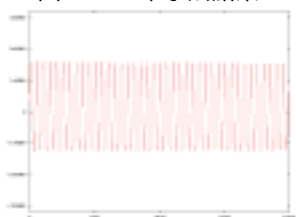


図 5.6 クラリネット原音

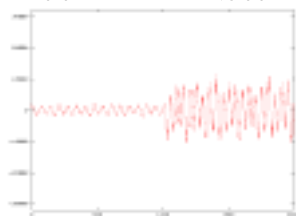


図 5.7 従来手法結果 1

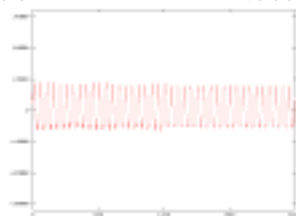


図 5.8 従来手法結果 2

図より、本手法が従来手法よりも正確な分離ができていることがわかる。

## 6. 音源分離実験

### 6.1 実験条件

サンプリングレート 44.1kHz, 16 ビットで量子化した MIDI 音源の音を WAVE フォーマットにしたものを使用した。音の長さはそれぞれ 1 秒とした。用意した音はギター 330Hz, クラリネット 880Hz, フルート 880Hz, サックス 659Hz, バイオリン 196Hz の 5 つの楽器音から 2 つの音を選択してミックスし、入力音とした。

### 6.2 実験結果

一例としてギターとクラリネットの混合音における本手法による分離結果と従来手法による分離結果を混合音の波形とそれぞれの楽器の原音の波形とともに図 5.3 から図 5.4 に示す。

その他の結果については、大体の混合音はほぼ分離ができたが、次のような時に分離が完全ではなかった。ギターとサックス, クラリネットとフルートの組み合わせ等、周波数パターンの類似した、あるいは倍音成分の重なりが大きい混合音では分離が完全ではなかった。これは基本周波数あるいは倍音成分におけるパワーの重なりが発生したためにその成分におけるパワーが分離結果のどちらかに偏ってしまったためである。

## 7. まとめ

本手法で 2 楽器音からそれぞれの楽器に分離することが可能なことが確認できた。これから 3 楽器以上が含まれている音についても実験を行っていく。また、特に倍音成分重なりの問題に対処するために、楽器ごとの周波数パターンを考慮した検討、もしくはモデリングを行わない手法についての検討[4]も進めていく。

## 参考文献

- [1] 杉山道則, 大内康裕, 天田武, 山崎芳男, “一般調和解析による音響信号の分離”, 日本音響学会講演論文集 3-7-23 p 607-608(1996)
- [2] 杉山道則, 天田武, 大内康裕, 山崎芳男, “一般調和解析による周波数成分の時間変化に着目した音源の分離” 日本音響学会講演論文集, 2-3-9 p521-522(1997)
- [3] 鶴木裕史, 赤木正人, “聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法”, 電子情報通信学会論文誌 A Vol.J82-A No.10 pp1497-1507(1999)
- [4] 亀岡弘和, 西本卓也, 篠田浩一, 嵯峨山茂樹, “ハーモニック・クラスタリングによる多重音の基本周波数推定アルゴリズム”, 情報処理学会研究報告, MUS-50 p27-32(2003)