

倍音成分に生じるうなりを利用した二種楽音の分離

3G-8

松村 衣里子

法政大学大学院工学研究科システム工学専攻

長坂 建二

法政大学工学部システム制御工学科

1 はじめに

コンピュータによる信号処理の一つに「自動採譜」があるが、複数楽器による混合音に対する処理は不可能であった。そこで最近、混合音から入力音階と演奏楽器の同定を行う音源分離が求められるようになってきている。しかし、今までに考案されている音源分離の方法は、どれも楽音の特性を十分に利用したものではなかった。そこで本研究では、楽音の特性に着目し、楽音ならではの分離方法を考案する。

楽音にはその特徴ともなる倍音成分が含まれている。また、人間の耳が多種の楽音を分離する要因には、微妙な周波数のずれから生じる「うなり」があげられる。従って、まず、楽音信号を Wavelet 変換により倍音成分 (wavelet 成分) に分離し、各倍音成分での微妙な周波数のずれ (うなり) を Fourier 変換により検知することで、二波形を分離する方法を検討する。

ここで、周波数解析には離散 Fourier 変換を用いる。そのため、「ある微小区間において、周波数の時間変化はない」という前提条件をおくことにする。

2 離散 Wavelet 変換

Polyphony Separation from Two Musical Instruments' Sound Using the Howl in the Harmonic Components

Eriko MATSUMURA; Kenji NAGASAKA
Hosei Univ., Grad. School, Eng. Division, System Eng.,
3-7-2, Kajino-cho, Koganei-shi, Tokyo, 184-8584, Japan.
Tel/Fax. 0423-87-6261

2.1 Gammatone filter

基本 wavelet には、中耳内の基底膜の特徴を模倣した Gammatone filter を用いる。

f_0 を中心周波数に持つ Gammatone filter のインパルス応答 $gt(t)$ は

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

と表される。(人間の場合 $n = 4$) また、Gammatone filter の周波数特性は、 $f_0 \gg b_f$ のとき、

$$GT(f) \approx \left[1 + \frac{j(f - f_0)}{b_f} \right]^{-n}, \quad (0 < f < \infty) \quad (2)$$

となる。これは、中心周波数を f_0 とする帯域通過フィルタの形態を表している。

ここで wavelet 関数には、基本 wavelet を $gt(t) = \Psi(t)$ と置いた時に $\Psi_{ij}(m) = \frac{1}{\alpha^i} \Psi\left(\frac{(m-j)T}{\alpha^i}\right)$ と離散化したものを用いる。

2.2 スケールパラメータ: α^i

先の wavelet 関数は、基本 wavelet の中心周波数 f_0 に対して $1/\alpha^i$ 倍の中心周波数をもつフィルタを設計していることに相当する。ここで、楽音の音階は基準音 (真ん中の A の音) 440 Hz から $2^{i/12}$ 倍になっているため、 i が整数を動き $\alpha^i \rightarrow 2^{i/12}$ とすることですべての音階をカバーできる。そこで、 $\alpha = 2^{1/12}$ と設定する。

3 周波数刻み幅の問題

うなりを利用する方法では微妙な周波数のずれを検知しなければならない。そのため、それだけの周波数刻み幅が必要となる。

ここで、離散 Fourier 変換における周波数の刻み幅: f_{cut} は、サンプリング周波数: f_p , データ

数: N とすると $f_{cut} = f_p/N$ となっている。よって、周波数刻み幅を小さく取るには、サンプリング周波数にあわせてデータ数を十分に大きく設定しなければならない。しかしそうすると周波数の時間変化には対応できなくなる。そこで今回、対称拡張処理をほどこすことでこれを解決した。

4 同一音高混合音の音源分離実験

試料には、トランペットとピアノの混合音(同一音高・440 Hz)をMIDI音源により作成し、waveファイルに変換したものを用いる。ファイルフォーマットは、"モノラル音・サンプリング周波数 44.100 kHz・8 bit 量子化"とする。

4.1 実験結果

例として、抽出された3番目の倍音成分(1320 Hz)の波形を示す。

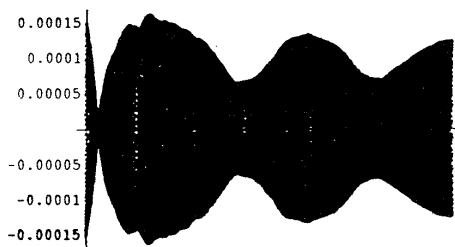


図 1: 倍音成分 (1320 Hz)

図 1 にブロックごとの DFT を施した結果、以下のようなエネルギースペクトルを得た。ある 1 ブロック (約 0.0045 秒分) の結果を示す。(横軸: 周波数 (Hz))

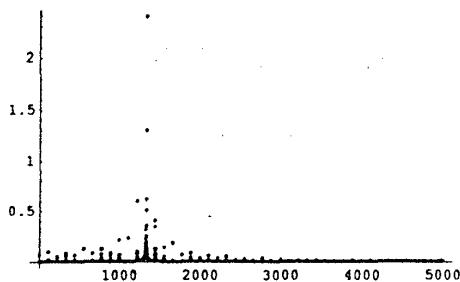


図 2: 倍音成分 (1320 Hz) の DFT 結果

4.2 考察

- 図 1 から、予想通り倍音成分にうなりが生じていることが分かる。また、それを Fourier 変換した結果 (図 2) から、倍音の周波数の近辺でとても近い値の二つのパワースペクトルが確認できる。
- 分離された音はとても機械的な音となっている。この原因としては、各倍音成分において二つの周波数を取り出し、それを単純におのおの合成しているためと思われる。これは、音の合成をしたことに一致する。

5 まとめ

今までに考案されてきた音源分離のアルゴリズムは、雑音から対象となる信号を分離する、という方法が多かった。そのため、全体的にとっても性質が似ている楽音の音のみによる複合音の分離には適していなかった。そこで今回は逆に、楽音の性質を十分利用しようという試みとなった。その結果、音の合成に近い処理を施したため、機械的な音色になってしまったが、採譜を前提とした音源分離という目的に対しては、一つの有効な手法になると考えられる。

参考文献

- [1] 井川 博行, 赤木 正人. 「位相の変化に着目した 2 波形分離法」. 日本音響学会講演論文集 3-4-15 (1994)
- [2] 野田 宗裕. 「人間の聴覚特性を考慮した波形分離法に関する基礎的研究」. JAIST 修士論文 (1995)
- [3] 鶴木 祐史. 「位相とスペクトルに着目した聴覚の情景解析に関する基礎的研究」. JAIST 修士論文 (1996)