

楽音の位相情報に基づいた音高推定困難和音の音高推定

伊藤正宏, 夏井雅典, 田所嘉昭
豊橋技術科学大学

あらまし 楽音を STFT (Short Time Fourier Transform) することによって得られる位相情報に注目して, 振幅スペクトルが重なる音同士の和音 (音高推定困難和音) の音高推定を実現する手法を提案する. 従来, 楽音の振幅特性は広く利用されてきたが, 位相情報はあまり注目されていなかった. STFT で得られる位相情報の時間変化より音高推定困難和音の各音高を推定する. 本手法で実楽器音 (RWC 研究用音楽データベース) の音高推定困難和音 (2 和音) の音高を推定できることを確認した.

Pitch Estimation Method for Difficult Consonance Using Phase Characteristics of Musical Instrument Sounds

Masahiro Ito, Masanori Natsui, Yoshiaki Tadokoro
Toyohashi University of Technology

Abstract This paper presents a pitch estimation method for difficult consonance whose frequency components overlap. The key idea is to use the phase characteristics obtained by the STFT of musical instrument sounds. From the time changes of the phase characteristics of the musical sounds, we can determine if the frequency components of the sound overlap or not and which frequency components are overlapping. Experimental results show that our method can estimate the pitches of the difficult consonance.

1. まえがき

マイクなどの入力装置, あるいは CD 再生などによって入力された音響信号を楽譜に直すシステム (採譜システム) として, これまでいくつかのシステムが報告されている [1][2][3]. また, いくつかの楽器メーカーによって商品化されている. しかし, 商品化されているシステムのほとんどは独唱や単独の楽器によって演奏される単音を対象にしたものであり, 周波数特性の異なる複数の音源によって構成された和音までを扱うことが可能な本格的なものは, いまだ実用化されていない. この大きな理由は, 2 和音を例に述べると, ユニゾン (基本周波数の等しい複数音によって形成される和音, 例: A_3+A_3), オクターブ (基本周波数の比率が 1:2 の関係にある和音, 例: A_3+A_4) や 3 倍音 (基本周波数の比率が 1:3 の関係にある和音, 例: A_3+E_5) といった振幅スペクトルが重なる和音 (図 1 参照) の

処理に関する技術が未だ確立されていないことがあげられる. 本稿ではこのような和音のことを音高推定困難和音と呼ぶ.

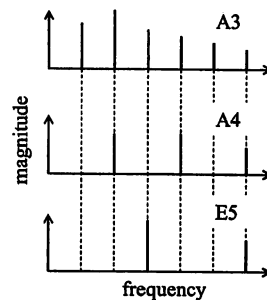


図 1. 音高推定困難和音になるスペクトル

音高推定困難和音は, 一般的な音楽演奏における最も基本的な和音として多く用いられるが, 離散フーリエ変換 (DFT: Discrete Fourier Transform) による周波数

解析に代表される振幅スペクトルに注目する従来の手法[4]では、和音において生じる周波数成分の重複を短時間で精度良く分離することが困難である。これは音高の認識精度の低下、あるいは分離された音の品質の低下へとつながり、高精度な採譜システムを実現する上で大きな妨げとなる。したがって、和音を含む楽音の音高を精度良く処理することが可能な高性能な採譜システムを構築するためには、音高推定困難和音の音高を精度よく推定する音高推定システム技術が必須となる。

和音の採譜に関しては研究段階であり、本研究で対象としている音高推定困難和音に関しては、遺伝的アルゴリズム (GA: Genetic Algorithm) による方法[5]、スペクトルの和を仮定した方法[6]、音楽規則等に従った推定法[7]がある。GA による方法は前もって各楽器の時間波形を記憶しておく必要があり、この論文では必ずしも音高推定困難和音を解決する手法を意図していない。また、他の二つの方法についても、位相の関係でスペクトルの和は必ずしも成立しないこと、また従来の音楽規則に従わない現代音楽等などには対応できないなどの問題がある。

本稿では、楽音の位相情報により音高推定困難和音の音高を推定するアルゴリズムを示す。解析対象として RWC 研究用音楽データベース[8]のうち、トランペット、ホルン、アルトサクソ、クラリネット、ピアノ、エレキギターの6楽器の単音についてはほぼすべての音高 (サンプリング周波数 $f_s=44.1$ [kHz]) と、それらを計算機上で加算して生成した2和音を使用する。また、和音の最低音高については我々が既に提案しているくし形フィルタ利用した音高推定システム[1]で既知とする。

2. 位相特性の測定法

本研究では、楽音の位相特性を得るために STFT を用いる。このときのフレームサイズは、対象の楽音の (和音の場合は低い方の音の) 基本周波数成分5周期分とする。このときの楽音の基本周波数は、A4=440 [Hz] を基準とする平均律から算出したものとする。

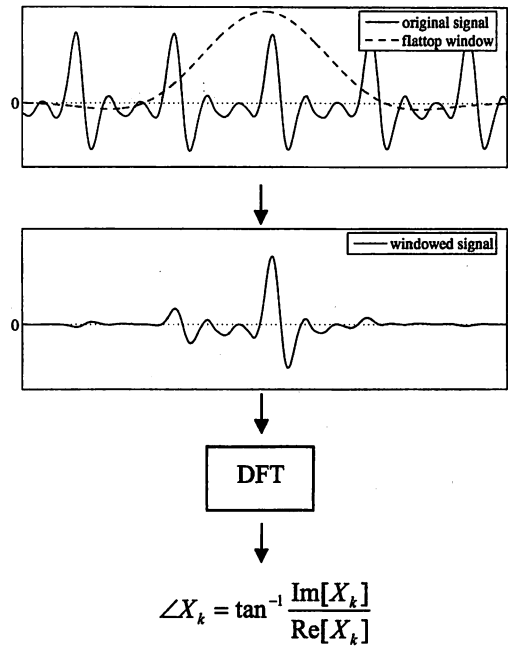


図2. 位相特性の測定法

また、時間窓としてフラットトップ窓[9]を使用する (図2)。この操作によって得られた k 倍音成分のベクトル X_k より各倍音成分の位相角を求めるが、このままではフレームがシフトすることにより位相角は変化する。この影響を考慮して位相を評価するために、基本周波数成分の位相を基準として k 倍音成分の位相 θ_k がどれだけ進んでいるのかを次式により得て、その時間変化を測定する。

$$\theta_k = \angle X_k - k\angle X_1 \quad (1)$$

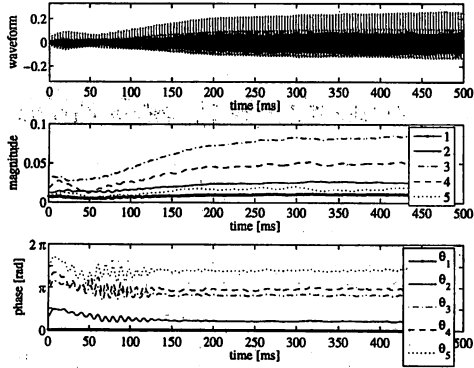
3. 楽音の位相特性

2. で示した測定法において、フレームシフト幅 (ホップサイズ) を約 1 [ms] (サンプル数 44 点) として測定した位相特性の結果を、時間波形、振幅特性とともに図3、図4に示す。

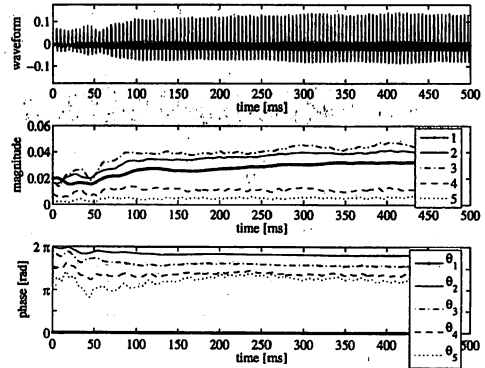
3.1. 単音

各楽器の単音の測定結果のうち代表的なものを図3に示す。基本周波数成分の位相 θ_1 は常に0であり、横軸と一致している。この結果から、トランペット、ホルン、アルトサックスなどの管楽器の位相はおおよそ

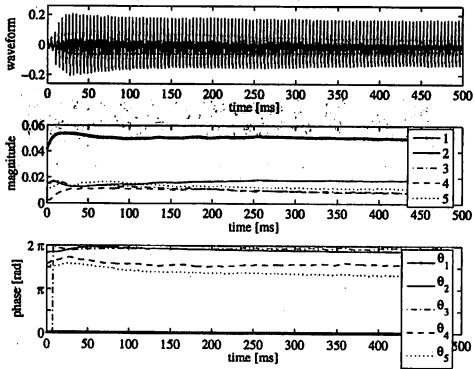
一定であり、弦楽器の位相は時間とともに変化してることがわかる。弦楽器の位相が時間とともに変化するのは、その楽音の倍音周波数が基本周波数の整数倍に完全に一致しない非調和性[10]に起因している。



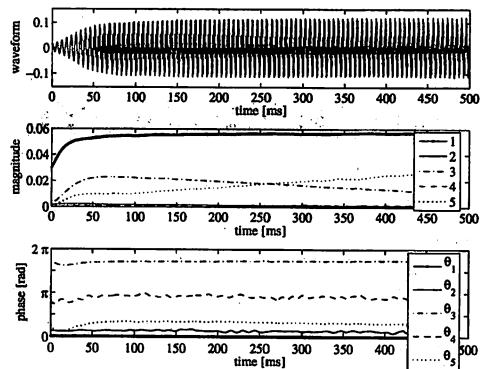
a. トランペット D4



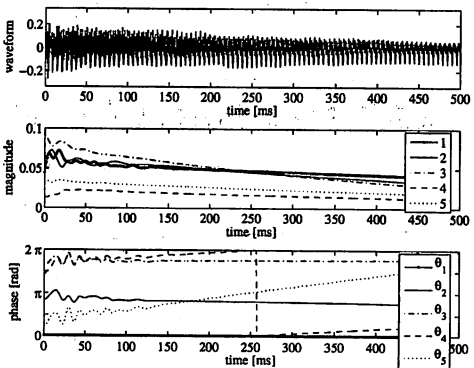
b. ホルン A3



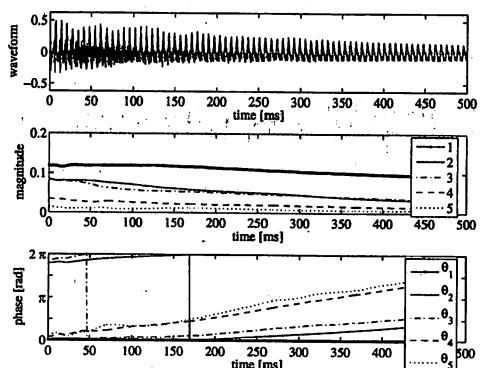
c. アルトサックス Bb3



d. クラリネット F3



e. ピアノ F3



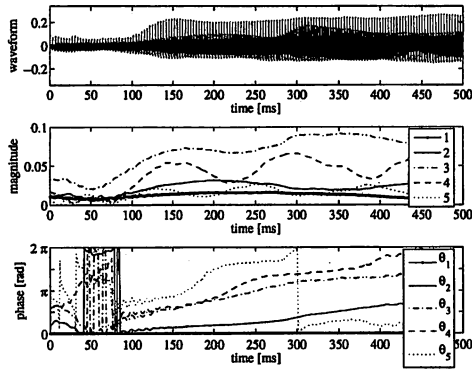
f. エレキギター D3

図3. 単音の測定結果 (上から時間波形, STFTの振幅特性, 位相特性)

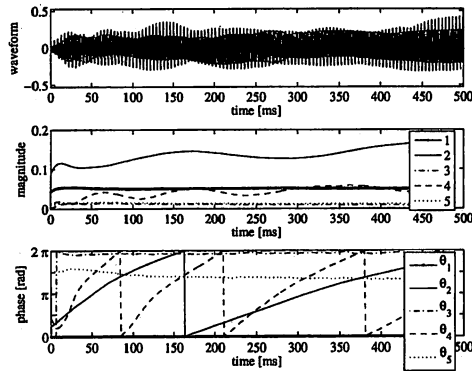
3.2. 和音

図 3. に示した音に異メーカーの音を足して作成した音高推定困難和音の代表的な測定結果を図 4 に示す。同音ユニゾンの例である a. のトランペットでは、すべての倍音成分の位相が時間とともに変化している。b. ホルンと c. アルトサックスはどちらもオクターブ違いの音を足したものであり、重なっている成分 (2 倍音

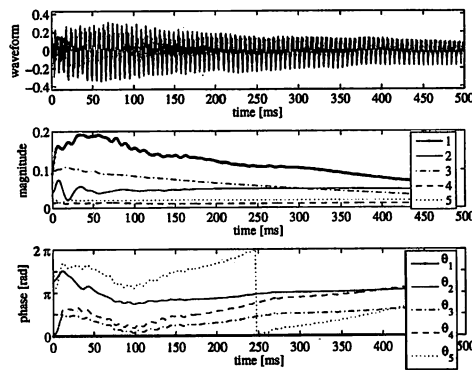
と 4 倍音) の位相は時間とともに大きく変化している。3 倍音の和音の例である d. クラリネットでは 3 倍音成分の位相のみが時間とともに変化している。e. ピアノと f. エレキギターは単音の場合は時間とともに多くの θ_k が増加傾向にあったが、和音になると単音とは違った変化を示している。



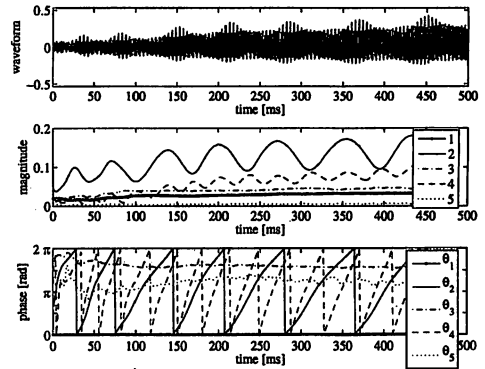
a. トランペット D4+D4 (同音ユニゾン)



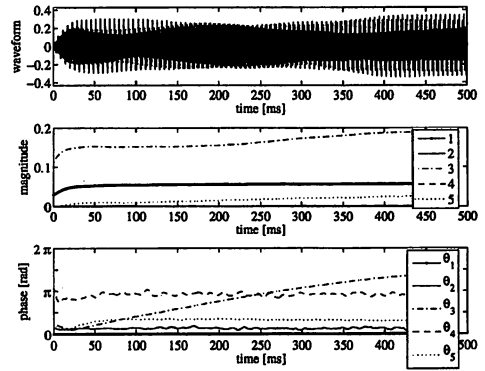
c. アルトサックス Bb3+Bb4 (オクターブ)



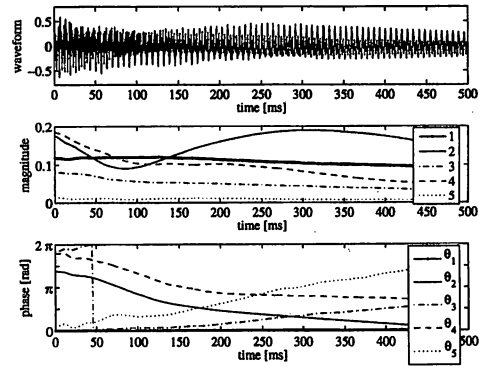
e. ピアノ F3+F3 (同音ユニゾン)



b. ホルン A3+A4 (オクターブ)



d. クラリネット F3+C5 (3倍音)



f. エレキギター D3+D4 (オクターブ)

図 4. 和音の測定結果 (上から時間波形, STFT の振幅特性, 位相特性)

4. 音高推定困難和音の音高推定

振幅スペクトルが重なっている成分の位相が時間とともに変化することを利用し、音高推定困難和音の音高推定を実現する手法を示す。ここでは、先に示した管楽器のような、単音で一定位相特性を持つ楽音を対象としている。

まず、単位時間当たりの位相変化量 p_k を次式のように定義する。

$$p_k = \frac{|\theta_k(t + \Delta t) - \theta_k(t)|}{\Delta t} \quad (2)$$

ここで Δt は使用したデータの時間長である。このとき、0 [rad]での折り返しによる影響が出ないように、あらかじめ $\theta(n)$ を修正して滑らかにしておく。これにより得られた測定結果の代表例として、ホルンの複数の音高 (F2-Eb5) の単音と1オクターブ違い和音のものを図5に示す。測定時間範囲は、楽音の立ち上がりから50 [ms] 後を測定開始点とし、そこから100 [ms] 後を終了点とした。よって、これらはそれぞれ式(2)における t と Δt に対応する。フレームシフト幅は10 [ms] とした。時間窓の種類やフレームサイズなどは前の位相特性測定と同じである。この範囲の単音の p_k のほとんどは 10π [rad/s]より小さいのに対して、和音の場合は p_2 と p_4 (重なっている成分の位相変化量) のみが大きくなっていることがわかる。この傾向は3.2の結果と一致している。よって、適切なしきい値を設定し、 p_k がそれよりも大きい小さいかを判断することで、その成分が単一のものか重なっているのかを検出できる。図5の結果において、 $thd=10\pi$ [rad/s]で重複成分の検出を試みたところ、約90%の精度で検出できた。

どの倍音成分が重複しているかがわかれば、音高推定困難和音を構成している音の高さを推定することができる。例えば、2, 4倍音成分が重なっていれば1オクターブ違い、3倍音成分が重なっていれば3倍音の和音、全ての成分が変化していれば同音ユニゾンである。また、2倍音成分が重なっていれば、4倍音成分の位相変化量が大きくなるとも1オクターブ違いの音が重なっているとみなすといった補正により、音高推定困難和音の音高推定精度を向上できる可能性がある。

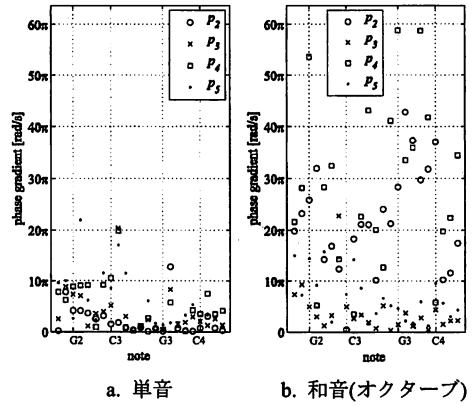


図5. ホルン位相変化量

5. 考察

前述(3.1)のように、位相の時間変化は倍音成分の周波数が基本周波数の整数倍に完全に一致しない非調和性に起因している。すなわち位相変化により音高推定をする本手法は非調和性を検出することと等価である。つまり、和音を構成する2つの音のスペクトルが完全に重なっている場合は本手法での音高推定ができない。これに該当するケース(クラリネット G3+D5, D5はG3の3倍音に該当)を図6に示す。この例では全成分の位相が一定であることがわかる。実際に足し合わせる前の単音をそれぞれDFT(周波数分解能1 [Hz])してピークを持つ周波数を測定すると、D5音の基本周波数は594 [Hz]、G3音の基本周波数成分は198 [Hz]であった。

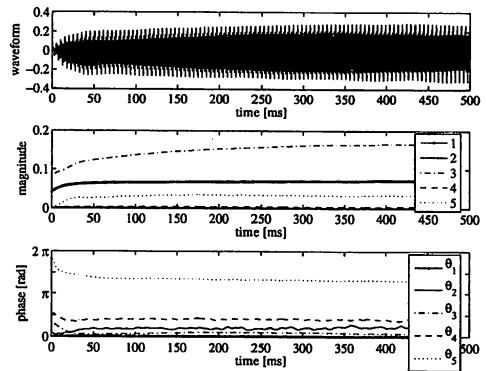


図6. クラリネット G3+D5 (3倍音)

この場合、 p_3 は小さな値となり、本手法はこの成分での重複はないと判断してしまう。しかし、このようなケースは少なく、4.の結果からもわかるように、実際に人の手で演奏される楽音により構成された和音は非調和であることが多いので本手法が有効であると考えられる。

ところで、重複成分の有無検出だけでなく、得られた p_k から重複成分の周波数を求めることもできる。低い方の音の基本周波数を f_1 とすると、 k 倍音の重複成分の周波数 kf_1' は次式で求まる。

$$kf_1' = kf_1 + \frac{p_k}{2\pi} \quad (3)$$

また、単音で非調和性を持つ楽音に対して本手法をそのまま適用することはできないが、単音の倍音周波数は非調和性係数から算出される規則性を持ったものとなるので、その規則性から外れていることを位相変化から得られる可能性がある。

6. まとめ

楽音の位相時間変化を測定する方法を示し、その結果から音高推定困難和音の音高を推定する手法を提案し、実際に一定位相特性を持つ楽器音同士の音高推定困難和音の音高推定が可能なことを確認した。今後は、非調和性を持つ楽音に対する音高推定法の検討をする予定である。

参考文献

- [1] 山口満, “くし形フィルタに基づく歌唱音の自動採譜に関する研究,” 豊橋技術科学大学博士学位論文, pp. 66-72, Jan 2005.
- [2] 森田健夫, 山口満, 田所嘉昭, “並列構成くし形フィルタの出力値に注目した採譜のための音高推定法,” 信学論, vol.J87-D-II, no.12, pp.2271-2279, 2004
- [3] 後藤 真孝, “実世界の音楽音響信号を対象としたメロディーとベースの音高推定”, 情報処理学会音楽情報科学研究会 研究報告 99-MUS-31-16, Vol.99, No.68, August 1999.
- [4] Curtis Roads, “コンピュータ音楽 歴史・テクノロジー・アート”, 東京電機大学出版局, p. 421, January 2001.
- [5] 小野, 齋藤, 小沢, “自動採譜のためのGAを用いた混合音推定法,” 計測自動制御学会論文集 Vol.33, No.5, pp.417-423, 1997
- [6] 植田, 橋本, “音源分離のためのブラインドコンポジションアルゴリズム,” 情報処理学会論文誌, Vol. 38, No.1, pp.146-157, 1997
- [7] 柏野, 木下, 中臺, 田中, “音楽情景分析の処理モデルOPTIMAにおける和音の分析,” 信学論, Vol.J79-D-II, No.11, pp.1762-1770, 1996
- [8] 後藤真孝, 橋口博樹, 西村拓一, “RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース,” 情処研報, 2002-MUS-45, pp.19-26, 2002.
- [9] 城戸健一, “デジタル信号処理入門,” 丸善, p.134, 1985.
- [10] N.H.フレッチャー, T.D.ロッシング, “楽器の物理学,” シュプリンガー・フェアラーク東京株式会社 p. 362, 2002.