

スペクトログラムの二階偏導関数を用いた調波成分のリッジ推定と 正弦波モデリングへの応用*

☆丸林諒, 赤石夏輝, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

正弦波モデリングは、振幅・周波数変調正弦波の和によって信号を表現する手法である。これによって信号を少数のパラメータで扱えるため、信号圧縮、音声分析、音声合成など様々な音響信号処理に利用することができる。音声信号に含まれる調波成分はスペクトログラム上では時間方向に連続したリッジとして現れるため、正弦波モデリングにおいてスペクトログラムのリッジを高精度に推定することが望まれる。しかし、通常のリッジ推定はスペクトルピークの逐次探索によって実現されており、スペクトルピークの見落としや不連続的な変化に対応できない場合がある。そこで、スペクトルピークの大域的推定法と、対数振幅スペクトログラムの二階偏導関数から構成されるヘシアンに着目し、固有値および固有ベクトルによるリッジ推定法を提案する。さらに、提案手法で推定したリッジの応用として、正弦波モデルによる音声信号の解析および合成を行った。

2 正弦波モデリング

音響信号 $x(t)$ を振幅・周波数変調正弦波の和

$$x(t) = \sum_{k=1}^K a_k(t) \cos(\phi_k(t)) + r(t) \quad (1)$$

によって表現することを正弦波モデリングと呼ぶ。ただし、 $a_k(t) (\geq 0)$ および $\phi_k(t) ((d\phi_k/dt)(t) > 0)$ は第 k 成分の瞬時振幅および瞬時位相であり、時間的に滑らかに変化するものとする。また、 $r(t)$ は残差を表す。全ての瞬時振幅および瞬時位相 (または瞬時周波数) を推定することで信号をパラメータ表現でき、各成分を個別に解析・処理することができる。

正弦波モデリングの代表的な手法の一つに、スペクトログラムのリッジ推定がある。リッジとは山の尾根を意味し、スペクトログラムにおいてはスペクトルピークの時間的に繋がった曲線を表す。スペクトログラムのリッジを推定できれば、その各点における振幅と位相 (または周波数) を読み取ることで、振幅・周波数変調正弦波のパラメータを得ることができる。そのため、高精度なリッジ推定法は高精度な正弦波モデリングを実現する手段となる。

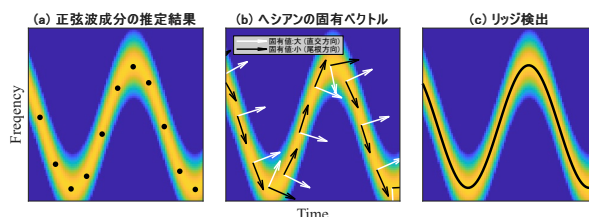


図-1 提案手法によるリッジ推定の例. 左のようにスペクトルピークを検出し、中央のように各点におけるヘシアンの固有ベクトルからリッジ方向 (黒矢印) を求めることで、右のように連続したリッジを推定する。ただし、視認性のためにスペクトルピークは間引いて表示した。

通常のリッジ推定は、スペクトログラムのピークを貪欲法などで探索するが、その精度は以下の二つの理由で制限される。一つ目は、スペクトルの離散化によるピーク位置の誤差である。本来スペクトルは連続関数なので、そのピークは導関数の零点として連続関数の意味で与えられるが、離散フーリエ変換が周波数軸を離散化する影響で、スペクトルのピーク位置に誤差が生じる。二つ目は、リッジの探索における推定ミスである。与えられたピーク位置を時間的に接続することでリッジを推定するが、その接続には貪欲法などのアルゴリズムが用いられており、誤りなく適切にリッジを接続するのは容易ではない。

3 提案手法：偏導関数に基づくリッジ推定

我々は以前、各時間フレームにおける正弦波成分の全てのピークを同時かつ高精度に計算する手法を提案した [1]。離散信号のスペクトルは三角多項式で表される連続関数なので、多項式の根を求めることで零点を計算できる。また、三角多項式の各項は解析的に微分できるので、微分したスペクトルの零点を固有値分解により求めることで、全てのスペクトルピークを同時に計算することが可能である。

本稿では、さらに二階の偏導関数を計算することで、対数スペクトルのピークの二次関数近似を行い、各ピーク点におけるリッジ方向の推定を行う。

3.1 ヘシアンを用いたリッジ推定

本稿では、対数振幅スペクトログラムの二階偏導関数が解析的に計算可能であることに着目する。すなわち、対数振幅スペクトログラム $S(f, t) = \log |X(f, t)|$

*Ridge estimation of harmonic components using second-order partial derivative of spectrogram and its application to sinusoidal modeling. By Ryo MARUBAYASHI, Natsuki AKAISHI and Kohei YATABE (Tokyo University of Agriculture and Technology).

の二階偏導関数から構成されるヘッセ行列

$$H(f, t) = \begin{bmatrix} \frac{\partial^2 S}{\partial t^2}(f, t) & \frac{\partial^2 S}{\partial t \partial f}(f, t) \\ \frac{\partial^2 S}{\partial f \partial t}(f, t) & \frac{\partial^2 S}{\partial f^2}(f, t) \end{bmatrix} \quad (2)$$

が解析的に計算できる．ただし各偏導関数は，窓関数の時間・周波数領域微分によって与えられる．ヘシアン固有値および固有ベクトルは，その点における変化方向および変化量を与えるので，ヘシアンを固有値分解することでリッジ方向を計算できる．

ヘシアンの固有値分解によって計算したリッジを図-1に示す．まずスペクトルピークを導関数の零点によって求め[1]，その各点におけるヘシアンの固有ベクトルを計算する．すると，小さい方の固有値に対応する固有ベクトルは図-1 (b) のようにリッジ方向を向いているので，その方向にスペクトルピークを接続することで，図-1 (c) のような連続的に接続されたリッジを推定できる．各時間フレームにおいて全てのスペクトルピークを同時に検出でき，その各点における多項式の微分によってヘシアンの計算ができるので，提案法は離散化誤差の影響を受けず，スペクトルピークを取りこぼすこともない．なお，ピーク同士の接続の際にリッジの急峻な変化を捉えきれないことがあるが，その対処については今後の課題である．

4 正弦波モデリングへの応用

提案手法の応用例として，音声の解析と合成を行った．音声信号に対してリッジ推定を行い，各調波成分をパラメータ表現することで，そのパラメータを解析し，調波成分を再合成できる．本稿ではJVS corpus [3]内の女性音声を用いた．

解析の一例として，各調波成分の位相の変化を図-2に示す．音声信号の第10倍音までのリッジを求め，そのそれぞれの成分について，基本周波数成分との位相差を算出した¹．推定されたリッジにおける位相差は図-2の下段に示している．右下の拡大図を見ると，矢印で示した付近で40度ほど位相差が変化しており，上段のスペクトログラムと見比べると，音声の共振と対応していることが見て取れる．

推定された振幅・周波数変調正弦波に加えて，ボコーダの非周期成分を足し合わせることで音声合成を試みた．非周期成分の推定にはWORLDに含まれるD4C [2]を利用した．合成結果を図-3に示す．高周波帯域はSN比が低いので，正弦波モデリングできる調波の数には上限があるが，非周期成分を足し合わせたことで不足していた高周波帯域が補われ，良好な

¹成分ごとに瞬時周波数が異なるので，各成分の瞬時周波数が基本周波数と一致するように整数倍して差を計算した．

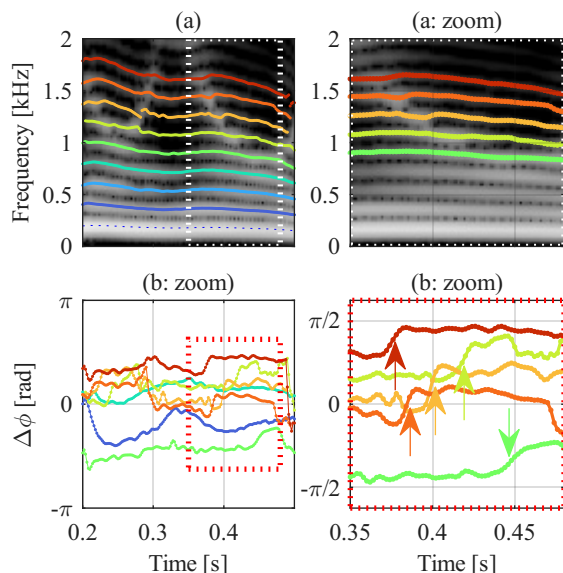


図-2 リッジ推定を用いた調波成分の位相差．(a) 推定された第10倍音までのリッジと，(b) 基本周波数成分との位相差を示している．左上では基本周波数成分を青破線で示しており，白破線の四角部分の拡大図を右上に示した．左下では各成分の位相差を示しており，赤破線の四角部分を拡大の拡大図を右下に示した．ただし，拡大図では見やすさのために第5倍音から第9倍音までを示している．また，位相差が変化する部分に矢印を表示した．

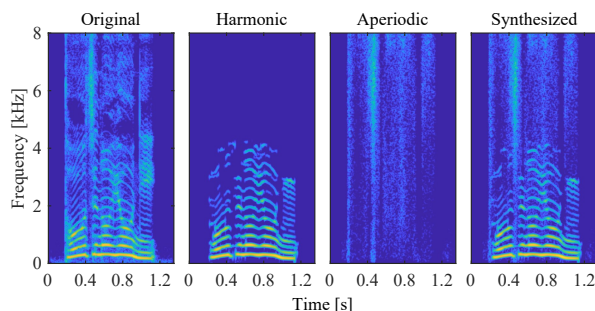


図-3 音声合成結果．左から，元信号，調波成分，非周期成分，再合成信号を示している．

合成結果を得ることができた．合成音の周波数分布を再現するように非周期成分にフィルタをかけることで，さらに自然な音声に近付けることができる．

5 むすび

本稿では，対数振幅スペクトログラムの二階偏導関数を用いたリッジ推定を提案し，正弦波モデリングに応用した．今後は，窓関数のパラメータや位相の微分との関係について調査する所存である．

参考文献

- [1] 丸林諒, 赤石夏輝, 矢田部浩平, “複素スペクトログラムの時間・周波数偏導関数の零点を用いた正弦波成分の推定,” 日本音響学会講演論文集, pp. 209–210 (2025.9).
- [2] M. Morise, “D4C, a band-aperiodicity estimator for high-quality speech synthesis,” *Speech Commun.*, **84**, 57–65 (2016).
- [3] S. Takamichi, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji and H. Saruwatari, “JVS corpus: free Japanese multi-speaker voice corpus,” *arXiv* (2019).