声門の閉鎖期の音声信号によるIAIFに基づく声門流推定*

☆河内太一，モクタリ パーハム，森川大輔（富山県立大）

1 はじめに

有声音声の音源である声門体積流波形（声門流）は，音声生成のメカニズムを解明する上で重要である。声門流は直接測定することが困難だが，声門逆フィルタリング（GIF）を用いることで推定されている。代表的なGIFとしてIterative Adaptive Inverse Filtering（IAIF）[1]がある。また，IAIFの改良手法としてIterative Optimal Pre-emphasis（IOP）[2]と組み合わせた IOP-IAIF などが提案されている。しかし，正確な声門流の推定はまだできていない。

IAIF や IOP-IAIF における声道特性を表すスペクトル包絡の分析区間は，声門の開放期と閉鎖期が複数含まれる音声信号である。その信号を自己相関法によって線形予測分析（LPC）することで，スペクトル包絡モデリングを行う。しかし，声門の開放期が含まれる音声信号では IAIF のスペクトル包絡モデリングに気管の影響が含まれているので，問題になると考えられる[3]。一方，閉鎖期のみの信号を用いてスペクトル包絡モデリングを行う方法にClosed Phase（CP）分析がある[4]。IAIFのスペクトル包絡モデリングにCP分析を組み合わせることで，開放期がスペクトル包絡モデリングに与える問題を解決できると考えられる。そこで本稿では，声門流推定精度の向上を目的とし，IAIF と CP 分析を組み合わせたIAIF-CP と IOP-IAIF-CP を提案し，従来法と推定誤差を比較した結果について述べる。

2 提案方法

提案方法はIAIFにおける2ヶ所のスペクトル包絡モデリングのステップで，従来のLPCを共分散によってLPCを行うCP分析に置き換えたものである。閉鎖期は真の声門流の最大振幅の5 %以下の区間として求める。その他は IAIF に基づく声門流推定と同様に処理し，声道特性をモデル化するための線形予測係数であるM_v，声門特性のモデル化するための線形予測係数M_g，口唇放射係数bのパラメータを用いる。

3 評価方法

3.1 音声データ

音声データにはOPENGLOTのRepositoryIIの男性モデル[5]を用いた。このデータは音声生成の計算物理モデルで生成された音声で，声門流（真の声門流）のデータが含まれる。サンプリング周波数を8 kHzにダウンサンプリングして用いた。音声データは母音4種，基本周波数(F0)4種，声質3種の合計48個のうち40個を用いた。8個の音声データは声門の閉鎖期の区間が短く, CP分析によるモデル化ができないため使用しなかった。それぞれの音声データの0.75 sから50 msの1つのフレームに対して声門流推定を行った。

3.2 パラメータ調整

声門流の推定結果はパラメータに依存するため[6]，パラメータの調整が重要である。パラメータはM_vは6～10の3段階，M_gは3～6の4段階，bは0.750～0.999の250段階を用いる。真の声門流と推定した声門流のRMSEが最小となるようにパラメータを調整する。また，提案法では分析区間に複数の閉鎖期が存在するが，RMSEが最小となる閉鎖期を用いて声門流推定を行う。

3.3 声門流推定の評価指標

声門流推定誤差には，時間領域の指標であるRMS, 声門流の最大振幅と微分声門流波形の最小振幅の比であるNormalized Amplitude Quotient（NAQ）と，周波数領域の指標である声門流のスペクトルの基本周波数の振幅と高調波の振幅の総和の比 Harmonic Richness Factor（HRF），声門流スペクトルのF0と第2次高調波の振幅の差のH1-H2を用いた。それぞれの指標を用いて真の声門流と推定した声門流の誤差を算出する。また，指標誤差を用いて声門流推定精度を調べるため，それぞれ

* IAIF based glottal flow estimation using the speech signal in the glottal closed phase, by KAWAUCHI Taichi, MOKHTARI Parham and MORIKAWA Daisuke (Toyama Prefectural University).

の中央値を求め，それらの割合の値の平均をとることで声門流推定誤差（GIFerror）を算出する。

4　結果と考察

IAIF，IOP-IAIF，IAIF-CP，IOP-IAIF-CP の声門流推定法をそれぞれの声門流推定の指標誤差の結果を Fig. 1 に示す。十字は外れ値，星印は中央値が低い手法を示す。RMS は IOP-IAIF，NAQ は IAIF-CP，HRF，H1-H2 は IOP-IAIF-CP が最も低い値になった。NAQ, HRF, H1-H2 では, IAIF より IAIF-CP, IOP-IAIF より IOP-IAIF-CP の方が優れている。これらの指標は声門流から声質を推定する指標であり，提案法は声質の指標に対して有効であると考えられる。

Fig. 2 (a) は音声データ全体，(b) から (d) はそれぞれ母音，F0，声質別の声門流推定誤差の結果を示す。(a) より，提案法の方が従来法よりも誤差が小さく，IOP-IAIF-CP が最も小さくなった。(b) から (d) より母音は/i/, F0 は 110 Hz 以下，声質は pressed で提案法が優れている。低い F0，閉鎖期の長い声質の音声信号では CP 分析が有効であり，先行研究と同様の結果となった[3]。母音/i/において従来法では第 1 フォルマント（F1）が低いことで全極モデリングの推定に偏りを与えることにより音源と声道の分離が困難になるため[7]，声門流推定に影響したと考えられる。また，母音/i/の F2 と F3 が従来法で 1 つのピークとしてモデリングされていることが要因になり，提案法がよくなったと考えられる。

5　まとめ

本研究では IAIF に基づく声門流推定法におけるスペクトル包絡モデリングを CP 分析と組み合わせた方法を提案し，従来法と比較した。その結果，母音/i/，低い F0，閉鎖期の長い声質の音声信号で提案法が有効であった。今後は F0 や声質を細かく変化させた音声データで調べる。また，声門流推定に用いるパラメータ調整方法も課題である。

謝辞

本研究の一部は，カシオ科学振興財団の支援を受けた。

参考文献

[1] Alku, Sp. Comm., 11, 109-118, 1992.
[2] Mokhtari & Ando, Proc. Interspeech, 1044-1048, 2017.
[3] Yegnanarayana & Veldhuis, IEEE Trans. SAP, 6(4), 313-327, 1998.
[4] Wong *et al.*, IEEE Trans. ASSP, 27(4), 350-355, 1979.
[5] Alku *et al.*, Sp. Comm., 107, 38-47, 2019.
[6] Mokhtari *et al.*, Sp. Comm., 104, 24-38, 2018.
[7] Alku *et al.*, JASA, 125(5), 3289-3305, 2009.

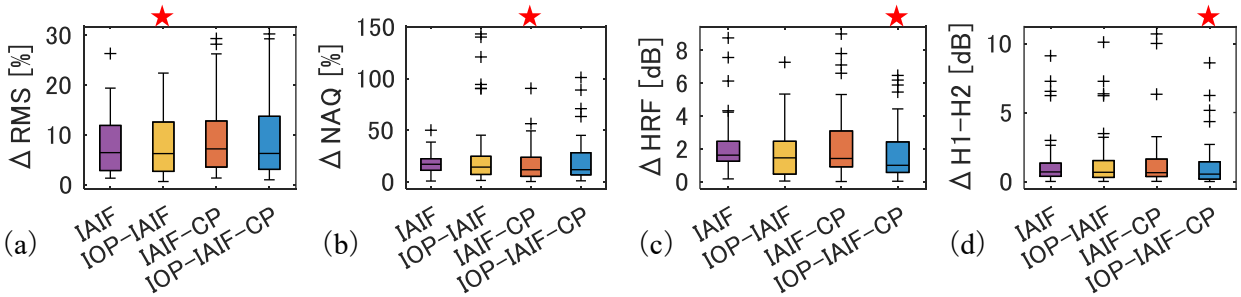

Fig. 1 声門流推定の指標誤差

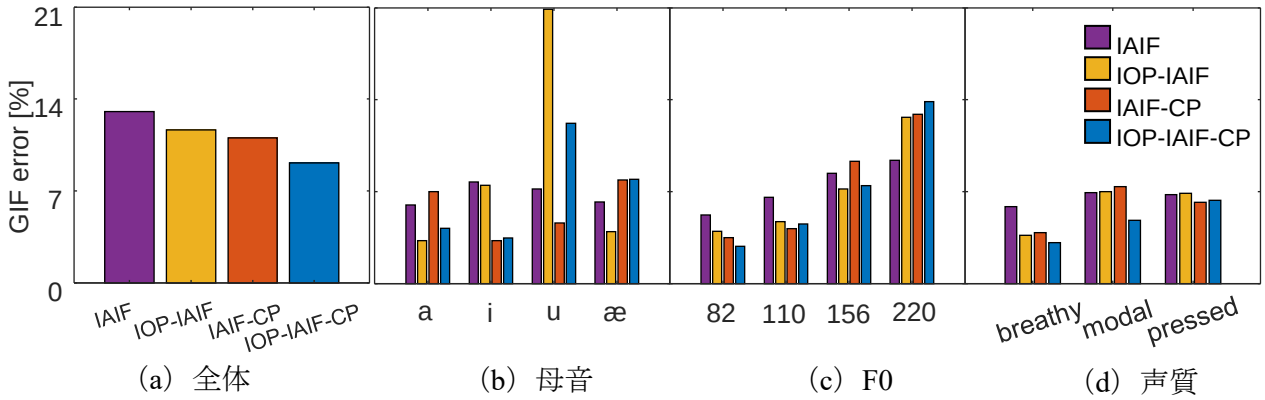

Fig. 2 声門流の推定誤差