

プリエンファシスの適用回数がIAIFに基づく声門流推定の精度に与える影響*

☆平野雄大, モクタリ・パーハム, 森川大輔 (富山県立大)

1 はじめに

有声音の音源である声門体積流波形 (声門流) は, 直接測定することは困難だが, 声門逆フィルタリング (GIF) での推定が検討されている. Iterative Adaptive Inverse Filtering (IAIF) [1]は GIF に広く使用されているアルゴリズムであり, 線形予測 (LP) 分析で声道共鳴と口唇放射をモデル化し, フィルタリングでこれらの特性を除去することで声門流を推定する. IAIF の改良手法として Iterative Optimal Pre-emphasis (IOP) [2]や Glottal Flow Model (GFM) [3]が提案されたが, 必ずしも正確な声門流の推定はできない. IAIF とこれらの手法は, 分析する音声のスペクトルの傾きを減らすプリエンファシスの適用回数が異なり, IAIF では 1 回, GFM では 3 回, IOP ではスペクトル傾斜が平坦になるまで繰り返す. IAIF と IOP, GFM の評価・比較により, スペクトル傾斜が声門流の正確な推定に重要であることが示唆された[4]. しかし, 声門流推定に最適なプリエンファシスの回数やそのスペクトル傾斜への影響は明らかになっていない.

そこで本稿では, IAIF の声門流推定精度の向上を目的として, スペクトル傾斜が声門流推定に与える影響を明らかにするため, プリエンファシスの適用回数を変化させて推定した声門流の推定誤差について述べる.

2 プリエンファシス

音声分析では, 高次のフォルマントを強調して, スペクトルの全体的な傾きを減らすために $1 - \mu z^{-1}$ の形式のフィルタ (プリエンファシスフィルタ) でプリエンファシスを 1 回音声データにかける[5]. μ の値は音声データに 1 次の LP 分析を行うことで算出し, 有声音の場合は 1 に近い値となり, 音声信号の近似的な微分をとることになる.

IOP は, 音声スペクトルの全体的な傾斜を完全に平坦にするために, 係数 μ がほぼ 0 に等しくなるまでプリエンファシスを繰り返す

手法で, 声門流を推定するための声道共鳴の正確なモデル化のために提案された[2][4].

3 実験方法

プリエンファシスの適用回数を 1~15 回まで変化させて IAIF を実行した. IAIF を自動化するには, 声道共鳴をモデル化するための LP 次数 M_v と声門流スペクトルをモデル化するための LP 次数 M_g , 口唇放射係数 b の値を決める必要があり, 声門流の推定結果はこれらのパラメータ値に大きく依存する[4].

そこで, 音声生成の計算モデルによって基準声門流 g_{ref} がわかっている音声データベース[4] (母音 6 種, 基本周波数 10 種, 声質 11 種, 合計 660 個) を用いて 3 つのパラメータ値を変化させて推定した声門流 g_{est} と g_{ref} の差が最も小さくなるように M_v , M_g , b の値を決定した. g_{est} と g_{ref} の差の評価指標は以下の 4 つ誤差の平均である.

最初の誤差は, g_{est} および g_{ref} をそれぞれ微分した微分推定声門流と微分基準声門流の二乗平均平方根誤差 (Δ_{RMSE}) とした.

第二の誤差は, 時間領域のパラメータ Normalized Amplitude Quotient (NAQ)[6]に基づき, g_{est} と g_{ref} のそれぞれの NAQ を算出し, g_{ref} の NAQ からの偏差の絶対値 (Δ_{NAQ}) とした.

第三の誤差は, 周波数領域のパラメータ Harmonic Richness Factor (HRF)[7]に基づき, g_{est} と g_{ref} のそれぞれの HRF の差の絶対値 (Δ_{HRF}) [dB] とした.

第四の誤差は, 周波数領域のパラメータ H1-H2 に基づき, g_{est} と g_{ref} のそれぞれの H1-H2 の差の絶対値 (Δ_{H1-H2}) [dB] とした.

これらの 4 つの誤差について, 標準 IAIF (プリエンファシス 1 回) の誤差 $\Delta_{Standard}$ と比較してプリエンファシス適用回数による相対的な誤差の割合 $\delta\Delta$ [%] を算出した.

$$\delta\Delta = \frac{\Delta - \Delta_{Standard}}{\Delta_{Standard}} \quad [\%] \quad (1)$$

最後に 4 つの $\delta\Delta$ の平均値が最小となる M_v , M_g , b の値の組み合わせを決定した.

* Effect of the number of times pre-emphasis is applied on the accuracy of glottal flow estimation based on IAIF, by HIRANO Yudai, MOKHTARI Parham and MORIKAWA Daisuke (Toyama Prefectural Univ).

4 結果

複数回のプリエンファシスフィルタのスペクトル傾斜の合計を総スペクトル傾斜 ST_{total} [dB/oct] とする。660 の音声のそれぞれの IOP と誤差が最小のとき (Best) の ST_{total} の関係を Fig. 1 に示す。Fig. 1 の横軸は、IOP によって決められた ST_{total} 、縦軸は、Best の ST_{total} であり、各点の色は Best のプリエンファシス適用回数を示している。すべての音声で IOP に対して Best の ST_{total} は同じか小さかった。また、Best のプリエンファシス適用回数は 1 回よりも 2 回以上が多く、2 回が最多であった。

声質別の Best の ST_{total} の分布を Fig. 2 に示す。Fig. 2 の横軸は声質を示し、弱く気息性 (1) から強く圧迫した発声 (11) まで 11 段階に分けられており、縦軸は ST_{total} である。各箱は四分位数と中央値が示している。 ST_{total} は最も弱く気息性の発声 (1) では中央値が約 9.3 dB/oct で比較的大きい値に分布しており、強く圧迫した発声になるにつれて分布する値は小さくなり、最も強く圧迫した発声 (11) では、中央値が約 5.3 dB/oct となった。

5 考察

Best の ST_{total} は、全ての音声で IOP の ST_{total} 以下になった。したがって、IOP によって Best の ST_{total} の上限を決定できると考えられる。対角線上に分布している音声は、IOP によって Best の ST_{total} が得られたことを示す。気息性から圧迫した発声になるにつれて ST_{total} の分布する値が小さくなる。このことからスペクトルの傾斜が大きい声質 (1~6) の音声に複数回プリエンファシスを適用することは声門流推定において有効であると考えられる。Fig. 1 および Fig. 2 のグレーで示す領域は、1 回のプリエンファシスフィルタのスペクトル傾斜の範囲である。多くの音声の ST_{total} がグレーの領域よりも大きい値に分布していることや、Best のプリエンファシス適用回数が 1 回よりも 2 回以上が多くなっていることから、声門流推定精度の向上には複数回のプリエンファシスが必要であると考えられる。

6 まとめ

プリエンファシスの適用回数を変化させて IAIF を実行し、声門流の推定誤差を算出した。1 回のプリエンファシスに比べて複数回のプリエンファシスは声門流推定精度を向上させ、

IOP によって誤差が最小のとき (Best) の ST_{total} の上限を決定できることがわかった。

本稿では、基準声門流が分かっている音声データベースによって最適なプリエンファシスの適用回数を決定することができた。今後の課題は、実際の音声に対して最適な適用回数を決める方法を開発することである。

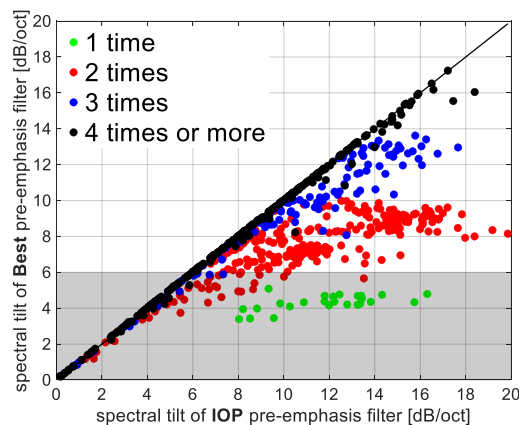


Fig. 1 IOP と Best の ST_{total} の関係

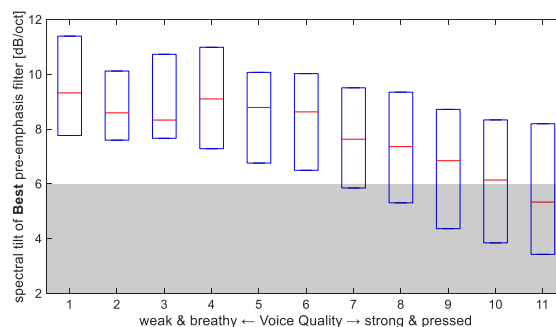


Fig. 2 声質別の Best の ST_{total} の分布

謝辞 本研究の一部は、カシオ科学振興財団の助成金の支援を受けた。音声データを使用させていただいた Brad Story 氏に感謝する。

参考文献

- [1] P.Alku, *Speech Communication*, 11(2-3), 109-118, 1992.
- [2] P.Mokhtari & H.Ando, *Proc. Interspeech*, 1044-1048, 2017.
- [3] O.Perrotin & I.V.McLoughlin, *arXiv preprint arXiv:1712.08034*, 2017.
- [4] P.Mokhtari *et al.*, *Speech Communication*, 104, 24-38, 2018.
- [5] J.D.Markel & A.H.Gray, *音声の線形予測*, コロナ社, 1980.
- [6] P.Alku *et al.*, *J. Acoust. Soc. Am.* 112(2), 701-710, 2002.
- [7] D.G.Childers & C.K. Lee, *J. Acoust. Soc. Am.* 90(5), 2394-2410, 1991.