

時間領域と周波数領域のスパース性を自己調整する LPC*

◎ 鳥居美月, 矢田部浩平 (農工大)

1 はじめに

線形予測分析 (LPC) はソース・フィルタモデルに基づき, 音声を励起信号と声道特性に分離する手法として知られている. 現在でも広く用いられている, 最小二乗法に基づく古典的 LPC [1, 2] は, 予測残差に対してガウス性を仮定している. しかし, 有声音の励起信号はパルス状の信号でありガウ斯的ではないため, 声道特性の推定性能に限界があった. そこで, 予測残差のスパース性を仮定して, ℓ_1 ノルムを用いるスパース LPC [3] が提案された. ただし, ピッチが高い音声の分析においては, 励起信号がスパースでないため, 性能が低下する可能性がある.

これに対し我々は, 音声のピッチが高いほど, 周波数領域では励起信号がスパースになることに着目し, 周波数領域でのスパース性を仮定する周波数領域スパース LPC を提案した [4]. さらに, ピッチに対して頑健な推定を行うため, 時間領域と周波数領域でのスパース性を同時に仮定するハイブリッドスパース LPC も提案した [4]. しかし, 本手法は単一の予測残差を両領域でスパースに促すため, 音声のピッチなどの影響でいずれかの領域でスパースでなくなると, 仮定に矛盾が生じ推定性能が低下する恐れがあった.

そこで本稿では, 時間・周波数の各領域でスパースな成分の和として予測残差をモデル化し, infimal convolution を用いて予測残差を各成分へと配分し分析する手法を提案する. これにより, 一方の領域でのスパース性の低下が他方に及ぼす影響を抑制し, 推定性能の向上を図った. 実験では, 高ピッチの音声や実音声に対して, フォルマント周波数推定の面で提案手法が有効であることが示唆された.

2 線形予測分析

予測残差のガウス性を仮定して, 観測信号 $\mathbf{x} \in \mathbb{R}^{N-P}$ に対して最小二乗問題

$$\min_{\mathbf{a} \in \mathbb{R}^P} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_2^2$$

を解き, 声道特性を表す予測係数 $\mathbf{a} \in \mathbb{R}^P$ を求める LPC [1, 2] は現在も広く用いられている. ここで, $\|\cdot\|_2$ は ℓ_2 ノルム, N は信号長, P は予測次数を表し,

$$\mathbf{X} = \begin{pmatrix} x_P & x_{P-1} & \cdots & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-P} \end{pmatrix} \in \mathbb{R}^{(N-P) \times P}$$

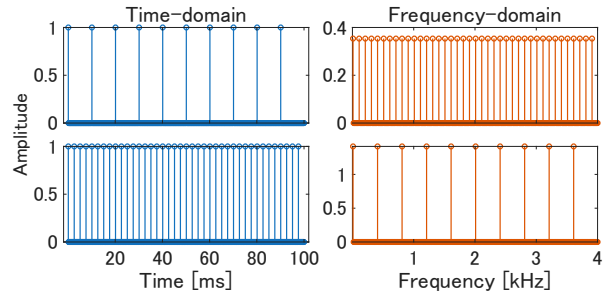


図-1 基本周波数 100 Hz (上段), 400 Hz (下段) における, パルス列で模擬した励起信号の時間信号 (左) と振幅スペクトル (右)

である. ただし, この手法は二乗誤差を用いるため, 外れ値に弱いという性質がある. したがって, 周期的なパルス列を励起信号とする有声音の分析においては, 声道特性の推定に限界がある.

そこで, 励起信号のスパース性を仮定し, 予測残差の ℓ_1 ノルム $\|\cdot\|_1$ を最小化するスパース LPC

$$\min_{\mathbf{a} \in \mathbb{R}^P} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1$$

が提案された [3]. しかし, 本手法が仮定する励起信号のスパース性は基本周波数 f_0 に依存する. 図-1 にパルス列でモデル化した, f_0 が異なる励起信号の時間信号と振幅スペクトルを表す. 時間信号 (図-1 左) を見ると, 100 Hz より 400 Hz の音声の方が, 周期が短く非零要素が多くなっている. このように, 音声のピッチが高い場合, 励起信号はスパースでなくなるため, スパース LPC の性能は低くなる可能性がある.

一方振幅スペクトル (図-1 右) は, f_0 が高いほど調波の間隔が広がり, よりスパースになっている. この点に着目し, 我々は周波数領域でのスパース性を仮定する周波数領域スパース LPC

$$\min_{\mathbf{a} \in \mathbb{R}^P} \|\mathbf{F}(\mathbf{x} - \mathbf{X}\mathbf{a})\|_1$$

を提案した [4]. $\mathbf{F} \in \mathbb{C}^{(N-P) \times (N-P)}$ はフーリエ変換行列を表す. この手法は, ピッチが高い音声に対してのみ良い性能を示した. ピッチに対する頑健性を高めるため, 時間と周波数の両領域のスパース性を同時に仮定するハイブリッドスパース LPC

$$\min_{\mathbf{a} \in \mathbb{R}^P} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1 + \lambda \|\mathbf{F}(\mathbf{x} - \mathbf{X}\mathbf{a})\|_1 \quad (1)$$

も提案した [4]. $\lambda > 0$ は各項のバランスを調整するパラメータである. ハイブリッドスパース LPC は,

*Linear predictive coding with adaptive adjustment of sparsity between time and frequency domains. By Mitsuki TORII and Kohei YATABE (Tokyo University of Agriculture and Technology).

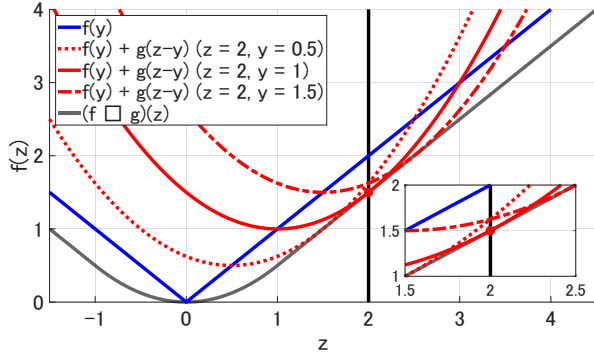


図-2 $f = \|\cdot\|_1$, $g = (\cdot)^2/2$ に対する infimal convolution (黒線). 赤線は $z = 2, y = 0.5, 1, 1.5$ のときの $f(y) + g(z-y)$ である. また, 右下は点 $(2, 1.5)$ 周りの拡大図である.

単一の予測残差 $\mathbf{x} - \mathbf{X}\mathbf{a}$ を 2 つの領域でスパースに誘導する. しかし, 図-1 に示したように, 音声のピッチに応じて励起信号は片方の領域でのスパース性を損う. そのため, 単一の予測残差を 2 つの領域でスパースに促すこの手法は, スパースでない領域での仮定に矛盾が生じ, 性能が低下する恐れがあった.

3 提案手法

前述の課題を解決するためには, 予測残差を時間領域でスパースな成分と周波数領域でスパースな成分の和として扱い, それぞれを時間領域, 周波数領域で最小化すればよい. ただし, 音声のピッチによって各領域のスパース性は異なるため, ピッチに応じた配分が重要となる. そこで本稿では, infimal convolution を用いて, 予測残差を各領域でのスパース成分に適応的に配分して分析する手法を提案する.

下半連続な真凸関数 $f, g: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{\infty\}$ に対し, infimal convolution $f \square g: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ は,

$$(f \square g)(z) = \inf_{\mathbf{y} \in \mathbb{R}^N} f(\mathbf{y}) + g(z - \mathbf{y}) \quad (2)$$

と定義される. これは, 入力変数 z を $\mathbf{y}$ と $z - \mathbf{y}$ の和と捉え, 各関数値の和 $f(\mathbf{y}) + g(z - \mathbf{y})$ の下限を新たな関数として与える. 例として, $f = \|\cdot\|_1$, $g = (\cdot)^2/2$ に対する infimal convolution を図-2 に示す. 下限値を達成する $\mathbf{y}$ は, 目的関数が微分可能な場合は, その勾配が 0 となる点で与えられる. 例えば $z \geq 1$ において, $y + (z - y)^2/2$ の微分がゼロとなるとき, $y = z - 1$ となる. そのため図-2 において, $f(y) + g(2 - y)$ (赤線) と $z = 2$ の交点に注目すると, 確かに下限を与える y は 1 となっている. このように, infimal convolution は入力複数成分の和として扱い, 目的関数値が最小となるように値を配分する性質を持つ.

提案手法では, infimal convolution が持つこの性質を利用する. f, g を時間領域, 周波数領域のスパース性を評価する関数とし, 各領域でスパースな成分の和

Algorithm 1 提案手法のアルゴリズム

Input: $\mathbf{x}, \mathbf{X}, \mathbf{a}^0, \boldsymbol{\eta}_1^0, \boldsymbol{\eta}_2^0, \mathbf{u}_1^0, \mathbf{u}_2^0, \rho, \lambda$

Output: $\mathbf{a}^K$

```

1: for  $k = 0, 1, \dots, K-1$  do
2:    $\mathbf{y}^{k+1} = \frac{1}{2} ((\mathbf{x} - \mathbf{X}\mathbf{a}^k - \boldsymbol{\eta}_1^k + \mathbf{u}_1^k) + \mathbf{F}^H(\boldsymbol{\eta}_2^k - \mathbf{u}_2^k))$ 
3:    $\boldsymbol{\eta}_1^{k+1} = \text{prox}_{(1/\rho)\|\cdot\|_1} [\mathbf{x} - \mathbf{X}\mathbf{a}^k - \mathbf{y}^{k+1} + \mathbf{u}_1^k]$ 
4:    $\boldsymbol{\eta}_2^{k+1} = \text{prox}_{(\lambda/\rho)\|\cdot\|_1} [\mathbf{F}\mathbf{y}^{k+1} + \mathbf{u}_2^k]$ 
5:    $\mathbf{a}^{k+1} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{x} - \mathbf{y}^{k+1} - \boldsymbol{\eta}_1^{k+1} + \mathbf{u}_1^k)$ 
6:    $\mathbf{u}_1^{k+1} = \mathbf{u}_1^k + \mathbf{x} - \mathbf{X}\mathbf{a}^{k+1} - \mathbf{y}^{k+1} - \boldsymbol{\eta}_1^{k+1}$ 
7:    $\mathbf{u}_2^{k+1} = \mathbf{u}_2^k + \mathbf{F}\mathbf{y}^{k+1} - \boldsymbol{\eta}_2^{k+1}$ 
8: end for

```

として予測残差をモデル化する. 領域ごとにスパースな成分に予測残差を配分し, 対応する関数で評価することで, ハイブリッドスパース LPC における仮定の矛盾が解消され, より正確な推定が期待できる.

3.1 定式化

式 (2) において, $f(\cdot) = \lambda \|\mathbf{F} \cdot\|_1, g(\cdot) = \|\cdot\|_1, z = \mathbf{x} - \mathbf{X}\mathbf{a}$ と置くことで, 以下のモデルを提案する.

$$\min_{\mathbf{a} \in \mathbb{R}^P} (\|\cdot\|_1 \square \lambda \|\mathbf{F} \cdot\|_1)(\mathbf{x} - \mathbf{X}\mathbf{a})$$

式 (2) に基づき上式を書き換え, ハイブリッドスパース LPC (式 (1)) を infimal convolution の枠組みで再定義した提案手法を以下のように定式化する.

$$\min_{\mathbf{a} \in \mathbb{R}^P, \mathbf{y} \in \mathbb{R}^{N-P}} \|\mathbf{x} - \mathbf{X}\mathbf{a} - \mathbf{y}\|_1 + \lambda \|\mathbf{F}\mathbf{y}\|_1$$

この最適化問題に, 交互方向乗数法を適用して得られる提案手法のアルゴリズムを Alg. 1 に示す. $(\cdot)^H$ は複素共役転置, $\boldsymbol{\eta}_1 \in \mathbb{R}^{N-P}, \boldsymbol{\eta}_2 \in \mathbb{C}^{N-P}$ は補助変数, $\mathbf{u}_1 \in \mathbb{R}^{N-P}, \mathbf{u}_2 \in \mathbb{C}^{N-P}$ は双対変数, $\text{prox}_{\rho\|\cdot\|_1}$ は $\rho > 0$ をパラメータとするソフト閾値作用素

$$(\text{prox}_{\rho\|\cdot\|_1}(\mathbf{v}))_i = \max\left(1 - \frac{\rho}{|v_i|}, 0\right) v_i$$

である. また, 本稿では予測係数 $\mathbf{a}$ の初期値には自己相関法に基づく従来の LPC [1] の解を用いた.

4 実験

自己相関法に基づく従来手法 [1] (Conv), スパース LPC (T), 周波数領域スパース LPC (F), ハイブリッドスパース LPC (H), 提案手法 (Prop) における, フォルマント周波数の推定性能を比較する実験を行った. 4.1, 4.2 節では, ランダムに生成した第一〜第四フォルマント周波数 (F1-F4) を用いて, 声道特性を模したフィルタを設計し, 励起信号にこのフィルタをかけて合成した音声を分析した. 4.3 節では, vowels database: adults and children [5] に含まれる実音声を分析した. 全ての実験において, サンプリング周波数 24 000 Hz, $\rho = 1000$, 最大反復回数 10 000

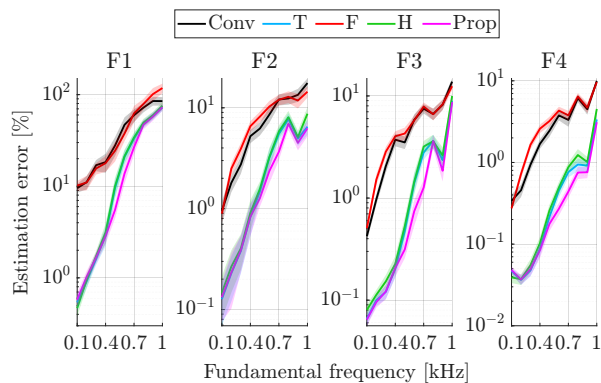


図-3 パルス列を励起信号として合成した音声に対する、基本周波数の変化に対するフォルマント周波数の推定誤差の推移。横軸は基本周波数、縦軸は真のフォルマント周波数に対する推定誤差の相対量 (単位: %) を表す。実線は 100 回の試行で得られた結果の平均、その周りの薄い色は 95 % 信頼区間を表す。なお、推定誤差を見やすくするために縦軸を対数スケールで示した。

回とした。分析の際には、音声に対しプリエンファシス処理 ($\alpha = 0.97$) を行った後、窓長 600 サンプル (25 ms) のハミング窓をかけ、正規化を行った。

4.1 パルス列を用いた音声の分析

パルス列を励起信号として合成した音声に対し、各手法で推定されたフォルマント周波数と、合成に用いた真のフォルマント周波数の誤差を評価した。 f_0 の上昇に伴う推定誤差の変化を確認するため、 f_0 を 100–1000 Hz の範囲で変化させ、各 f_0 に対し、100 組のフォルマント周波数を用いて 100 種類の音声を合成した。 $P = 20$, $\lambda = 0.1$ (H), 5 (Prop) とした際の、フォルマント周波数の推定誤差を図-3 に示す。

全手法、全フォルマントにおいて、 f_0 の上昇に伴い推定誤差が増加する傾向が見られた。励起信号の特徴が仮定と乖離している Conv と F は、他手法より大幅に推定誤差が大きくなった。 f_0 が 500–800 Hz の音声に対しては、全フォルマントにおいて Prop の誤差が H よりも小さくなった。この結果から、音声ごとに柔軟に成分を分割する提案手法のアプローチが、高ピッチの音声の分析において有効に機能することが示された。

4.2 声帯振動モデルを用いた音声の分析

Liljencrants-Fant (LF) model [6] で生成した声帯振動を励起信号として合成した音声に対する、フォルマント周波数の推定性能を評価した。 $P = 20$, $\lambda = 0.3$ (H), 2.7 (Prop) とした際の結果を図-4 に示す。

先ほどの実験結果 (図-3) と同様に、Conv と F の誤差が他 3 手法と比較して大きくなった。 f_0 が 200, 300 Hz の音声に対しては、全フォルマントにおいて Prop の推定誤差が最小となった。パルス列を用いた実験では、 f_0 が 500 Hz 以上の音声に対して、

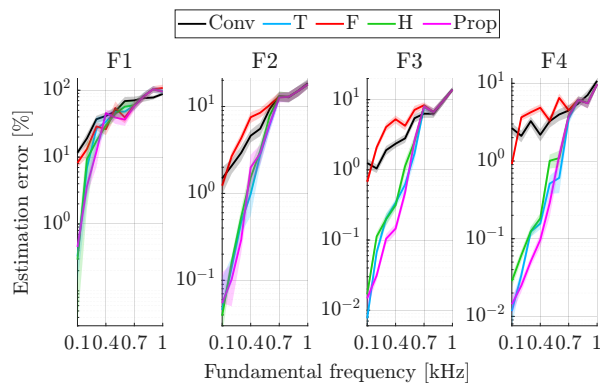


図-4 LF モデルで生成した声帯振動を励起信号として合成した音声に対する、基本周波数の変化に伴うフォルマント周波数の推定誤差の推移。図の見方は図-3 と同様である。

表-1 推定されたフォルマント周波数のフレーム間の差の絶対値和。表中の数値の単位は Hz である。各フォルマントで最もよい値 (最小値) を太字で示した。

手法	男声				女声			
	F1	F2	F3	F4	F1	F2	F3	F4
Conv	70	1698	2560	2313	94	847	1396	776
T	186	626	1370	1311	302	1395	2093	774
F	90	1937	2257	2169	64	608	1086	723
H	126	738	1405	1463	105	794	1307	724
Prop	166	520	1032	1568	113	560	988	688

T および H と Prop の間に有意な差が見られたが、より実際の声帯振動に近い LF model を用いた本実験では、実音声で一般的な基本周波数において、他手法を上回る精度を示した。このことから、実音声に対する提案手法の有効性が示唆された。

4.3 実音声の分析

成人男性と成人女性の音声を分析した。これらの音声の真のフォルマント周波数は未知であるため、本実験では各フレームで推定されたフォルマント周波数の軌跡の滑らかさを評価した。短時間でフォルマント周波数が急激に変化することは考えにくいいため、フォルマント周波数の軌跡の推移が滑らかなほど、推定に安定性があると判断できる。男声の実験において $P = 24$, $\lambda = 0.9$ (H), 1.2 (Prop), 女声の実験において $P = 22$, $\lambda = 3$ (H), 0.4 (Prop) とした際の、音声のスペクトログラムと、推定されたフォルマント周波数の軌跡を図-5 に、フレーム間における、推定されたフォルマント周波数の差の絶対値和を表-1 に示す。

男声の分析では、図-5 上段の青枠の区間において、Prop 以外の手法では軌跡に急峻な変動が見られる一方、Prop は滑らかに推移したことから、表-1 における男声の結果の F2, F3 の値が全手法の中で最小となった。なお、黄色枠で示された終端フレームの F4 は、直前のフレームの結果から急激に低下したが、これは全手法で共通して見られたことから、原因は分析

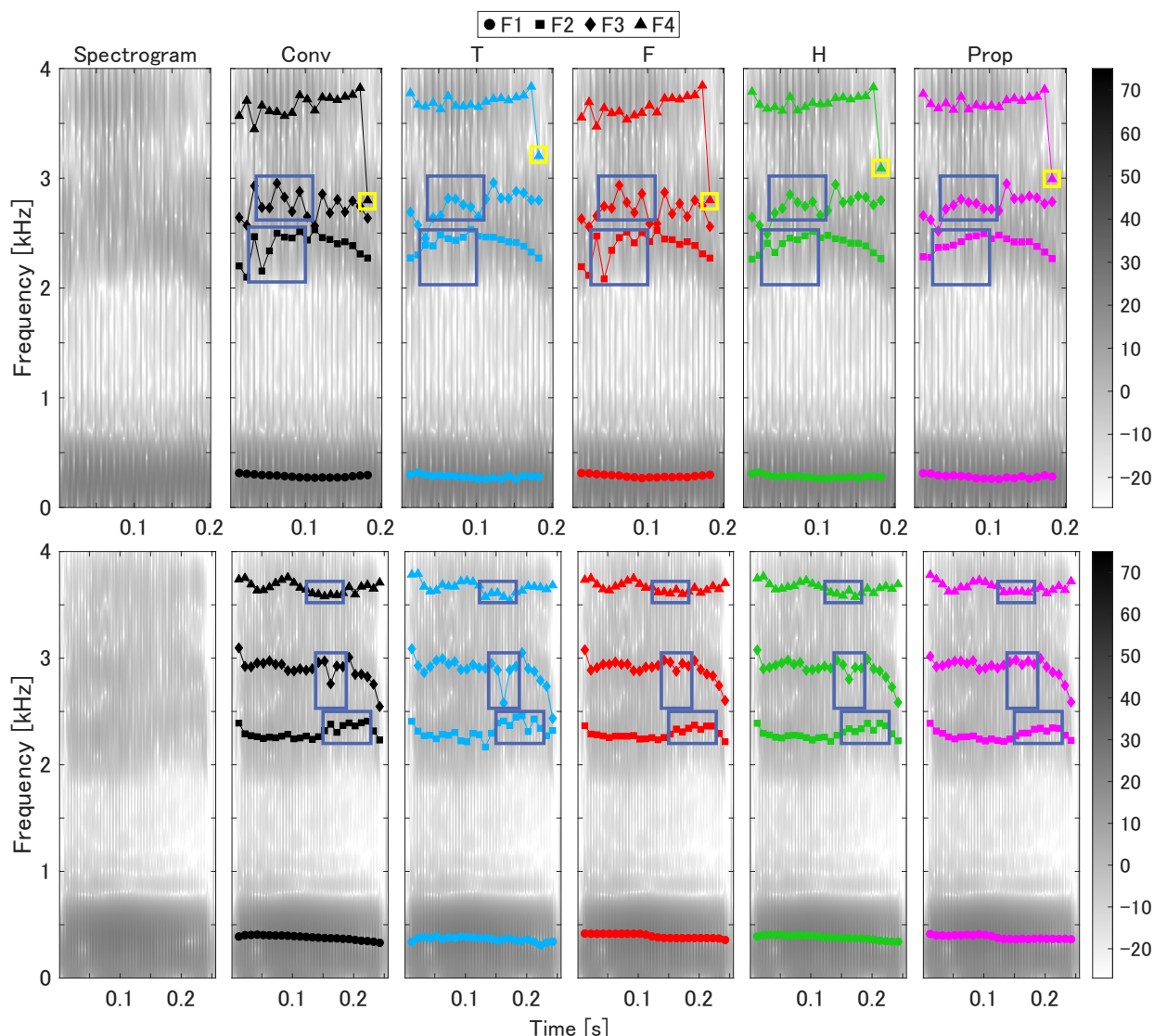


図-5 音声のスペクトログラムと、推定したフォルマント周波数の軌跡. 短時間フーリエ変換では窓長 120 サンプル (5 ms), シフト長 60 サンプル (2.5 ms) とした. LPC 分析ではフレーム長 600 サンプル (25 ms), シフト長 240 サンプル (10 ms) とした. 上段は男声, 下段は女声の分析結果であり, 全手法で共通して見られた推定値の変化が急峻な点を黄色い枠で, 手法間で結果に大きな差が生じた区間を青枠で囲んだ.

手法ではなく音声にあると考えられる. 女声の分析では, f_0 の上昇により男声の実験結果と比較して, T より F の方が滑らかな推移となった. また, Prop は F2, F3, F4 で安定性が最も高くなった. 特に F3 で H の結果との間に大きな差が生じたが, 4.1, 4.2 節の実験結果にも同様の傾向が見られた. このことから, 提案手法は高次フォルマントの推定に対して, 他手法より優位な可能性が示唆された.

5 むすび

本稿では, 時間領域でスパースな成分と周波数領域のスパースな成分の和として予測残差をモデル化し, infimal convolution を用いて予測残差を各成分に配分して分析を行う LPC 分析手法を提案した. フォルマント周波数の推定性能を比較する実験により, ピッ

チが高い音声や実音声の分析において, 提案手法は他手法より優れた推定が行える可能性があることが示唆された. 今後は スパース化が困難な成分にも柔軟に対応するため, ℓ_2 ノルムを用いる手法を検討する.

参考文献

- [1] 板倉文忠, 斎藤収三, “統計的手法による音声スペクトル密度とフォルマント周波数の推定,” 信学論, **53A**, pp. 35–42 (1970).
- [2] B. S. Atal and S. L. Hanauer, “Speech analysis and synthesis by linear prediction of the speech wave,” *J. Acoust. Soc. Amer.*, **50**, pp. 637–655 (1971).
- [3] E. Denoël and J.-P. Solvay, “Linear prediction of speech with a least absolute error criterion,” *IEEE Trans. Acoust., Speech, Signal. Process.*, **33**(6), 1397–1403 (1985).
- [4] 鳥居美月, 矢田部浩平, “周波数領域でのスパース性に基づく線形予測分析,” 日本音響学会講演論文集, pp. 1149–1152 (2025.9).
- [5] P. F. Assmann and W. F. Katz, “Time-varying spectral change in the vowels of children and adults,” *J. Acoust. Soc. Amer.*, vol. 108, no. 4, pp. 1856–1868 (2000).
- [6] G. Fant, J. Liljencrants and Q. Lin, “A four-parameter model of glottal flow,” *STL-QPSR*, **26**(4), pp. 1–13 (1985).