IIR ヒルベルト変換器によるリアルタイムポップノイズ低減*

☆照沼卓磨, 高津航輝, 矢田部浩平 (農工大), 泉悠斗, 高橋祐, 近藤多伸 (ヤマハ株式会社)

1 まえがき

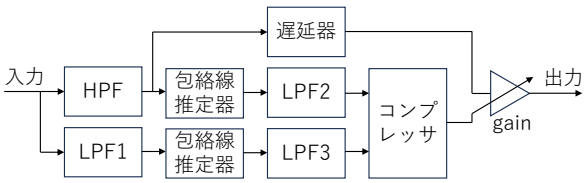
ポップノイズは，話者の発声時の息がマイクロホンに直接かかることで発生し，音質劣化の要因となるため，ポップノイズ低減が必要となる．従来のポップノイズ検出手法 [1–3] を用いる場合，録音データに対して手法を適用し，検出箇所のゲインを調整することでポップノイズを低減できる．しかし，講演会のようなリアルタイムで音声が入出力される状況には適用が困難である．ポップノイズは低周波成分を多く含むという特徴 [1] を持つことから，本研究では，信号の低域と高域それぞれの包絡線の比に注目することでポップノイズを検出し，低減する手法を提案する．実験では，提案手法のポップノイズ低減効果とリアルタイムへの拡張性の両面で有効性が確認できた．

2 ポップノイズ検出

ポップノイズは低周波成分を多く含むことが知られている [1]．従来手法では，その性質を利用して信号を短時間フーリエ変換し，低域のエネルギー変動からポップノイズを検出している [1, 2]．また，機械学習を用いて，スペクトログラムからポップノイズを検出する手法も提案されている [3]．しかし，これらの手法では信号を短時間フーリエ変換する必要があるため，リアルタイムでの運用をする場合には，窓長に応じた遅延が生じる．例えば，音声の分析において典型的な 32 ms の窓を用いた場合，32 ms とフーリエ変換の計算時間を合わせた分のシステム遅延が生じる．

3 提案手法

本研究では，従来手法に比べてより少ない遅延でポップノイズを検出し，低減することを目的とする．本研究の提案手法の概要を図–1 に示す．まず，信号の低域と高域それぞれの包絡線の比を求める．その比が閾値を超えた場合は，ポップノイズの性質 [1] より，ポップノイズが発生したとしてコンプレッサーを適用し，ゲインを調整する．手法内で使用するローパスフィルタ (LPF) 及びハイパスフィルタ (HPF) は 1 次 IIR フィルタとなっている．図–1 における HPF と LPF1 のカットオフ周波数は 100 Hz，LPF2 のカットオフ周波数は 10 Hz，LPF3 のカットオフ周波数は 200 Hz である．包絡線推定には，IIR ヒルベルト変換器を用いた．



図–1 提案手法の概要図

3.1 ヒルベルト変換器による包絡線推定

解析信号 $s(t)$ は，実信号 $x(t)$ とそのヒルベルト変換 $\hat{x}(t)$，虚数単位 j を用いて，式 (1) から求まる．

$$s(t) = x(t) + j\hat{x}(t) \tag{1}$$

フーリエ変換を $\mathcal{F}$ とすると，ヒルベルト変換 $\hat{x}(t)$ は，$x(t)$ の周波数領域表現 $X(f)$ と，ヒルベルト変換の伝達関数 $H_{\mathrm{HT}}(f)$ を用いて式 (2) から求まる．

$$\hat{x}(t) = \mathcal{F}^{-1}[H_{\mathrm{HT}}(f)X(f)] \tag{2}$$

H_{HT} は，サンプリング周波数を f_{s} とすると式 (3) で定義される．

$$H_{\mathrm{HT}}(f) = \begin{cases} -j & (0 < f < \frac{f_{\mathrm{s}}}{2}) \\ j & (\frac{f_{\mathrm{s}}}{2} < f < f_{\mathrm{s}}) \end{cases} \tag{3}$$

式 (3) より，実信号 $x(t)$ とヒルベルト変換 $\hat{x}(t)$ の間には 90 度の位相差があると分かる．実信号を解析信号として扱うことで，式 (4) から包絡線 (瞬時振幅) $A(t)$ を求めることができる．

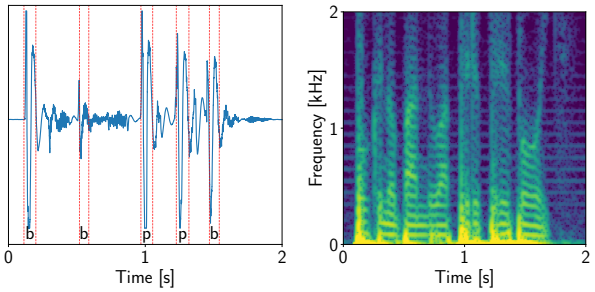
$$A(t) = |s(t)| = \sqrt{x(t)^2 + \hat{x}(t)^2} \tag{4}$$

本研究では，包絡線を得るために IIR ヒルベルト変換器を設計する．このヒルベルト変換器は，実信号から解析信号の実部 $x_{\mathrm{real}}(t)$ と虚部 $x_{\mathrm{imag}}(t)$ それぞれを作るフィルタ対であり，$x_{\mathrm{real}}(t)$ と $x_{\mathrm{imag}}(t)$ の間の位相差が 90 度となるように設計されている．ヒルベルト変換器により得られた解析信号の実部と虚部を用いて式 (5) から推定包絡線 $A_{\mathrm{est}}(t)$ が求まる．

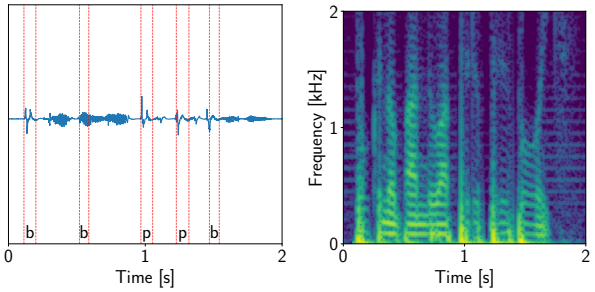
$$A_{\mathrm{est}}(t) = \sqrt{x_{\mathrm{real}}(t)^2 + x_{\mathrm{imag}}(t)^2} \tag{5}$$

理想的には，$x_{\mathrm{real}}(t) = x(t)$ であるが，本研究で設計したヒルベルト変換器では，この等式は実現できていないことに注意が必要である．

*Real-time pop noise reduction using an IIR Hilbert transformer. By Takuma TERUNUMA, Koki TAKATSU, Kohei YATABE (Tokyo University of Agriculture and Technology), Yuto IZUMI, Yu TAKAHASHI and Kazunobu KONDO (YAMAHA Corporation).



図–2　左にポップノイズ低減前の時間波形，右にそのスペクトログラムを示す．色の範囲は 80 dB とした．



図–3　左にポップノイズ低減後の時間波形，右にそのスペクトログラムを示す．色の範囲は 80 dB とした．

3.2　コンプレッサによるポップノイズ低減

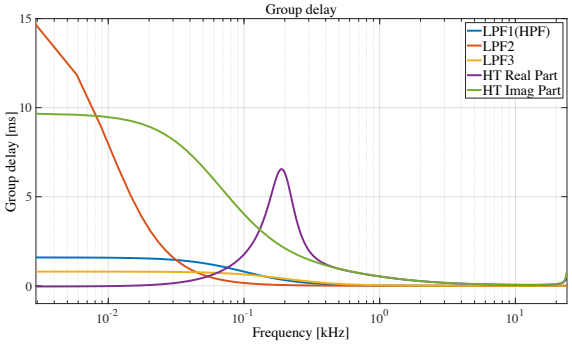
HPF 通過後の信号 $x_{\mathrm{HPF}}(t)$ に対して，ポップノイズ検出時にコンプレッサを適用することで，ポップノイズを低減する．本研究では，元信号 $x(t)$ の低域の包絡線が高域の包絡線の 4.2 倍を超えたときにポップノイズが発生しているとみなし，コンプレッサにより信号の大きさを 5 ms かけて指数関数的に 1/5 に抑制する．信号の大きさが抑制された状態でポップノイズが無くなった際には，100 ms かけて指数関数的に信号の大きさを徐々に元に戻す．なお，遅延器による遅延は 0 ms に設定した．

4　実験

ポップノイズを発生させやすい破裂音 (/b/や/p/) を含む文章を発話し，それを音声データとして記録した．その音声データに対して，提案手法によるポップノイズ低減処理を施した．

図–2 と図–3 にそれぞれポップノイズ処理前後の時間波形とそのスペクトログラムを示す．時間波形には，ポップノイズが生じた区間とそのときの発声内容の発音記号を示している．時間波形を見比べると，ポップノイズ低減後は低減前に比べて，振幅の変化が抑えられている．また，スペクトログラムを見ると，ポップノイズ低減後は低減前に比べて，低域のパワーが小さくなっていることが分かる．以上のことから，提案手法によりポップノイズを検出し，低減ができることがわかる．

提案手法のリアルタイム性を検証するために，使用したフィルタの，サンプリング周波数 48 kHz における群遅延特性を確認する．図–4 に各フィルタの群遅延特性，表–1 に各フィルタの群遅延の最大値を示す．提案手法のシステム遅延は HPF の群遅延のみである．HPF の最大群遅延は約 1.6 ms で，これは音声の分析で典型的な窓長 32 ms の 1/20 以下となる．このことから提案手法は，短時間フーリエ変換が必要となる従来法に比べて大幅に低遅延化できているため，大いにリアルタイム性を有しているといえる．ただし，包絡線の推定には最大約 27.30 ms の遅延が発生し，これによりコンプレッサのゲイン調整が間に合わない場合がある．この場合には遅延器で更に遅延を追加する必要があり，その分追加でシステム遅延が発生する．



図–4　提案手法で使用している各フィルタの群遅延特性

表–1　各フィルタの群遅延の最大値

フィルタ	群遅延の最大値 [ms]
LPF1 (HPF)	1.592
LPF2	15.92
LPF3	0.7957
ヒルベルト変換器 (実部用)	6.552
ヒルベルト変換器 (虚部用)	9.793

5　むすび

本研究では，IIR ヒルベルト変換器を用いたリアルタイムポップノイズ低減手法を提案した．実験の結果，従来法に比べ，より少ない遅延でポップノイズの検出と低減が可能であると分かった．今後は聴取実験を通して，音量の変化と遅延両方の面から違和感の少ないポップノイズ低減ができているか調査する．

参考文献

[1] S. Shiota, F. Villavicencio, J. Yamagishi, N. Ono, I. Echizen, T. Matsui, “Voice liveness detection algorithms based on pop noise caused by human breath for automatic speaker verification,” in *Proc. Interspeech 2015*, pp. 239–243, 2015.

[2] K. Akimoto, S. P. Liew, S. Mishima, R. Mizushima, and K. A. Lee, “POCO: A voice spoofing and liveness detection corpus based on pop noise,” in *Proc. Interspeech 2020*, pp. 1081–1085, 2020.

[3] K. Khoria, A. T. Patil, and H. A. Patil, “On significance of constant-Q transform for pop noise detection,” *Comput. Speech Lang.*, vol. 77, 101421, 2023.