

窓 Fourier 変換と 2 次元変換を用いた音声ノイズ除去

Denoising Audio Signals Using Windowed Fourier Transform and 2-D Transforms

船渡 由梨 (Yuri Funato)¹, 鈴木 俊夫 (Toshio Suzuki)², 石渡 恵美子 (Emiko Ishiwata)³

¹ 東京理科大学大学院 (Graduate School, Tokyo University of Science),

² 神奈川大学 (Kanagawa University), ³ 東京理科大学 (Tokyo University of Science)

e-mail : 1424525@ed.tus.ac.jp

1 はじめに

近年, 元の音声や音楽を保持したまま, 速くノイズを除去する様々な手法が提案されている. その一つに, ノイズ付き音声信号に何らかの変換を施し, そのスパース性を前提としてノイズを除去する方法がある. [1] では, 信号に施す変換として 1 次元離散ウェーブレット変換を用いている. また, [2] では, 窓 Fourier 変換による 2 次元信号化とカーブレット変換を組み合わせた手法が提案されている. カーブレット変換とはカーブレットフレームを用いた積分変換であり, 曲線の特徴を捉えやすいため, 昨今, 画像処理分野で注目されている. しかし, この手法は 2 次元信号処理を含むため, 実行に時間がかかることが課題である. 本研究では他の 2 次元変換, 特にカーブレット変換より計算量が少ない 2 次元離散ウェーブレット変換とシアレット変換に着目し, カーブレット変換の代わりに時間周波数平面に適用することを試みる.

2 ノイズ除去の手順

窓 Fourier 変換とカーブレット変換を用いた [2] のノイズ除去は, 以下の 5 つの手順で行われる.

- 1) ノイズ付き信号に窓 Fourier 変換を適用して, 時間と周波数をパラメータとする 2 次元信号を得る. この 2 次元信号の絶対値のカラーマップをスペクトログラムと呼ぶ.
- 2) 1) の 2 次元信号にカーブレット変換を適用して, カーブレット係数を得る.
- 3) カーブレット係数にハード閾値処理を行う. ハード閾値処理とは, 閾値よりも小さい値を 0 に置き換える処理である.
- 4) 閾値処理したカーブレット係数に逆カーブレット変換を適用する.
- 5) 逆窓 Fourier 変換を適用し, ノイズ除去信号を得る.

本研究では, 2) のカーブレット変換を 2 次元離散ウェーブレット変換, またはシアレット変換に置き換え, それに対応する逆変換を 4) で適用する.

カーブレット変換は, フレームと呼ばれる正規直交基底を一般化した解析手法の一つである. 回転パラメータを用いて関数を表現するため, 曲線を捉えやすく, 画像解析などに用いられる. 本研究で用いるシアレット変換もフレームの一つである. 回転行列の代わりに剪断行列を用いるため, 計算量を抑えつつ画像の斜め方向の特徴を捉えやすい変換となっている.

2 次元離散ウェーブレット変換は, 正規直交基底を用いた解析手法であり, フレームを用いた上記の 2 つの変換よりも計算量の少ない変換である.

3 数値実験

従来法 [2] と, カーブレット変換の代わりに 2 次元離散ウェーブレット変換やシアレット変換を用いた手法を, ノイズ除去後の SN 比と実行時間で比較する. SN 比とは, 元の音声信号とノイズの大きさ

の比を表す指標である．SN 比が大きいほど，音声信号に含まれるノイズが少ないことを表す．本実験では，サンプリング周波数 24kHz，約 9 秒間のスピーチ音声 [3] と，これにホワイトノイズを人工的に加えた音声を用いる．元のスピーチ音声とノイズ付き音声のスペクトログラムを図 1 に示す．ホワイトノイズを加えたことで，元の音声のスペクトログラムで濃い青色で表示されていた成分が水色に変化したことがわかる．このように，ホワイトノイズは，時間周波数平面上の各成分で同程度の強度を持つ．本実験で使用した PC の CPU は 2.6 GHz 6-core Intel Core i7 であり，OS は macOS 14.6.1，メモリは 16GB である．アプリケーションは MATLAB2024a を使用した．

表 1 に実験結果を示す．表 1 の時間は，各手法について上記の 5 つの手順を 100 回試行した平均実行時間である．SN 比は，100 回の試行における最大値を示す．表 1 より，実行時間は，カーブレット変換を用いた手法と比べて，2 次元離散ウェーブレット変換を用いた場合には 99.2%，シアレット変換を用いた場合には 93.4% 削減された．SN 比は，カーブレット変換を用いたとき，最も大きくなったが，2 次元離散ウェーブレット変換やシアレット変換を用いたときも同程度まで上昇した．2 次元離散ウェーブレット変換やシアレット変換を用いても，従来法 [2] と同程度のノイズ低減を達成できると考えられる．

カーブレット変換の代わりに 2 次元離散ウェーブレット変換やシアレット変換で用いることで，大幅に実行時間を短縮でき，2 次元信号処理を用いた音声ノイズ除去手法の実用性の向上が期待できる．ノイズ除去音源のスペクトログラムの特徴や違いについては当日の講演時に述べる．また，スピーチ音声以外の音源に対するノイズ除去結果についても，当日に紹介する．

謝辞 本研究にあたり，佐賀大学の皆本晃弥教授には高速カーブレット変換のプログラム CurveLab の情報をご提供いただきました．心より御礼申し上げます．

参考文献

- [1] Donoho, D. L. and Johnstone, I. M., Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81, 3 (1994), 425-455.
- [2] Chiba, M., Yatabe, K. and Oikawa, Y., Redundant representation of acoustic signals using curvelet transform and its application to speech denoising, *Acoustical Science and Technology*, 36, 5 (2015), 457-458.
- [3] Takamichi, S., Mitsui, K., Saito, Y., Koriyama, T., Tanji, N. and Saruwatari, H., JVS corpus: free Japanese multi-speaker voice corpus, arXiv preprint, 1908.06248, 2019.

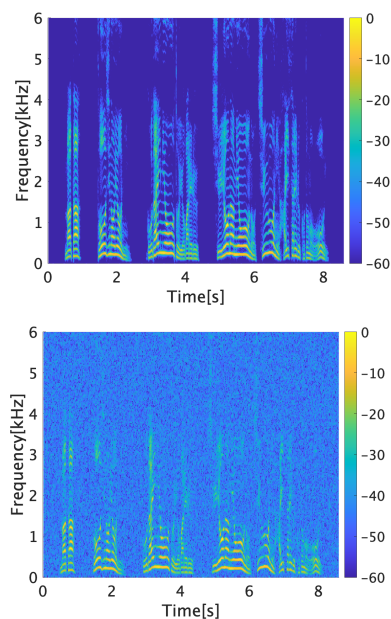


図 1. 元のスピーチ音声 [3] (上) とホワイトノイズを付与した音声 (下) のスペクトログラム

表 1. 実行時間と SN 比

手法	時間 [秒]	SN 比 [dB]
ノイズ付き信号		9.98
カーブレット	18.57	18.92
ウェーブレット	0.14	18.70
シアレット	1.24	17.75

重み付き正則化項を用いた Shearlet システム上でのリングアーティファクト除去について

Ring Artifact Removal on the Shearlet System Using Weighted Regularization Terms

藤井克哉¹(Katsuya Fujii), 鈴木俊夫²(Toshio Suzuki), 藤ノ木健介²(Kensuke Fujinoki)

¹ 畿央大学 (Kio University), ² 神奈川大学 (Kanagawa University)

e-mail: ¹k.fujii@kio.ac.jp

1 概要

CT 画像再構成とは, X 線を利用して断層画像を得る手法であり, 工学, 医学など様々な分野への応用を持つ. しかしながら X 線の照射状況により再構成された画像にはアーティファクトとよばれるノイズが発生する. 本研究では, Shearlet 変換の持つノイズ抑制の性質に加えて, 重み付き正則化項, 特に重み付き核ノルム正則化を組み合わせた手法を提案する. 特にリングアーティファクト (図 1 参照) へ適用し, 一定の画質改善を得た.

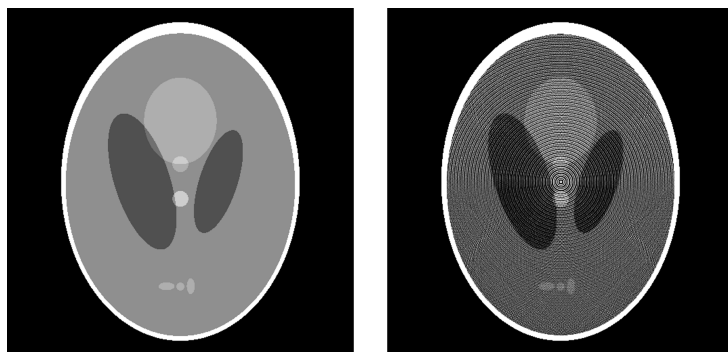


図 1. 左: 元画像, 右: リングアーティファクト画像

2 CT 画像再構成の数学的枠組み

CT 画像再構成では以下のラドン変換と呼ばれる積分変換が重要な役割を果たす [1].

$$\mathcal{R}f(t, \alpha) = p(t, \alpha) = \int_{\mathbf{R}} f(t\theta(\alpha) + s\theta^\perp(\alpha)) ds.$$

ここで, $\theta(\alpha) = (\cos \alpha, \sin \alpha)$ は単位ベクトルであり, $\theta^\perp(\alpha) = (-\sin \alpha, \cos \alpha)$ はそれに直交する単位法線ベクトルを表す. また $\mathcal{R}f$ の集合をサイノグラム (投影データ) と呼ぶ. CT 画像再構成は, 適当なサイノグラムから密度関数 f を再構成する問題である. しかしながら, 一般には完全なサイノグラムを得ることは困難であるため, 再構成には工夫を要する. さらに, 不完全なサイノグラムからの再構成はアーティファクトが発生する場合がある [2]. そこで, Shearlet 変換のもつスパース性 [3] に着目して, 正則化項を組み込んだ最適化問題を提案する. 密度関数 f を Shearlet 展開した $f = \sum_{n \in \mathbf{Z}^3} c_n \psi_n$, $n = (j, k, m)$ において, 最適化問題を以下の式で定義する:

$$\hat{c} = \arg \min_{c = \{c_n\}} \left\{ \frac{1}{2} \left\| \mathcal{R} \left[\sum_{n \in \mathbf{Z}^3} c_n \psi_n \right] - y^\delta \right\|_{L^2(\mathbf{S}^1 \times \mathbf{R})}^2 + \alpha \Phi(c) \right\}.$$

ここで Φ は正則化項であり, y^δ は不完全な投影, α はステップサイズパラメータである. 特に本稿では, 正則化項として重み付き核ノルムを扱う [4]:

$$\|X\|_{*,w} = \sum_{i=1}^{\min(m,n)} w_i \sigma_i(X).$$

ここで, $\sigma_i(X)$ は Shearlet 係数から作られる行列 X の特異値であり, w_i は重みである.

3 数値実験

リングアーティファクトを含む画像 (サイズは 512×512) に対しての本手法のノイズ除去結果を, Shearlet 変換単独のデノイジング結果も含めて図 2 に示す. 実験結果より, Shearlet 変換の持つデノイジングの性質で取り切れなかったノイズがある程度改善されている. Shearlet 変換単独のデノイジングはエッジ強調によるものが大きく, 複雑なアーティファクトの除去には正則化項を用いた手法が有効であることが示唆された.

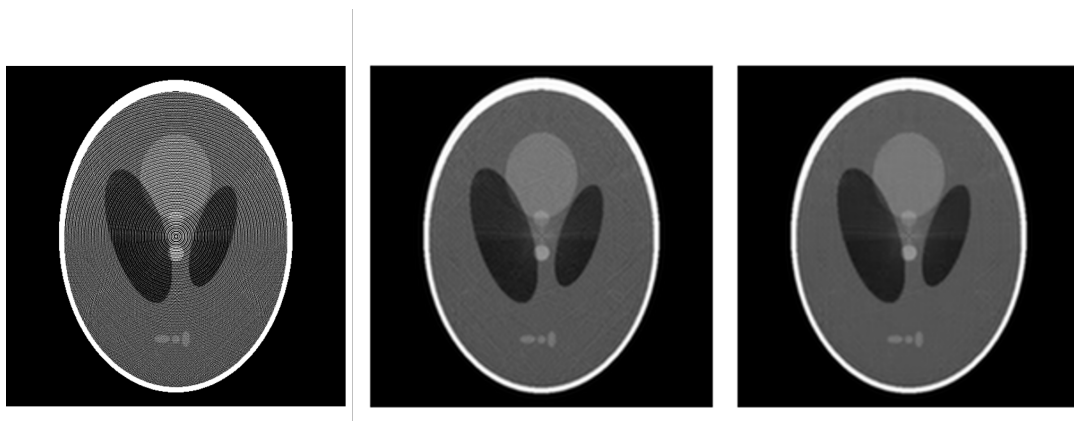


図 2. 左: リングアーティファクト, 中央: Shearlet 変換単独, 右: 提案手法

参考文献

- [1] F. Natterer, The Mathematics of Computerized Tomography, Wiley, New York, 1986.
- [2] F. Edward Boas, D. Fleischmann, CT artifacts: causes and reduction techniques, Imaging Med 4.2 : 229–240, 2012.
- [3] D. Labate, W. Lim, G. Kutyniok, and G. Weiss, Sparse multidimensional representation using shearlets, SPIE Proc. 5914 (2005), SPIE, Bellingham, WA, 254–262, 2005.
- [4] S. Gu, L. Zhang, W. Zuo, and X. Feng, Weighted nuclear norm minimization with application to image denoising, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2862–2869, 2014.