

非負値行列因子分解および回帰型ニューラルネットワークによる 音響信号の可視化と信号予測

輪島 幸治[†]

[†] ロジカル・アーツ株式会社

1. はじめに

情報社会の発展に伴う情報機器の普及で、オンラインミーティングを始めとしたコミュニケーションが増加している。オンラインミーティングにおける議事録作成では、音声から自動でテキストデータを書き起こす Amazon Transcribe¹などの Web サービスも用いられてきている。

ビジネスだけでなく、コンシューマによるオンラインの配信活動も盛んになったことから、音を対象としたメディアは重要インフラの電話やラジオ放送と同様に、研究でも着目されるようになった。一方で、音は、雑音を始めたとした多くのノイズ信号によって、単語や内容が認識できないなどの問題が発生して、オンラインの業務環境に影響が発生する場合もある。本研究では、オンラインコミュニケーションで特に重要とされている音に着目して、音響信号の可視化および予測を試みた。本研究で分析した音響信号を図 1 に、ノイズ信号を図 2 に示す。図 2 のノイズ信号は、NHK Radio English の音響信号にノイズ信号を合成して可視化した。

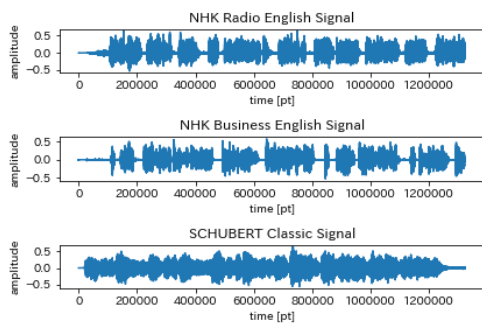


図 1: 分析対象の 3 種類の音響信号

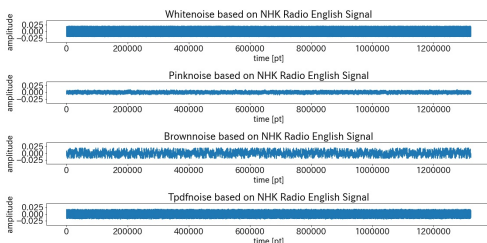


図 2: 分析対象の 4 種類のノイズ信号

さて、音を対象とした研究は、スピーカーなどのハードウェアやプログラムなど盛んに研究が行われている。また、特許や実用新案が検索できる J-PlatPat²では、音源検索、言語識別、シーン抽出など放送分野を中心に出版され権利化多くの特許が確認できる。特許や実用新案だけでなく、音については、2015 年から商標が、音楽的要素からなる音商標を登録の対象とした。

ここで、音声信号処理を始めたとした信号処理では、雑音を始めたとしたノイズ信号の影響で、コミュニケーション内容が相手に十分に伝わらないことも少なくない。図 2 のノイズ信号をヒストグラムにて可視化した例を図 3 に示す。

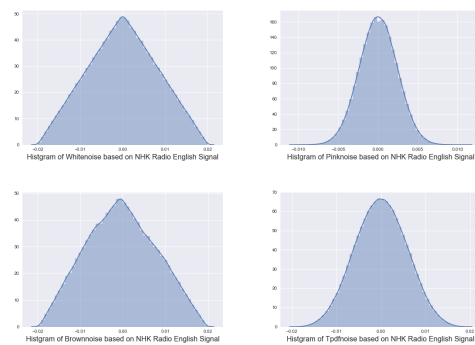


図 3: Noise based on NHK Radio English Signal

本研究では、ホワイトノイズ、ピンクノイズ、ブラウンノイズ、三角形確率密度関数ノイズという 4 種類のノイズ信号を分析対象とした。図 2 および図 3 から明らかのように、ノイズ信号はノイズ種類に応じて波形が異なる。このため、ノイズ除去には多くの方法が必要とされている。

本研究では、音響信号など信号処理の分野で有効な非負値行列因子分解 (NMF: Nonnegative Matrix Factorization) [1] による可視化に用いる。また、回帰型ニューラルネットワーク (RNN: Recurrent Neural Network) を用いた信号予測で信号を予測できるかの評価を行った。

2. 評価実験

本研究の実装は、NMF の実装に scikit-learn³、回帰型ニューラルネットワークの実装には、Keras⁴を用いた [2]。また、音声データから数値行列への変換には SoundFile⁵、ノイズ信号の生成および信号の合成については、SoX⁶を用いている。

1 Amazon Transcribe: <https://aws.amazon.com/jp/transcribe/>

Visualization of acoustic signals and signal prediction using Non-negative Matrix Factorization and Recurrent Neural Network

Koji Wajima[†] (wajimak@nict.go.jp, kwajima@ce.slis.tsukuba.ac.jp)

[†] Logical Arts Corporation, Osaka 541-0054, Japan, wajima@logical.co.jp

2 特許情報プラットフォーム J-PlatPat: <https://www.j-platpat.inpit.go.jp/>

3 scikit-learn : <https://scikit-learn.org/stable/>

4 Keras : <https://keras.io/>

5 SoundFile : <https://pysoundfile.readthedocs.io/>

6 SoX: Sound eXchange : <http://sox.sourceforge.net/>

2.1 非負値行列因子分解を用いた可視化

NMF を用いて音響信号を観測行列 Y として、基底行列 H と観測行列 Y に行列分解を行い、ホワイトノイズが含まれる音響信号を可視化した結果を図4および図5に示す。図4および図5は時系列グラフであり、縦軸は各基底、横軸が時間である。また、色が明るいほど係数が大きい。図4と図5の違いは、行列分解を行う際の乖離規準である。図4では最小二乗法、図5では Kullback Leibler ダイバージェンス規準を用いている。

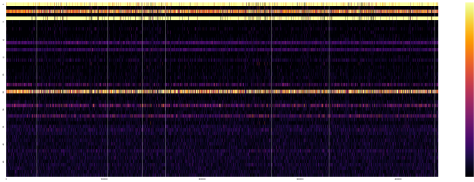


図 4: Whitenoise - SCHUBERT Classic - 最小二乗法

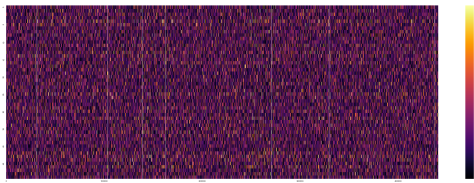


図 5: Whitenoise - SCHUBERT Classic - Kullback Leibler

結果から、図5と比較して、図4では明確な結果得られた。また、本研究では、係数の大きい重要な基底を明らかにする場合における有効な乖離度規準は、最小二乗法が適していることが明らかとなった。このため、雑音の混合信号においても、NMF を用いて乖離度規準や基底数を最適化することで、音響信号の可視化できることが明らかとなった。NMF が適用できれば不要な基底の除去と信号の再構成で、ノイズ信号を除去することが期待できる。

2.2 回帰型ニューラルネットワークを用いた信号の予測

次に、RNN を用いてブラウンノイズとの混合信号の予測およびピンクノイズとの混合信号の予測を図6および図7に示す。図6および図7は、左側が混合信号の波形である。右側が前半部分の青色で着目されているのが混合信号の波形で、後半部分の赤色で着目されているのが混合信号を元に RNN で信号を予測した波形である。

結果から、RNN の適用は、ブラウンノイズ混合信号の予測においては、図6から、RNN によって目的の値に近い値を予測できた。一方で、ピンクノイズ混合信号を予測した図7では、傾向としては類似の波形だが、ブラウンノイズ混合信号の予測と比較して、不明瞭な予測結果となった。このことから、RNN による予測で、ブラウンノイズよりもピンクノイズの影響が大きいという結果が得られた。

3. ま と め

本稿では、音響信号の可視化および信号予測を目的に、NMF と RNN を用いて、ノイズを混合した信号の分析を試みた。結果、NMF では、有効な乖離度規準することにより音響信号の可視化が行えた。また、RNN では、予測に影響のあるノイズ信号が明らかとなり、ブラウンノイズのようにノイズ信号との混合信号でも期待された予測が得られる結果も得られた。

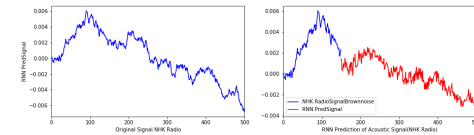
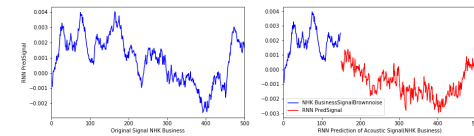
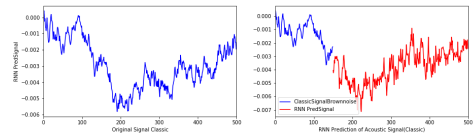


図 6: ブラウンノイズ混合信号の RNN による予測

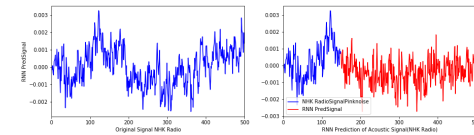
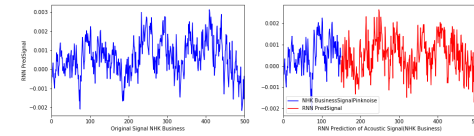
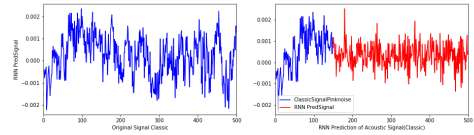


図 7: ピンクノイズ混合信号の RNN による予測

また、本稿では、テレワークの広がりから音響信号を対象としたが、デジタル世代と呼ばれる世代はツール、メディア、コミュニティがアナログ世代と異なり多くの課題が指摘されている[3]。今後の研究課題としては、コミュニティにおける表現、アナログとデジタルとの比較、研究成果を働き方改革などのビジネスに適用するなど、多くの課題解決のために、NMF や RNN など時系列データに効果的な分析手法を検討していきたい。

謝 辞

本稿の研究成果だが、実際に音声信号を体験して、研究の着想を頂いた。ここに記して、国立研究開発法人 情報通信研究機構の皆様へ謝辞を示します。合わせて、2021 年 4 月 30 日に博士（情報学）の学位を授与頂いた筑波大学、筑波大学の博士後期課程における連携大学院の株式会社電通の木暮 啓教授⁷、株式会社電通の皆様、博士後期課程時に在籍していた株式会社セールスフォース・ジャパン（投稿時点：株式会社セールスフォース・ドットコム）の皆様、エンバーポイント株式会社の皆様（所属時：エクスペリアンジャパン株式会社）、横河ソリューションサービス株式会社の皆様へ謝辞を示します。

文 献

- [1] 亀岡弘和. 非負値行列因子分解. 計測と制御, Vol. 51, No. 9, pp. 835–844, sep 2012.
- [2] François Chollet, et al. Keras, 2015.
- [3] 信田和宏. 電通鬼十則の記憶 電通事件批判の中で. C&R 研究所, 2018.