様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月　 ２ 日現在

機関番号：１２６０１
研究種目：若手研究
研究期間：2020～2022
課題番号：２０Ｋ１９８１８

研究課題名（和文）時間領域深層学習と多重解像度解析を融合した音響情景分析の研究

研究課題名（英文）Research on acoustic scene analysis by integrating time-domain deep leraning and multiresolution analysis

研究代表者
中村　友彦（Nakamura, Tomohiko）

東京大学・大学院情報理工学系研究科・特任助教

研究者番号：５０８６６３０８

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：本研究では，時間領域で直接音源分離を行う波形領域深層音源分離モデルと多重解像度解析との間のアナロジーを基に新たな音源分離手法を提案した．具体的には，離散ウェーブレット変換に基づくダウンサンプリング（プーリング）層を提案し，従来の波形領域音源分離手法に比べ，高精度に分離できることを示した．また，提案層を拡張し，事前に定めたウェーブレットだけでなく，深層学習モデルと同時にウェーブレットも学習できることを示した．この拡張により，タスクに応じてより適したウェーブレットを学習により得る方法を確立した．さらに，多チャネル音源分離や重唱分離へと拡張・適用し，提案法の導入により分離性能が向上することを示した．

研究成果の学術的意義や社会的意義
本研究では，時間領域で直接分離を行う深層音源分離モデル（時間領域深層学習）と，信号処理・ウェーブレット解析で培われてきた多重解像度解析を融合する分野横断的方法論を創出した．時間領域深層学習では，高性能な音源分離を実現するように各構成要素のパラメータが学習されるため，各構成要素の機能は明確ではなかった．一方，多重解像度解析は，音源によって適切に設計する必要があるものの，機能が明確な構成要素を用いている．本研究成果は，両者を統合することで深層学習の高性能性と信号処理の高い解釈性を両立する第一歩となるものである．

研究成果の概要（英文）：In this study, we proposed an audio source separation method, multiresolution deep-layered analysis. It comes from our finding that a waveform-domain audio source separation model, Wave-U-Net, resembles multiresolution analysis in downsampling (DS) architecture. Inspired by the resemblance, we developed a DS layer using the discrete wavelet transform. Music source separation experiments showed that the proposed method achieves higher separation performance than conventional waveform-based methods. We also extended the proposed layer so that its wavelets can be trained together with the other components of a deep neural network. This extension paves the way for obtaining suitable wavelets for target tasks in an end-to-end manner. Finally, we applied the proposed methods to monaural vocal ensemble separation and multi-channel audio source separation tasks and demonstrated the effectiveness of the proposed methods through experiments on these tasks.

研究分野：音響信号処理，音楽信号処理

キーワード：音響情景分析　時間領域深層学習　多重解像度解析　音源分離　音響信号処理　深層学習　機械学習

１．研究開始当初の背景

音響情景分析は，音響信号から周囲の状況やいつどこで何が起きているかを推定する技術である．これは，動画・音楽・音声コンテンツの検索をはじめ，監視システム，高齢者の見守りシステム，自動運転などの周囲の環境の把握が必須となる様々なシステムへ応用が可能である．そのため，音響情景分析の要素技術は活発に研究されており，学術的にも産業的にも大きな関心が向けられている．

音響情景分析の起点となる技術の一つは，音源分離（混合音を各音源に分離する技術）である．多くの音源分離手法では，時間周波数変換を用いて得られたスペクトログラム領域で分離を行う．これに対し，近年，スペクトログラム領域での音源分離手法を実現するために導入された仮定（信号の区分定常性，瞬時混合仮定など）を用いずとも高性能な分離を達成する方法として，時系列信号を扱う End-to-End 深層学習の１種である時間領域深層学習が注目されている．

時間領域深層学習では，スペクトログラムを介さず直接音響信号波形を入出力できるため，各種変換操作に起因する歪みや劣化を抑制できる可能性がある．一方で，与えられた学習データに対して，高性能な音源分離を実現するよう各構成要素のパラメータが学習されるため，各構成要素の機能は学習の中で定まる．その際に，必ずしも人間にとって意味のある機能が特定の構成要素で実現されるとは限らないため，学習済みモデルから構成要素の機能を同定することは難しい．この特徴から，何をどのように変更すれば性能が向上するか把握するのは困難であり，発見的に研究が行われているのが実情である．

これに対し，音響信号処理では機能が明確な構成要素を用いて全体の機能を設計することが多い．例えば，多重解像度解析は，入力信号を低周波，高周波帯域の信号に分割する離散ウェーブレット変換（DWT）を構成要素として，複数の時間周波数解像度で信号を解析する．そのため，解釈性は高く各要素に対してどのような変更をすべきかの指針は比較的立てやすい．しかしながら，多くの場合様々な音源に対して適切に要素を選択する必要がある．

２．研究の目的

時間領域深層学習と多重解像度解析を比較すると，時間領域深層学習は大局的な機能から局所的な機能を学習するトップダウン的アプローチであり，多重解像度解析は局所的な機能から大局的な機能を構成するボトムアップ的アプローチとみなせる．これらはいずれも音源分離に有用であり，両者を適切に融合できれば，音源分離に有用な機能を構成要素に導入しつつ，高性能な音源分離を実現できるはずである．そこで，本研究では，時間領域深層学習と多重解像度解析を融合した新たな音源分離手法（多重解像度深層分析）を構築し，分野横断的な音響情景分析技術の創出を目指す．

３．研究の方法

（1）時間領域深層学習と多重解像度解析を融合した深層学習モデルの構築

時間領域深層学習の１つである Wave-U-Net は，繰り返しダウンサンプリング，アップサンプリングを行う構造を持つ．同様に，多重解像度解析も繰り返しダウンサンプリング，アップサンプリングすることで信号を分析・合成する（図１右参照）．この類似性に着眼し，多重解像度解析の特徴を時間領域深層学習へ導入する方法を検討する．具体的には，ダウンサンプリング操作として多重解像度分析で用いられる DWT を利用した深層学習モデルの構築に取り組む．

Wave-U-Net では，ダウンサンプリングを時間方向に入力を間引くことで実現している．しかし，この操作は高周波成分が間引き後に雑音として現れる現象（エリアシング）の発生原因となるだけではなく，分離に必要な情報も捨ててしまう可能性がある．これらの問題はある程度学習により軽減できる可能性はあるものの，その程度は学習データや学習手法に依存する．

一方，多重解像度解析のダウンサンプリングは DWT で実現されているため，低周波に対応する周波数成分のエイリアシングを低減し，生じたエイリアシングを打ち消すフィルタを構成することもできる．また，DWT は可逆な変換であるため入力の情報を全て保持できる．したがって，ダウンサンプリングに DWT を用いることで，Wave-U-Net の抱えるダウンサンプリングに関する問題を解決できる可能性がある．

（2）DNN とウェーブレットの同時学習

DWT のウェーブレットとしては，これまでに信号処理，ウェーブレット解析分野で様々なウェーブレットが提案されてきた．しかし，これらは所望の要請に応じて設計されたものであり，時間領域深層学習を用いた音源分離に適している保証はない．また，深層学習モデルの構造やタスクに応じて適切なウェーブレットは変わりうるため，その都度設計するのは現実的でない．そこで，他の深層学習モデルのモジュールと同時に学習する方法についても検討する．

４．研究成果

本研究課題では，以下の 3 項目に関する成果を得た．

（1）多重解像度深層分析の構築

時間領域深層学習モデル Wave-U-Net と多重解像度深層分析を融合した音源分離手法，多重解像度深層分析を構築した（図 1 参照）．提案手法の深層学習モデルは，Wave-U-Net においてダウンサンプリング・アップサンプリング操作をそれぞれ DWT・逆 DWT へと置き換えたものである．これにより，Wave-U-Net で生じていた，特徴量領域でのエイリアシングや情報の欠落の問題を一挙に解決した．楽音分離実験（異なる種類の楽器音からなる混合音を各楽器音へと分離）を通じて，提案法を導入することにより，Wave-U-Net を含む従来の時間領域深層学習モデルに比べ分離性能が向上することを示した．また，主観評価実験においても同様の傾向があることを示した．

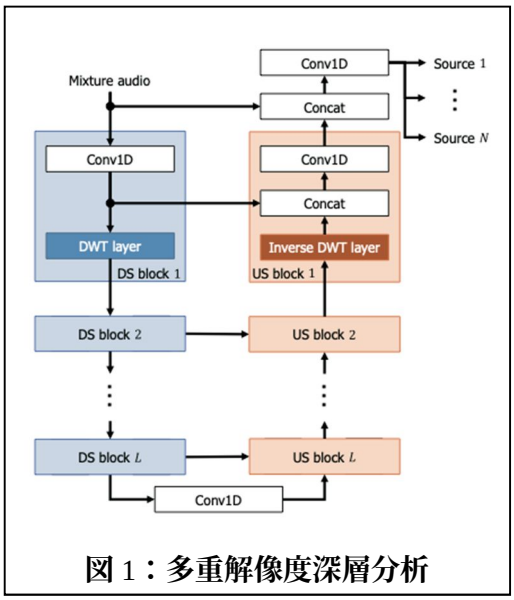

図 1：多重解像度深層分析

DWT の実装では，リフティングスキームと呼ばれる技法を用いた．リフティングスキームでは，ウェーブレットを予測・更新作用素によって間接的に定義できる．これらの作用素を適切に変更することで，様々な有限長のウェーブレットを実現できる．これにより，様々なウェーブレット間での性能比較が容易にできるようになった．ウェーブレット間の性能比較実験により，既存のウェーブレット間では分離性能に大きな差がないことを確認し，実装が簡易で計算速度も速い Haar ウェーブレットを用いれば十分であることを実験的に確認した．また，実験的に長いタップ長のウェーブレットを用いると学習が数値的に不安的になることも確認した．この場合，ウェーブレットの周波数特性は特定周波数帯域において高いゲインを持ちやすく，周波数応答と学習の安定性に関連があることが示唆された．

（2）深層学習モデルとウェーブレットの同時学習方法の構築

リフティングスキームを用いた DWT では，予測・更新作用素として有限長のフィルタを用いることができる．このフィルタを単純にパラメータとみなして深層学習モデルと同時に学習することもできるが，その場合初期値によって分離性能が大幅に低下ことを実験的に発見した．また，予測・更新作用素のパラメータをランダムに定めた際には，完全再構成性は満たされるものの DWT が低域通過フィルタを持たない場合があることを確認した．

そこで，低域通過フィルタを持つための予測・更新作用素に対する制約条件を導出し，その制約条件を満たしつつこれらの作用素のパラメータを学習することが可能であることを示した．具体的には，各作用素のフィルタ係数の和に対する制約として記述できるため，当該作用素による演算を行う前にフィルタ係数を所望の和の値となるように正規化すれば良いことを示した．これにより，フィルタ係数の初期値をランダムに定めても分離性能が大きく低下せず，安定的に学習できるようになった．

楽音分離実験を通じて，ウェーブレットを同時学習することで，特にモデルサイズが小さい場合に性能が向上することを示した．ここまでの研究成果に関して，音響信号処理分野のトップジャーナルである IEEE/ACM Trans．ASLP に採録された．また，当該論文において第 17 回日本音響学会・独創研究奨励賞板倉記念を含む 2 件の賞を受賞した．

（3）他タスクへの多重解像度深層分析の適用

上述の多重解像度深層分析の検討では，モノラルかつ異なる楽器同士の音源分離を対象としてきた．しかし，提案手法はこのタスク以外にも利用できるため，多チャネル音源分離（複数のマイクロホンで録音された混合音を対象とした音源分離）と重唱分離(重唱を各パートの歌唱へと分離するタスク)へと適用し，その性能を検証した．

多チャネル音源分離への適用では，周波数間の相関を考慮した多チャネル音源分離モデルを実現するため，パワースペクトログラム領域の分離モデルと時間領域での分離モデルを組み合わせる方法を提案した．時間領域の分離モデルとして多重解像度深層分析を用いることで，他チャネル音源分離においても性能が向上することを示した．当該成果は音響信号処理分野の国際会議 EUSIPCO 2021 にて発表を行った．

重唱分離は，これまで扱ってきた楽音分離に比べ同一楽器（歌声）同士の分離となるため，より困難なタスクである．これに対し，多重解像度深層分析は従来の重唱分離と同程度の分離性能を示しており，様々な音源分離タスクにおいて高い性能を示すことを確認した．このタスクに関して実装やデータの公開も行った．当該成果は，音響信号処理分野の国際会議 ICASSP 2023 で発表を行った．

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　0件／うちオープンアクセス　3件）

１．著者名 Koichi Saito、Tomohiko Nakamura、Kohei Yatabe、Hiroshi Saruwatari	４．巻 30
２．論文標題 Sampling-Frequency-Independent Convolutional Layer and its Application to Audio Source Separation	５．発行年 2022年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 2928～2943
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2022.3203907	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Nakamura Tomohiko、Kozuka Shihori、Saruwatari Hiroshi	４．巻 29
２．論文標題 Time-Domain Audio Source Separation With Neural Networks Based on Multiresolution Analysis	５．発行年 2021年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 1687～1701
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2021.3072496	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Nakamura Tomohiko、Kameoka Hirokazu	４．巻 29
２．論文標題 Harmonic-Temporal Factor Decomposition for Unsupervised Monaural Separation of Harmonic Sounds	５．発行年 2021年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 68～82
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2020.3037487	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計11件（うち招待講演　0件／うち国際学会　4件）

１．発表者名 Tomohiko Nakamura、Shinnosuke Takamichi、Naoko Tanji、Satoru Fukayama、Hiroshi Saruwatari
２．発表標題 jaCappella corpus: A Japanese a cappella vocal ensemble corpus
３．学会等名 IEEE International Conference on Acoustics, Speech, and Signal Processing（国際学会）
４．発表年 2023年

１．発表者名
中村 友彦，高道 慎之介，丹治 尚子，深山 覚，猿渡 洋

２．発表標題
jaCappella コーパス:重唱分離・合成に向けた日本語アカペラ歌唱コーパス

３．学会等名
日本音響学会第148回(2022年秋季)研究発表会

４．発表年
2022年

１．発表者名
中村 友彦、猿渡 洋

２．発表標題
多重解像度深層分析を用いた楽音分離の実験的評価

３．学会等名
音学シンポジウム2021（第131回 音楽情報科学研究会）

４．発表年
2021年

１．発表者名
齋藤 弘一、中村 友彦、矢田部 浩平、猿渡 洋

２．発表標題
周波数領域でのフィルタ設計に基づくサンプリング周波数非依存畳み込み層を用いたDNN音源分離

３．学会等名
音学シンポジウム2021（第131回 音楽情報科学研究会）

４．発表年
2021年

１．発表者名
Koichi Saito、Tomohiko Nakamura、Kohei Yatabe、Hiroshi Saruwatari

２．発表標題
Sampling-frequency-independent Audio Source Separation Using Convolution Layer Based on Impulse Invariant Method

３．学会等名
European Signal Processing Conference 2021（国際学会）

４．発表年
2021年

１．発表者名
Naoki Narisawa、Rintaro Ikeshita、Norihiro Takamune、Daichi Kitamura、Tomohiko Nakamura、Hiroshi Saruwatari、Tomohiro Nakatani

２．発表標題
Independent Deeply Learned Tensor Analysis for Determined Audio Source Separation

３．学会等名
European Signal Processing Conference 2021（国際学会）

４．発表年
2021年

１．発表者名
成澤 直輝、池下 林太郎、高宗 典玄、北村 大地、中村 友彦、猿渡 洋、中谷 智広

２．発表標題
ヘビーテイル生成モデルに基づく独立深層学習テンソル分析

３．学会等名
日本音響学会 2021年秋季研究発表会

４．発表年
2021年

１．発表者名
齋藤 弘一、中村 友彦、矢田部 浩平、猿渡 洋

２．発表標題
サンプリング周波数非依存音源分離モデルを用いた楽音分離の実験的評価

３．学会等名
日本音響学会 2021年秋季研究発表会

４．発表年
2021年

１．発表者名
Shihori Kozuka, Tomohiko Nakamura, Hiroshi Saruwatari

２．発表標題
Investigation on Wavelet Basis Function of DNN-based Time Domain Audio Source Separation Inspired by Multiresolution Analysis

３．学会等名
49th International Congress and Exposition on Noise Control Engineering（国際学会）

４．発表年
2020年

１．発表者名
齋藤 弘一，中村 友彦，矢田部 浩平，小泉 悠馬，猿渡 洋

２．発表標題
潜在アナログフィルタ表現に基づく畳み込み層を用いたサンプリング周波数非依存なDNN音源分離

３．学会等名
日本音響学会2021年春季研究発表会

４．発表年
2021年

１．発表者名
齋藤 弘一，中村 友彦，矢田部 浩平，小泉 悠馬，猿渡 洋

２．発表標題
アンチエイリアシング機構を導入したサンプリング周波数非依存な畳み込み層を用いた音源分離

３．学会等名
情報処理学会 第130回音楽情報科学研究会

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
分離音デモページ
https://tomohikonakamura.github.io/Tomohiko-Nakamura/demo/MRDLA/
提案手法コード（GitHub）
https://github.com/TomohikoNakamura/dwtls

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関