

研究終了報告書

「音メディア処理のための標本化周波数非依存深層学習」

研究期間：2021 年 10 月～2024 年 3 月

研究者：中村 友彦

1. 研究のねらい

混合音を各音源信号に分離する音源分離は音メディア処理の基盤技術であり、深層ニューラルネットワーク(DNN)の導入により性能が飛躍的に向上しつつある。実環境では目的音以外の音も不可避に観測音に混合される。多くの音メディア処理手法ではクリーンな目的音や目的音が支配的な観測音が入力されることを前提としており、音源分離を用いて観測音から所望の音源信号を抽出することで、各手法の性能を向上させられる。そのため、**様々な音メディア処理システムの前処理として汎用的に使用できる高性能な音源分離(汎用音源分離)の実現は、音メディア処理における究極の目標の 1 つである。**

後段に配置される音メディア処理システムはそれぞれの目的で処理を行う。そのため、汎用音源分離は分離性能の高さだけでなく、各システムが要請する多様な音響的条件を扱える汎用性も備えるべきである。代表的な音響的条件の 1 つが、音響信号の単位時間あたりの標本数(標本化周波数)である。例えば、人間が分離音を聴取するシステム(音楽のリミキシングや楽音加工など)では、人間の可聴域をカバーするため標本化周波数として 44.1, 48 kHz が用いられる。一方、音に含まれる内容を認識するシステム(音声認識や自動採譜など)では、高周波帯域の成分は必須ではなく性能低下要因にもなりうる。この場合はより低い 8, 16, 22.05 kHz などが頻繁に用いられる。そのため、DNN を用いた汎用音源分離の実現には、**様々な標本化周波数を単一の DNN で扱う必要がある。**しかし、従来の深層学習を用いた音メディア処理手法(音源分離手法も含む)では、標本化周波数が変化しないものとして扱われており、学習していない標本化周波数の音響信号を直接扱えなかった。

本研究では、汎用音源分離の実現のため、**未学習の標本化周波数を含む様々な標本化周波数の音響信号を扱える深層学習(標本化周波数非依存深層学習)手法を構築し、それを用いた音源分離手法を提案する。**具体的には、標本化周波数が信号処理で確立された概念であることに着眼し、DNN の構成要素である層を信号処理の観点から再解釈する。この解釈に基づき、標本化周波数に非依存な層を構築する。

2. 研究成果

(1) 概要

本研究の主要な成果は、以下の 3 点である。

[テーマA] 標本化周波数非依存(SFI)畳み込み層の提案

[テーマB] 非整数カーネルサイズ・ストライド処理アルゴリズムの提案

[テーマC] DNN を用いた潜在アナログフィルタ表現方法の提案

テーマ A では、本研究の核となるアプローチを用いて、畳み込み層を標本化周波数に非依存な形に拡張した。本テーマでは、畳み込み層がデジタルフィルタの集合とみなせることに着眼し、層のパラメータが標本化周波数と不可分であることを発見した。そこで、連続時間・周波

数領域で定義された関数(潜在アナログフィルタ), すなわち SFI な関数から, 入力標準化周波数に応じて層のパラメータを生成する過程を畳み込み層に導入することで, 様々な標準化周波数に対して一貫したパラメータを生成できることを示した. この生成機構を導入した畳み込み層を SFI 畳み込み層として提案した.

テーマ B では, SFI 畳み込み層のカーネルサイズ・ストライドが非整数の場合でも計算できるアルゴリズムを提案した. この成果により, SFI 畳み込み層を DNN の一部にのみ適用し, SFI でない従来の層と混在させながら利用することがより容易になった. これは, 従来の層と混在させながら利用する際には, 入力標準化周波数に応じて SFI 畳み込み層のカーネルサイズとストライドを適切に調節する必要があり, その結果として非整数となる場合があったからである.

テーマ C では, SFI 畳み込み層の潜在アナログフィルタの関数形状を学習により獲得できる方法を構築した. SFI 畳み込み層を用いる際には潜在アナログフィルタを定める必要があるものの, 連続時間・周波数に関する関数であるため直接パラメータとして扱うのは難しく, 関数形を層に合わせて手動で与え, その関数のパラメータを学習によって獲得する方法を採っていた. 本テーマでは, この連続時間・周波数の関数を DNN により表現することで, この問題を解消し, DNN の全体の学習と同時に潜在アナログフィルタの関数形も同時に学習する方法を構築した.

(2) 詳細

[テーマ A] 標準化周波数非依存(SFI)畳み込み層の提案

本テーマでは, 本研究課題の核となる, 標準化周波数に非依存な畳み込み層(SFI 畳み込み層)を提案した. この層は, DNN の主要な構成要素である畳み込み層を信号処理の観点から解釈することで導出されたものである. 畳み込み層は, そのパラメータと入力特徴量の相互相関を計算する. 相互相関は, 層のパラメータを時間方向に反転させると信号処理での畳み込み演算と一致する. そのため, 畳み込み層はデジタルフィルタの集合とみなせる. デジタルフィルタは離散時間領域で定義されるため, 層のパラメータも離散時間領域で定義されており, 標準化周波数と不可分である.

この発見に基づき, 層のパラメータを標準化周波数と独立な形で定義するためのアプローチを構築した. 具体的には, 信号処理分野で確立されたアナログフィルタからのデジタルフィルタ設計手法にヒントを得て, 層のパラメータが連続時間・周波数領域で定義された関数(潜在アナログフィルタ)から入力標準化周波数に応じて生成される過程を導入した. この生成過程を導入した畳み込み層が SFI 畳み込み層である(図 1 参照). SFI 畳み込み層のパラメータは, 潜在アナログフィルタのパラメータであるため SFI な領域で定義される. また, 様々な標準化周波数に対して同一の潜在アナログフィルタを用いるため, 未学習の標準化周波数に対しても一貫した畳み込み層のパラメータを生成できる. この生成方法を用いることで, 未学習の標準化周波数用の DNN のパラメータが得られ, リサンプリングを用いずとも様々な標準化周波

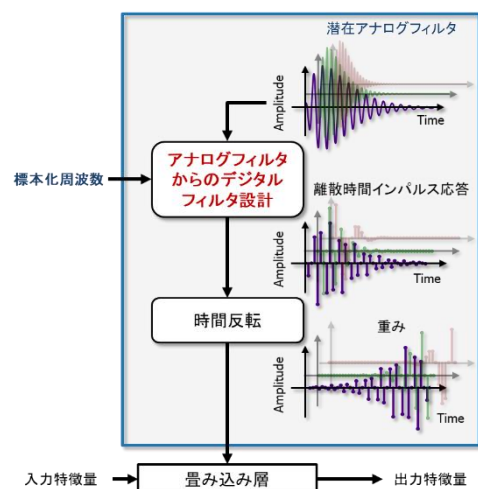


図 1 : SFI 畳み込み層の概要図

数に対応できる。

SFI 畳み込み層の効果は楽音分離実験で検証した。SFI 畳み込み層を代表的な波形入力型音源分離用 DNN である Conv-TasNet へと適用した。結果として、単一の標準化周波数で学習したとしても、従来の楽音分離用 DNN に比べ、提案層を用いた DNN は未学習の標準化周波数を含む様々な標準化周波数で一貫した性能で動作することを示した。また、入力音響信号を学習した標準化周波数にリサンプリングした場合に、従来の楽音分利用 DNN では大きく性能が低下する場合があることを示した。その場合でも、提案する未学習の標準化周波数への適応法は、学習した標準化周波数と同程度の性能で動作することを示した。本テーマの研究成果は音メディア処理分野のトップ国際論文誌に採録された(5. (1)論文 1 参照)。

[テーマ B] 非整数カーネルサイズ・ストライド処理アルゴリズムの提案

本テーマでは、より多様な標準化周波数にも SFI 畳み込み層を適用するための畳み込み層の計算アルゴリズムを提案した。SFI 畳み込み層を既存の DNN に導入する場合、当該 DNN の一部の畳み込み層にのみ適用するだけでも全体を SFI 化できることがあり、最小限の改変で標準化周波数への頑健性を向上させられる。このようなネットワーク構造では通常の層と SFI 畳み込み層が混在することになり、通常の層に対する入力の解像度を学習時と同一にするために、入力標準化周波数に応じて SFI 畳み込み層のカーネルサイズとストライドを適切に調節する必要がある。調節の際に、カーネルサイズやストライドが非整数となる場合があるものの、畳み込み層の演算は整数のカーネルサイズやストライドにしか定義されておらず、畳み込み層の処理アルゴリズムをそのまま利用することはできなかった。単純には、非整数値を整数値に丸めることで対処できるが、ストライドを丸めると特徴量の時間解像度が学習時から変化してしまい、分離性能が低下することを実験的に確認した。

そこで、窓付き sinc 補間を用いることで、非整数のストライドを扱える畳み込み層の処理アルゴリズムを提案した(図 2 参照)。具体的には、畳み込み層の処理アルゴリズムを相互相関の計算処理と間引き処理に分割し、間引きの際に窓付き sinc 補間を用いることで離散時間の特徴量の連続時間表現を与えた。この連続時間表現から所望のストライドで標準化することで、任意のストライドに対処できる。同様に、転置畳み込み層に対しても拡張し、その SFI 版である SFI 転置畳み込み層においても非整数のストライドを扱えるアルゴリズムを提案した。楽音分離実験により、整数値へ丸める場合や、入力信号を学習した標準化周波数やストライドが整数となる最も近い標準化周波数へリサンプリングする場合に比べ、提案アルゴリズムは高い分離性能を与えることを示した。

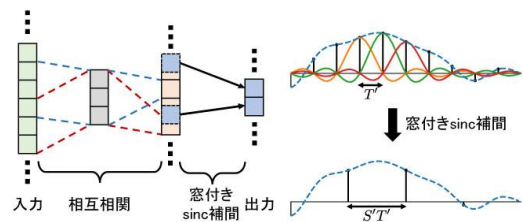


図 2 : SFI 畳み込み層に対する非整数ストライド処理アルゴリズムの概要図

本テーマは研究を進める中で生まれた計画外のものではあるものの、これまでの培ってきた信号処理と深層学習を融合した研究から着想を得て解決できた。この成果は、信号処理分野のトップ国際会議の 1 つである EUSIPCO に採録された(5. (1)論文 2 参照)。

[テーマ C] DNN を用いた潜在アナログフィルタ表現方法の提案

本テーマでは、DNN を用いた潜在アナログフィルタの表現方法を提案し、潜在アナログフィル

タそのものを学習できる方法を構築した。SFI 畳み込み層を用いる際には、潜在アナログフィルタを学習前に定める必要がある。潜在アナログフィルタは層のパラメータの原型となるため、ネットワーク構造内での位置やタスクに応じて適切に設定する必要がある。テーマ A, B で提案した手法では、連続時間・周波数に関する関数であるため直接パラメータとして扱うのは難しく、導入対象の層が時間周波数変換と対応することからヒントを得て設計していた。具体的には、関数形を層に合わせて手動で与え、その関数のパラメータを学習する方法を採っていた。しかし、ネットワーク構造内の任意の層では信号処理での対応物があるとは限らず、適切な関数形を決定するのは難しい。

そこで、潜在アナログフィルタそのものを DNN で表現することで、関数形を事前に定めずとも学習できる方法を提案した。具体的には、連続時間インパルス応答や連続周波数応答がそれぞれ連続時間と連続周波数の関数とみなせることに着眼し、その関数自体を DNN で表現する。これにより、潜在アナログフィルタの関数形を手動で設計した場合に比べ、特に学習した標本化周波数よりも低い標本化周波数での分離性能が向上することを実験的に示した。この成果のベースとなる国内発表([3] 参照)で、研究補助者が優秀発表賞を受賞した((a)参照)。

3. 今後の展開

本研究では、深層学習を信号処理の観点から見直すことで、デジタルフィルタ設計と畳み込み層を融合した手法を構築しその有効性を示した。一方、標本化周波数による性能低下の原因については未解明な部分が多い。リサンプリングの精度が性能低下と関連することは判明したものの、どのような条件であればリサンプリングにより性能低下が起こるかは判明していない。また、提案法が有効な条件に関しても部分的にしか判明していない。そのため、SFI 畳み込み層を様々な DNN へと適用する際には、実際に学習して所望の標本化周波数で評価することが必要であり、計算コストが高い。これは実用に向け解決が必要な課題であり、今後はリサンプリングや提案法が有効な条件を解明するための方法を構築する必要がある。

4. 自己評価

【研究目的の達成状況】

本研究の核である SFI 畳み込み層と DNN を用いた潜在アナログフィルタ表現方法について達成できており、研究期間中順調に研究が進んだと言える。計画当初予期していなかった課題も発見し、一つのテーマとして成り立つほどに昇華できた。また、アドバイザーからのコメントから国際会議論文へと発展したものもあり、計画外の成果も出た点は重要な成果であった。

【研究の進め方(研究実施体制および研究費執行状況)】

代表者として主導して研究を行い、期間中計 2 名の研究補助者と本研究課題に取り組んだ。また、領域会議やサイトビジットによるアドバイザーからのコメントが、計画外の課題の発見や全体の進捗向上に繋がった。また、領域会議での他の研究者からの発表に刺激されて、当該研究課題とは別の研究で国際会議論文へと繋がった成果もあった。

研究費に関しては、研究期間前半は COVID-19 の感染拡大もあり主に計算機に利用でき、研究速度の向上に寄与した。期間後半は国内・国際会議での発表のための旅費が主であり、発表を通して関連分野の研究者との議論を深めることができた。

【研究成果の科学技術および社会・経済への波及効果】

本研究課題に関する成果は、音メディア処理分野のフラグシップ国際論文誌および国際会議で発表した。また、当該内容で研究補助者が国内で受賞もしており、国内外で高い評価を得ていると考えている。本研究の成果は、様々なデバイスや環境で音源分離手法を一貫した性能で動作させることにも寄与するため、音源分離手法を用いた製品・サービスを実環境へ実装する際の一助になると考えている。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 2件

1. Koichi Saito, Tomohiko Nakamura, Kohei Yatabe, and Hiroshi Saruwatari, “Sampling-frequency-independent convolutional layer and its application to audio source separation,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 2928–2943, Sep. 2022.

本論文では、未学習の標準化周波数も扱える畳み込み層の拡張(SFI 畳み込み層)を提案し、それを用いた楽音分離手法を構築した。提案層は、入力特徴量の標準化周波数に合わせて通常の畳み込み層のパラメータ(重み)を生成することで、未学習の標準化周波数に対応できる。楽音分離実験により、提案法が入力信号をリサンプリングするよりも高性能に動作することを、客観・主観指標の両方で示した。

2. Kanami Imamura, Tomohiko Nakamura, Norihiro Takamune, Kohei Yatabe, and Hiroshi Saruwatari, “Algorithms of sampling-frequency-independent layers for non-integer strides,” in Proceedings of European Signal Processing Conference, pp. 326–330, Sep. 2023.

本論文では、SFI 畳み込み層に関する、非整数ストライドを扱う計算アルゴリズムを提案した。SFI 畳み込み層を通常の層と混合して用いるには、SFI 畳み込み層のストライドを入力標準化周波数に応じて調節する必要がある。しかし、その際にストライドが非整数となってしまうことがあり、直接計算できなかった。そこで、入力の連続時間表現を介することで非整数のストライドを扱えるアルゴリズムを提案し、その有効性を楽音分離実験により示した。

(2) 特許出願

研究期間全出願件数: 0 件(特許公開前のものは件数にのみ含む)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

● 査読なし国内会議

- [1] 今村 奏海, 中村 友彦, 高宗 典玄, 矢田部 浩平, 猿渡 洋, “サンプリング周波数非依存畳み込み層における非整数ストライド処理アルゴリズム,” 日本音響学会 2023 年秋期研究発表会, pp. 157–160, Sep. 2023.
- [2] 中村 友彦, 矢田部 浩平, “Universal Sound Separation へのサンプリング周波数非依存畳み込み層の適用,” 日本音響学会 2023 年秋期研究発表会, pp. 161–164, Sep. 2023.

- [3] 今村 奏海, 中村 友彦, 矢田部 浩平, 猿渡 洋, “サンプリング周波数非依存畳み込み層のための時間領域ニューラルアナログフィルタ,” 日本音響学会 2022 年秋期研究発表会, pp. 187–190, Sep. 2022.
- [4] 中村 友彦, 矢田部 浩平, 猿渡 洋, “ニューラルアナログフィルタを用いたサンプリング周波数非依存畳み込み層とモノラル音源分離への適用,” 日本音響学会 2022 年春期研究発表会, pp. 181–184, Mar. 2022.

● 受賞

- (a) 今村 奏海, 日本音響学会 第 25 回学生優秀発表賞, 2022 年 11 月
- (b) 中村 友彦, 日本音響学会 第 55 回粟屋潔学術奨励賞, 2024 年 3 月