

音源分離を用いた耳コピー・アドリブ練習支援アプリケーション『MuSep』の開発

久野 文菜^{†1,a)} 濱川 礼^{†2,b)}

概要: 本論文では GAN を用いた音源分離手法と、その音源分離技術を用いた耳コピー・アドリブ練習支援アプリケーションの開発について述べる。様々な楽器で演奏されている多重奏から各楽器の音を分離する音源分離技術を利用し、多重奏から選択した楽器の音のみ抽出することで耳コピーの支援を行う。また反対に選択した楽器の音を消去し、消去した箇所に合わせて自分で演奏することでアドリブの練習を行うことも可能にする。また音源分離の精度を向上させるため GAN を用いた音源分離手法の提案を行う。音をスペクトログラムに変換し、多重奏のスペクトログラムと各楽器のスペクトログラムをペアで GAN で学習させ分離モデルを作成する。分離させたい多重奏のスペクトログラムにその分離モデルを適用させることで各楽器のスペクトログラムを生成する。

1. 背景及び目的

図 1 は 2 つともピアノの楽譜である。左がクラシックで有名なフレデリック・ショパンによる『幻想即興曲』、右がジョー・ガーランドによる『In The Mood』の楽譜である。ジャズにはコードに沿って即興で自由に演奏する『アドリブ [1]』が存在するためミュージシャンによって弾く音や弾き方が違う。そのためジャズの楽譜はクラシックの楽譜と比較すると、演奏中に弾くべき音符や強弱記号が楽譜中にほぼ記されていない。

ジャズを演奏するミュージシャンは他のミュージシャンがどのようにその曲を演奏しているのかを耳で聴き取ることがよくある。これを『耳コピー [2]』という。

しかしアドリブの練習を行ったり耳コピーをする上でそれぞれ問題点が出てくる。

- (1) アドリブ練習をする際、自分で演奏する楽器以外による伴奏が必要である。しかし自分の演奏する曲において自分の楽器の音のみが消えている（マイナスワン）音源があるとは限らない。
- (2) 耳コピーをする際、図 2 に示すビッグバンドジャズ隊形 [3] のように、演奏が大人数により複数種類の楽器



図 1 左：クラシックのピアノ譜、右：ジャズのピアノ譜

(主にピアノ・ベース・ドラム・ギター・トランペット・トロンボーン・サックスの 7 種類) で構成されている場合、自分の聴き取りたい楽器とそれ以外の楽器の音が混ざり、耳コピーしにくいことが多々ある。

これら 2 つの問題を解決するために本論文では、様々な楽器で演奏されている多重奏から各楽器の音を分離する音源分離技術を利用した、アドリブ練習・耳コピー支援アプリケーション『MuSep (ミューゼップ)』を開発した。

2. 音源分離に関する従来研究

近年ではディープラーニングを用いて音源分離を行う手法が数多く登場している。Daniel らは、U-Net の入力にス

^{†1} 現在、中京大学大学院工学研究科

^{†2} 現在、中京大学

a) kuno.ayana.312@gmail.com

b) rei-hama@sist.chukyo-u.ac.jp



図 2 ビッグバンド隊形

ステレオ音源の音の波形を用いることで音源分離を行う手法『Wave-U-Net[4]』を提案している。U-Netとは全層畳み込みネットワークであり、その入力に音の波形を用いることで音声信号の位相情報とその定位を考慮しながら音を分離する。Chuangらは自己教師あり学習を利用し、楽器を演奏する複数人の動きを映像から分析することで個々の楽器の音を分離する手法を提案している[5]。Fabian-RobertらはスペクトログラムをBLSTM (Bidirectional Long Short-Term Memory) で学習させることによって音源分離を行う手法『Open-Unmix[6]』を提案している。BLSTMを多重奏のスペクトログラムから各楽器のスペクトログラムへの変換に利用し、多重奏のスペクトログラムから各楽器のスペクトログラムになる為のマスクを生成している。Open-Unmixは自分でGithubのコードをクローンすることで自分の選択した曲を音源分離することが可能だが、ソフトウェアやアプリケーションにはなっていないので簡単に利用することは不可能である。

3. 耳コピ支援に関する関連ソフトウェア・アプリケーション

3.1 関連ソフトウェア

耳コピを支援するソフトウェアは昔から多く開発されてきている。図3に示したTranscribe! [7]やWave Tone [8]は音源を入力すると、ある時刻にどの周波数が鳴っているのかを示すソフトウェアである。多重奏を入力すると全ての楽器の周波数が表示されるため、自分の耳コピしたい楽器で演奏されている周波数がどれに当たるのかは不明確である。Band Producer 5 [9]は音源分離により耳コピを支援するソフトウェアであるが、周波数領域で楽器の音を分離している（例えばベースのように基本周波数の低い楽器を耳コピしたい場合、周波数の低域のみを聴く）ため、ベースの他に周波数の低い音を鳴らしている楽器（ピアノやバストロンボーンなど）が曲中に存在するとそれも含めて分離されてしまう。

3.2 関連スマホアプリケーション

耳コピを支援するスマホアプリはこれまでに沢山あり



図 3 耳コピ支援関連ソフトウェア

スされている。中でも使いやすく評判が高いのが図4に示したmimiCopy [10]とAudipo [11]である。mimiCopyでは曲中で自分の聴きたい箇所をループ再生でき、その部分の速度を落としたりピッチを変更したりが可能である。Audipoでは自分が聴き始めたい箇所に印をつけてそこから聴き始めることができ、曲の全体を掴みつつどこから聴き取りたいのかの把握や、速度を落として聴くことが可能である。この2つのアプリをはじめとして従来のアプリでは、曲中で自分の好きな箇所にクリップを付け、その間を繰り返したり、ゆっくり再生することで聴き取れるようにするといった支援方法を取っていた。しかし、やはり多重奏となるとゆっくり繰り返して聴いたところで耳コピできないのが現状である。



図 4 耳コピ支援関連スマホアプリケーション

3.3 アドリブ練習支援

図5に示したiReal Pro [12]ではアプリ内に内蔵されている曲を選択するとその曲のコードを表示し、自動でその曲にドラム・ベース・ピアノによる伴奏を付けることでアドリブの練習ができるアプリである。曲の中で何回ソロを演奏するかの指定や曲全体のピッチなどを変更することは可能だが、アプリ内蔵の曲ではあるが指定したCDのアレンジで演奏を行いたい場合や、そもそもアプリ内に内蔵されていない曲のアドリブを練習することは不可能である。

4. 提案手法

2章で述べた関連研究では、ステレオ音源や映像が必要であったり、アプリケーションやソフトウェアには応用されていない為ユーザーが気軽に音源分離をすることができな



図 5 アドリブ練習支援従来スマホアプリケーション

いものが多く挙げられた。また 3 章で述べた耳コピの支援ソフトウェアやアプリケーションは、主に再生方法に工夫を凝らすことで耳コピを支援しており、楽器ごとに音を分離することで耳コピを支援する方法はとっていない。そこでステレオ音源や映像を必要とせず音源分離を行うことができる手法と、その音源分離技術を使用しユーザが気軽に音源分離を使用してアドリブ練習や耳コピに使用できるスマートフォンアプリケーションの提案を行う。

4.1 音源分離

本手法では図 6 に示すように、スペクトログラムを画像として扱い、多重奏のスペクトログラムから各楽器のスペクトログラムを GAN によって生成することで音源分離を行う。GAN とはディープラーニングを使用しデータから特徴を学習することで、実在しないデータを生成することのできる敵対的生成ネットワークの一種である。ベースの分離における提案手法の例を図 7 に示す。多重奏のスペクトログラムからベースのスペクトログラムを生成するモデルを作成し、ベースを分離させたい多重奏のスペクトログラムに作成したモデルを適応させることでベースのスペクトログラムを生成する。これを楽器ごとに行う。2 章で示した音源分離の研究ではスペクトログラムを生成する際に短時間フーリエ変換を使用しているが、本手法では音という連続性あるデータを 1 枚 1 枚画像として切り離して学習するため時間領域の情報が失われてしまう。そこで、時間領域の情報を残すことが可能なウェーブレット変換を使用することで GAN による音源分離を可能にした。

4.1.1 連続ウェーブレット変換

フーリエ変換では、時間幅を短くすると周波数の情報がハッキリと分からなくなり、逆に周波数の情報をきめ細かく知ろうとすると、時間幅を大きく取らなければならないという不確定性関係が時間と周波数の間にでてくる。そこで低周波領域では周波数の分解能を優先し、高周波領域では時間の分解能を優先するように変換することのできる連続ウェーブレット変換を使用する。

4.1.2 GAN による学習

連続ウェーブレット変換によって変換されたスペクトロ

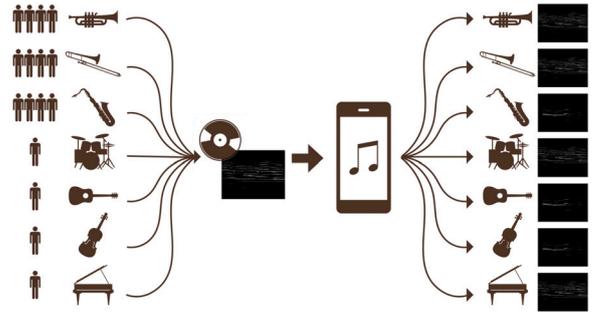


図 6 GAN を使用した音源分離手法

グラムを画像として GAN で学習させる。スペクトログラムは 8 層畳み込み処理を行い、活性化関数は leaky relu を用いた。データセットは 290 時間分の楽曲を用いた。学習は 2 つのペア画像から画像間の関係を学習してモデルを生成し、入力画像に対して学習したモデルを利用して新たな画像を生成することができる、GAN の一種である Pix2Pix[13] で行う。

4.1.3 データセット

2 章で述べた関連研究ではデータセットに MUSDB[14] が用いられている。MUSDB に用意されているのはドラム・エレクトリックベース・ボーカル・ピアノの 4 種類の楽器である。今回は 1 章で述べたビッグバンドジャズ隊形に含まれる楽器も分離させたいため、打ち込みキーボード KRONOS[15] と Mac に標準で搭載されている DTM ソフト GarageBand[16] を使用しウッドベース・トランペットの 2 種類の楽器の音源を追加で作成した。ドラム・ボーカル・ピアノ・ウッドベース・トランペットの楽器で演奏された各音源 300 時間分と、各音源を合成し多重奏とした音源 300 時間分を学習データとし、連続ウェーブレット変換を用いてスペクトログラムへと変換する (図 8)。スペクトログラムの大きさは 4096×2048 である。

4.2 スマートフォンアプリケーション

MuSep は iOS アプリケーションである。図 9 がアプリの基本画面であり、ユーザは iPhone 内の曲 (DTM 保護を除く) と分離したい楽器を選択することで、選択した楽器の音を抽出して聴く、もしくは曲から選択した楽器の音が消えた伴奏を聴くことができる。楽器選択後、図 10 のように耳コピ支援かアドリブ練習支援かを選択できる画面になる。

アドリブ練習支援を選択した場合、図 11 に示すアドリブ練習支援画面へと遷移する。ユーザは楽器を選択した後、選択した楽器の音のみが消えた音源を流すことで、それに合わせてアドリブ練習ができる。またユーザは Playback position bar によって曲中で練習したい箇所に飛ぶこともできる。

耳コピ支援を選択した場合、図 12 に示す耳コピ支援画

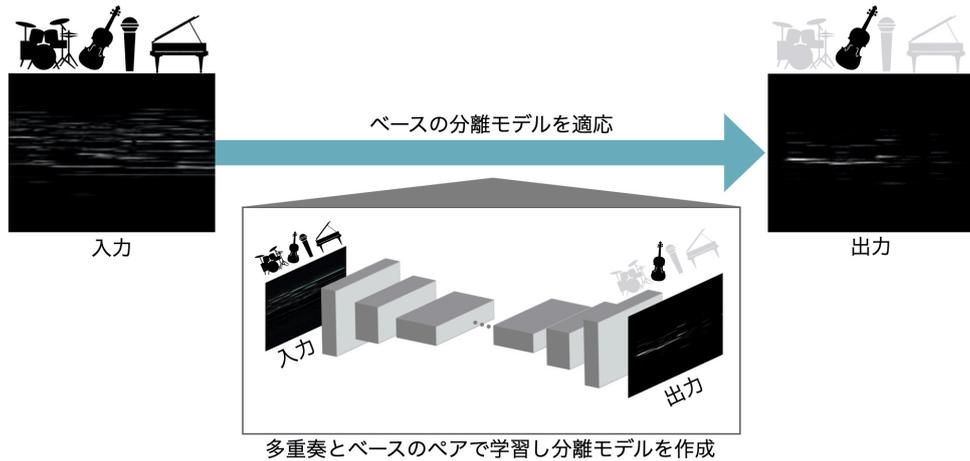


図 7 ベースの音分離の例

表 1 アンケート調査結果

番号	質問	1	2	3	4	5	平均
Q1	アドリブ練習は大変か？	0	0	0	2	3	4.6
Q2	何も使用せずアドリブ練習した場合と MuSep を使用してアドリブ練習した場合を比較して MuSep はどうだったか？	0	0	2	0	3	4.2
Q3	普段使用しているアプリでアドリブ練習した場合と MuSep を使用してアドリブ練習した場合を比較して MuSep はどうだったか？	0	1	1	1	2	3.8
Q4	耳コピは大変か？	0	0	0	0	5	5.0
Q5	何も使用せず耳コピした場合と MuSep を使用して耳コピした場合を比較して MuSep はどうだったか？	0	0	2	2	1	3.8
Q6	普段使用しているアプリで耳コピした場合と MuSep を使用して耳コピした場合を比較して MuSep はどうだったか？	1	0	2	2	0	3.0

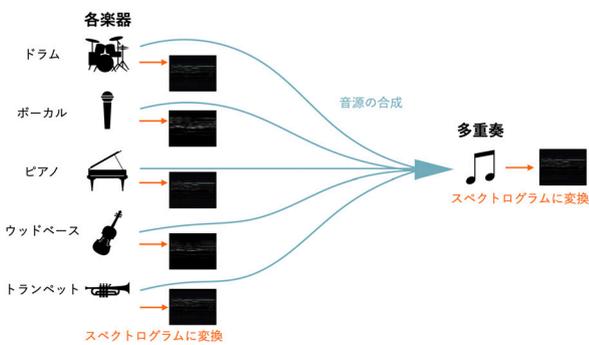


図 8 データセット作成方法

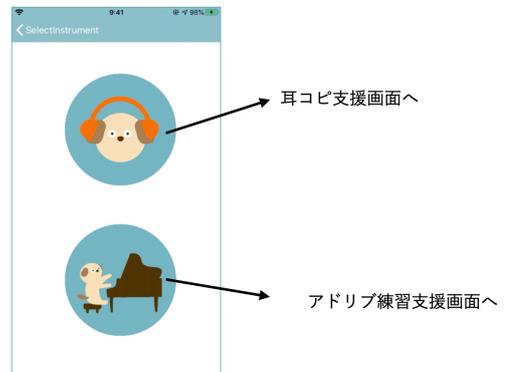


図 10 耳コピ支援・アドリブ練習支援の選択画面



図 9 アプリケーション基本画面

面へと遷移する。separate の Volume bar を上げ、original の Volume bar は 0 にすることで、分離した選択した楽器

の音のみを聴くことができ、「今どこを演奏しているのか」がわからなくなってしまったら original の Volume bar を上げることで元の音源を聴くことが可能である。ユーザは Playback position bar によって曲中で聴き取りたい箇所へ飛んだり、Tempo adjustment bar によってゆっくり再生したりなどができる。

5. アプリケーションの評価実験

5 人のアマチュアジャズミュージシャンに、自分で選択した曲のアドリブ練習と、自分がいつも演奏している楽器の耳コピを行ってもらい、アンケートを行った。アンケー

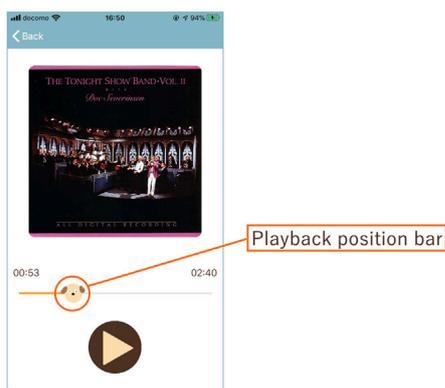


図 11 アドリブ練習支援画面

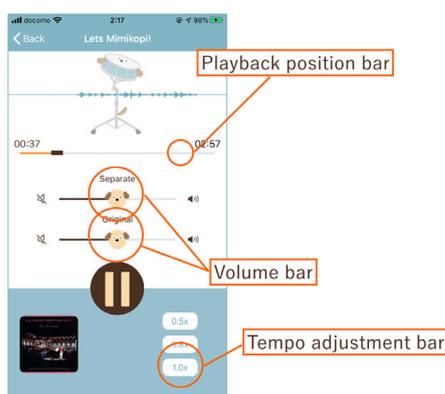


図 12 耳コピ支援画面

トの設問を表 1 に示す。設問に対して対称型 5 段階評価を用意し、Q1 と Q4 では 1 が『そうでもない』～5 が『とても大変』、それ以外は 1 が『良くない』～5 が『とても良い』として回答してもらった。また上記の設問以外に MuSep に関してのコメントを自由記述で書いてもらった。

6. 結果・考察

今回の評価ではユーザーに自由にアドリブ練習・耳コピの曲を選んでもらった。アンケートの自由記述では「自分の演奏したい曲を練習できてよかった」という意見が多い中、ユーザーの選んだ楽器・曲によっては分離精度が悪く「分離精度が高かったらもっと使いたくなる」という意見もあった。分離精度が悪いのは特にピアノで、音源で演奏されているそもそもの音量が小さい為ではないかと考えられる。また普段使用しているアプリにも繰り返し再生などのいいところがあるため、MuSep にもそのような機能があると嬉しいとの意見もあった。

7. まとめ

連続ウェブレット変換と GAN による音源分離の有用性を示した。またその手法を取り入れた iOS アプリを開発し、ユーザーから高い評価を得た。

8. 展望

MuSep は 2021 年 3 月に開催される SXSW の Trade Show[17] で展示予定である。そこでさらなるフィードバックを得てより使いやすいアプリにしていきたい。また評価実験にてもらった意見をもとに繰り返し再生機能の実装などを行っていきたい。

参考文献

- [1] Wikipedia: adlib (2020). <https://ja.wikipedia.org/wiki/adlib>.
- [2] Wikipedia: 耳コピ (2020). <https://ja.wikipedia.org/wiki/>
- [3] Wikipedia: bigband (2020). <https://ja.wikipedia.org/wiki/bigband>.
- [4] Daniel, S., Sebastian, E. and Simon, D.: Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation, *19th ISMIR* (2018).
- [5] Gan, C., Huang, D., Zhao, H., Tenenbaum, J. B. and Torralba, A.: Music Gesture for Visual Sound Separation (2020).
- [6] Stöter, F.-R., Uhlich, S., Liutkus, A. and Mitsufuji, Y.: Open-Unmix - A Reference Implementation for Music Source Separation, *Journal of Open Source Software*, Vol. 4, No. 41, p. 1667 (online), DOI: 10.21105/joss.01667 (2019).
- [7] seventhstring: Transcribe! (2019). <https://www.seventhstring.com/xscribe/download.html>.
- [8] Sound, A.: Wave Tone (2015). <http://ackiesound.ifdef.jp/download.htmlwt>.
- [9] KAWAI: Band Producer 5 (2019). <https://cm.kawai.jp/products/bp/>.
- [10] Apple: mimiCopy (2020). <https://apps.apple.com/us/app/mimicopy-slow-down-player/id405894824>.
- [11] Apple: Audipo (2020). <https://apps.apple.com/us/app/audipo-audio-speed-changer/id607971056>.
- [12] Apple: iReal pro (2020). <https://apps.apple.com/us/app/ireal-pro/id298206806>.
- [13] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks (2016).
- [14] Rafii, Z., Liutkus, A., Stöter, F.-R., Mimitakis, S. I. and Bittner, R.: The MUSDB18 corpus for music separation (2017).
- [15] Korg: KRONOS (2020). <https://www.korg.com/jp/products/synthesizers/kronos2/>.
- [16] Apple: GarageBand (2020). <https://www.apple.com/jp/ios/garageband/>.
- [17] SXSW: SXSW 2020 (2020). <https://www.sxsw.com/>.