ハモリ練習のための歌声入力による二声旋律楽曲からの主旋律削除

小西沙幸[1,a)]　井尻敬[1]

概要：ポップスなどのボーカルを含む楽曲には，メインボーカルに加えて，ハーモニーをつくるハモリパートを含むものが多い．このハモリパートを正しく歌うことができれば，カラオケなどの歌を歌うアクティビティをより楽しむことができる．しかし，一般的な楽曲音源では，ハモリパートはメインボーカルに比べて音量が小さいことが多く，ハモリパートの旋律がわかりにくいという課題がある．そこで本研究では，ハモリパートの練習支援を目的とし，ボーカル楽曲からメインボーカルを削除しハモリパートのみを再生可能にする手法を提案する．具体的には，メインボーカルは比較的簡単に歌えるという仮定に基づき，ユーザは対象となる楽曲音源に加え，そのメインボーカルを自身で歌った音源を入力する．すると，提案手法は，歌声の基本周波数とその倍音を楽曲音源から削除することで，ハモリパートのみの音源を作成する．提案手法の有用性を確認するため，ユーザスタディを実施した．その結果，多くの実験参加者が歌声の入力によりおおむね良好な主旋律削除を行えたことを確認した．また，参加者から提案手法を利用したハモリ練習に肯定的なコメントを得ることができた．

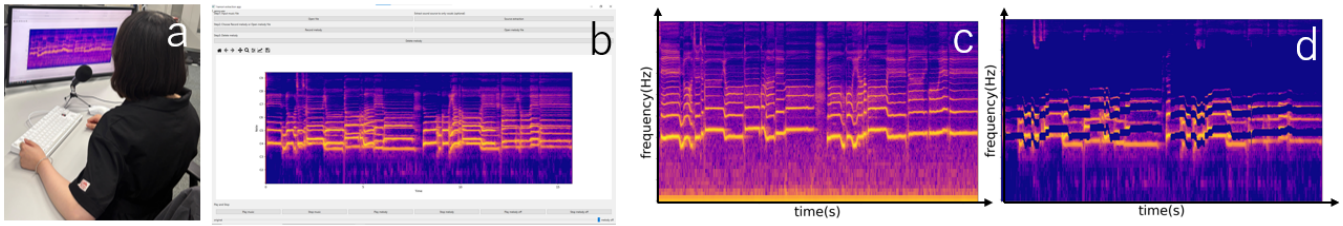

図 1: （a）提案手法を利用している様子，（b）アプリケーション画面，（c）ユーザが歌声により入力したメインボーカルのスペクトログラム，および，（d）メインボーカルを削除したスペクトログラム．

1. はじめに

ポップスなどのボーカルを含む楽曲には，『メインボーカル（主旋律）』とハーモニーをつくる『ハモリパート（副旋律）』を含むものが多い．このハモリパートを正しく歌うことができれば，カラオケなどの歌を歌うアクティビティをより楽しむことができる．しかし，一般的な楽曲音源ではハモリパートはメインボーカルに比べて音量が小さいことが多く，また，ハモリパートのみを聴く機会は限られる．そのため，ハモリパートの旋律はわかりにくく，その練習は難しいという課題がある．

音楽の分析や応用のため，音源データからのメロディ抽出や楽器分離を行う手法が研究されている．例えば，スペクトラルピークをもとにメロディを検出する手法 [1,2]，半教師あり学習によりメロディを推定する手法 [3]，機械学習により楽曲を楽器別の音源に分離する手法 [4,5,6,7] が提案されている．しかし，これらは旋律が単一であると仮定して推定するものや音源をボーカルや楽器に分離するものである．そのため，複数ボーカルを含む音源からメインボーカルとハモリパートを分離することは難しい．また，歌唱練習支援手法も研究されている．例えば，楽曲の音源のピッチとユーザの歌声のピッチを同時に可視化することで歌唱練習の支援をする手法 [8] や，基準音と目標音の 2 つを聴きながら目標音の音程を歌うことで相対音感の獲得を目指す手法 [9] が挙げられる．しかし，これらの手法は目標の音程を正しく歌う練習を行うためのものであり，ハモリパートの練習を支援するものではない．

本研究では，ポップスなどのボーカル楽曲におけるハモ

[1] 芝浦工業大学
[a)] al21113@shibaura-it.ac.jp

リパートの練習支援を目的とし，楽曲のメインボーカルを削除することでハモリパートのみを再生できる手法を提案する．具体的には，楽曲におけるメインボーカルは比較的簡単に歌えるという仮定に基づき，削除したい旋律をユーザの歌声を用いて入力し，この歌声の基本周波数とその倍音を元音源から削除することで，ハモリパートのみの音源を作成する．提案手法を用いると，ハモリパートのみを強調した練習用の音源を手軽に作成できる．加えて，メインボーカルとハモリパートの区別が明確ではない（両方が主旋律となり得る）楽曲において，歌声入力により削除したい旋律を選択的に指定できる．

提案手法の有用性を評価するため，大学生を対象にユーザスタディを実施した．実験では，機械音声の楽曲・人間の音声の楽曲・商用楽曲の 3 曲を対象に，提案手法を用いてメインボーカルを削除しハモリパートのみの音源を作成してもらい，これを利用してハモリ練習をしてもらった．結果，多くの参加者がメインボーカルを削除した音源を用いてハモリ練習ができた．実験後に実施したアンケートでは，提案手法を用いたハモリ練習に関して，肯定的な意見が得られた．

2. 関連研究

2.1 多声音楽データからの歌声分離

ボーカルや複数楽器を含む多声音楽（polyphonic music）データから，メロディライン（主旋律）を抽出する手法が研究されている．Goto [1] は，音源データより顕著なスペクトラルピークを複数取得し，それらを時間的連続性を考慮して追跡することでメロディラインを抽出する手法を提案した．Yao ら [2] も同様に，顕著性の高いスペクトラルピークを抽出し，連続フレーム間で類似したピッチを持つピークを接続することでメロディラインを抽出した．しかし，これらの手法は，複数のピークの追跡を行うものの，旋律が単一であることを仮定し，単一のメロディラインを推定する．そのため，メインボーカルとハモリパートの両方を推定することはできない．また，Zhang ら [3] は，半教師あり Extreme Learning Machine を利用し，注目フレーム周辺の音楽データの特徴ベクトルからそのフレームにおけるピッチを推定する手法を提案した．この手法も単一の旋律を仮定しており，ハモリパートの推定は行えない．

多声音楽データを，ボーカル・ベース・ドラムといった楽器ごとの音源に分離する手法も発表されている．Uhlich ら [5] は，3 層の feed-forward ネットワークと双方向 LSTM を統合した音源分離手法を提案した．Nugraha ら [4] は，ステレオ音源のようなマルチチャネル信号を入力に利用し，深層ニューラルネットワークを用いて注目フレームのスペクトル情報より各楽器のパワースペクトル密度を予測した．この手法では，さらに EM アルゴリズムを用いて各楽器の空間共分散を計算することで分離精度の向上を実現した．一方，画像領域分割に広く利用される U-Net により入力音源のスペクトログラムから楽器ごとのスペクトログラムを推定する手法 [6] や，畳み込みニューラルネットワークを用いて入力音源のスペクトログラムからボコーダ合成用のパラメータを推定することでボーカル音源を分離する手法 [7] が提案されている．これらの音源分離では，多声音楽をボーカルや楽器ごとの音源に分離することが可能であるが，複数ボーカルが入った音源をメインボーカルとハモリパートに分離することは難しい．

2.2 歌唱練習支援

歌唱の練習支援に関する手法も複数発表されている．Nakano ら [8] は，ユーザの歌声の基本周波数とビブラートをリアルタイムに視覚化し，音楽 CD 音源から推定されるボーカルパートの基本周波数と併せて表示する手法を提案した．福本ら [9] は，相対音感の習得を目的とし，基準音と目標音の 2 つを同時に聴きながら，目標音の音程を歌う練習を繰り返す練習ツールを提案した．この手法では，目標音と自身の歌声のピッチをリアルタイムに画面上で確認することも可能である．しかし，これらの手法は音源を加工しないことや，対象が単一の旋律に限られているため，実際の楽曲でのハモリパートの練習には向いていない．

Shiraishi ら [10] は，楽曲の副旋律を生成することでハモリ練習を支援する手法を提案した．この手法では，入力された楽曲の MIDI ファイルからハモリパートをルールベースの手法で自動生成し，主旋律とハモリパートを再生しながら練習することが可能である．この手法はハモリパートを持たない楽曲における副旋律生成とその練習に主眼を置くものであり，もともと存在するハモリパートの練習を支援するものではない．

3. 提案手法

本研究では，ボーカル楽曲におけるハモリパートの練習支援を目的とし，既存の楽曲からメインボーカルを手軽に抽出・削除し，ハモリパートのみを再生できるツールを提案する．本研究では，メインボーカルは比較的簡単に歌えるという仮定に基づき，削除したい旋律を歌声により入力するユーザインタフェースを提案する．提案手法は，入力された歌声の基本周波数を推定し，この倍音成分を入力音源から削除することでハモリパートのみを含む音源を作成する．提案手法を用いると，メインボーカルを選択的に削除し，ハモリパートのみを強調したハモリ練習用の音源を手軽に作成できる．

3.1 ユーザインタフェース

提案手法を利用する様子と提案手法のスクリーンショットを図 1a，b に示す．まずユーザは，ハモリを練習したい楽曲音源ファイルを用意する．提案手法では，練習した

いハモリパートは特定の短いフレーズであると仮定し，10秒程度の音源が入力されるものとする．なお，楽曲音源にボーカル以外の楽器音が含まれる場合は，事前に音源分離 [5] を実施しボーカルのみを抽出しておく．ユーザが，提案システムを起動し楽曲音源ファイルを読み込ませると，アプリケーション画面上にその楽曲のスペクトログラムが表示される（図 2a）．

続いてユーザは削除したい旋律の歌声を録音する．ユーザが『録音』ボタンを押すと，4 拍のカウント後に楽曲音源が再生され，この音源を聴きながらメインボーカルを歌うとその歌声が録音される（図 1c）．一度録音した歌声は wav ファイルとして記録されるため，2 回目以降はこのファイルを歌声として読み込むことも可能である．歌声が入力されると，そのピッチ（基本周波数）の時間変化が楽曲音源のスペクトログラムに重ねて表示される（図 2b 青線）．ユーザはこの可視化結果を確認し，歌声の録音をやり直すことが可能である．

最後に，ユーザが『主旋律削除』ボタンを押すと，楽曲音源から歌声の基本周波数の倍音成分を削除したデータ（ハモリ音源）（図 1d）が生成され，そのスペクトログラムが表示される（図 2c）．ハモリ音源を作成後，『再生』ボタンを押すとその音源が再生される．このとき，ユーザはスライダ操作により，メインボーカルの音量を調整できる．これにより，メインボーカルを完全に削除した音源や，その音量を半分程度に調整した音源を聴きながらハモリパートを練習できる．

3.2 主旋律の削除

提案手法は，主旋律と副旋律を含む楽曲音源 $m(t)$ とユーザが主旋律を歌った歌声音源 $v(t)$ を入力とし，楽曲音源から主旋律を削除したハモリ音源 $h(t)$ を出力する．まず我々は，$m(t)$ のスペクトログラム $M(n,f)$ と，$v(t)$ のスペクトログラム $V(n,f)$ を計算する．ここで，n は時間方向のインデックス，f は周波数ビンのインデックスである．続いて我々は，歌声音源の基本周波数の追跡を行う．ある時間フレーム n において，歌声のパワースペクトル $|V(n,f)|^2$ が閾値以上の周波数ビンのうち，そのインデックスがもっとも小さいものをフレーム n の基本周波数インデックス $f_0(n)$ とする．なお，フレーム n にて閾値以上の周波数ビンがない場合は，$f_0(n) = Null$ として，そのフレームでは後述する倍音削除は行わない．

次に，楽曲音源のスペクトログラム $M(n,f)$ から，歌声音源より推定した基本周波数を削除する．ここで，人間の聴覚は，音声信号から基本周波数のみが欠落していても，残っている倍音成分から基本周波数を補完して知覚する特性を持つ [11]．そのため，倍音成分を多く含む人間の声から基本周波数のみを削除しても，音が削除されたようには認識されにくい．そこで我々は，楽曲音源のスペクトログラム $M(n,f)$ より，歌声音源の基本周波数とその倍音成分を削除したハモリ音源のスペクトログラム $M'(n,f)$ を作成する．具体的には，下記の通りマスク処理を行う，

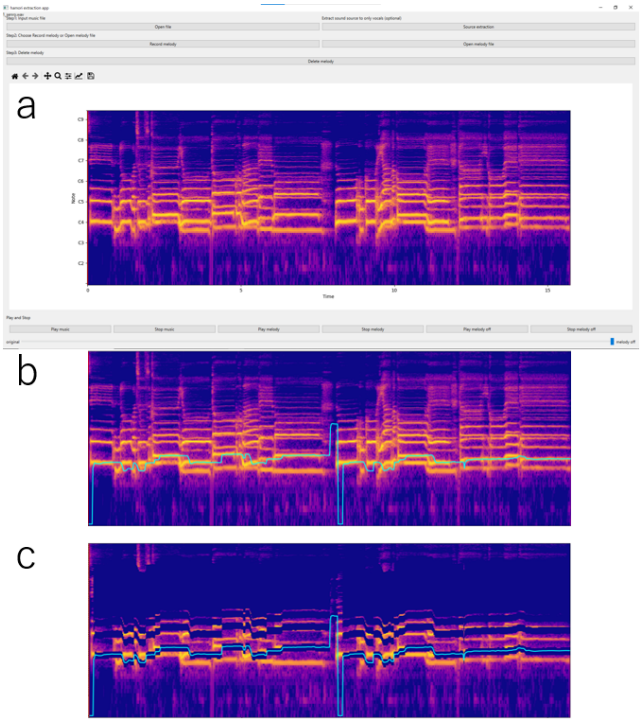

図 2: （a）楽曲音源を読み込んだ際に表示されるスペクトログラム，（b）歌声音源の基本周波数の時間変化の表示（青線），（c）ハモリ音源のスペクトログラム表示．

$$M'(n,f) = c(n,f)M(n,f) \tag{1}$$

$$c(n,f) = \begin{cases} 0 & if\ kf_0(n)2^{-\frac{2}{12}} \leq f \leq kf_0(n)2^{\frac{2}{12}} \\ 1 & otherwise. \end{cases} \tag{2}$$

ここで，k は倍音の番号で $k = 1, 2, ..., 20$ とする．つまり，歌った基本周波数 $f_0(n)$ および第 20 倍音までの倍音成分に対して，その上下 2 度までの範囲の音域を削除することとなる．上下 2 度の範囲を削除する理由は，メインボーカルに対するハモリパートの音は最低でも 3 度以上離れていると考えられるためである．なお本研究では，入力音源のサンプリング周波数は $22050Hz$ とし，スペクトログラムの計算では窓長を $W = 2048$，ホップ長を $S = 512$ とした．

4. ユーザスタディ

提案手法の有用性を評価するため，楽曲音源を聴きながらハモリ練習を行う『従来手法』と，メインボーカルを削除した音源を聴きながらハモリ練習をする『提案手法』をユーザスタディにより比較する．本研究では，ハモリパートを含んだ『ちょうちょう』と『線路は続くよどこまでも』を課題曲として利用する．なお，ちょうちょうは機械音声 [12] にて，線路は続くよどこまでもは歌声を実際に録音することで作成する．ここで，実際の楽曲ではハモリパー

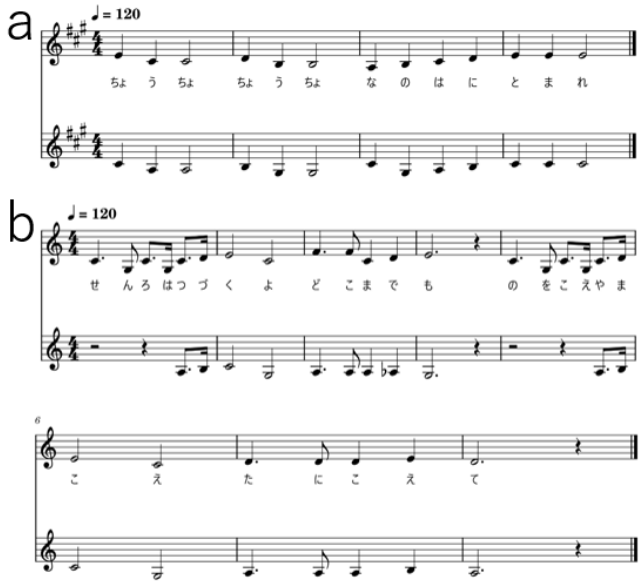

図 3: (a)『ちょうちょう』の楽譜，(b)『線路は続くよどこまでも』の楽譜.

トの音量はメインボーカルと比較して小さいことが多いため，ハモリパートの音量をメインボーカルの半分程度となるようにこれらの音源を作成する．両者の楽譜を図 3 に，楽曲音源のスペクトログラムを図 4a，d に示す．さらに，実際の利用場面における提案手法の有効性を確認するため商用楽曲も用意する．この商用楽曲は，男性ボーカルの有名なバラード曲で，メインボーカルの 3 度下のハモリパートを含むものである．

4.1 タスク

実験参加者は，3 曲分のハモリ練習を実施する．1 曲目は，ふたつの課題曲のどちらかについて，従来手法によりハモリ練習を行う．2 曲目は，先に利用しなかった課題曲について，提案手法を利用してハモリ練習を行う．最後に，3 曲目は，商用楽曲について提案手法でハモリ練習を行う．実験参加者は，3 曲分の練習終了後，アンケートに回答する．なお，周囲を気にせず自由に歌唱できるよう実験は参加者が一人になれる個室で実施する．また，楽曲の提示順による影響を考慮し，1 曲目と 2 曲目に使用する課題曲の順序は参加者ごとに変更する．

従来手法を利用した練習では，メインボーカル・ハモリパートを含む音源を提案手法の音源再生機能を用いて再生し，その音源を聞きながら自由にハモリ練習を行う．一方，提案手法を利用した練習では，まず最初に単純なチュートリアル用音源を用いて提案手法によるメインボーカル削除方法を練習したのち，課題曲のメインボーカルを削除し，作成したハモリ音源を聴きながらハモリ練習をする．

4.2 結果と考察

大学生 9 名の協力の下，前述の実験を実施した．参加者のうち 5 名はアカペラサークルに所属し普段からハモリパートを練習しており，残り 4 名は特別な歌唱経験を持たない学生であった．

ちょうちょうと線路は続くよどこまでもの 2 曲に対して，楽曲音源・入力された歌声音源・提案手法により生成されたハモリ音源のスペクトログラムの一例を図 4 に示す．また，図 4 の矩形領域の拡大図を，図 4A，C，D，F に示す．拡大図において，水色の矩形がメインボーカルを，赤色の矩形がハモリパートを示す．図 4C，F を見ると，出力されたハモリ音源ではメインボーカルのスペクトル（水色枠）が除去されていることがわかる．著者が，実験参加者の作成した音源を試聴したところ，どの参加者・どの楽曲においてもおおむね良好にメインボーカルの削除が行えていることを確認した．特に，商用楽曲においてもハモリ音源では 3 度下のハモリパートを聞き取りやすい音源が生成されていた．

アンケートの結果を図 5 に示す．アンケートでは，3 曲それぞれについて以下に示す 2 項目の質問に 5 段階のリッカート尺度で回答してもらった．

- ハモリパートの音程がわかりやすかった
- ハモリパートの練習がしやすかった

アンケート項目について対応のある両側 t 検定を行ったところ，ハモリパートの練習がしやすかったかどうかについて，有意差が認められた（$p = 0.017$）．なお，楽曲ごとの内訳について確認したところ，ちょうちょう，線路は続くよどこまでもの両方において，提案手法を用いた練習を肯定的に評価する傾向は同様であった．また，商用楽曲のハモリ練習を提案手法を用いて行ったケースへのアンケート結果においても，参加者が肯定的な評価をしており，提案手法が商用楽曲へも適用できる可能性が示唆された．

自由形式でのアンケートでは，商用楽曲の結果について『ハモリをうまく抽出できて，練習しやすかった』といった肯定的なコメントが得られた．この曲は，サビにロングトーンを含み，その部分では歌声のピッチが比較的安定するため，特にハモリパートの音程を確認しやすかったと思われる．これより，提案手法はバラードのような音の移り変わりが少なく，ロングトーンを持つ楽曲に特に有用である可能性が高いと考えられる．ちょうちょう及び線路は続くよどこまでもの結果に対しても『主旋律の音が小さくなるだけでハモリの練習がしやすかった』『主旋律もある状態だとハモリが聞こえなかったが，主旋律を削除することで音程が分かり，練習しやすかった』といった肯定的なコメントが得られた．他にも，一部の参加者からは『消したい音の大きさを好きに変化させられるのはとても使いやすい』といった，スライダを用いて段階的に元の音源に近づけていく練習に対しても肯定的なコメントが得られた．

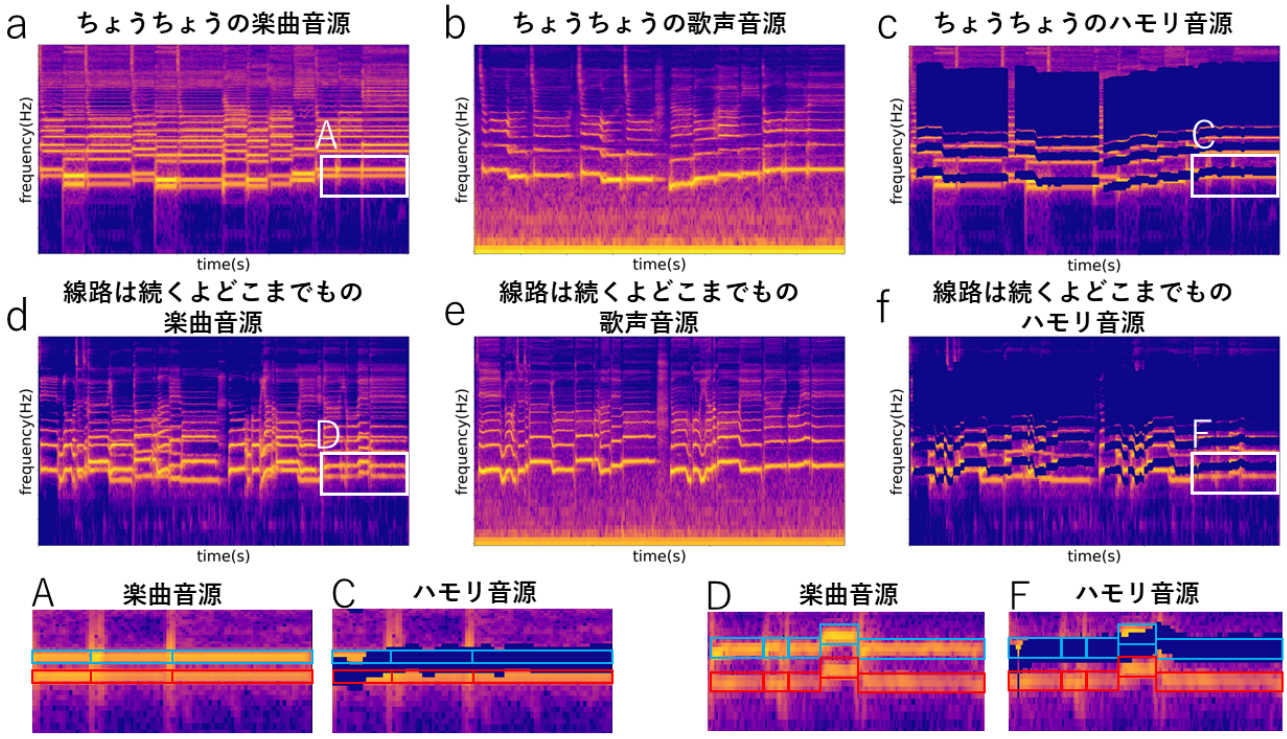

図 4: 『ちょうちょう』の（a）楽曲音源，（b）歌声音源，（c）ハモリ音源，『線路は続くよどこまでも』の（d）楽曲音源，（e）歌声音源，（f）ハモリ音源のスペクトログラムおよびその拡大図（A,C,D,F）.

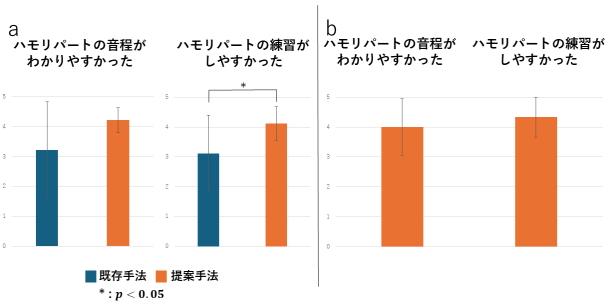

図 5: （a）『ちょうちょう』と『線路は続くよどこまでも』のアンケート結果，（b）商用楽曲のアンケート結果.

一方，スペクトログラムの拡大図（図 4C，F）をよく見ると，歌声の細かなピッチのずれによりハモリパートの一部が削除されてしまっていることが確認できる．メインボーカルの音高が変化する部分においては，歌声のピッチが安定しないことも多いため，このようなエラーが発生したと考えられる．また，提案手法に対して，『主旋律の音程を取るのに時間がかかってしまった』という否定的なコメントがあった．これは，歌唱に対して不慣れであった一部の参加者が主旋律削除に手間取ってしまったと考えられる．提案手法の利用には，正確なピッチで歌う必要があるため，今後は GUI を用いて細かなピッチの修正を手動で行う機能を追加することで，結果が改善される可能性がある．

5. まとめと展望

本研究では，ハモリパートの練習支援を目的とし，楽曲からメインボーカルを手軽に抽出・削除し，ハモリパートのみの音源を再生できるツールを提案した．特に本研究では，メインボーカルは比較的簡単に歌えるという仮定に基づき，削除したい旋律をユーザの歌声を用いて入力するユーザインタフェースを提案した．提案手法の有用性を確認するためユーザスタディを実施した結果，多くの実験参加者が歌声の入力によりおおむね良好なメインボーカル削除を行えたことを確認した．また，多くの参加者より，提案手法に対して肯定的な評価が得られた．

提案手法の将来課題のひとつはユーザビリティの向上である．テンポの速い曲の主旋律の正確な入力のため，歌った音源のピッチを手動で修正する機能および歌詞表示機能・メトロノーム機能を実装したい．また，ハモリ音源生成の高精度化のため，単純なマスク処理ではなく機械学習を導入することも将来課題のひとつである．

参考文献

[1] Goto, M.: A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals, *Speech Communication*, Vol. 43, No. 4, pp. 311–329 (2004).
[2] Yao, G., Zheng, Y., Xiao, L., Ruan, L. and Li, Y.: Efficient Vocal Melody Extraction from Polyphonic Music

Signals, *Electronics and Electrical Engineering*, Vol. 19, pp. 103–108 (2013).
[3] Zhang, W., Wang, R., Zhang, Q. and Fang, S.: A joint pitch estimation and voicing detection method for melody extraction, *Applied Acoustics*, Vol. 166, p. 107338 (2020).
[4] Nugraha, A. A., Liutkus, A. and Vincent, E.: Multichannel music separation with deep neural networks, *EUSIPCO*, pp. 1748–1752 (2016).
[5] Uhlich, S., Porcu, M., Giron, F., Enenkl, M., Kemp, T., Takahashi, N. and Mitsufuji, Y.: Improving music source separation based on deep neural networks through data augmentation and network blending, *ICASSP*, pp. 261–265 (2017).
[6] Jansson, A., Humphrey, E. J., Montecchio, N., Bittner, R. M., Kumar, A. and Weyde, T.: Singing Voice Separation with Deep U-Net Convolutional Networks, *ISMIR* (Cunningham, S. J., Duan, Z., Hu, X. and Turnbull, D., eds.), pp. 745–751 (2017).
[7] Chandna, P., Blaauw, M., Bonada, J. and Gomez, E.: A Vocoder Based Method for Singing Voice Extraction, *ICASSP*, IEEE, p. 990–994 (2019).
[8] Nakano, T., Goto, M. and Hiraga, Y.: MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data, *ISMW 2007*, pp. 75–76 (2007).
[9] 福本愛由星，橋田光代，片寄晴弘：RelaPitch: 歌って鍛える相対音感習得支援システム，エンタテインメントコンピューティングシンポジウム 2016 論文集， Vol. 2016, pp. 332–333 (2016).
[10] Shiraishi, M., Ogasawara, K. and Kitahara, T.: HamoKara: A System that Enables Amateur Singers to Practice Backing Vocals for Karaoke, *Journal of Information Processing*, Vol. 27, pp. 683–692 (2019).
[11] Moore, B. C. J.: 聴覚心理学概論，誠信書房 (1994). 大串健吾 監訳.
[12] 飴屋／菖蒲：UTAU (2008). Version 0.4.19. [Software]. Available from: `https://utau-synth.com/`.