

スペクトル差異に着目した 単独歌唱音声の合唱への最適化の検討

稲津 遥太郎^{†1,a)} 齋藤 大輔^{†1,b)} 峯松 信明^{†1,c)}

概要：合唱歌唱では、通常時の歌唱と比較して発声面で差異が生じることが知られている。本研究ではその差異に着目し、斉唱歌唱も含めた複数人歌唱で条件ごとにどのようなスペクトル上の特徴が現れるのかを検証した。また、メルケプストラム差分フィルタを用いた単独歌唱音声から重唱・斉唱への最適化を検討した。単独歌唱音声にメルケプストラム差分フィルタを用いることで生成した合唱音声に対して主観評価実験を行ったところ、二声部合唱の副旋律声部歌唱に関しては合唱音声の自然さが向上する傾向があることがわかった。また、メルケプストラム差分フィルタによって合唱音声が向上する可能性が専門の見地からも確認された。

1. はじめに

歌唱における合唱は、複数の人間が同時に歌唱を行うことで実現される音楽の表現形式である。また、その中でも全く同じメロディを複数人でなぞる歌唱を特に斉唱という。合唱や斉唱は、複数の歌声が重なり合い調和して、メロディや場合によってはハーモニーを奏することで、単独での歌唱とはまた異なる音色を得ることができるとされている。

近年需要が増加している音声認識・音声合成技術において通常の発話のみならず歌声に関する研究も盛んになってきており、VOCALOID[1]をはじめとする歌声合成技術が音楽制作の中で用いられることも一般的になっている。その中で歌唱の中でも一般的な部類に入ると考えられる合唱歌唱の研究課題としての重要性は工学的にも高まっている。

合唱では複数の歌唱者が同時に歌うことで声道の状態や声帯振動が異なる歌唱が混在し特有の音色として表出すると考えられる。それに加えて複数の歌唱が同時に起こることで生じる歌唱者への相互作用も音色に影響する要因となりうる。この相互作用に関する要因の具体的な影響については研究が未だ十分にされていない。本研究では、周辺歌唱が存在するときに歌唱者にどのような影響を与えるのかについてメルケプストラム分析の手法を用いて明らかにし、歌声の音色に現れる影響の度合いを評価・検討する。また単独歌唱時と合唱歌唱時のメルケプストラムから差分

フィルタを作成して歌唱音声を合唱に最適化する手法を構築し、実験的に評価・検討する。これによって単独での歌唱音声から自然な合唱音声の生成が実現でき、また分析の過程で合唱自体の知覚・生成機構への知見を深めることに繋がる。

本稿は全6章によって構成される。第2章でソースフィルタモデルに基づくスペクトル分析の手法について述べ、第3章では関連する先行研究について述べる。第4章では行った歌唱実験の詳細と分析結果について述べる。第5章ではメルケプストラム差分フィルタによる合唱音声の主観評価実験について述べ、第6章では本研究の課題と今後の展望を述べる。

2. ソースフィルタモデルに基づくスペクトル解析の手法

2.1 スペクトル包絡

スペクトル包絡は音声の音色を表すパラメータであり、発声時の声道の状態に対応する。これはソースフィルタモデルという、人間の発声を声帯振動の基本周波数のパラメータと声道によって付与される音色を説明するパラメータでモデル化する考え方に基づいている。後者のパラメータが、スペクトル包絡にあたる。

このモデルでは、音声のスペクトル $Y(\omega)$ はソースに由来するスペクトル $G(\omega)$ とフィルタ由来のスペクトル $H(\omega)$ によって以下のように記述する。

$$Y(\omega) = G(\omega)H(\omega) \quad (1)$$

またこれは時間領域においては、出力信号たる音声信号

^{†1} 東京大学大学院工学系研究科

a) inatsu@gavo.t.u-tokyo.ac.jp

b) dsk_saito@gavo.t.u-tokyo.ac.jp

c) mine@gavo.t.u-tokyo.ac.jp

$y(t)$ は声帯がもたらす周期振動に起因した信号 $g(t)$ とフィルタに起因する信号 $h(t)$ の畳み込み積分の形で表されるということを意味する。

$$y(t) = g(t) * h(t) \quad (2)$$

ソースフィルタモデルによれば歌唱音声はこのようにソース成分の周期信号とフィルタ成分の組み合わせとして考えて紐解くことができる。そのため分析においては、この2成分をなんらかの方法で分離して別個に捉え直し、声帯の振動と声道の状態を分けて観察することが重要な意味を持つ。歌唱においては、基本周波数の動きは歌唱される楽句にほぼ対応されるものであり、スペクトル包絡は歌唱者自身の特性や母音の特性などを合わせた音色の特徴量に対応する。

2.2 ケプストラム分析

ケプストラム分析は、スペクトル包絡を推定するためによく用いられている手法である。ケプストラムとは、音声の対数振幅スペクトルを逆フーリエ変換して得られるパラメータである。前述のモデルにおけるフィルタ成分であるスペクトル包絡を表すことのできるパラメータであり、音声分析において重要なものである。式 (1) の両辺の対数をとれば、

$$\log |Y(\omega)| = \log |G(\omega)| + \log |H(\omega)| \quad (3)$$

となり、スペクトル領域の関数の積で記述されていたものが和へと変換できる。これによりソース成分とフィルタ成分の分離が行いやすくなる。これに逆フーリエ変換を施したものがケプストラム $c(\tau)$ であるので、

$$c(\tau) = \mathcal{F}^{-1} \log |G(\omega)| + \mathcal{F}^{-1} \log |H(\omega)| \quad (4)$$

となる。ケプストラムの時間軸単位はケフレンシーと呼ばれる。

2.3 メルケプストラム

このケプストラムに非線型変換を施して人間の知覚特性により適合させたものは、メルケプストラムと呼ばれる。これは数十次元程度でスペクトル包絡を近似的に表現するパラメータである。人間は高周波成分よりも低周波成分の方により敏感になるため、近似の標本を取る時に低周波数領域の方をより密に標本化するような変換をすることで、より知覚特性に沿った精度の高い近似が実現できる。

メルケプストラムは歌唱時の声道の状態・声質を表すパラメータとなるため、これにより複数の歌唱音源における歌唱者の声質の差異を定量的に論じることができる。また、メル対数スペクトル近似 (mel log spectrum approximation, MLSA) フィルタという手法を用いることで、音源波とメルケプストラム係数から音声を合成することができる [2]。

2.4 差分スペクトル補正

ソースフィルタモデルを用いて音声を変換する手法の一つに、差分スペクトル補正と呼ばれる手法がある [3]。入力音声 $y_1(t)$ に対して逆フィルタをかけて得られるインパルス列をソース成分として、そこに異なるフィルタ $h_2(t)$ を畳み込み新たな音声を合成することを考える。これをスペクトル領域で表すと求める合成音声は以下のように表せる。

$$Y(\omega) = \frac{Y_1(\omega)}{H_1(\omega)} H_2(\omega) = Y_1(\omega) \frac{H_1(\omega)}{H_2(\omega)} \quad (5)$$

すなわち、合成音声は入力音声に対して $\frac{H_1(\omega)}{H_2(\omega)}$ をかけることで得られる。これは適用したいフィルタを入力音声のフィルタで割った形になるが、ケプストラム領域に直すと差分の足し合わせという簡単な構造で音声合成が実現できる。これはスペクトル領域で分数だった部分が対数変換によって差分の形に表すことができ、そこから逆フーリエ変換をしてもこれは線形変換であるため差分の関係性が保存されることによるものである。メルケプストラムにおいても同様にこの関係は成り立つ。

3. 関連研究

本章では、合唱歌唱分析に関連する先行研究を説明する。合唱とソロ歌唱の歌声の違いについて、ソプラノ歌手が合唱に声を混ぜ歌唱した場合と単独歌唱をした場合の両方で歌声比較を行ったところ、そのスペクトル分析から単独歌唱時の方がより高音が大きくまた大きな声量であった [4]。また、男性歌手がソロとして歌唱した際と合唱として歌唱した際に、同じ楽句の歌唱であっても平均スペクトル包絡に差が現れた。具体的には、ソロでの歌唱の方が低い周波数領域の音が比較的弱くなり、500Hz より高周波である領域の歌声フォルマントが少し強い傾向があった [5]。この実験は合唱歌手・ソロ歌手の両方の経験を持つ歌手を被験者としたものであり、合唱とソロ歌唱においては声の使い方が異なると考えられる。

四声部からなる混声合唱では、合唱時に望ましい声として「明るい声」が要求されるとされ、この知覚には音響特徴量として長時間平均スペクトルのスペクトル包絡と微細構造の高周波域に現れる倍音構造の規則性が関連することが明らかになっている [6]。また、単独で歌唱したペアのスペクトル包絡をソースフィルタモデルに基づいて合成した合唱音声について、聴者が感じる調和度はペア内のメルケプストラム歪み (Mel-cepstral distortion; MCD) が 0.75–0.5 倍程度になるよう双方向にスペクトル包絡を近づけるとより調和する傾向があった [7]。

4. 周辺歌唱フィードバックの影響の調査実験

合唱における周辺歌唱がもたらす歌唱への影響を定量的に考察するため、音声フィードバックとして周辺歌唱の存在する状態で歌唱音源を録音し、それを周辺歌唱の音声

表 1 実験ごとの周辺歌唱条件と収録歌唱条件

	実験での収録歌唱
実験 1：斉唱実験	主旋律
実験 2：主旋律声部実験	主旋律
実験 3：副旋律声部実験	副旋律

フィードバックのない状態で収録した歌唱と比較し差異を検討する。

4.1 概要

合唱歌唱は、斉唱と呼ばれる歌唱者全員が同一の旋律を重ねて歌う形式と、斉唱でない複数の声部からなる合唱の二つに大別することができる。後者の場合歌唱者は聴こえてくる周辺歌唱とは異なる旋律を歌唱するのに対し、前者の斉唱の場合周辺歌唱と全く同じ旋律を歌唱することになる。このことを鑑みて、本研究では斉唱の場合とそれ以外の合唱の場合を区別して扱い、両方の条件で実験を実施することとした。また斉唱でない合唱において、曲の主旋律を担う声部と主旋律でない旋律（以下、副旋律と呼称する）についても区別して実験を実施した。

以上のことから本実験では表 1 に示すように、斉唱の環境を再現し周辺歌唱として与えられるのと同一の主旋律を被験者に歌唱させる実験 1、副旋律の声部の歌唱条件を再現し周辺歌唱として主旋律を聴かせて副旋律を歌唱させる実験 2、主旋律の声部の歌唱条件を再現し周辺歌唱として副旋律を聴かせて主旋律を歌唱させる実験 3 を行うこととした。それぞれの実験条件で周辺歌唱ありでの歌唱と周辺歌唱なしでの歌唱を収録し、メルケプストラム分析からスペクトル上にどのような差が現れるのかを検証した。

4.2 共通の実験条件

音声収録に参加した被験者は 20 代の男性 4 名で各々 4 年、7 年、9 年、16 年の合唱経験を有している。以下、被験者 A、B、C、D とする。

歌唱音声の収録にあたっては、マイクはコンデンサーマイクの MXL V67G-HE を、オーディオインターフェースは Scarlett 2i2 3rd Gen を使用した。また音源の録音と伴奏音源の作成に際してソフトウェアは Logic Pro X を使用した。音声の提示に用いたヘッドホンは、OneOdio Pro10 を使用した。

歌唱には、日本の童謡で文部省唱歌である『かたつむり』を用いた。周辺歌唱のフィードバック音声として、データセットである JVS-MuSiC[8] 内に収録されている音声から選定した jvs001、jvs098 の歌唱音声を用いた。以降の実験ではそれぞれ周辺音声 1、2 として ID を割り振る。データセット JVS-MuSiC は、100 人の歌唱者がそれぞれ単独歌唱した音声で収録されているもので、『かたつむり』は全員が歌う共通曲である。データセットの歌唱音声は音声編集ソフトウェア Melodyne を用いて音高・テンポが修正

かたつむり

文部省唱歌



図 1 使用したかたつむりの歌唱用楽譜

され、曲のキーは D フラット、テンポが 100bpm となるよう統一された音声を周辺歌唱音声とし、実験条件によっては適宜 Melodyne を用いて音声の音高を再度修正して用いた。

本実験では歌唱に際して、歌い出しや曲中の音高・テンポの指標になるものとして、簡単なピアノ伴奏とメトロノーム音を合わせた音源を用意した。ピアノ伴奏音源やメトロノーム音は Logic Pro X で作成した。音声収録の際のサンプリング周波数は 44100Hz、また全てモノラル形式での収録である。

4.3 実験 1：斉唱実験

周辺歌唱と同一の旋律を被験者に歌唱してもらうことで、斉唱合唱における周辺歌唱の影響を調べる。各条件で 3 回ずつの歌唱収録を行う。被験者はピアノ伴奏音源をヘッドフォンから提示されている状態でキー D フラット、テンポが 100bpm で『かたつむり』を歌唱する。その後ピアノ伴奏音源に加えて周辺歌唱 1 も聞こえる状態で周辺歌唱に合わせた斉唱歌唱を行い、これを収録する。周辺歌唱 2 でもこれを行い歌唱を収録する。

実験に用いた楽譜を、図 1 に示す。

4.4 実験 2：主旋律声部実験

二声部合唱の副旋律の周辺歌唱を音声フィードバックとして用意し、被験者には主旋律の声部を歌唱してもらうことで、合唱における主旋律声部への周辺歌唱の影響を検証する。各条件で 3 回ずつの歌唱収録を行う。

周辺歌唱 1、2 の音高を修正し、図 2 における下声部の旋律の歌唱音声を用意する。この譜面は二声部からなる合唱譜面であり、上声部が『かたつむり』の通常的主旋律となり、対して下声部は副旋律を担う。被験者はピアノ伴奏音源をヘッドフォンから提示されている状態でキー D フラット、テンポが 100bpm で『かたつむり』を歌唱する。その後ピアノ伴奏音源に加えて、音高修正を施した周辺歌唱 1 も聞こえる状態で周辺歌唱に合わせた主旋律声部の合唱歌唱を行い、これを収録する。周辺歌唱 2 でもこれを行い歌唱を収録する。



図 2 かたつむりの主旋律声部歌唱用二声部合唱譜

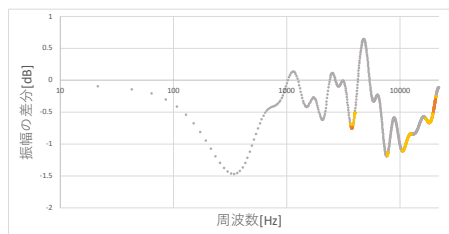


図 3 斉唱歌唱の差分スペクトルの被験者平均

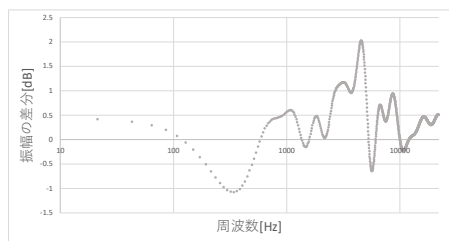


図 4 主旋律声部歌唱の差分スペクトルの被験者平均

4.5 実験 3：副旋律声部実験

二声部合唱の主旋律の周辺歌唱を音声フィードバックとして用意し、被験者には副旋律の声部を歌唱してもらうことで、合唱における副旋律声部への周辺歌唱の影響を検証する。各条件で 3 回ずつの歌唱収録を行う。

歌唱者はピアノ伴奏をヘッドホンから提示された状態で二声部合唱譜の下声部の旋律を歌唱する。この二声部合唱の楽譜は実験 2 で用いた図 2 の楽譜のキーを E に変更したものである。伴奏音源もキーを E に変更したものを、周辺歌唱もキー E の主旋律となるよう音高を修正したものをを用いる。これらの変更は被験者の歌唱が平均的な男性の地声音域から大きく離れるのを防ぐためである。

その後、ピアノ伴奏音源に加えて音高修正後の周辺歌唱 1 も聞こえる状態で周辺歌唱に合わせた副旋律声部の合唱歌唱を行い、これを収録する。周辺歌唱 2 でもこれを行い歌唱を収録する。

4.6 音声分析結果

SPTK[9] を用いて得られた収録歌唱音声のメルケプスト

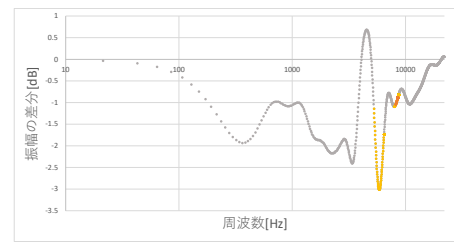


図 5 副旋律声部歌唱の差分スペクトルの被験者平均

ラム係数を算出し、歌唱の音色を示すスペクトル包絡の成分を取り出した。メルケプストラム係数は、次数が 24 次、 $\alpha = 0.54$ として算出した。窓長は 2048 サンプル、シフト長は 5.4ms に設定した。得られたメルケプストラムを時間平均することで平均スペクトル包絡を算出し、周辺歌唱のある状態での歌唱の平均スペクトル包絡から周辺歌唱のない状態での平均スペクトル包絡を引いた差分を用いて、合唱歌唱でのスペクトル包絡の変化分を取り出す。なおここでのスペクトル包絡の値は対数値を用いており、ケプストラム領域での差分フィルタに対応している。

斉唱歌唱条件における被験者 4 名の差分スペクトル包絡の平均を図 3 に示す。ここでは $n = 4$ で両側 t 検定を用い、有意水準 10% で有意傾向、有意水準 5% で有意差があると判定している。

斉唱歌唱の差分スペクトル包絡は、3800Hz, 20000Hz 周辺に有意な振幅の低減があることがわかった。

主旋律声部実験に関して、被験者 4 名の差分スペクトル包絡の平均を図 4 に示す。算出にあたっての手順や検定の条件は実験 1 と同様である。主旋律声部の歌唱に関しては、被験者平均のスペクトルではどの周波数領域でも有意な差が出なかった。

副旋律声部実験に関して、被験者 4 名の差分スペクトル包絡の平均を図 5 に示す。算出にあたっての手順や検定の条件は実験 1, 2 と同様である。結果をみると、6000Hz の周波数低減に有意傾向が、8300Hz 周辺に有意な振幅の低減があった。

5. メルケプストラム差分フィルタによる合唱歌唱への最適化の評価

前章での実験から得られた単独歌唱時のメルケプストラムと合唱時のメルケプストラムを用いて差分フィルタを構成し、単独歌唱音声から合唱に最適化する手法を実験的に評価する。メルケプストラム差分フィルタによる最適化の評価は、2 段階の実験に分けることで手法の効能と汎化性の評価を行う。まず第一の段階ではメルケプストラム差分フィルタを歌唱者個人の特性に最適化して構成して行う。第二に、異なる歌唱被験者のデータで最適化されたメルケプストラム差分フィルタで変換して生成される合唱音声の自然さの主観評価を行う。

5.1 歌唱者最適化型フィルタ

前章の実験1における周辺歌唱なしの単独歌唱音声の一つを被験対象音声とし、それを除いた残り2つの単独歌唱音声と6つの周辺歌唱ありの音声から、メルケプストラム差分フィルタを構成する。具体的な手順としては、周辺歌唱ありの歌唱音声の時間平均メルケプストラム6つの平均を取ったメルケプストラム係数ベクトルから、単独音声2つの時間平均メルケプストラムをそれぞれ求め平均を取ったメルケプストラム係数ベクトルを差し引くことで、メルケプストラム差分フィルタが得られる。このようにして得られたこのメルケプストラム差分フィルタを用いてMLSAフィルタリングの手法で被験対象音声にフィルタ変換を施すことができる。

周辺歌唱1を声質変換を施した被験対象音声と足し合わせることで、歌唱被験者A個人に最適化された斉唱歌唱音声を作成した。こうして得られた音声と、周辺歌唱1とフィルタリングを施していない被験対象音声の単純な足し合わせによって合成される斉唱合唱音声とを対にした。また、聴取実験における被験者の負担感の軽減のため、斉唱合唱音声を『かたつむり』の一番と二番で切ってペアを二つに分けた。

これらを周辺歌唱2に対しても同様に行った。このようにして対になる合唱音声を得たのち、それぞれの対ごとにどちらの斉唱合唱音声の方がより自然であるかを選択させるABテストの聴取実験を実施し主観評価を行った。聴取者はクラウドソーシングサービスを用いて募り、各ペアそれぞれ25人からの評価を受けるようにした。歌唱被験者B,C,Dでも同様に、それぞれの個人性に最適化した斉唱歌唱変換を行った。

また、前章の実験2の主旋律声部の合唱歌唱実験のデータからは同様の方式で主旋律声部の合唱へ最適化したフィルタを作成し、周辺歌唱1,2の音高を修正した副旋律声部の周辺歌唱音源と足し合わせることで二声部合唱音声を作成し同様に主観評価実験を実施した。前章の実験3の副旋律声部の合唱歌唱実験のデータからも同様にフィルタ作成と主観評価実験を実施した。

5.2 歌唱者汎化型フィルタ

歌唱者個人に最適化したメルケプストラム差分フィルタを作成し、そのうち1人を除いた3人のフィルタを平均して歌唱者汎化型フィルタを作成した。これを用いたMLSAフィルタリングを、あらかじめ除いていた被験者の歌唱した単独歌唱音声に施した。その他の手順は同様にして対になる合唱音声を得たのち、クラウドソーシングを用いて集めた聴取被験者に対してABテストの主観評価実験を行った。各ペアそれぞれ25人ずつから評価を受けた。

また、この歌唱者汎化型フィルタを用いた合唱音声に関しては、合唱歌唱指導の経験がある声楽家2人に対してア

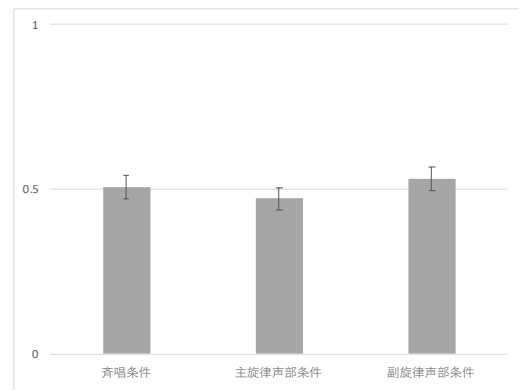


図6 個人最適化型フィルタの被験者全体での主観評価値

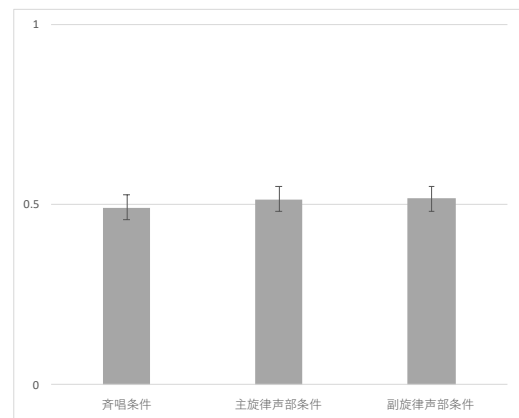


図7 歌唱者汎化型フィルタの被験者全体での主観評価値

ンケート評価を実施した。評価に当たって、フィルタによる変換後の音源と変換前音源を比較してどのような変化があるか、そしてその他気になった点に関して自由に記述してもらう形式で行った。斉唱音声、主旋律歌唱音声、副旋律歌唱音声の3通りの条件に歌唱被験者4人それぞれの歌唱に関して実施したため、評価に当たって12組のペア音源からなる聴取実験となった。また参照音源として、被験者に周辺音声を聴かせた状態で歌唱させた収録音源を用いて合唱音声を作成した。フィルタ変換前の音源は周辺歌唱を聞かない単独歌唱であるので、周辺歌唱をきいて歌うという通常の合唱時を模した収録条件である参照音源とは差異があると思われる。

5.3 実験結果

歌唱者個人に最適化したメルケプストラム差分フィルタを用いた合唱音声の主観評価結果を図6に示す。エラーバーは95%信頼区間である。この主観評価値が0.5を超えていればメルケプストラム差分フィルタによって合唱音声の自然性が向上することが示される。

被験者全体での平均的な傾向として、副旋律声部へのフィルタリングがもっともよく合唱の自然性の向上を示した。主旋律声部のフィルタリングは合唱の自然性の向上という点では比較的低い結果であった。

次に、歌唱者に対して汎化型のメルケプストラム差分フィルタを用いた合唱音声の主観評価結果を図7に示す。エラーバーは95%信頼区間である。

二声部合唱歌唱の主旋律声部と副旋律声部の歌唱者汎化フィルタが高い傾向があると言える。有意な差ではないが、斉唱でない合唱歌唱に関するスペクトル包絡の変化に汎化性があると思われる。斉唱歌唱に最適化したフィルタについては本実験では有効性は見られなかった。

歌唱者汎化型フィルタに対する専門家へのアンケート調査の結果としては、12通りの条件全てにおいて、評価者両者ともがフィルタによる変化によって合唱に改善が見られたとする評価が得られた。条件ごとに異なるものの、音量のバランス・合唱自体のバランスの向上、声の「柔らかさ」・「円やか」さの向上、聞き取りやすさの改善、などがフィルタ変換自体への評価としてよく挙げられていた。特にバランスの向上に関して言及する評価は多い傾向にあった。なお、フィルタによる音声加工での合唱へのネガティブな変化が起きたとする評価は本実験では全く得られなかった。

参照音源に関しては条件ごとに評価が大きく分かるところとなったが、傾向としてフレーズのおわりなどが揃っていて聞き取りやすいという評価がよく見られた。

6. まとめ

本研究では、合唱における周辺歌唱を再現した音声フィードバックを与えた時、歌唱がそれによってどのような影響を受けるのかを検証した。歌唱音源のスペクトル包絡の様子をメルケプストラム係数を用いて分析し、斉唱合唱や二声部の合唱の際にスペクトルに現れる差異を調べた。得られたデータを元に、斉唱合唱歌唱や二声部の合唱歌唱へ最適化されたメルケプストラム差分フィルタを用いて歌声変換を行った。これを元に作成した合唱音声に対して聴取実験を実施し主観評価を行った。その結果、二声部合唱の副旋律声部に対するメルケプストラム差分フィルタを用いて生成された合唱音声は自然さが向上する傾向があることがわかった。また、歌唱者汎化型のフィルタ変換によって合唱音声のバランスの向上が見られる可能性が専門的な見地からも確認された。

本研究で扱った合唱歌唱への最適化について、主旋律声部条件や斉唱条件のABテストでは合唱音声変換の自然性の有意な向上は見られなかった。声質変換による合唱音声のバランスの向上は専門家のアンケート調査からも確認されているものの、それ以上にフィルタをかける変換自体によって損なわれる歌唱音声の自然性の部分が大きく反映されたことが原因として考えられる。合唱音声変換の自然性に関して今後の課題も残るものの、声質変換技術によって合唱自体が向上することが専門的な見地から示されたことは大きく、今後のさらなる合唱音声変換技術の研究が期待

される。一方、本研究では基本周波数に関して、与えられた楽句によってほぼ決定されることから歌唱者・歌唱条件間で差はないと考えて扱わなかった。しかし実際には歌唱者によってフレーズの歌い回しの際に微妙に異なるような基本周波数の変動成分が存在することが知られており、今後の課題である[10]。また本研究で明らかにできた合唱歌唱でのスペクトルの傾向は、被験者ごとの個人性が強く現れた結果であった。より一般的な傾向を議論するには、被験者を増やした大規模な実験が今後必要である。

参考文献

- [1] 剣持秀紀, 大下隼人. 歌声合成システム VOCALOID. 情報処理学会研究報告. [音楽情報科学], Vol. 72, No. 102, pp. 25–28, 2007.
- [2] 今井聖, 住田一男, 古市千枝子. 音声合成のためのメル対数スペクトル近似 (MLSA) フィルタ. 電子情報通信学会論文誌 A, Vol. J66-A, No. 2, pp. 122–129, 1983.
- [3] 小林和弘, 戸田智基, 中村哲. 差分スペクトル補正による統計的歌声声質変換とパラメータ生成法 (音声). 電子情報通信学会技術研究報告 信学技報, Vol. 115, No. 253, pp. 7–12, 2015.
- [4] A. Goodwin. Acoustic study of individual voices in choral blend. *Journal of Research in Music Education*, Vol. 28, No. 2, pp. 119–128, 1980.
- [5] T. D. Rossing, J. Sundberg, and S. Ternstrom. Acoustic comparison of voice use in solo and choir singing. *The Journal of the Acoustical Society of America (JASA)*, Vol. 79, No. 6 pp. 1975–1981, 1986.
- [6] 田和明洋, 田中利幸. 合唱において望ましいとされる声に関する音響特徴分析. 電子情報通信学会技術研究報告 信学技報, Vol. 114, No. 475, pp. 301–306, 2009.
- [7] 山内孔貴, 須田仁志, 齋藤大輔, 峯松信明. ソースフィルタ分解に基づく複数歌唱者の調和制御に関する検討. 研究報告音声言語情報処理 (SLP) 2020-SLP-132, 35, pp. 1–6, 2020.
- [8] H. Tamaru, S. Takamichi, N. Tanji, and H. Saruwatari. JVS-MuSiC: Japanese multispeaker singing-voice corpus. *arXiv:2001.07044 [cs.SD]*, Jan 2020.
- [9] Speech signal processing toolkit (SPTK). SPTK Working Group <http://sp-tk.sourceforge.net>, 2009
- [10] T. Saitou, M. Unoki, M. Akagi. Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis, *Speech Communication*, Vol. 46, No. 3, pp. 405–417, 2005.

付 録

各サンプルに対する声楽家評価者のコメントを表 A.1 – A.6 に示す。

表 A-1 声楽家評価者1のアンケート評価結果：斉唱条件

	フィルタによる変化	(その他・備考)
歌唱者 a	歌唱者 a の声が多少聴きやすくなった	参照音源が一番よく合っていた 歌唱に一部歌詞間違いがあった
歌唱者 b	より 2 声が密接に重なり合ったと感じた	りへ、るへ、せへ、などの長音が伸びすぎないようにしたほうが 重唱としてより良いと思う
歌唱者 c	元音源では歌唱者 c の声がやや小さく感じたが より聞こえるように修正された	参照音源は声質がややバラバラな感じが、 メロディの縦線はあっていて元気に歌っている印象を受けた
歌唱者 d	2 声がより混じり合うようになり 歌唱者 d の歌唱が目立たなくなっていた	歌唱者 d の方は、いい声をしている こぶし回しや子音の強さなどは合唱の時には注意した方がいいが 個性としてはとても良いものだと思う

表 A-2 声楽家評価者2のアンケート評価結果：斉唱条件

	フィルタによる変化	(その他・備考)
歌唱者 a	よりまとまって聞こえるようになった	参照音源はフレーズの終わりや声質が揃っていて聴きやすい
歌唱者 b	歌唱者 b の声が落ち着いて聞こえるようになった	参照音源はフレーズの終わりや声質が揃っていて聴きやすい
歌唱者 c	音量、声の質ともに まとまって聞こえるようになった	参照音源はフレーズの終わりや子音の立て方が より揃っていて聴きやすい
歌唱者 d	元音源より音量のバランスが 整ったと感じる	装飾的な発声技巧 (小節) を時折感じたが 独唱以外はないほうが望ましいのでは 声はとても良いと思う

表 A-3 声楽家評価者1のアンケート評価結果：主旋律声部条件

	フィルタによる変化	(その他・備考)
歌唱者 a	元音源ではやや沈んでいた旋律が 聴きやすく調整されていたと感じた	つのだせ、から盛り上がっていく歌い方はいい
歌唱者 b	元音源では主旋律が強く聞こえていたが、 バランスが良くなっていたと思う	なし
歌唱者 c	元音源では主旋律メロディの後半が やや弱くなってしまっていたが、 最後まで聴きやすくなっていたと感じた	なし
歌唱者 d	元音源ではやや副旋律が聞き取りづらかったが、 主副のバランスが良くなりハーモニーが聴きやすくなった	なし

表 A-4 声楽家評価者2のアンケート評価結果：主旋律声部条件

	フィルタによる変化	(その他・備考)
歌唱者 a	元音源に比べて主旋律が聞こえて 重唱としてバランスがいいと感じる	参照音源は副旋律に引っ張られる影響なのか 音程が定まらないところがある 変換元音源は音程がしっかりしていて良いと感じる
歌唱者 b	元音源と参照音源も共にしっかり主旋律が聞こえ 重唱として良いと感じるが、 変換後の音源の方がバランスを感じて若干抑えめに聞こえる 変換後の方が重唱として好ましく感じる	フレーズの終わりがやや揃わないところがあるので ない方が望ましい
歌唱者 c	元音源に比べ主旋律が印象に残り 重唱として好ましく感じる	歌詞をもっとレガートで歌った方が 重唱として縦の線が揃いやすいのでは
歌唱者 d	元音源に比べ声の質が明るい印象に聞こえる 発声法的には好感が持てる	参照音源は副旋律と縦の線やフレーズが揃っており、 変換後の音源より重唱として望ましいと感じる

表 A-5 声楽家評価者1のアンケート評価結果：副旋律声部条件
(フィルタによる変化 (その他・備考))

	フィルタによる変化	(その他・備考)
歌唱者 a	主旋律と混ざりやすい声質になったと思う	全体的に2声のハモリはきれい
歌唱者 b	より主旋律と副旋律が混ざりあう声質になった	参照音源は歌唱の縦線がよく合っている バランスはやや副旋律が強気味 変換元音源は副旋律が抑え気味で主旋律が立つ歌い方になっていた
歌唱者 c	元音源より、より張りのある声質になっていたと感じた 副旋律の後半もよく聞こえていた	なし
歌唱者 d	元音源では副旋律が強く聞こえていたが、 変換後滑らかな声質でやや強さが抑えられ、 主旋律が聞きやすくなったと感じた	なし

表 A-6 声楽家評価者2のアンケート評価結果：副旋律声部条件
(フィルタによる変化 (その他・備考))

	フィルタによる変化	(その他・備考)
歌唱者 a	元音源に比べ音量のバランスが良く聴こえ、 まとまった印象を受ける	参照音源より他の音源の方が 歌詞を丁寧に歌っていて 主旋律とフレーズも揃い、重唱として好ましい
歌唱者 b	副旋律が変換後の方がまろやかに聴こえる 重唱として主旋律を跳び出さずに支えバランス良く聴こえた	参照音源は主旋律と声の質や フレーズの終わりの長さが揃っていて重唱として好ましい 変換元音源は歌詞の縦の線が揃わない所が若干ある 声質も（周辺歌唱に合わせるからか）統一感の面で参照音源に劣る
歌唱者 c	元音源に比べて副旋律の声質が柔らかく聴こえた 結果、鋭さが無くなりバランスが良くなったと感じる	参照音源は他の音源に比べ歌詞の縦の線や フレーズの終わりが揃っていて重唱として好ましい
歌唱者 d	元音源に比べ声の質や音量がやや円やかに聴こえて 重唱の副旋律として良いと思う	持ち声が良いので、重唱を歌う場合に子音の立て方を減らした方が 周辺歌唱と揃いやすいのでは