Vol.2020-MUS-127 No.35
Vol.2020-SLP-132 No.35
2020/6/7

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

山内 孔貴[1,a)]　須田 仁志[2,b)]　齋藤 大輔[2,c)]　峯松 信明[2,d)]

概要：本稿では，ソースフィルタ分解に基づいて複数歌唱者の歌唱音声を分析することで，合唱において複数歌唱者の調和を定量的に扱う指標を検討する．二人での重唱における調和の傾向を実験的に検証することで，調和に関する定量的な指標の妥当性を確認する．加えて，その指標をもとに合唱をより調和するよう制御する方法を検討する．

1. はじめに

近年，音声合成技術の需要は増加している．通常の発話にとどまらず，歌声を合成するシステムも盛んに研究・開発されており，VOCALOID [1] をはじめ，CeVIO[*1]，UTAU[*2]などが一般向けに公開され，創作の手段として広く利用されている．

音声を様々に操作および合成する際，その音声が扱いやすいようにモデル化されていることが望ましい．音声をその生成過程モデルに基づいて音源とフィルタの畳み込みで表現したモデルがソースフィルタモデルである．ソースフィルタモデルに基づいてパラメータを操作することにより，音声を効果的に制御することができる例として，楽譜から歌声音声を合成するシステムである Sinsy [2] がある．Sinsy では，楽譜の情報と音声波形の間の関係としてソースフィルタモデルに基づいたパラメータ列を隠れマルコフモデル（HMM）でモデル化して学習し，合成時には与えられた楽譜と HMM から合成に必要なパラメータを推定し，歌声を合成する．また，VOCALOID はスペクトル包絡に相当するパラメータを調整することで歌声合成の声色を変化させることができるが，一般にこれを曲に合わせて調整することは難しい．VocaListener [3] は，既存の歌声合成ソフトウェアで歌声を合成する際に，パラメータ調整が困難であることへの解決法として，合成される歌声が入力として受け取った歌声に近くなるように音高や音量のパラメータを調整することができる．VocaListener2 [4] では，それに加えて声色もまねるようにパラメータの調整ができる．

上述のように，ソースフィルタモデルで発声をモデル化することで，合成音声をパラメータで効果的に制御することができる．しかし，従来の研究では単一の歌唱者による歌声は研究されてきたものの，複数の歌唱者による複数音源の歌声については定量的な研究があまりなされてこなかった．一方，ソースフィルタモデルに基づいたボコーダによる音声合成では人間が出しえない音源と生成モデルを組み合わせることができるため，複数の音高を有する音源と組み合わせることで，複数の歌唱者による歌声を効果的にモデル化することが可能であると期待できる．また，合唱においては，合唱の調和に関して歌唱者間の「最適な距離」が存在すれば，その声に近づくように歌うことで合唱としての調和がとれるようになることが期待できる．さらに，個々の歌声を評価することにより合唱として調和がとれる相手がどんな声であるのかの判断をするための指標を得ることもできると考えられる．そこで本研究では，合唱における歌声の調和に対する定量的な扱いについて検討する．

2. 合唱音声のソースフィルタ分解

本研究では，ソースフィルタモデルに基づくボコーダを利用して，人間では出しえない，複数の高さの音声を同時に発するような音源による音声合成の手法を検討する．ホーミーと呼ばれるモンゴルの伝統的な歌唱法 [5] では，高さが異なる 2 音が一人の歌手から同時に知覚されるが，通常，人一人がある時刻に発声できる音の高さは一種類の

[1] 東京大学 大学院情報理工学系研究科電子情報学専攻，
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
[2] 東京大学 大学院工学系研究科電気系工学専攻，
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
[a)] kyamauchi@nae-lab.org
[b)] hitoshi@gavo.t.u-tokyo.ac.jp
[c)] dsk_saito@gavo.t.u-tokyo.ac.jp
[d)] mine@gavo.t.u-tokyo.ac.jp
[*1] http://cevio.jp/.
[*2] http://utau-synth.com/.

みであり，複数の高さの音が一人から同時に聞こえてくることはない．

2.1 フィルタ成分が共通の場合

一人の人が二つの異なる高さで同じ音素の有声音を発声したとする．このとき，同じ人が同じ音素を発声しているためフィルタ成分は共通であると仮定してどちらも $h(t)$ とすると，二つの音声信号をそれぞれ $y_1(t)$，$y_2(t)$ とすれば，次のように表すことができる．

$$y_1(t) = g_1(t) * h(t) \tag{1}$$

$$y_2(t) = g_2(t) * h(t) \tag{2}$$

ただし，$g_1(t)$，$g_2(t)$ はそれぞれ基本周波数の異なるインパルス列，$*$ は畳み込みである．これらの 2 音が同時に聞こえるとき，その信号は

$$\begin{aligned} y_1(t) + y_2(t) &= g_1(t) * h(t) + g_2(t) * h(t) \\ &= (g_1(t) + g_2(t)) * h(t) \end{aligned} \tag{3}$$

と表される．したがって，音源を $g_1(t) + g_2(t)$ とすれば，同時に 2 つの高さの音声を合成することのできるボコーダを設計することができ，その信号も，2 つの音声を同時に鳴らしたものと全く同一になる．

2.2 フィルタ成分が異なる場合

2 音のフィルタ成分が異なる場合を考える．すなわち，

$$y_1(t) = g_1(t) * h_1(t) \tag{4}$$

$$y_2(t) = g_2(t) * h_2(t) \tag{5}$$

という状況を仮定する．これら 2 音が同時に聞こえるとき，その信号は

$$y_1(t) + y_2(t) = g_1(t) * h_1(t) + g_2(t) * h_2(t) \tag{6}$$

である．もっとも簡単な近似として，音源を $g_1(t) + g_2(t)$，フィルタを $h_1(t) + h_2(t)$ とするような方法で合成すると，合成音声は

$$\begin{aligned} &(g_1(t) + g_2(t)) * (h_1(t) + h_2(t)) \\ =\ & g_1(t) * h_1(t) + g_1(t) * h_2(t) \\ & + g_2(t) * h_1(t) + g_2(t) * h_2(t) \\ =\ & y_1(t) + y_2(t) + g_1(t) * h_2(t) + g_2(t) * h_1(t) \end{aligned} \tag{7}$$

となるため，単純に 2 音が聞こえるのとは異なる音声となってしまう．

そこで，式 (6) を周波数領域で考える．スペクトルは次式のように表される．

$$\begin{aligned} Y_1(\omega) + Y_2(\omega) &= G_1(\omega)H_1(\omega) + G_2(\omega)H_2(\omega) \\ &= G(\omega)H(\omega) \end{aligned} \tag{8}$$

ただし，

$$G(\omega) = G_1(\omega)\frac{H_1(\omega)}{H(\omega)} + G_2(\omega)\frac{H_2(\omega)}{H(\omega)} \tag{9}$$

$$H(\omega) = \sqrt{H_1(\omega)H_2(\omega)} \tag{10}$$

とする．これは，励起信号のスペクトルが $G(\omega)$，スペクトル包絡が $H(\omega)$ の信号とみなすことができる．このようにとらえることで，第三者の声道特性を持つ歌唱者の音声として，複数の音声を表現できると考えられる．ここで，$H(\omega)$ について考える．

$$\begin{aligned} &\mathcal{F}^{-1}\left[\log H(\omega)\right] \\ =\ & \mathcal{F}^{-1}\left[\frac{\log|H_1(\omega)| + \log|H_2(\omega)|}{2}\right] \\ =\ & \frac{\mathcal{F}^{-1}\left[\log|H_1(\omega)|\right] + \mathcal{F}^{-1}[\log|H_2(\omega)|]}{2} \end{aligned} \tag{11}$$

であり，$H(\omega)$ はメルケプストラム領域ではメルケプストラムの平均として現れる．同様に，$\frac{H_1(\omega)}{H(\omega)}$ や $\frac{H_2(\omega)}{H(\omega)}$ もメルケプストラム領域ではメルケプストラムの加減として現れる．

2.3 合唱を表すパラメータ

複数歌唱者の重唱音声 $Y_1(\omega) + Y_2(\omega)$ を単一歌唱者の場合に倣ってソースフィルタ分解すると，前項で述べたように $G(\omega)$ をソース成分，$H(\omega)$ をフィルタ成分とみなした歌声ととらえることもできる．このとき，重唱の立場に戻り，それぞれの成分が重唱のどのような特徴を表しうるのかを考えると，ソース成分は重唱の旋律についての情報を表している．ただし，複数歌唱者間の音色の違いが各音高に対して重み付けされている．また，フィルタ成分は各歌唱者のフィルタ成分の平均であり，重唱自体の音色を表していると考えられる．

したがって，歌唱者それぞれについて，その歌唱者の音色としての個性はフィルタ成分である $H_1(\omega)$ や $H_2(\omega)$ に現れるのに対して，合唱音声の場合は合唱団としての個性は $H(\omega)$ に現れていると考えられる．ここで，合唱音声のソース成分に現れる重み成分は，各個人の音色であるフィルタ成分と合唱団としての音色に相当する平均フィルタとの距離である．合唱音声の中での各個人の個性はこの重み成分に相当していると考えられる．

3. 合唱音声の調和制御と評価

従来の研究においては，単一歌唱に厚みを持たせるためのダブルトラッキング音声の合成 [6] や，合唱というジャンルにおいて適した声質にまつわる特徴量の考察 [7] は行われてきているものの，合唱音声の調和については定量的に扱われてこなかった．そこで，本研究では，どのような組み合わせでの合唱が調和しているのかという観点から，合唱音声を定量的に取り扱い，歌唱者間のフィルタ成分の

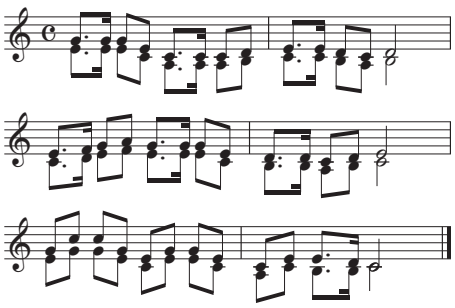

図 1 実験で使用した，童謡「かたつむり」の楽譜
Fig. 1 Score of "Katatsumuri"

距離と調和の関係を明らかにすることを目的とした実験を段階的に行う．

以下では，$H(\omega)$ を固定した状態で $\frac{H_1(\omega)}{H(\omega)}$ と $\frac{H_2(\omega)}{H(\omega)}$ を変化させたときに重唱の調和性がどのように変化するのか，実験的に検討する．

ここで，以降の実験に先んじて，共通の実験条件について記す．

実験ではデータセットとして JVS-MuSiC [8] を用いた．JVS-MuSiC には 100 人の歌唱データが収録されており，全員が歌っている共通曲として日本語童謡「かたつむり」がある．本実験ではこのデータを使用した．

しかし，JVS-MuSiC には主旋律となる歌唱データは収録されているが，副旋律のデータは収録されていない．そこで，本実験では，JVS-MuSiC に収録されている歌唱データの F0 を「かたつむり」の楽譜に合わせて変換した合成音声を用意し，それを使用した．下声部の作曲は，合唱伴奏の経験が豊富なプロのピアニストに依頼した．図 1 にその楽譜を示す．

JVS-MuSiC には，男性 49 人，女性 51 人の歌唱データが収録されており，それぞれが自由なキーとテンポで歌ったものと，それらを最も近いキーとテンポに揃えたもの，そしてそれらのテンポを BPM100 に揃え，男女それぞれ 3 種類のキーに揃えてグループ分けしたものがある．本実験では，これらのうち BPM が 100 に揃えられ，キーも男女それぞれ 3 種類に揃えられたデータのうち，女性の音声を使用した．実験では，主旋律の歌唱音声として，上記の歌唱音声を分析して得られた F0 を一曲全体にわたりスケーリングすることで楽譜の主旋律と同じキーに合わせ，再合成することで得られた音声を使用した．しかし，副旋律の F0 は主旋律の F0 を一律でスケーリングすることでは得られない．そこで，以下のような流れで合成した音声を使用した．

まず，JVS-MuSiC のデータはサンプリング周波数が 24kHz であり，これを 16kHz にダウンサンプリングした．次に，Julius [9] [10] を用いて強制アライメントをとり，さらに筆者が手動でそれを調整し，時系列音素ラベルを得た．次に，楽譜と得られたラベルデータをもとに，ステップ関数の重ね合わせとして F0 系列を生成した．しかし，この F0 系列をそのまま用いて再合成すると，人間の歌唱としての自然性が低い合成音声となってしまう．そこで，これを人間の歌唱の特徴を反映させた F0 系列に変換する必要がある．人間の歌声には，歌唱者や歌唱スタイルに依存しない F0 動的変動成分があることが知られている [11]．本研究では，先行研究に則った方法でステップ状の F0 系列に動的成分を付与し，人間の歌唱の特徴を反映した F0 系列を得たのちそれを用いて副旋律を合成し，実験で使用した．

また，実験の際に重唱のペアとした歌唱者は，メルケプストラム歪み（Mel-cepstral Distortion; MCD）を基準に選定した．本実験では合唱団の音色が変わらないように歌唱者のペアを決めるため，データの母集団である 51 人全体の平均のメルケプストラムとペアに選んだ 2 歌唱者間の平均のメルケプストラムの MCD が大きく変化しないような歌唱者のペアとして，図 2 の中で 51 人のメルケプストラムの平均と 2 人ペアのメルケプストラムの平均との MCD が 3.91dB から 4.02dB の間で選定し，かつペア内の MCD ができる限り大きいペアから小さいペアまで含まれるように 10 ペア選定した．図 2 は，JVS-MuSiC でグループに分けられているグループ内でのみペアを組むという制約のもとで組んだ全ペアにおける，51 人のメルケプストラムの平均と 2 人ペアのメルケプストラムの平均との MCD と，2 人ペア内の MCD との関係である．また，表 1 は実験を行ったペアのペア内での MCD である．

上記の他の実験条件として，音声の分析合成には WORLD [12](D4C edition [13]) を用いた．音声は「かたつむり」の 1 番のみ使用し，音声の長さは 15 秒である．窓長は 512 サンプル，シフト長は 5ms，メルケプストラムの次数は 24 次とした．聴取実験は Web 上のクラウドソーシングサービスにて AB テストを行った．被験者数は各対の評価に対して 25 人とし，一種類の比較についてどちらを先に聞くかの順序を考慮して各人 2 回評価した．

3.1 実験 1：パート分けに関する評価

3.1.1 実験条件

本実験では，異なる 2 人の歌唱者が主旋律と副旋律をそれぞれ入れ替えて歌った重唱音声に対して，どちらがより調和するのかについての聴取実験による主観評価を行った．

聴取実験では，ペア内で主旋律と副旋律をそれぞれ入れ替えて歌った 2 つの重唱音声に対して，主観評価でより自然だと感じた音声を選択する AB テストを行った．

3.1.2 実験結果

図 3 に主観評価の結果を示す．ここで，ペア内での主旋律と副旋律の入れ替えのうち，51 人全体のメルケプストラムの平均との MCD が大きいほうの歌唱者が主旋律を歌っている重唱音声の方が自然だと評価された場合に 0，その逆が自然だと評価された場合に 1 とした．この主観評価

Vol.2020-MUS-127 No.35
Vol.2020-SLP-132 No.35
2020/6/7

表 1　選定したペアの ID と，ペアを組んでいる歌唱者の JVS-MuSiC での番号，そのペア内の MCD [dB]

Table 1　IDs of selected pairs, the singers' number in JVS-MuSiC, and the MCD between singers of the pair.

A	B	C	D	E	F	G	H	I	J
065,095	084,090	002,007	085,093	007,053	036,095	004,030	008,035	035,092	010,030
5.99	6.51	6.96	7.16	7.40	7.71	7.95	8.16	8.50	10.02

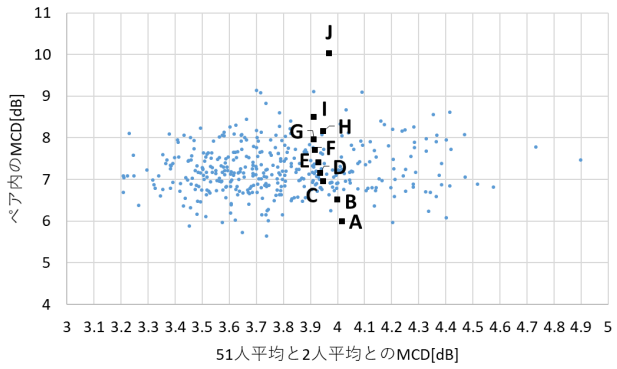

図 2　51 人のメルケプストラムの平均と 2 人ペアのメルケプストラムの平均との MCD と 2 人ペア内の MCD の分布．実験で用いたペアには A から J の ID のラベルが振ってある．

Fig. 2　Distribution of pairs of singers. Abscissa denotes MCD between the average of 51 singers and the one of a pair. Ordinate denotes MCD between singers in a pair. The IDs from A to J are assigned to the pairs which are evaluated in this experiment.

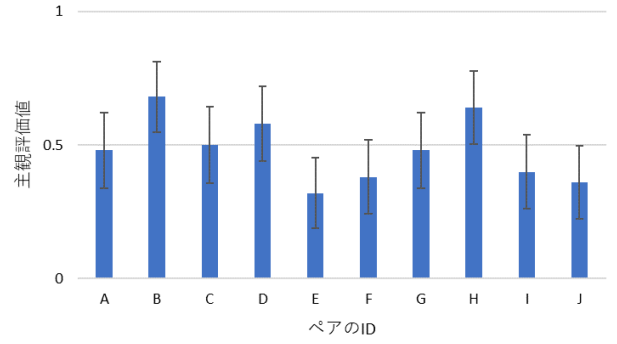

図 3　実験 1 の各ペアへの主観評価結果 (エラーバーは 95%信頼区間)．ペアの ID は MCD が小さい順にソートされている．

Fig. 3　Subjective results; the IDs are sorted by MCD.

結果に対して t 検定により 95%信頼区間を求めたところ，パートを入れ替えたときに片方の分け方がもう片方の分け方よりも有意に自然だと評価されているペアは 4 ペアで，それらのペア内 MCD，すなわち歌唱者間のフィルタ成分の距離には依存していなかった．

3.2　実験 2：ペアになる歌唱者と合唱の調和に関する評価

3.2.1　実験条件

本実験では，重唱のペアを組む歌唱者を変えたときに重唱の調和にどのような影響があらわれるのかについて検討した．重唱のペアは実験 1 と同一のペアに加え，ペアとなる二人が全く同じ歌唱者となりペア内 MCD が 0dB となるサンプルとして，51 人の平均メルケプストラムとの MCD が実験 1 のサンプルの 3.91dB から 4.02dB に最も近い 4.64dB である jvs016 のダブルトラッキング音声（ID を K とする）を選定し，計 11 ペアとした．ペア内でのパート分けは実験 1 にてより自然性が高いと評価された回数が多かった分け方とした．ただし jvs002 と jvs007 のペアであるペア C については，どちらの歌唱者を主旋律としても自然性の評価に差がなかったため，実験 1 全体で 51 人の平均メルケプストラムとの MCD がより大きい歌唱者を主旋律とした場合の方が評価値が高い傾向があることから，51 人の平均メルケプストラムとの MCD がより大きい歌唱者である jvs002 が主旋律を歌っている場合でのパート分けとした．11 ペアの重唱音声に対し，より自然だと感じた音声を選択する AB テストで聴取実験による総当たりの主観評価を行った．他の実験条件は前項に記したとおりである．

3.2.2　実験結果

聴取実験により得られた知覚スコアをもとに，サーストンの一対比較法によって間隔尺度を計算した．図 4 にペア内の MCD とそのペアの重唱音声の間隔尺度を示す．ダブルトラッキング音声が他の重唱音声に比して低いスコアを示した．これは他の重唱音声は二人の間で発声のタイミングが完全に同期してはいないのに対して，本実験のダブルトラッキング音声はタイミングが完全に同期しているため，重唱として聞いたときに違和感を与えていた可能性がある．図 4 についてペア内の MCD を横軸にとった図を図 5 に示す．ダブルトラッキング音声以外の重唱音声に関して，ペア内の MCD が 6.5 から 7.5 付近で最も重唱が調和している一方，その区間においても他に比べて調和していないと評価された重唱音声もあった．しかし，MCD が 7.5 を超えると重唱は調和がとれなくなっていく傾向が見られた．ダブルトラッキング音声を除いた重唱音声についてペア内の MCD と間隔尺度の相関係数は-0.69，p 値は 0.026 であり，この MCD の区間において相関が見られた．

そして，実験 1 のパート分けに関する評価での 0.5 からの距離，すなわちパートを入れ替えることでどれだけ重唱の自然性に差が出るかの指標との関係を図 6 に示す．実験 1 では信頼区間が 0.14 ほどの幅を持っていたため関連性を

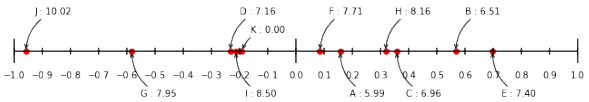

図 4　実験 2 の AB テストでのスコアから求めた間隔尺度（アノテーションはペア名と MCD）

Fig. 4　Distance measure calculated by the subjective results of experiment 2.

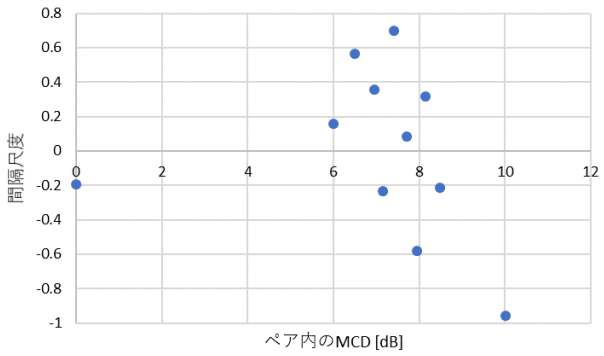

図 5　ペア内の MCD と一対比較の間隔尺度の関係

Fig. 5　The relation between MCD in the pair and distance measure.

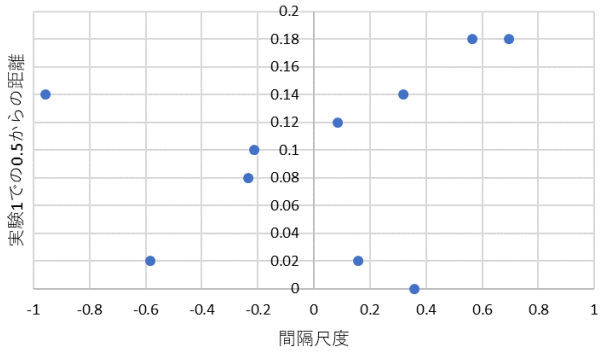

図 6　一対比較の間隔尺度と実験 1 のパート分けに関する評価での 0.5 からの距離の関係

Fig. 6　The relation between distance measure and distance from 0.5 in the result of experiment1.

議論するにはデータが少ないと思われるが，0.5 からの距離が大きくなるほど適切なパート分けで歌えばその重唱の調和はとれているという傾向が見られた．この結果から，重唱において調和がとれているためにはどちらが主旋律を歌うかについても考慮する必要があり，さらにそれぞれのパートとしての理想的な声があり，歌唱者がその声に近くなるほど重唱の調和がとれると考えられる．

3.3　実験 3：歌唱者のフィルタ成分の制御による合唱の調和制御に関する検討

本実験では，歌唱者のフィルタ成分を制御することで，重唱としての調和がどのように変化するかについて検討する．重唱音声のスペクトルは式 (8) のようにあらわされ，各個人の個性は $\frac{H_1(\omega)}{H(\omega)}$ や $\frac{H_2(\omega)}{H(\omega)}$ に関係していると考えられる．そのため，重唱が調和しているかどうかはこれらの項の関係で記述できると推測される．今，重唱音声として

$$Y_U(\omega, u) = G_U(\omega, u)H(\omega) \tag{12}$$

ただし，

$$\begin{aligned} &G_U(\omega, u) \\ &= G_1(\omega)\left(\frac{H_1(\omega)}{H(\omega)}\right)^u + G_2(\omega)\left(\frac{H_2(\omega)}{H(\omega)}\right)^u \end{aligned} \tag{13}$$

を考え，u を制御することで重唱音声 Y を変化させる．このとき，$u = 0$ で両方の歌唱者のフィルタ成分が $H(\omega)$ のダブルトラッキング音声に，$u = 1$ で元の重唱音声になる．これをメルケプストラム領域であらわすと，歌唱者 1 と 2 のメルケプストラムはそれぞれ $\frac{1+u}{2}mc^{(1)} + \frac{1-u}{2}mc^{(2)}$，$\frac{1+u}{2}mc^{(2)} + \frac{1-u}{2}mc^{(1)}$ となり，u を 0 から大きくしていくことで歌唱者のメルケプストラムが歌唱者間の平均からそれぞれの元のメルケプストラムの方向へと変化し，重唱音声を変化させることができる．

3.3.1　実験条件

前項の式 (12) において u の値を $u = 0, 0.25, 0.5, 0.75, 1, 1.25$ と変化させたときに，そのペアの調和がどのように変化するのかについて聴取実験による主観評価を行った．重唱のペアとして E，J，A の 3 ペアを選定し，パート分けは実験 1 にてより自然だと評価されたパート分けとした．なお，これらのペアは，ペア間の調和に関する間隔尺度を基準として，E が最も調和しているペア，J が最も調和していないペア，A が中程度調和しているペアで，かつペア内 MCD は E が中程度，J は最も大きく，A が最も小さいペアとなっている．各ペアについて，同一のペア内での総当たりの AB テストを行った．他の実験条件は前項に記したとおりである．

3.3.2　実験結果

聴取実験で得られた知覚スコアをもとに求めた，サーストンの一対比較法による間隔尺度と u の値との関係を図 7 に示す．まず，$u = 0$ についてはスコアが低く，これはペア間の比較においてダブルトラッキング音声のスコアが低かったことと共通している．ただし，こちらは別の歌唱者同士のペアであるため，実験 2 のペア間の比較で用いたサンプルと異なり，歌声のタイミングにわずかな違いがある．そのような条件でも同様に知覚スコアが低かったため，知覚スコアの低下は極端なタイミングの同期によるものだけではなく，タイミングが違っていてもそもそも 2 歌唱者間の特徴が近接しすぎている場合に発生することが示唆された．また，$u = 1.25$ についてはどのペアでも他の u の値と比べて知覚スコアが低かった．これは MCD が大きくなってしまったことに加え，歌唱者同士のメルケプストラムの

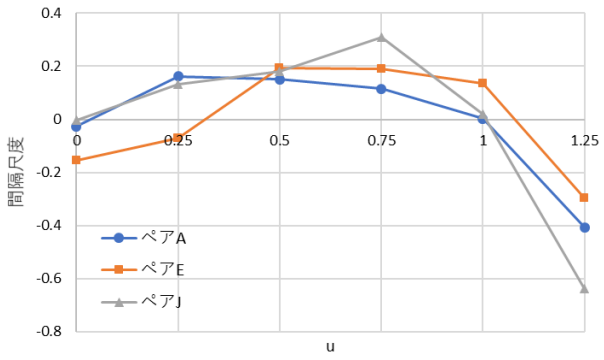

図 7　u と一対比較の間隔尺度の関係
Fig. 7　The relation between u and distance measure.

内分である $|u| < 1$ と異なり $u = 1.25$ では外分となるため，そもそも人の声として不自然であると感じられた可能性がある．$u = 1$ については，元の重唱音声であるが，実験 2 のペア間の比較において最もスコアの高かったペアの重唱音声も含めて，どのペアも u を小さくすることで知覚上の自然性が高くなった．したがって，少なくともこの実験で選んだ歌唱ペアでは，歌唱者間の MCD にかかわらず，歌唱者本人それぞれの歌唱より互いに声が似るようにして歌うことで，より調和のとれた重唱となることがわかる．

4. まとめ

本研究では，複数歌唱者による合唱音声をソースフィルタ分解することで，歌唱者の調和について定量的に扱う手法を提案した．具体的には，合唱団の個性をあらわすパラメータと合唱中での個々人の個性をあらわすパラメータを各歌唱者の音声を分析して得られる特徴量によって記述することで，それらを制御することにより調和を制御することを試みた．

その結果，重唱においてペア内の MCD とペアの調和の間に相関がみられた．また，重唱においてペアがより調和する方法を定量的に検討し，歌唱者のメルケプストラムをペア内の MCD が 0.75 から 0.5 倍になるように双方同じだけ近づければより調和する傾向があることが示唆された．

今後の展望として，主旋律をどちらが歌うかによって歌唱者の調和が有意に変化するペアの組み方が存在したことを考慮して，合唱のモデルを調整することが考えられる．本研究では式 (8) のように合唱団の個性がフィルタの平均であらわされるモデルのみを評価の対象としたが，主旋律側と副旋律側で重みに偏りを持たせた，

$$Y_1(\omega) + Y_2(\omega) = G_A(\omega) H_A(\omega) \tag{14}$$

ただし，

$$G_A(\omega) = G_1(\omega) \frac{H_1(\omega)}{H_A(\omega)} + G_2(\omega) \frac{H_2(\omega)}{H_A(\omega)} \tag{15}$$

$$H_A(\omega) = H_1(\omega)^{1-\alpha} H_2(\omega)^{\alpha} \tag{16}$$

のようなモデルも考えられ，そのモデルによる合唱音声での評価に関しても検討の余地がある．

今後の展望として，合唱における調和を的確に定量化できれば，そのパラメータで学習した歌声合成モデルを構築することで，歌声を入力としてよく調和するような副旋律や主旋律を合成するシステムを作ることが期待できる．

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] 大浦圭一郎，間瀬絢美，山田知彦，徳田恵一，後藤真孝：Sinsy:「あの人に歌ってほしい」をかなえる HMM 歌声合成システム，情報処理学会研究報告. [音楽情報科学], Vol. 86, No. 1, pp. 1–8 (2010).
[3] T. Nakano and M. Goto: VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation, *Proc. SMC*, pp. 343–348 (2009).
[4] T. Nakano and M. Goto: Vocalistener2: A singing synthesis system able to mimic a user's singing in terms of voice timbre changes as well as pitch and dynamics, *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 453–456 (2011).
[5] 村岡輝雄，武田昌一，糸賀昌士：モンゴル歌唱法「ホーミー」の音響的特徴の解析，日本音響学会誌，Vol. 56, No. 5, pp. 308–317 (2000).
[6] H. Tamaru, Y. Saito, S. Takamichi, T. Koriyama and H. Saruwatari: Generative moment matching network-based random modulation post-filter for DNN-based singing voice synthesis and neural double-tracking, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7070–7074 (2019).
[7] 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).
[8] H. Tamaru, S. Takamichi, N. Tanji and H. Saruwatari: JVS-MuSiC: Japanese multispeaker singing-voice corpus, *arXiv:2001.07044 [cs.SD]* (2020).
[9] A. Lee, T. Kawahara and K. Shikano: Julius - An Open Source Real-Time Large Vocabulary Recognition Engine, *Proceedings of European Conference on Speech Communication and Technology*, Vol. 3, pp. 1691–1694 (2001).
[10] A. Lee and T. Kawahara: Recent Development of Open-Source Speech Recognition Engine Julius, *em Proceedings of the 2009 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (2009).
[11] T. Saitou, M. Unoki and M. Akagi: Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis, *Speech Communication*, Vol. 46, pp. 405–417 (2005).
[12] M. Morise, F. Yokomori and K. Ozawa: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, *IEICE transactions on information and systems*, Vol. E99-D, No. 7, pp. 1877–1884 (2016).
[13] M. Morise: D4C, a band-aperiodicity estimator for high-quality speech synthesis, *Speech Communication*, Vol. 84, pp. 57–65 (2016).