Vol.2015-MUS-107 No.62
2015/5/24

アカペラ演奏支援のための歌声に対する楽譜追跡手法の検討

森 大毅[1,a)]　上田 新[1]

概要：本研究では，アカペラ (ここでは，無伴奏で行う重唱を指す) の練習を行う際にグループに欠員が出た場合を想定し，その欠員の担当パートを他のメンバーの演奏と同調しながら演奏するシステムの実現を目指す．アカペラ演奏支援には，音高が不安定という問題と，歌詞の影響を受けてスペクトル特徴が変化するという問題があり，従来の楽器演奏を対象とした楽譜追跡手法をそのまま適用しても十分な性能が得られない．本報告では，アカペラ演奏に対する楽譜追跡の精度向上を目的とし，特徴量，チャンネルの増加および最適パス探索手法に関する改善手法を提案する．まず，従来の楽譜追跡手法で用いられている特徴量が母音の違いにより異なる様相を示す問題に対処するため，特徴量として基本周波数を併用した．次に，アカペラ演奏では同時に複数パートの歌声が利用できる利点を生かし，複数チャンネルを入力とした楽譜追跡手法を検討した．さらに，早すぎる状態遷移に対処するため，パスの傾斜制限を導入した．本報告では，3 曲のアカペラ演奏に対し楽譜追跡実験を行い，これらの手法の有効性について検討した結果について述べる．

1. はじめに

自動伴奏とは，与えられた総譜の下で，人間の演奏する旋律に対し自動的に同期して伴奏を行うことを指す．人間の演奏への同期はこの自動伴奏問題の部分問題であり，楽譜追跡と呼ばれる．楽譜追跡の問題は状態推定問題の一種であり，HMM(隠れマルコフモデル) やカルマンフィルタのような統計的手法と親和性が高い．Eurydice [1] は，MIDI 出力可能な楽器に対し HMM に基づく自動伴奏を行うシステムであり，弾き飛ばしや弾き直し，弾き間違い等への対応を実現している．[2], [3] は，これをクラリネット演奏の音響入力に拡張した研究である．また，Ryry [4] ではセミマルコフ条件付確率場とカルマンフィルタにより多声楽器演奏の音響入力から楽譜追跡を行っている．

本研究では，アカペラ (ここでは，無伴奏で行う重唱を指す) の練習を行う際にグループに欠員が出た場合を想定し，その欠員の担当パートを他のメンバーの演奏と同調しながら演奏するシステムの実現を目指す．大学生等のアマチュアによるアカペラサークルは非常に人気が高く，宇都宮大学においても数多くのグループが活動している．しかしながら，就職活動等のため，メンバー全員が練習に参加できる時間を見つけることは案外難しい．人数の多い合唱とは異なり，アカペラの場合，欠員はパート全体の欠落となるため影響が大きい．以上のことから，アカペラグループの練習時における欠員の補完は，多くの演奏者が必要としているものの一つであると言える．

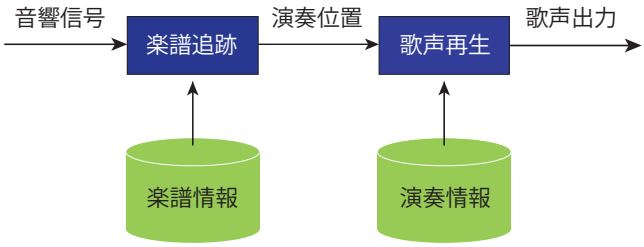

図 1　アカペラ演奏支援システムの概要

本研究が実現を目指す，自動で複数の人間の歌唱に同調しながら欠員パートを演奏することのできるシステムの概要を**図 1** に示す．システムにはあらかじめ楽譜上の音符の長さや音高，テンポなどの「楽譜情報」と，事前に収録した欠員パートの演奏データや，そこから抽出した音響特徴量などの「演奏情報」の二つを与えておく．システムは図 1 に示すように，大きく楽譜追跡部と歌声再生部の二つからなる．楽譜追跡部ではマイクロフォンから入力された演奏者の歌声 (音響信号) と楽譜情報から瞬間ごとの楽譜上の演奏位置を推定し，歌声再生部では楽譜追跡部において得られた演奏位置に対応する歌声を演奏情報から生成し，リアルタイムで再生する．

アカペラ演奏支援は，自動伴奏問題の変種と考えられ，その部分問題である楽譜追跡は自動伴奏におけるそれと同じ問題であると位置付けられる．しかしながら，アカペラ

[1] 宇都宮大学大学院工学研究科
Graduate School of Engineering, Utsunomiya Univeristy, 7-1-2, Yoto, Utsunomiya, 321-8585 Japan
[a)] hiroki@speech-lab.org

表 1 収録曲

呼称	テンポ [bpm]	調性	曲調
楽曲 A	108	ト長調	ポップス
楽曲 B	116	イ長調	ポップス
楽曲 C	80	ハ長調	ポップス

には楽器演奏の自動伴奏にはない独特の問題がある．

- 音高が不安定．アカペラは基準音を演奏する楽器を伴わないため，絶対音感を持たない演奏者では演奏のたびごとに音高が少しずつ異なってしまう．さらに，1曲の演奏中でさえも音高が徐々に変化する (ドリフト) 現象が見られることがある．
- 歌詞が存在するために，同じ音高の音符であってもスペクトル特徴が母音によって異なる．

これらの問題のために，楽器演奏を対象にした従来の楽譜追跡手法をそのままアカペラ演奏に対して適用しても十分な性能が得られないことが予想される．

本研究の目的は，アカペラ演奏に対する楽譜追跡の精度向上である．この目的を達成するため，以下の改善手法を検討した．まず，従来の楽譜追跡手法で用いられている特徴量が母音の違いにより異なる様相を示す問題に対処するため，特徴量として基本周波数を併用した．次に，アカペラ演奏では同時に複数パートの歌声が利用できる利点を生かし，複数チャンネルを入力とした楽譜追跡手法を検討した．さらに，早すぎる状態遷移に対処するため，パスの傾斜制限を導入した．本報告では，3 曲のアカペラ演奏に対し楽譜追跡実験を行い，これらの手法の有効性について検討した結果について述べる．

2. 評価データ

楽譜追跡実験システムの構築および評価のために，アカペラ演奏を収録した．演奏者は宇都宮大学のアカペラサークル「U-MiC」に所属する男子大学生 6 人組であり，リードボーカル 1 名，コーラス 3 名，ベース 1 名，ボイスパーカッション 1 名からなる．収録楽曲は日本のプロアカペラグループ「RAG FAIR」の楽曲であり，全 3 曲である．**表 1** に各楽曲の情報を示す．収録は防音室にて行い，6 人による同時演奏を手持ちマイクロフォン (Shure SM58) を用いて 6 チャンネル同時にサンプリング周波数 44100 Hz で PCM 録音した．

以降，収録したチャンネルのうち，3 曲それぞれのコーラスパート 3 名分 (計 9 曲分に相当) のみを用いる．なお，収録した演奏の中に，弾き飛ばしや弾き直し，ジャンプに相当するものは確認されていない．

3. 楽譜追跡アルゴリズム

本研究の楽譜追跡は，隠れマルコフモデル (HMM) を用いた方法 [2] に基づいている．この手法では，楽譜上の一つの音符 (休符) を HMM の一つの状態で表現し，演奏の進行を状態間の遷移で表現する．本稿では問題を簡単にするため，自己遷移と 1 つ次の状態への遷移のみが存在するものとしてモデル化を行った．

特徴抽出は 20 ms ごとに行った．演奏位置は実時間での動作を考慮し，分析フレームごとに Viterbi アルゴリズムによって最適パスを算出し，その時点での尤度最大の状態を演奏位置と見なしてその都度求める．遷移確率は音符の長さに関わらず，自己遷移確率を 0.96，一つ次の状態への遷移確率を 0.04 として与えた．

学習データには男性 3 名によるアカペラ演奏曲 3 曲 9 パート中のコーラス部分のみを用い，楽譜追跡の対象は，学習データとして用いた楽曲のうち 2 曲から，歌詞を歌う部分を含まないように一部分を切り出したものを用いた．

4. 音響特徴量の検討

[2], [3], [4] では，音響特徴量としてクロマベクトル [5] を用いている．クロマベクトルは 12 次元のベクトルであり，その各要素は「C」「C#」「D」などの音階名で表される半音刻みの音高に対応する周波数のパワースペクトルを，オクターブをまたいで全て足し合わせた値である (式 (2))．N 点の離散入力信号 $x(n)$ $(n = 0, 1, \cdots, N-1)$ の DFT スペクトル $X(k)$ $(k = 0, 1, \cdots, N-1)$

$$X(k) = \sum_{n=0}^{N-1} e^{-j\frac{2\pi kn}{N}} \cdot x(n) \tag{1}$$

のパワーを音階ごとに足し合わせたものがクロマベクトルの各要素 $C(p)$ $(p = 0, 1, \cdots, 11)$ である．

$$C(p) = \sum_{l \text{ s.t. } M(l)=p} |X(l)|^2 \tag{2}$$

$$M(l) = \begin{cases} -1 & (l = 0) \\ \mathrm{round}\left(12\log_2\left(\left(f_\mathrm{s}\cdot\frac{l}{N}\right)\Big/ f_\mathrm{ref}\right)\right) \bmod 12 & \\ & (l = 1, 2, \cdots, N/2-1) \end{cases} \tag{3}$$

f_s はサンプリング周波数を表し，$f_\mathrm{s}\cdot\frac{l}{N}$ は l 点目のスペクトルに対応する周波数の値を表す．音階 p は 0 から 11 までの 12 個の値を持ち，f_ref は $p = 0$ とする基準の音階の周波数 (例えば，A4 なら 440 Hz) を表す．

歌声から抽出したクロマベクトルを**図 2**，**図 4** に示す．縦軸は音階 $(C = 1, C^\# = 2, \cdots, B = 12)$，横軸は時間 (サンプル点) を表す．図 2 は母音/u/の地声で C4 から C3 まで半音ずつ降下させながら歌った声から抽出したクロマベクトルを，図 4 は母音/a/の地声で同じように歌った声から抽出したクロマベクトルを表す．**図 3** は図 2 中の C4 部分のスペクトル，**図 5** は図 4 中の C4 部分のスペクトルを表している．図 2 と図 4 を比較すると，同一音階であるにも

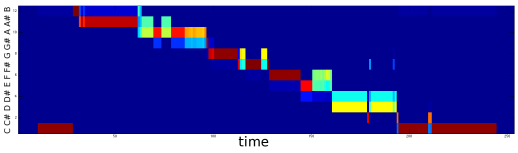

図 2 C4〜C3 地声/u/のクロマベクトル

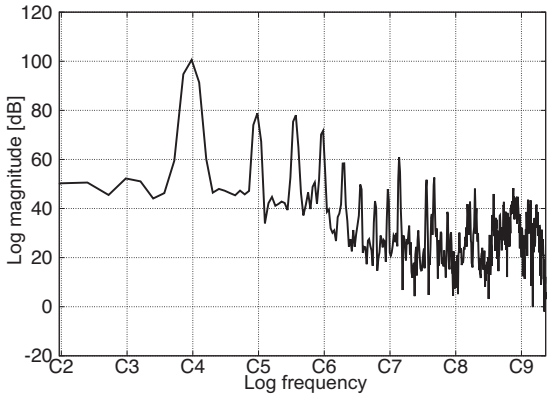

図 3 C4 地声/u/の対数振幅スペクトル

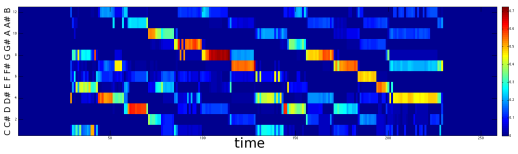

図 4 C4〜C3 地声/a/のクロマベクトル

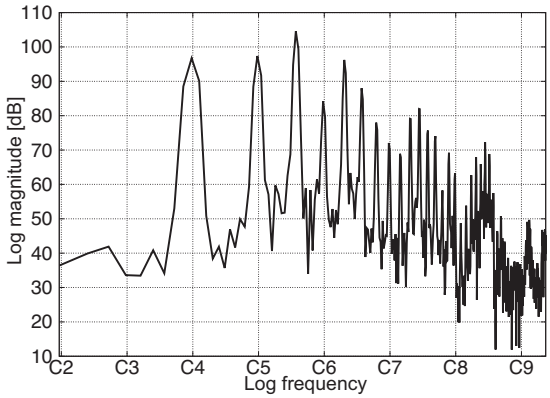

図 5 C4 地声/a/の対数振幅スペクトル

関わらず，クロマベクトルの様相が異なっていることが確認できる．このことから，歌声のクロマベクトルは，母音の違いに対しては十分に頑健とは言えないことがわかる．

この問題に対処するため，特徴量として基本周波数を使うことを考える．音階 n における基本周波数の分布を 1 次元 (有声) と 0 次元 (無声) の多空間分布 HMM[6] でモデル化すると，出力確率は次式で与えられる．

$$b_n(\boldsymbol{x}) = \begin{cases} w_{n1} \cdot \dfrac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\dfrac{(x-\mu_n)^2}{2\sigma_n^2}\right) & \text{(有声)} \\ w_{n2} & \text{(無声)} \end{cases} \tag{4}$$

ただし，$w_{n1/2}$ は音階 n における有声/無声フレームの存在比率，μ_n, σ_n^2 は正規分布の平均と分散をそれぞれ表す．

表 2 音響特徴量の違いが楽譜追跡精度に与える影響

音響特徴量	楽曲 A[%]	楽曲 B[%]
クロマベクトル	61.7	26.5
基本周波数	64.7	27.3
クロマベクトル+基本周波数	62.0	29.5

また，特徴量としてクロマベクトルと基本周波数を併用することも考えられる．この場合，音階 n における出力確率 $f_n(\boldsymbol{x})$ は次式で与えられる．ここでは，$k = 13$ の特徴量を基本周波数としている．

$$b_n(\boldsymbol{x}) = \begin{cases} w_{n1} \cdot \displaystyle\prod_{k=1}^{13} \frac{1}{\sqrt{2\pi\sigma_{nk}^2}} \exp\left(-\frac{(x_k-\mu_{nk})^2}{2\sigma_{nk}^2}\right) & \text{(有声)} \\ w_{n2} \cdot \displaystyle\prod_{k=1}^{12} \frac{1}{\sqrt{2\pi\sigma_{nk}^2}} \exp\left(-\frac{(x_k-\mu_{nk})^2}{2\sigma_{nk}^2}\right) & \text{(無声)} \end{cases} \tag{5}$$

ただし，$w_{n1/2}$ は音階 n における有声/無声フレームの存在比率，μ_{nk}, σ_{nk}^2 は特徴量の k 次元目の平均と分散をそれぞれ表す．

表 2 に，特徴量としてクロマベクトル，基本周波数，クロマベクトルと基本周波数の両方をそれぞれ用いた場合の楽譜追跡精度を示す．表中の数値は正解率を表す．正解率 C は以下で定義する．

$$C = \frac{n_c}{N} \tag{6}$$

ただし，N はテストデータのフレーム数，n_c は推定された演奏位置と正しい演奏位置とが一致したフレームの数を表す．この実験条件では，どの特徴量を用いた場合でも性能に大きな差は見られなかった．しかしながら，様々な実験条件での検討から，全体としてクロマベクトルよりも基本周波数の方が精度が高い傾向があった．特に，Viterbi アルゴリズムで逐次得られる最適パスの最終状態ではなく遅延幅 α 秒過去の演奏位置を用いて楽譜追跡を行う方法 [4] では，基本周波数を用いた場合の性能が顕著に高い結果が観察された．

5. 複数チャンネルの利用

これまでに行われてきた楽譜追跡の研究は，一人の人間による演奏に対して追跡を行うものを想定していた．しかし，本研究では複数人によるアンサンブルの中でシステムが人間と同期しながら演奏する状況を想定している．同期しながら演奏している複数の人間の歌声を同時にシステムへの入力とすることで，楽譜追跡に用いる情報量が増し，より精度よく演奏位置を推定できることが期待できる．

出力確率は，各チャンネルそれぞれの出力確率の積により計算する．本研究のように対角共分散正規分布モデルを仮定している場合，これは複数のチャンネルが持つ特徴ベクトルをすべて繋げた 1 つの特徴ベクトルとして用いることと同じである．

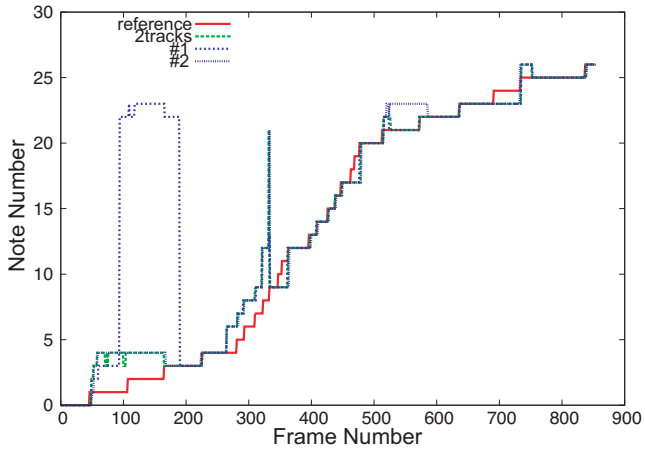

図 6 2 名の歌唱による楽譜追跡結果

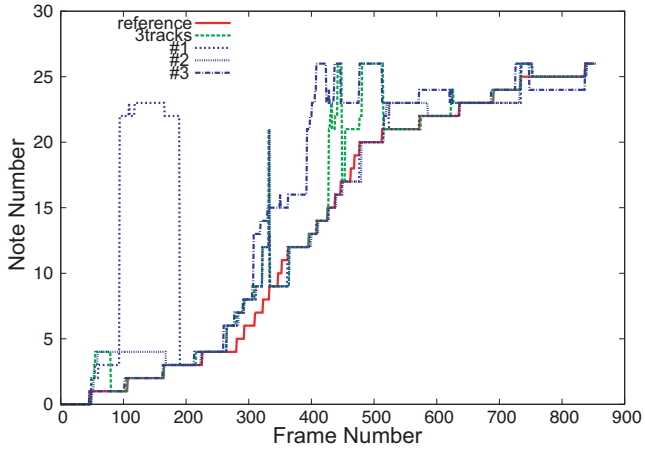

図 7 3 名の歌唱による楽譜追跡結果

表 3 楽曲 A に対する歌唱者ごとの楽譜追跡の正解率

#1 [%]	#2 [%]	#3 [%]
62.0	57.2	42.0

表 4 楽曲 A に対する複数歌唱者による楽譜追跡の正解率

#1 [%]	#1,#2 [%]	#1,#2,#3 [%]
62.0	64.6	70.6

図 6 に 2 名分の歌唱を入力とした場合の楽譜追跡の結果を示し，**図 7** に 3 名分の歌唱を入力とした場合の楽譜追跡の結果を示す． また，**表 3** には歌唱者一人ずつのデータを入力とした場合の楽譜追跡の正解率を示し，**表 4** には 2 名および 3 名のデータを入力とした場合の楽譜追跡の正解率を示す． #3 の歌唱者は他の 2 名に比べて，正しい音高からややずれて歌う傾向が強い．

図 6，図 7 を見ると，複数チャンネルを用いた場合のパスは，個別チャンネルを用いたパスのうち，正解に近い位置が推定できているチャンネルのパスに選択的に沿う傾向が見られ，複数チャンネルを入力とすることが有効であることを示している．このことは，表 4 に示した正解率の変化からも理解できる．

興味深いのは，音程が不正確な#3 のデータの併用も有効であったことである．表 3 に示されているとおり，#3 のみを用いて楽譜追跡を行った結果は比較的正解率が低い．しかし，複数チャンネルの利用時にはこれが悪影響を及ぼすことはなく，比較的正確な部分だけが選択的に利用されることで，全体の精度向上に貢献できていることがわかった．

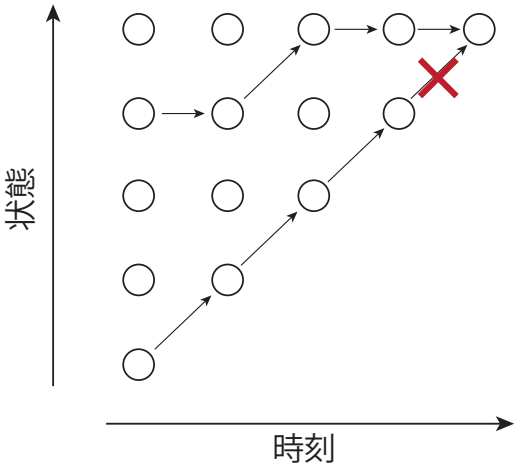

図 8 許容傾斜 3/4 の傾斜制限

6. 傾斜制限

アカペラ演奏に対して HMM を用いた楽譜追跡を行うと，音程の不正確さのために，最適パス上の各状態での停留時間が極端に短くなり，本来の演奏位置よりもどんどん先の音符へと誤って行くことがしばしばある．本研究で用いている HMM では，状態停留時間を明示的にモデル化することができない．そこで，モデルで解決するかわりに，最適パスの探索経路に制限を加えることで，極端に早い状態遷移を抑制することを試みた．

次式のように許容傾斜 a を設定する．

$$a = \frac{\text{状態間の距離 } n}{\text{遡るフレーム数 } t} \tag{7}$$

HMM の遷移確率を計算する中で，ある状態への遷移を考える際，その状態と t フレーム過去の演奏位置が n より離れた位置であった場合，その経路は計算しないものとすることで遷移を制限する．**図 8** は許容傾斜 $a = n/t = 3/4$ とした場合の例である．図 8 の右上の状態へ繋がる経路は矢印で示した二通り存在したとして，$\times$ 印のついた経路は時刻を $(t =)4$ さかのぼった状態との距離が 4 であり，$(n =)3$ を上回っているため，計算しないものとする．

2 または 3 チャンネルの歌声を入力とした楽譜追跡に傾斜制限を導入した場合の結果を，**図 9**，**図 10** に示し，正解率を**表 5** に示す．許容傾斜の値は実験的に求め，$d = 4$，$t = 20$ とした．

2 チャンネルの入力に対して傾斜制限を行った結果，傾斜制限を行う前の正解率とまったく同じ 64.6 %を示し，傾斜制限による効果は確認されなかった．一方で 3 チャンネ

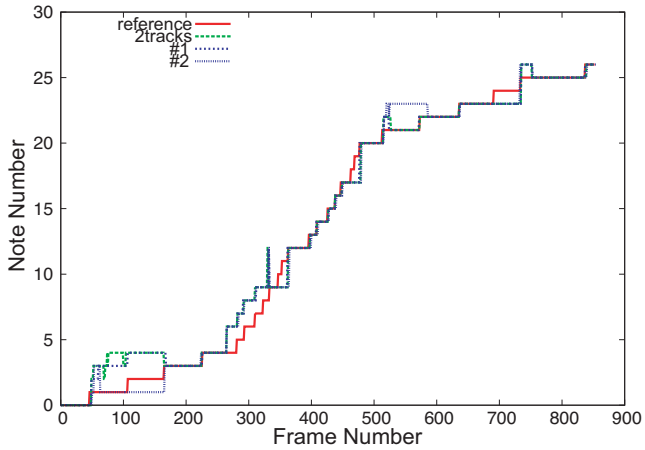

図 9 楽曲 A #1,#2 + 傾斜制限

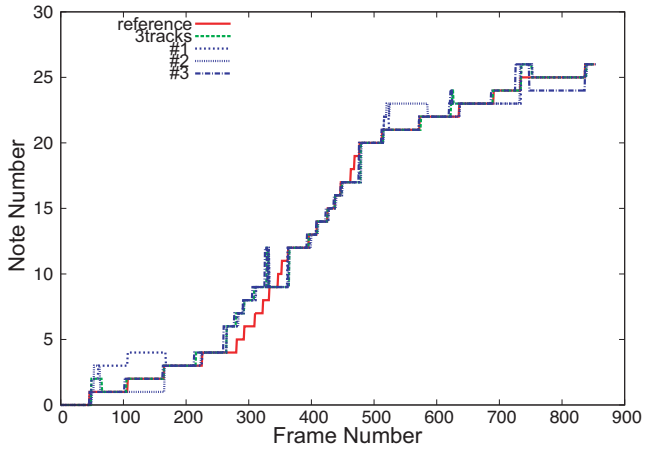

図 10 楽曲 A #1,#2,#3 + 傾斜制限

表 5 楽曲 A に対する複数チャンネルの入力と傾斜制限を用いた楽譜追跡の正解率

#1 [%]	#1,#2 [%]	#1,#2,#3 [%]
64.7	64.6	79.5

ルの入力に対して傾斜制限を行った結果，79.5 %と非常に高い正解率を示した．これは，元々の正解率が低かった歌唱者#3 が傾斜制限の影響を強く受けたためであると考えられる．

7. 聴覚的評価

本研究が目標とするアカペラ演奏支援における上記楽譜追跡手法の有効性を検証するために，アカペラ演奏の経験を有する著者 1 名による聴覚的評価を行った．具体的には，楽譜追跡の結果に従うような歌声データを作成し，仮に欠員補完システムとして動いた場合どのように聴こえるかをテストした．

作成した歌声は，まずラベリングされた音符ごとに歌声データを分割し，それを楽譜追跡結果に登場する順に再配置する．その後，各音符の長さを PSOLA 法によって楽譜追跡結果に従うように伸縮させる，という手順によって作成した．

聴覚的評価を行った対象を以下に示す．

(1) 楽曲 B 1 チャンネル入力 工夫なし
(2) 楽曲 B 1 チャンネル入力 + 傾斜制限
(3) 楽曲 A 3 チャンネル入力 + 傾斜制限

まず傾斜制限の効果を聴覚的に評価するため，楽曲 B に対する傾斜制限前後の結果から作成した演奏データを聞き比べた．その結果，どちらもところどころうまくいっている，といった印象であり，傾斜制限を導入する前は音がめまぐるしく変化し非常に聞き苦しかった部分が，傾斜制限を導入することでやや抑制され，聞きやすくなっている．しかし，どちらの結果も音に違和感のある部分がほとんどであり，欠員パートの補完に用いるには不十分な結果であった．

次に，複数チャンネル入力と傾斜制限を併用する効果を評価するため，3 チャンネルを入力とし，傾斜制限をかけた結果 (図 10) に基づいて演奏データを作成した．その結果，やや音の歌い出すタイミングのずれが存在したが，音高の関する不都合はあまり感じられず，練習に用いる想定であればほぼ問題の無い品質のものであった．したがって，本研究の目的を達成するために十分な楽譜追跡が行えていると言える．

8. おわりに

本研究では，アカペラグループが練習を行う際の欠員を補完するシステムの実現のため，歌声に対する楽譜追跡に対する検討を行った．

楽器による演奏に対する楽譜追跡において用いられていた特徴量であるクロマベクトルを歌声から抽出すると，発する母音の違いによって異なる様相を示すという特性を明らかにし，歌声に対する楽譜追跡の中では問題になる可能性があることを指摘するとともに，基本周波数を特徴量として併用する方法を述べた．次に，楽譜追跡の精度を向上させるため，複数チャンネルの歌声を入力とする方法，およびパスに傾斜制限を導入する方法，の 2 つを提案し，楽譜追跡実験により性能評価を行った．これらの提案法を導入することにより，歌声に対しても高精度な楽譜追跡を実現することができた．

残された課題として，楽譜追跡システムによって得られた演奏位置からリアルタイムで歌声を再生する歌声再生部の実装が挙げられる．歌声再生部では，時々刻々変化するテンポやピッチの変化に追従しながら人間の歌唱と調和する歌声をリアルタイムで生成する必要があり，興味深い研究課題である．

参考文献

[1] 武田 春登, 西本 卓也, 嵯峨山 茂樹, “HMM による MIDI 演奏の楽譜追跡と自動伴奏,” 情報処理学会研究報告 [音楽

情報科学] 2006(90), pp. 109–116, 2006.
[2] 鈴木 孝輔, 上田 雄, 齋藤 康之, 小野 順貴, 嵯峨山 茂樹, “HMM を用いた音響演奏の楽譜追跡による弾き直しに追従可能な自動伴奏,” 情報処理学会研究報告, Vol. 2011-MUS-89, No. 29, pp. 1–6, 2011.
[3] 中村 栄太, 武田 晴登, 山本 龍一, 齋藤 康之, 酒向 慎司, 嵯峨山 茂樹, “任意箇所への弾き直し・弾き飛ばしを含む演奏に追従可能な楽譜追従と自動伴奏,” 情報処理学会論文誌, Vol. 54, No. 4, pp. 1338–1349, 2013.
[4] 山本 龍一, 酒向 慎司, 北村 正, “Ryry: 多声楽器に対応可能な音響入力自動伴奏システム,” インタラクション 2013, pp. 612–615, 2013.
[5] T. Fujishima, “Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music,” ICMC Proceedings, pp. 464–467, 1999.
[6] K. Tokuda, T. Masuko, N. Miyazaki, and T. Kobayashi, “Multi-space probability distribution HMM,” IEICE Trans. Inf. & Syst., Vol. E85-D, No. 3, pp. 455–464, 2002.