

アカペラ演奏支援のための歌声に対する楽譜追跡 Score Following for Supporting A Cappella Vocal Performance

上田 新[†]
Arata Ueda

森 大毅[†]
Hiroki Mori

1. はじめに

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。

人間の演奏に同期して演奏を行うシステムに、自動伴奏システムがある。[1]~[4]では楽器の演奏による音響入力を対象に確率モデルを用いた楽譜追跡を行っているが、歌声に対する検討はされていない。歌声への楽譜追跡では、演奏者が意図せずに不正確な音高によって演奏されるケースや、発する母音の違いによって異なる特徴を示すなどの問題に対応する必要がある。

本稿では確率モデルを用いた楽譜追跡システムによる、歌声に対する楽譜追跡について検討した結果について述べる。

2. 音響特徴量

2.1. クロマベクトル

[1]~[4]では、音色の変化に頑健であり、その低次元性から統計的学習が容易であることから、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 (1))。

$$V(p) = \sum_{l \text{ s.t. } M(l)=p} |X(l)|^2 \quad (p = 0, 1, 2, \dots, 11) \quad (1)$$

$$M(l) = \begin{cases} -1 & (l = 0) \\ \text{round} \left(12 \log_2 \left(\left(f_s \cdot \frac{l}{N} \right) / f_{\text{ref}} \right) \right) \bmod 12 & (l = 1, 2, \dots, N/2 - 1) \end{cases} \quad (2)$$

$X(k)$ は N 点 DFT スペクトル ($k = 0, 1, \dots, N-1$)、 f_s はサンプリング周波数を表し、 $f_s \cdot \frac{l}{N}$ は l 点目のスペクトルに対応する周波数の値を表す。ピッチクラス 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。

2.2. 歌声のクロマベクトル

歌声から抽出したクロマベクトルを図 1、図 2 に示す。これらの図を比較すると、同じ音高であっても発する母音の種類によってクロマベクトルの様相が異なることが確認できる。図 3、図 4 に、母音ごとのスペクトルを示す。図より、 $/a/$ の母音では基本周波数よりも高い周波数成分が強められてしまうため、 $/u/$ のときと様相の異なるクロマベクトルが得られていることがわかる。これは楽器の音には無い、歌声特有の特徴であると考えられ、楽譜追跡において問題となる可能性がある。

[†]宇都宮大学大学院 工学研究科 電気電子システム工学専攻

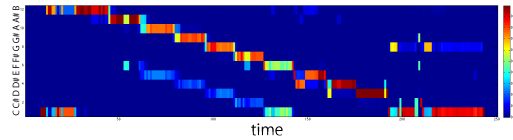


図 1: 母音 $/u/$ のコーラスで C4 から C3 まで半音ずつ下降したときのクロマベクトル

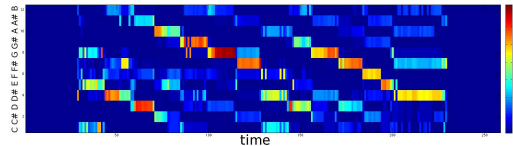


図 2: 母音 $/a/$ のコーラスで C4 から C3 まで半音ずつ下降したときのクロマベクトル

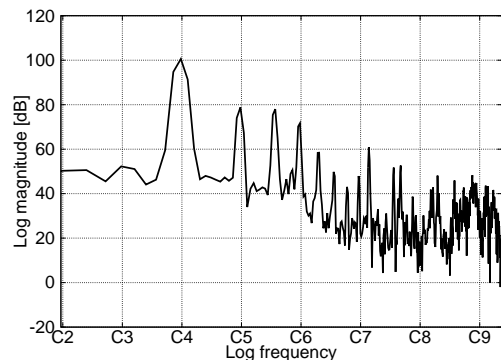


図 3: 母音 $/u/$ 、音高 C4 の対数振幅スペクトル

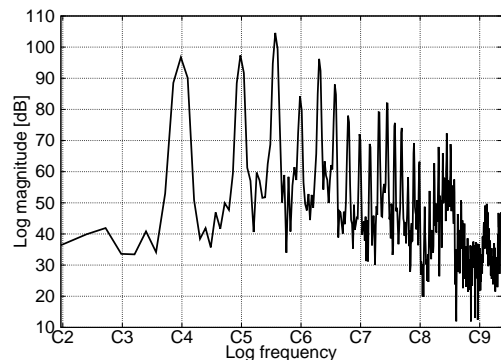


図 4: 母音 $/a/$ 、音高 C4 の対数振幅スペクトル

本稿では、この音響特徴量の様相が楽譜追跡に与える影響を明らかにするため、母音の違いを考慮せず、異なる母音から抽出されたクロマベクトルであっても同等のものとして扱う。

3. 楽譜追跡実験

3.1. アカペラ演奏曲データ

楽譜追跡実験システムの構築および評価のため、3曲のアカペラ演奏曲を収録した。演奏者は男性6人組であり、歌詞を歌うリードボーカル1名、コーラス3名、ベース1名、パーカッション1名からなる。収録は防音室にて行い、手持ちのマイクロフォンを用いて6チャンネルを同時に標本化周波数44.1 kHzでPCM録音した。以降の実験では3曲それぞれのコーラスパート3名分(計9曲に相当)を用いる。

3.2. 実験条件

前述のアカペラ演奏曲に対して、隠れマルコフモデル(HMM)を用いた楽譜追跡[1]を行った。ただし、本稿では問題を簡単にするため、演奏脱落や弾き飛ばし、弾き直しの無い演奏について考えることとし、自己遷移と一つ次の状態への遷移の二つの遷移のみが存在するものとしてモデル化を行った。特徴量であるクロマベクトルの分析は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ている。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。学習データには前述のアカペラ曲のコーラスパートのうち「uh-」や「ah-」のように単一の母音をのびして歌うコーラス部分のみを用い、半音刻みの12音階に休符を加えた13個のモデルを学習した。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曲の一部を切り出したものを用いた。

3.3. 実験結果

楽譜追跡の結果を図5、図6に示す。図5では前半部において誤推定が存在するものの、大部分において正確な追従が行われている。図5の100~200フレーム付近で発生している誤推定は音階が半音違いの部分で発生しており、演奏中に音高が変化し、半音違いの音に近付いてしまったために起きたものと考えられる。これは演奏位置を逐次的に最適パスを求めることで得ているために起きたものであり、音符の継続長を制御することで解決できると考えられる。300フレーム付近では同音の連続が演奏されているため音の区切りを認識できず、演奏位置のずれが発生している。図6も前半部分において誤りが見られるが、これは図5と同様の音高の変化や、息継ぎを行う際の無音区間において休符の状態の尤度が高まってしまふこと、また母音/a/でのコーラス部分が原因となって発生していると考えられる。

4. 結論

歌声に対して確率モデルを用いた楽譜追跡を行った結果、音高のずれや息継ぎ部分、発する母音の影響などによって誤推定が認められた。息継ぎ部分における誤推定は、一つの音符に対して演奏部と息継ぎ部の二つの状態を対応づけることで回避

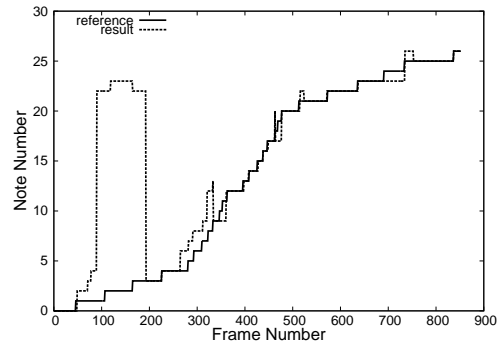


図5: 楽曲1に対する追跡結果(実線: 正しい音符の遷移, 破線: 推定された演奏位置)

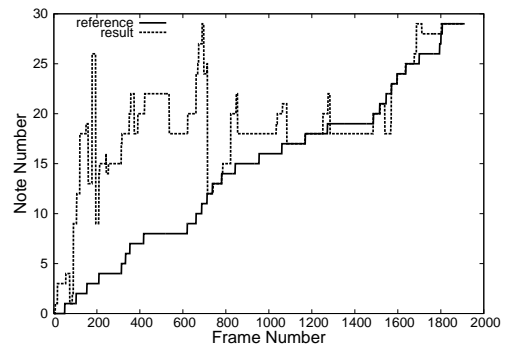


図6: 楽曲2に対する追跡結果(実線: 正しい音符の遷移, 破線: 推定された演奏位置)

することができると考えられる。また、フォルマントの影響を取り除いた残差信号からクロマベクトルの抽出を行うことで、母音の種類に影響されない、音高ごとにある程度共通なクロマベクトルを得られるのではないかと考えられる。

参考文献

- [1] 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.
- [2] 中村 友彦, 中村 栄太, 嵯峨山 茂樹, “誤り・任意の弾き直し・弾き飛ばしを含む演奏音響信号への高速な楽譜追跡,” 情報処理学会研究報告, Vol. 2013-MUS-99, No. 40, pp. 1-5, 2013.
- [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跡と自動伴奏,” 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,” インタラクシオン 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.