

メロディの調和を考慮した 合唱の歌声パートに対する多重音高推定

河合 彬弘^{1,a)} 酒向 慎司^{1,b)} 北村 正^{1,c)}

概要: 本研究では、合唱を構成する複数の歌声パートを対象とした多重音高推定を扱う。合唱には歌声パートが調和を保ちながら進行していく特有の性質があり、その性質を利用した手法を提案する。本研究では、調和の要素として歌声パート間の音高の調和と、拍位置における音高変化の有無の同期の2つを考える。提案法は、フレームごとのピッチ候補からパート別音高軌跡候補を構成するステップと、候補から考えられる組み合わせの中から歌声パート間の調和を考慮した最良なものを選出するステップに分けられる。実験結果から、比較的推定の困難な部分では、パート間の調和を考慮することで若干の推定精度改善が見られた。

キーワード: 合唱, 多重音に対する音高推定, 隠れマルコフモデル, 歌声パートの調和

1. はじめに

自動採譜等への応用を目的として音響信号からの音高推定が盛んに研究されている。音高推定は、音楽音響信号の分析分野のみ留まらず、音楽情報検索への応用も期待できる非常に有用な技術である。

しかしながら、特に複数の音が混合される多重音に対する音高推定は非常に複雑な問題である。音源数や楽器の種類といった未知パラメータが多く存在し、混合後の信号から混合前の信号を復元することは原理的に困難であることが主な原因である。しかし、多声音楽にも対応可能である実用的な音高推定を実現するためには、混合音への対応は避けて通れない問題である。これに対し、音響信号の性質や音楽的知見を利用した手法が多く提案され、有効性が示されている。

既存手法としては、スペクトルの性質に基づく制約により多重音スペクトルを分解するアプローチや、ピッチの候補を立て、最適化基準を用いて尤もらしい組み合わせを選択するアプローチなどが提案されている。前者では、スペクトルの倍音構造や音のエンベロープ [1] などが、後者ではスペクトルの滑らかさ [2, 3] やスペクトルの重心 [4] な

どが利用されている。一方で、多声楽曲の中でも特定のメロディラインに着目した研究では、ピッチの顕著さやメロディの時間的連続性を利用した手法が提案されている [5-7]。メロディの時間的連続性を利用することを目的とし、動的計画法ベースの手法が度々用いられる。

本研究では推定の対象として合唱を構成する歌声のメロディを考える。合唱とは複数の人間が複数の歌声パートに分かれて歌う音楽の形態であり、多くの場合は歌声パートの他に伴奏も含まれる。複数の歌声が重なる合唱の場合、歌声のスペクトルは平滑化されて観測される。これに伴奏の影響も加わることから、音高推定の難しさは独唱とは異なることが予測される。音高推定を考える際は、特に合唱特有の性質を利用することが求められる。

合唱の楽しみの1つに歌声パートの成すハーモニーが挙げられる。音高推定においても、歌声の成すハーモニーを利用することが望ましい。同時発音される音の調和を考慮した音高推定手法としては、[7, 8] などが挙げられる。[8] では同時発音される音の調和を表現するために和音進行が用いられている。しかし、合唱の歌声パートには音域や役割の違いが存在するため、このような合唱特有の関連性を、より具体的に利用することが望ましい。

本研究では、合唱の歌声パート間における調和として、音高の調和と音高変化の同期の2点を考える。前者は同時発音する音は調和する関係にあるという性質であり、高音パートと低音パートの音高の差を利用する。後者はメロディとして統一感を持たせるために歌声パートはある程度

¹ 名古屋工業大学
Nagoya Institute of Technology, Gokisho-cho, Showa-ku,
Nagoya, Aichi 466-8555, Japan

a) kawai@mmsp.nitech.ac.jp

b) sako@mmsp.nitech.ac.jp

c) kitamura@mmsp.nitech.ac.jp

同期して進行するという仮説に基づいており、言い換えれば音高変化パターンの類似性を扱うものである。提案法は大きく2つのステップに分けられる。第1のステップは各歌声パートのメロディ候補の抽出であり、ここではメロディの時間的連続性と歌声パートによる役割の違いが考慮される。第2のステップでは、メロディの候補から考えられる組み合わせの中から最適ものを選択する。ここでは、主に歌声パート間の調和を考慮する。

以降の構成は以下の通りである。まず第2節で本研究で扱う合唱の歌声パートの性質について述べる。第3節ではその性質を利用した多重音高推定的手法について述べ、第4節では実験と、それに対する考察を行う。最後に、第5節で本研究のまとめと今後の展望を述べる。

2. 合唱を構成する歌声パートの性質

2.1 合唱を構成する歌声のメロディの性質

人間の歌声のピッチは、楽譜が指示する音高の理想的な基本周波数と完全に一致せず、若干の揺らぎを持っている。独唱における基本周波数の変動としてはオーバーシュート、ビブラート、プレパレーションなどが知られている [9]。複数の人間が同時に歌う合唱の場合、歌声は平滑化されて観測されることから、独唱とは若干異なる変動があることが予測される。しかし、合唱が歌声であることは変わらないため、そのピッチの揺らぎは依然として存在する。

また、1つのメロディラインに着目したとき、隣接音符間における音高の変化量は多くの場合1オクターブ以内であり、それを超える急激な音高変化が起こることは稀であると言われている。1つの音高が継続する長さは、歌唱法や演奏指示による揺らぎを持つが、楽曲のテンポや音価によって大凡定まる。音価の連続であるリズムの出現傾向に偏りが見られる。

一方で、合唱を構成している歌声パートは各々が異なる役割を持っている。混声三部で構成される合唱の場合、ソプラノが主旋律を担当し、低音パートは主旋律を支える役割を持っている場合が多い。このような役割の違いは、メロディ遷移の傾向に影響を与えられられる。

2.2 歌声パート間の調和の関係

合唱の歌声パートは調和を保ちながら進行する。歌声パート間の調和としては、次のようなものが考えられる。
歌声パート間における音高の調和

合唱の趣きの1つとして、歌声の美しいハーモニーを楽しむことが挙げられる。音の組み合わせには調和するものと調和しないものがあるため、合唱曲を作曲する段階で、複数の歌声パートがよく調和するように意識されていると思われる。

一般に、2つの音がハーモニーを成すかどうかは、倍音の重なり具合によって決定されると言われている。倍音

表 1 音程の分類

Table 1 Classification of musical intervals.

分類	音程	基本周波数の比	2音間の音高の数
完全協和音程	完全1度	1:1	0
	完全8度	1:2	12
	完全5度	2:3	7
	完全4度	3:4	5
不完全協和音程	長3度	4:5	4
	短3度	5:6	3
	長6度	3:5	9
	短6度	5:8	8
不協和音程	長2度	8:9	2
	短2度	15:16	1
	減5度	32:45	6
	長7度	8:15	11
	短7度	9:16	10

の重なり頻繁さは基本周波数の比に依存していることから、基本周波数の比を利用することで調和の程度を測ることが可能であると考えられる。平均律では各音高の基本周波数は対数的に配置されているため、基本周波数の比は音高の差（音程）と捉えることもできる。音楽において、よく調和する音程を協和音程 [10] という。音程の分類を表1に示す。

歌声パート間における音高変化の同期

合唱は複数の歌声パートによって構成される。ただし、個々のパートがまとまりなく進行してしまつては統一感の欠ける不自然な曲となるため、歌声パートはある程度連動しながら進行させる必要がある。

歌声パートの連動の中でも、本研究では特に拍位置における音高変化の有無の同期について考える。これは、あるパートで音高が変化する拍位置では、他のパートにおいても音高変化が起きやすいというものである。ただし、この同期は常に一定の割合で起こるわけではなく、楽譜上の位置など様々の要因によって傾向が異なることも予測される。

3. 歌声パート間の調和を考慮した多重音高推定

3.1 提案法のアプローチ

合唱を構成する歌声パートは、パート間の調和を保ちつつ一定の独立性を持って進行していく。本研究で扱う歌声パートの性質には、単独の歌声パートに関するものとパート間の関連に関するものの2種類があり、これらを同時に最適化するの是非常に複雑な問題となる。そこで本研究では、音高推定の問題を次の2つに分割する。

音高軌跡候補の抽出

フレームごとに取得したピッチの候補をもとに、パート別の音高軌跡の候補（小節単位）を抽出する。ここ

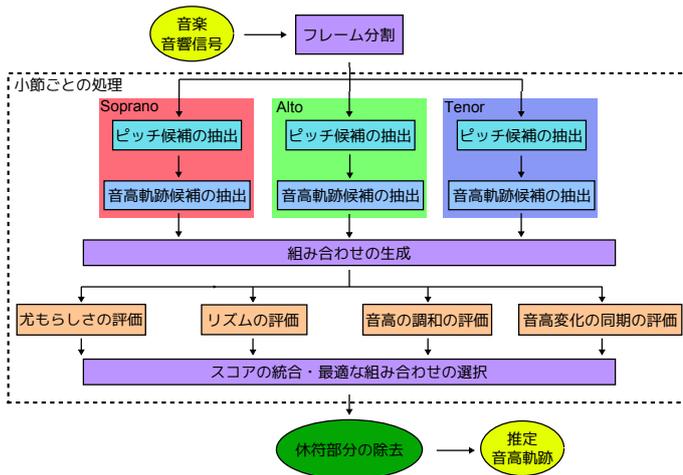


図 1 音高推定の流れ
 Fig. 1 System overview.

では、歌声のピッチの不安定性と時間的連続性、各歌声パートの役割を考慮する。

最適な音高軌跡の組み合わせの選択

パートごとに得られた音高軌跡の候補から取り得る全ての組み合わせの中から最適なものを選択する。ここでは、音高の調和、音高変化の同期の2点を含むスコアリングを行う。

本研究では合唱の中でも特に混声3部（ソプラノ、アルト、テナー）で構成されるものを扱う。音高推定全体の流れを図1に示す。

3.2 音高軌跡候補の抽出

ここでは、パート別の音高軌跡の候補（小節単位）を抽出する。

3.2.1 ピッチ候補の抽出

音高軌跡候補抽出の前処理として、フレーム分割された音響信号からピッチの候補（ピッチとその信頼度）を抽出する。音響信号に含まれる複数のピッチを推定する手法は数多く提案されており、本研究では音高推定のための特徴量である Modified Sub-Harmonic Summation (MSHS) を用いる。MSHS は Normalized SHS [7] の倍ピッチ・半ピッチ成分を抑圧した特徴量であり、これを用いることで倍ピッチ・半ピッチの誤りを軽減することが期待できる。MSHS の値は音響信号において各々のピッチがどの程度顕著に現れているかを示しており、ピッチ候補としては MSHS の値の大きいものを取り上げることが妥当である。本研究では、MSHS 上位 N_C 個を与えるピッチをピッチ候補とし、その信頼度には MSHS の値を用いる。

3.2.2 HMM に基づく歌声パートのモデル化

音高軌跡を構成するため、合唱の歌声パートを生成するモデルを考える。これにより、ピッチ候補から音高軌跡を構成する問題を、モデルがピッチ候補を生成するプロセス

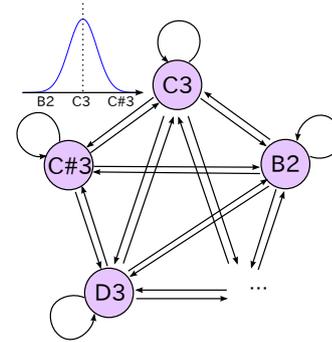


図 2 合唱の歌声パートのピッチを生成する HMM
 Fig. 2 HMMs to generate singing pitches.

の逆問題として解くことが可能になる。

歌声パートの性質を踏まえ、本研究では合唱の歌声パートを図2のような隠れマルコフモデル (HMM: Hidden Markov Model) によりモデル化する。図2の HMM では各状態は音高1つに対応している。出力確率には、各音高の理想的なピッチを平均とするような正規分布を与える。正規分布の分散により、歌声のピッチの揺らぎに対応できる。なお、各音高の理想的なピッチは対数的に分布していることを考慮し、ピッチは全て単位 cent で扱う。

HMM の各状態は音高に対応しているため、状態遷移確率はメロディの移り変わりを表現する役割を持つ。歌声パートの役割の違いを表現するため、遷移確率は合唱の楽譜からパート別で学習する。

3.2.3 Viterbi アルゴリズムの N Best サーチに基づく音高軌跡候補の抽出

フレーム番号を t 、小節を構成するフレーム数を T 、ピッチ候補を $\mathbf{c}_t = \{c_{t,i}\}_{i=1}^{N_C} (1 \leq t \leq T)$ 、HMM の状態系列 $\mathbf{n} = \{n_t\}_{t=1}^T$ としたとき、任意の音高軌跡 \mathbf{n} に対する尤もらしさ s_L は式 (1) により与えることができる。ここで、 a_{n_{t-1}, n_t} は状態 n_{t-1} から n_t への遷移確率を、 $b_{n_t}(\hat{c}_t)$ は音高 n_t からピッチ \hat{c}_t が出力される確率を、 $v(\hat{c}_t)$ はピッチ \hat{c}_t の信頼度 (MSHS の値) をそれぞれ意味している。

$$s_L(\mathbf{n}) = \sqrt[T]{\max_{\hat{c}_t \in \mathbf{c}_t, 1 \leq t \leq T} \prod_{t=1}^T a_{n_{t-1}, n_t} b_{n_t}(\hat{c}_t) v(\hat{c}_t)} \quad (1)$$

音高軌跡候補の抽出は、ピッチの候補を出力する HMM の状態系列の N Best として抽出する。ただし、極端に類似した軌跡が出現することを防ぐため、次のような枝刈り処理を行う。

- (1) t が増加するごとにバックトレースを行い、軌跡の途中経過を取得する。
- (2) 取得した軌跡を終端の音高で分類する。終端の音高が等しい軌跡は同一のクラスタに割り当てられる。
- (3) それぞれのクラスタの中で、次の処理を行う。
 - (a) 軌跡から2本を選び出す全ての組み合わせを考え、各組み合わせに対して軌跡の間の距離を計算

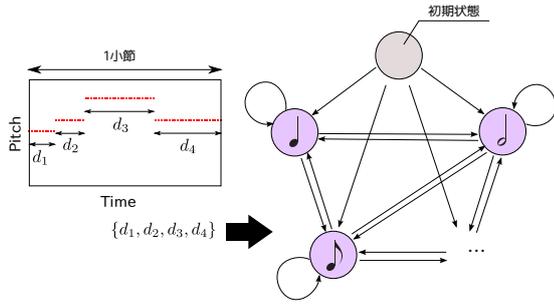


図 3 HMM によるリズムの評価

Fig. 3 Rhythm Evaluation by using HMMs.

する。距離の算出には DP マッチングを用いる。
(b) 距離が閾値未満のとき、尤もらしさが低い軌跡を無効化する。

3.3 スコアリングによる最適な組み合わせの選択

ここでは、パートごとに得られた音高軌跡から取り得る全ての組み合わせを 4 つの観点から評価し、最適なものを決定する。以降では、ソプラノの音高軌跡候補 $\{n_i^{Sp}\}_{i=1}^{N_P^{Sp}}$ 、アルトの音高軌跡候補 $\{n_j^{Al}\}_{j=1}^{N_P^{Al}}$ 、テナーの音高軌跡候補 $\{n_k^{Tn}\}_{k=1}^{N_P^{Tn}}$ から取り得る組み合わせ $n_{i,j,k} = (n_i^{Sp}, n_j^{Al}, n_k^{Tn})$ に対してスコアを付加することを考える。

3.3.1 組み合わせの尤もらしさに関する評価

ここでは、音高軌跡の組み合わせ自体の尤もらしさを評価する。式 (2) に示すように、評価値には 3 本の音高軌跡 $n_i^{Sp}, n_j^{Al}, n_k^{Tn}$ の尤もらしさ s_L の平均を用いる。

$$s_L(n_{i,j,k}) = \sqrt[3]{s_L(n_i^{Sp})s_L(n_j^{Al})s_L(n_k^{Tn})} \quad (2)$$

3.3.2 リズムに関する評価

ここでは、各々の軌跡のリズムがどの程度妥当であるかを評価する。3.2 節で述べた HMM (図 2) には音高の継続長に関する制約がなく、音楽的に不自然なリズムの軌跡が出現する可能性を否定できない。このため、不自然なリズムを排除する仕組みが必要となる。

リズムの評価では、軌跡を構成する音高 1 つ 1 つの状態継続長 $D = \{d_m\}_{m=1}^{N_D}$ を HMM で評価する (図 3)。各状態は 1 つの音価に対応しており、出力確率分布で各音価はどの程度の時間継続するのかを表現する。小節を構成するフレーム数は既知として与え、出力確率分布は正規分布とし小節を構成するフレーム数、拍子、音価から平均を定める。正規分布の分散には全状態共通のものを与える。状態遷移確率は音価遷移の傾向を表現する役割を持つ。遷移確率は合唱曲の楽譜より学習する。音高軌跡 n に対するリズムの評価値には式 (3) で得られる D の尤度を用いる。ここで、 V_m は音価に対応する HMM の各状態を、 a'_{V_{m-1}, V_m} 、 $b'_{V_m}(d_m)$ はそれぞれ遷移確率、出力確率を示している。

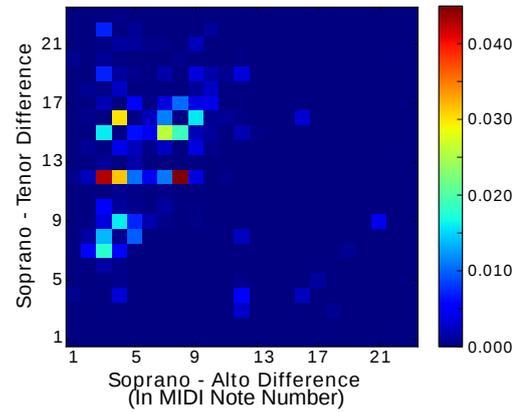


図 4 パート間における音高の差の出現頻度: 横軸、縦軸は音高を MIDI ノート番号に置き換えたときの差を示す。

Fig. 4 Frequency of appearance of $(\Delta^{Sp-Al}, \Delta^{Sp-Tn})$

$$s_R(n) = \sqrt[3]{\max_{V=\{V_m\}_{m=1}^{N_D}} \left(\prod_{m=1}^{N_D} a'_{V_{m-1}, V_m} b'_{V_m}(d_m) \right)} \quad (3)$$

また、音高軌跡の組み合わせに対するリズムの評価値には、式 (4) に示すように軌跡単独の評価値の平均を用いる。

$$s_R(n_{i,j,k}) = \sqrt[3]{s_R(n_i^{Sp})s_R(n_j^{Al})s_R(n_k^{Tn})} \quad (4)$$

3.3.3 音高の調和に関する評価

ここでは、音高軌跡間の調和を音高の側面から評価する。前述の通り、一般的に 2 つ以上の音が調和するか否かは音高の差に依存すると考えられるため、これを基準にした評価を考える。評価には、合唱曲の楽譜から学習した歌声パート間における音高の差の出現頻度の情報 (図 4) を用いる。

軌跡の組み合わせ $n_{i,j,k}$ に対し、音高の調和に関するスコアは式 (5) の形で算出する。 $p(\Delta^{Sp-Al}, \Delta^{Sp-Tn})$ はソプラノ-アルト間の音高差が Δ^{Sp-Al} 、ソプラノ-テナー間の音高差が Δ^{Sp-Tn} というパターンの出現頻度を意味している。

$$s_{H_n}(n_{i,j,k}) = \sqrt[T]{\prod_{t=1}^T p(\Delta_{i,j,t}^{Sp-Al}, \Delta_{i,k,t}^{Sp-Tn})} \quad (5)$$

3.3.4 音高変化の同期に関するスコアリング

ここでは、音高軌跡間の調和を音高変化の同期の観点から評価する。音高変化パターンの類似度を評価するためには、その表現法を定める必要がある。本研究では、音高の変化パターンを K 次元のビット列として扱う。 K 次元とは 1 小節を K 分割することを意味し、音高が変化する拍位置には 1、変化しない拍位置では 0 をセットする。2 つの音高変化パターン $X = \{x_e\}_{e=1}^K$ 、 $Y = \{y_e\}_{e=1}^K$ があつたとき、 $x_e = y_e$ であれば拍位置 e で音高変化の同期が起きていることを示している。逆に、 $x_e \neq y_e$ であれば拍位置 e では音高変化の同期が起きていないことを示す。類似した表現法が [11] で用いられている。

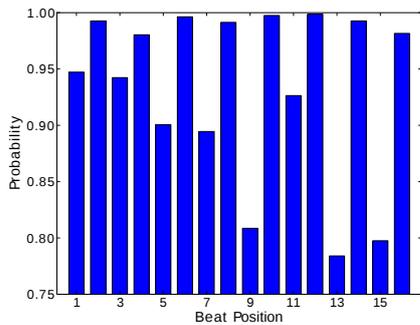


図 5 音高変化の同期が起こる確率: 横軸は 1 小節を 16 分割した拍位置を示す。

Fig. 5 Probability of occurring synchronization of existence of pitch change at each beat position.

音高変化がどの程度同期するかは様々な要因によって変化する。本研究では、音高変化の同期の程度は小節内における拍位置で異なると仮定し、 K 分割した各拍位置でどの程度音高変化の同期が起きやすいのかを合唱の楽譜から学習する。学習によって得られた結果を図 5 に示す。

音高変化の同期に関するスコアは、図 5 を利用して算出する。まず、2 つの音高変化パターン X, Y の類似度を式 (6) の形で与える。ここで、 ϕ_e は拍位置 e における音高変化の同期が起こる確率を示している。

$$r(X, Y) = \sqrt[16]{\prod_{e=1}^{16} \phi'_e}, \quad \phi'_e = \begin{cases} \phi_e & (x_e = y_e) \\ 1 - \phi_e & (x_e \neq y_e) \end{cases} \quad (6)$$

軌跡の組み合わせに対する音高変化の同期に関するスコア S_{H_r} は、 n_i^{Sp} と n_j^{Al} , n_i^{Sp} と n_k^{Tn} , n_j^{Al} と n_k^{Tn} のそれぞれの類似度の平均を用いる。

3.4 スコアの統合と最良な組み合わせの選択

このステップでは、前節までで述べた 4 種類の評価値を統合する。4 つの評価値は意味付けが大きく異なるため、優先度をつけることが望ましい。また、4 つの評価値はスケールが異なるため、正規化も必要になる。以上を踏まえ、4 種類の評価値に対して偏差値を求め、その重み付き平均を統合スコアとする。全組み合わせの中で統合スコアの最も高い組み合わせを推定音高軌跡とする。

3.5 後処理: 休符部分の推定

音高軌跡候補を抽出するステップでは全てのフレームで注目しているパートのピッチが存在することを前提としているため、後処理としてピッチの無い部分 (休符など) の推定し、該当部分を除去する。

休符部分を捉えるためピッチの信頼度に着目する。音高軌跡を抽出する際に、本来ピッチがない部分を探索すると、他のフレームに比べてピッチの信頼度が低くなる。図 6 は、実際に出現した 1 小節の音高軌跡を示す。各フレームの音高を上図に、その音高として選ばれたピッチ候補の

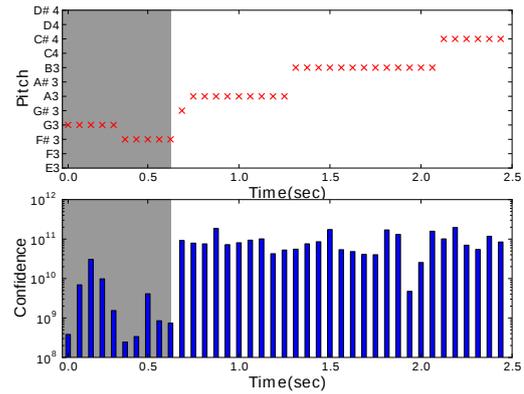


図 6 音高軌跡と対応する信頼度 (1 小節分)

Fig. 6 Pitch trajectory and its confidence at each frame.

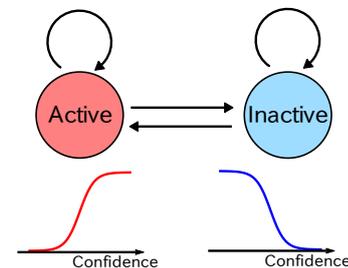


図 7 ピッチの信頼度の変化を表現する HMM

Fig. 7 HMMs that represent changes of confidence of pitch.

信頼度を下図に図示しており、灰色の部分は楽譜上休符の部分に相当している。休符の部分では、ピッチの信頼度が低下していることが確認できる。これを利用し、元々休符であった部分を推定する。

あるフレームが元々休符であるとき、次のフレームも休符である可能性が高い。逆にあるフレームが休符でなかった場合、次のフレームも休符でない可能性が高くなる。休符部分の推定では、このような時間的連続性を考慮することが必要である。これを踏まえ、信頼度の変化を図 7 のような HMM によってモデル化する。図 7 の HMM は Active と Inactive の 2 つの状態を持っており、Active は休符でない場合、Inactive は休符である場合をそれぞれ意味している。それぞれの状態には、sigmoid 関数の形で信頼度の出力確率を与える。曲全体の推定音高軌跡からフレーム単位の信頼度を取得し、それを出力する HMM の最尤状態系列を求める問題として休符部分の推定を扱う。

4. 評価実験

提案法の性能を確認するための評価実験を行った。モデルのパラメータには合唱曲 10 曲から学習したものをを用いた。スコアが正解率に与える影響を確かめるため、次の 4 パターンで実験を行った。

手法 A: $S_L, S_R, S_{H_n}, S_{H_r}$ を用いる。

手法 B: S_L, S_R, S_{H_n} を用いる。

表 2 実験条件

Table 2 Experimental Conditions.

サンプリング周波数	16 kHz
フレーム長 / フレーム周期	128 ms / 64 ms
窓関数	blackman 窓
N Best	500
ピッチ候補の数	1 フレームあたり 15 個
ピッチの出力確率の分散	50 cent
ピッチ候補を抽出する範囲	ソプラノ 261–880 Hz
	アルト 196–659 Hz
	テナー 82–349 Hz
音高の継続長の分散	0.01×1 小節のフレーム数
1 小節の分割数	16

表 3 実験結果

Table 3 Experimental Result.

	data1				data2			
	Sp	Al	Tn	Pr	Sp	Al	Tn	Pr
手法 A	67.0	47.6	40.1	59.2	57.6	34.8	29.8	45.5
手法 B	67.3	48.2	39.0	58.7	58.1	35.5	28.3	45.1
手法 C	67.0	48.7	39.5	59.2	58.9	34.1	26.4	43.7
手法 D	67.3	47.3	38.8	58.0	58.7	33.2	25.4	43.3
上限	79.0	68.0	69.0	–	65.0	54.0	59.0	–

手法 C: S_L, S_R, S_{H_r} を用いる.

手法 D: S_L, S_R を用いる.

4 つのスコアに対する重みは予備実験により $(S_L, S_R, S_{H_n}, S_{H_r}) = (0.88, 0.06, 0.04, 0.02)$ と定めた. なお, 各々のフレームの推定音高がソプラノ, アルト, テナーの順で高くなる制約を設ける. 実験データとしては [12] に収録されている混声 3 部の合唱曲中の 3 曲を用いた. また, 音響信号の複雑さによる精度の変化を確認するため, 各楽曲に対し, 歌声に似た音源で合成したデータ (data1), 実際の合唱 (data2) の 2 種類を用意した. 2 種類とも歌声パートと共にピアノ伴奏が含まれている. その他の実験条件を表 2 に示す. 精度の評価には DP マッチングを用いたパート別の正解率と, パートの区別を無視したときの推定音高の適合率を用いる. また, 音高軌跡候補抽出の性能を確認するため, 候補から考えられる組み合わせの中から最も正解率が高くなるものを選んだ場合の正解率も算出した. これは, パート別正解率の上限になる値である.

実験結果を表 3 に示す. Sp, Al, Tn をそれぞれソプラノ, アルト, テナーを, また Pr は適合率を意味している. 軌跡候補の抽出では, 実験データの双方において, パート別正解率がソプラノ, テナー, アルトの順で高くなる結果となった. アルトは中央のパートであり, ソプラノやテナーと区別が付き難いためであると考えられる. 全体の結果としては, 高音パートの影響や音響信号の複雑さにより音高軌跡の尤もらしさの信頼性が低下することが data2 の低音パートに対して調和の効果が見られた. また, S_{H_n} は低音

パート, S_{H_r} では data2 の精度が若干向上する傾向が見られた.

5. むすび

本研究では, 合唱における歌声パート間における調和の関係に着目した音高推定の手法を提案した. 提案法では, 音高推定の問題をメロディの連続性を考慮した音高軌跡候補の抽出と, 歌声パート間の調和を考慮した最適な組み合わせの選択の 2 つに分割した. 歌声パート間の調和については, 特に音高の調和と音高変化の同期に関して評価を行った.

実験において, 軌跡の尤もらしさの信頼度が低く, 推定が困難な場合について効果が確認できた. 今後の課題としては, 音域を考慮した音高軌跡候補の抽出や, 調和に関する評価で用いる分布の詳細化などが挙げられる.

謝辞

本研究の一部は, 市原国際奨学財団の支援を受けた.

参考文献

- [1] Kameoka, H., Nishimoto, T. and Sagayama, S.: A Multi-pitch Analyzer Based on Harmonic Temporal Structured Clustering, *IEEE Trans. on ASLP*, Vol. 15, No. 3, pp. 982–994 (2007).
- [2] Emiya, V., Badeau, R. and David, B.: Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle, *IEEE Trans. on ASLP*, Vol. 18, No. 6, pp. 1643–1654 (2010).
- [3] Pertusa, A. and Inesta, J. M.: Multiple fundamental frequency estimation using Gaussian smoothness, *Proc. ICASSP*, pp. II–1085–1088 (2004).
- [4] Benetos, E. and Dixon, S.: Polyphonic music transcription using note onset and offset detection, *Proc. IC-CASP*, pp. 37–40 (2011).
- [5] Goto, M.: A Predominant-F0 Estimation Method for Polyphonic Musical Audio Signals, *Proc. ICA*, pp. II–1085–1088 (2004).
- [6] 藤原弘将, 後藤真孝, 奥野 博: 歌声の統計的モデル化とビタビ探索を用いた多重奏のボーカルパートに対する音高推定手法, *情報処理学会論文誌*, Vol. 49, No. 10, pp. 3682–3693 (2008).
- [7] Hsu, C. L., Chen, L. Y., Jang, J. S. and Li, H. J.: Singing pitch extraction monaural polyphonic songs by contextual audio modeling and singing harmonic enhancement, *Proc. ISMIR*, pp. 201–206 (2009).
- [8] Raczynski, S. A., Vincent, E., Bimbot, F. and Sagayama, S.: Multiple pitch transcription using DBN-based musical models, *Proc. ISMIR*, pp. 363–368 (2010).
- [9] Saitou, T., Unoki, M. and Akagi, M.: 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).
- [10] 石冢真礼生, 丸太昭三, 金光威和雄, 末吉保雄, 飯沼信義: 楽典, 音楽之友社 (2008).
- [11] 中野拓帆, 深山 覚, 小野順貴, 嵯峨山茂樹: リズム木構造仮説に基づく楽曲解析とリズム木学習, *日本音響学会春季研究発表会講演集*, pp. 977–978 (2010).
- [12] 若松正司: クラス合唱 旅立ちの日に, ヤマハミュージックメディア (2010).