

ハモリパート練習支援システムのための 音響信号を対象とした副旋律生成の検討： 隠れマルコフモデルの導入

福田 康太^{1,a)} 北原 鉄朗^{1,b)}

概要：

本稿では、音楽知識が乏しい人でも簡単に副旋律を練習できる支援システムの開発を目指し、主旋律の基本周波数（F0）を基に副旋律を生成する新しい手法を提案する。隠れマルコフモデル（HMM）を用い、主旋律が変化しない区間でのピッチ安定性を向上させる遷移確率行列の工夫や、不協和音を避ける出力確率の調整を行うことで、副旋律の音楽的適切性を高め、特にロングトーン部分での安定性を実現した。今後は転調対応の精度向上やユーザインタラクション設計を進め、誰もが副旋律練習を楽しめるシステムを目指す。

1. はじめに

カラオケなどにおいて、副旋律を即興で考え、主旋律と一緒に歌うといった楽しみ方が存在する。そこで、様々な曲の副旋律を歌えるようにするためにある程度の練習をしたいと考えたとき、音楽知識に乏しい人にとって副旋律を練習することは難しい。それは、副旋律が存在していない場合、副旋律を自分で考えて作らなくてはならないからである。また、副旋律のみの音源をユーザーが入手しているかによって、副旋律を練習する際の習熟度は大きく影響する。

本研究では、音楽知識が乏しい人でも、自分が好きな曲の副旋律を練習することが出来る支援システムの開発を目指す。同様のコンセプトの研究に、白石らの HamoKara[1] がある。このシステムには、ユーザが指定した楽曲に対して副旋律を自動生成し、副旋律の歌唱練習を支援する機能がある。しかし、ユーザが練習したい楽曲の MIDI ファイルをあらかじめ準備する必要があり、MIDI ファイルが準備されていない楽曲に利用できない。カラオケから離れると、副旋律の自動生成に関する研究は数多く存在する（たとえば [2], [3]）。しかし、MIDI データや楽譜データを対象としているものが多い。ピアノの音響信号を対象とした研究も存在する [4] が、音高の変化が複雑な歌唱を対象としたものではなかった。

我々は以前も、ポピュラー楽曲の音響信号を与えると、副旋律の歌唱を自動生成するシステムを試作した [5]。このシステムでは、主旋律の基本周波数（F0）を推定し、フレームごとにその「短3度下」「長3度下」「完全4度下」のいずれかを選択して、ピッチシフトにより副旋律の歌唱を生成する。副旋律の F0 がフレームごとに決まるため、特にロングトーン部分のピッチが不安定であるという問題点があった。また、ロングトーン部分以外でも、音楽理論的に不適切な音が選ばれることがあった。

本稿では、この問題を解決するための手法を提案する。まず、主旋律に対する「短3度下」「長3度下」「完全4度下」を決める際に、2種類の遷移確率行列を持つ隠れマルコフモデル（HMM）を用いる。主旋律の F0 が変わらないフレームでは自己遷移の確率を極端に高くした遷移確率行列を用いることで、不用意にピッチが変わることを防ぐ。また、伴奏で際立つ音高と半音の関係になるものの出力確率を低くするなどの工夫を行い、伴奏と不協和な副旋律をできるだけ避けるようにする。

2. 副旋律の自動生成手法

本手法では、与えられた歌唱曲音源に対して、歌唱（主旋律）と伴奏を分離し、伴奏音源から得られるクロマグラムを利用して副旋律を生成する。クロマグラムには和声（コード進行）の情報が反映されており、これを基に主旋律と調和する副旋律の音高を決定する。副旋律生成には隠れマルコフモデル（HMM）を用い、音楽的に自然な音高遷移を実現する。

¹ 日本大学文理学部情報科学科

^{a)} fukuda@kthrlab.jp

^{b)} kitahara@kthrlab.jp

2.1 前処理

2.1.1 主旋律（歌唱）および伴奏のみの音源の取得

与えられた歌唱曲の音響信号に対して、音源分離ライブラリー「Spleeter」を適用することで、主旋律（歌唱）のみの音響信号と伴奏のみの音響信号を得る。

2.1.2 主旋律の音高推定

主旋律の音響信号（48kHz にリサンプリング）に対して、音声解析ライブラリー「PyWorld」を用いて 5ms ごとの基本周波数を推定する。時刻 t における主旋律の基本周波数を $x(t)$ とする。PyWorld によって調波構造がないと判定されたフレームは、以降の処理は行わず、副旋律も付与しないものとする。

2.1.3 伴奏に対するクロマグラム作成

伴奏の音響信号（48kHz にリサンプリング）に対して、Librosa を用いてクロマグラムを得る。クロマグラムは、12 音階の各音名の振幅を表す 12 次元ベクトル（クロマベクトルと呼ぶ）を時間軸上に並べたものである。主旋律の基本周波数と時間分解能（5ms）を合わせるため、シフト幅を 240 サンプルとした。

その後、局所的な非和声音などの影響を削減するため、数値解析ライブラリー「SciPy」を用いて、クロマグラムの時間軸方向の平滑化を行った。平滑化の窓幅を 100 フレームとした。

2.1.4 副旋律の候補の選定

現在の実装では、副旋律は主旋律に対して短 3 度下、長 3 度下、完全 4 度下のいずれかの状態とする。時刻 t における副旋律の音高の候補の集合を $C(t)$ とすると、

$$C(t) = \{2^{-3/12}x(t), 2^{-4/12}x(t), 2^{-5/12}x(t)\}$$

となる。

2.2 HMM を用いた副旋律生成

副旋律生成においては、主旋律に調和しつつ、音楽的に自然な音選択を行うことが重要である。しかし、音楽信号の特性は、時間的に変動する主旋律や伴奏との関係に強く依存しており、単純なルールベースのアプローチでは、これらの複雑な関係性を表現することが困難である。この課題を解決するため、本研究では隠れマルコフモデル（HMM）を用いた副旋律生成手法を提案する。

時刻 t における副旋律の候補 $C(t)$ を状態集合とする。副旋律は伴奏と和声的に協和する必要があるため、伴奏から抽出したクロマベクトルを $\mathbf{h}(t)$ とする。このとき、 $P(\mathbf{h}(1), \dots, \mathbf{h}(T) \mid c(1), \dots, c(T))$ を最大にする $[c(1), \dots, c(T)]$ を推定する ($c(t) \in C(t)$)。これは、状態遷移確率 $P(c(t) \mid c(t-1))$ 、出力確率 $P(\mathbf{h}(t) \mid c(t))$ が定義されれば、ビタビアルゴリズムによって求めることができる。

2.2.1 遷移確率の定義

前報の方法 [5] では、フレームごとに副旋律の音高を決めていたため、主旋律が同じ音高を伸ばしている箇所でも、副旋律が揺れ動くことがあった。それをふせぐために、主旋律が同じ音高を伸ばしているときは、副旋律の音高を決める状態遷移においても、同一状態への遷移確率を高くすることで、副旋律の音高が揺れ動くことを防ぐ。具体的には、主旋律の音高が直前フレームから変わらないときは表 1、変わるときは表 2 に示す状態遷移確率行列を用いる。

表 1 同一音高の遷移確率

F0 のピッチ	長 3 度	短 3 度	完全 4 度	none
長 3 度	0.7	0.15	0.1	0.05
短 3 度	0.15	0.7	0.1	0.05
完全 4 度	0.2	0.2	0.6	0.1
none	0.35	0.35	0.25	0.05

表 2 異なる音高の遷移確率

F0 のピッチ	長 3 度	短 3 度	完全 4 度	none
長 3 度	0.35	0.35	0.25	0.05
短 3 度	0.35	0.35	0.25	0.05
完全 4 度	0.35	0.35	0.25	0.05
none	0.3	0.3	0.3	0.1

2.2.2 出力確率の定義

音楽的調和を維持しつつ、副旋律を生成するには、旋律の音高がその時点での和音構成音に適合する必要がある。この適合性を評価するため、本研究ではクロマベクトルを用いて出力確率を定義する。具体的には、 $C(t)$ の各要素 $c(t)$ が表す音高が次の条件を満たすときに、その音高の出力確率を高くなるように設定する。

条件 1 $c(t)$ が、その時刻におけるコードの主要な構成音である可能性が高いとき

条件 2 $c(t)$ が、その時刻におけるコードの主要な構成音と半音でぶつかる可能性が低いとき

条件 3 $c(t)$ が、その楽曲のキーのスケールの構成音である可能性が高いとき

条件 1 と条件 2 は、時刻 t におけるクロマベクトル $\mathbf{h}(t) = (h_1(t), \dots, h_{12}(t))$ に基づいて定義する。副旋律の音高候補 $c(t)$ に対応する音名の添え字を i としたとき、

$$P_1(\mathbf{h}(t) \mid c(t)) = \frac{h_i(t)}{\sum_j h_j(t)}$$

は、時刻 t において音名 i が主要な構成音である可能性を表すと解釈できるので、これを「条件 1」に対応する出力確率とする。同様に、

$$P_2(\mathbf{h}(t) \mid c(t)) = 1 - \frac{h_{(i+1) \bmod 12}(t) - h_{(i-1) \bmod 12}(t)}{\sum_j h_j(t)}$$

を「条件 2」に対応する出力確率とする。

条件 3 については、キーのスケールの構成音をクロマベクトルの楽曲全体の時間平均 $\mathbf{s} = \text{mean}_t \mathbf{h}(t)$ で代用する。つまり、「条件 3」に対応する出力確率を

$$P_3(\mathbf{h}(t) | c(t)) = \frac{s_i}{\sum_j s_j}$$

と定義する (s_i は $\mathbf{s}$ の i 番目の要素)。この場合、楽曲の途中で転調がある場合に、適切にスケールの構成音を近似できなくなるが、転調に対する対応は今後の課題とする。

これらを求めた後、

$$P(\mathbf{h}(t) | c(t)) = P_1(\mathbf{h}(t)|c(t))^\alpha P_2(\mathbf{h}(t)|c(t))^\beta P_3(\mathbf{h}(t)|c(t))^\gamma$$

を出力確率とする。重み値 α, β, γ は実験的に定める。

2.3 副旋律の音響信号を生成する

前節の方法でフレームごとの副旋律の基本周波数が確定したら、PyWorld を用いてピッチシフトにより副旋律の音響信号を生成する。さらに、主旋律、副旋律、伴奏を重ねた音響信号も生成する。

3. 予備実験

3.1 実験方法

提案手法と従来手法を用いて、3つの音源を対象に副旋律生成を行った。従来手法とは、提案手法の 2.2 が含まれていない手法である。具体的には、伴奏音源から得られたクロマベクトルの値に基づき、副旋律の候補音高を選定し、最もクロマベクトルの値が大きい音高を直接採用した。

対象の音源は以下の 3 曲である。

- 秦基博 「ひまわりの約束」:
<https://www.youtube.com/watch?v=rKsQ-3N-Bks>
- テレサテン 「つぐない」:
<https://www.youtube.com/watch?v=PhAjY67NR9k>
- あいみょん 「マリーゴールド」:
<https://www.youtube.com/watch?v=0xSiBpUdW4E>

生成された副旋律に対して、副旋律の音高として選ばれた短 3 度、長 3 度、完全 4 度の割合を述べ、考察する。

3.2 実験結果

表 3 提案手法において副旋律の音高として選ばれた音名の分布

タイトル	長 3 度	短 3 度	完全 4 度	総数
秦基博 「ひまわりの約束」	16448	19722	7579	63372
テレサテン 「つぐない」	9710	17203	5112	45342
あいみょん 「マリーゴールド」	14341	16831	11216	64308

副旋律として選ばれた音程 (長 3 度、短 3 度、完全 4 度) の分布を表 3・表 4 に示す。集計せずに一部分を時間軸上

表 4 従来手法において副旋律の音高として選ばれた音名の分布

タイトル	長 3 度	短 3 度	完全 4 度	総数
秦基博 「ひまわりの約束」	7719	17692	18338	63372
テレサテン 「つぐない」	7365	14637	10023	45342
あいみょん 「マリーゴールド」	10309	11331	20748	64308

に色で図示したものを図 1 に示す。

従来手法を用いたすべての楽曲において、選ばれた音程は完全 4 度が一番多かった。対して、提案手法を用いたすべての楽曲においては、短 3 度が一番多く選ばれており、反対に完全 4 度が一番選ばれた数が少なかった。

ただし、従来手法において完全 4 度はどの楽曲でも 5000 フレーム以上選ばれており、特定の音程への過度な偏りは見られなかった。

提案手法で完全 4 度が一番選ばれた数が少なくなった理由として、遷移確率の値が長 3 度・短 3 度と比べて低いことが影響していると考えられる。

和声学的な観点で言えば、従来手法の際に発生していたロングトーン部分における不自然な音高の切り替わりが提案手法ではほとんど見られなかった。

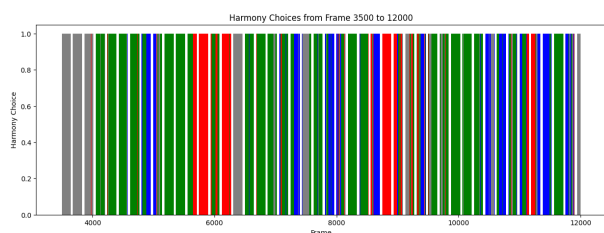
特に、ロングトーン部分を抽出した「つぐない」では、提案手法 (図 1 の (c)) では、短 3 度で安定したロングトーンを構成できていることがわかるが、同じ箇所の従来手法 (図 1 の (d)) では、短 3 度で安定せず、長 3 度や完全 4 度に切り替わってしまい不安定であることがわかる。

4. おわりに

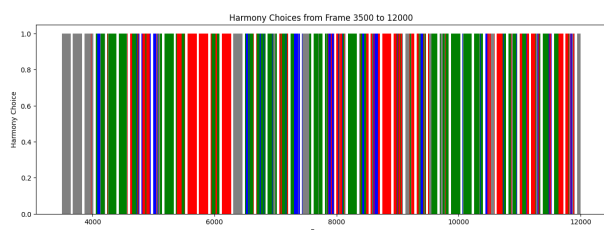
本稿では、ユーザが用意した歌唱音源を用いていわゆるハモリパートの歌唱練習ができるシステムの実現をめざし、その第 1 段階として、副旋律の自動生成における課題を解決する新しい手法を提案した。本研究では、主旋律の基本周波数 (F_0) を推定し、それに基づいて適切な副旋律を生成する際の精度向上を図った。

従来手法 [5] ではロングトーン部分のピッチ不安定性などの問題があったが、本稿では、隠れマルコフモデル (HMM) を用いることでこれらの問題に対処した。具体的には、主旋律の F_0 が変化しない区間におけるピッチの安定性を高めるための遷移確率行列の工夫や、伴奏音に不協和な副旋律を避けるための出力確率の調整を行った。これにより、生成された副旋律が音楽的に適切であり、かつロングトーン部分での安定性を確保することが可能となった。

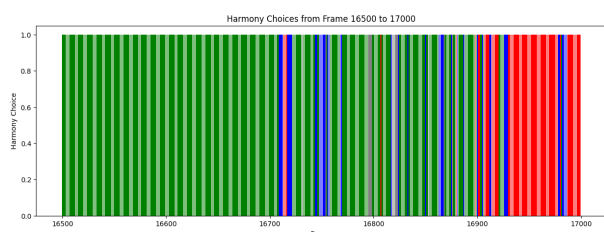
とはいえ、副旋律の音高を決める処理にはまだ改善の余地がある。現在の手法では、曲の転調に対応することが難しい。そのため、転調に対応しながら副旋律生成の質を上げることが出来る工夫が求められる。さらに、練習支援システムとしてのユーザインタラクションの設計も重要であ



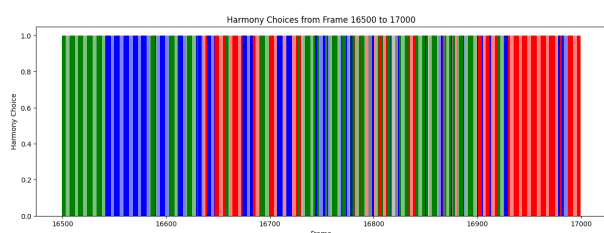
(a) 提案手法における「ひまわりの約束」



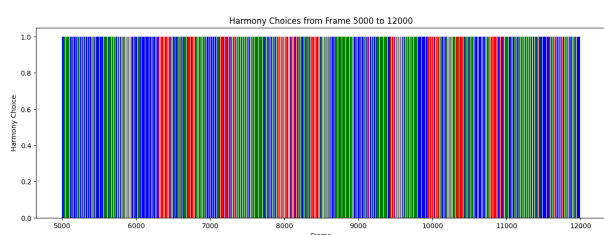
(b) 従来手法における「ひまわりの約束」



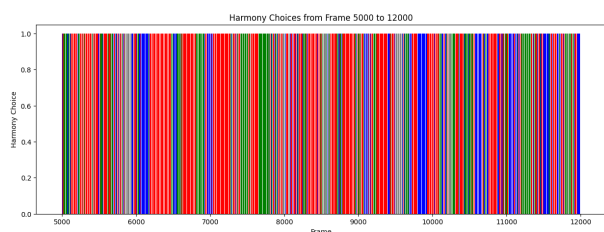
(c) 提案手法における「つぐない」



(d) 従来手法における「つぐない」



(e) 提案手法における「マリーゴールド」



(f) 従来手法における「マリーゴールド」

図 1 副旋律の音高として選ばれた副旋律に対する音程の候補（青：長3度，緑：短3度，赤：完全4度，灰色：どれにも該当せず）

る。主旋律につられないように副旋律を練習する方法などを検討していく必要がある。これらの検討を進め、誰もが気軽にハモリ練習ができるシステムを目指していきたい。

謝辞

本研究は、JSPS 科研費 23K24966, 24H00748 の助成を受けた。

参考文献

- [1] 白石 美南, 小笠原 梢, 北原 鉄朗, カラオケのためのハモリパート練習システム～ハモリパートの自動生成および練習支援システムの試作～, 情報処理学会 第 80 回全国大会講演論文集, vol.2018, No.1, pp.127–128, 2018.
- [2] 三浦 雅展, 青山 容子, 谷口 光, 青井 昭博, 尾花 充, 柳田 益造, ポップス系の旋律に対する和音付与システム:AMOR, 情報処理学会論文誌, vol.46, No.5, pp.1176–1187, 2005.
- [3] 山崎 健一, 坂 知樹, 鎌田 洋, 深層学習を用いた主旋律に基づく和音生成, 映像情報メディア学会誌, vol.77, No.1, pp135–140, 2023.
- [4] Jianhua Li, Automatic Piano Harmony Arrangement System Based on Deep Learning, Sensors, vol.2022, Article ID 76624432022, 2022.
- [5] 福田 康太, 川原 未波, 北原 鉄朗, ハモリパート練習支援システムのための音響信号を対象とした副旋律生成の検討, 情報処理学会 インタラクシオン 2024 論文集, 2024