

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

福田 康太^{1,a)} 川原 未波^{1,b)} 北原 鉄朗^{1,c)}

概要: 本稿では、音楽知識が乏しくても簡単に副旋律を練習することが出来るシステムの実現を目指し、与えられた音源に対して副旋律の歌唱を生成する処理を試行する。歌唱曲音源が与えられると、歌唱（主旋律）と伴奏とに分離し、伴奏音源からクロマグラムを取得する。その取得したクロマグラムから副旋律の候補の選定・候補の絞り込みを行い、音響信号を生成する。数曲の音源に対して副旋律の生成を試したところ、和声学的観点として、多くの箇所では妥当性のある副旋律が生成された。しかし、副旋律の音高が何度も入れ替わる箇所がノイズのように聞こえるという問題点も見られた。

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。

本稿では、このような支援システムを実現する第一歩と

して、音響信号として与えられた歌唱曲から歌唱のみの音響信号を抽出し、ピッチシフトにより副旋律の音響信号を生成する手法を検討する。

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）と伴奏とに分離する。次に、伴奏音源からクロマグラムを取得する。クロマグラムは和声（コード進行）が反映されるため、これを考慮して副旋律の音高を決めて、音響信号を生成する。

2.1 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「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。

2.2 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「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。

2.3 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

¹ 日本大学文理学部情報科学科

^{a)} fukuda@kthrlab.jp

^{b)} minami@kthrlab.jp

^{c)} kitahara@kthrlab.jp

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「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。時刻 t における平滑化後のクロマベクトルを $\mathbf{h}(t) = (h_1(t), \dots, h_{12}(t))$ で表す。

2.4 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。

2.5 副旋律の候補の絞り込み

副旋律の候補 $C(t)$ のうち、対応する音名のクロマベクトルの値が最も大きいものを、副旋律の最終候補として採用する。これは、クロマベクトルの値が大きい音名は、その時刻における和音の構成音である可能性が高く、そのような音名を選ぶことで、伴奏との不協和を避けられる可能性が高いと考えられるからである。

副旋律の候補 $C(t)$ の各要素 $c_i(t)$ ($i = 1, 2, 3$) に対して、対応する音名 $n_i(t)$ を次の式で求める。

$$n_i(t) = \text{mod}(12 \log_2(c_i(t)/440) + 9, 12)$$

ここで、 $\text{mod}(x, m)$ は x を整数に丸めた上で m で割った余りを求めることを意味する。その上で、 $h_{n_1(t)}(t)$, $h_{n_2(t)}(t)$, $h_{n_3(t)}(t)$ が最大となる $n_i(t)$ を求め、それに対応する $c_i(t)$ を時刻 t における副旋律の音高 $y(t)$ とする。

2.6 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、伴奏を重畳した音響信号も生成する。

3. 予備実験

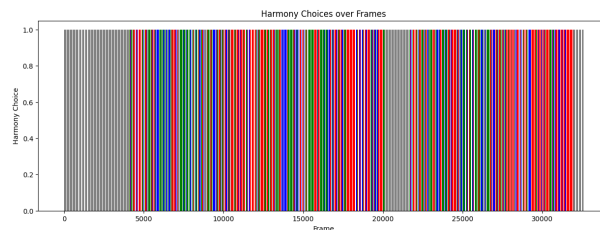
3.1 実験方法

本手法を用いて、3つの音源を対象に副旋律生成を行った。対象の音源は以下の3曲である。

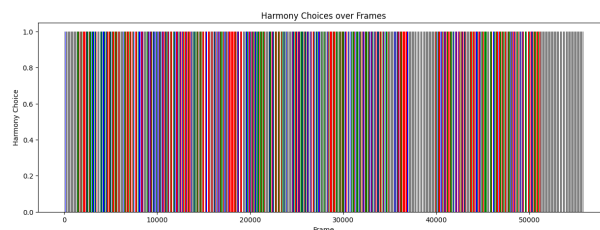
- となりのトトロ 「さんぽ」：
https://www.youtube.com/watch?v=k_a8vsgeg_k
- DISH// (北村匠海) 「猫」：
<https://www.youtube.com/watch?v=gsT6eKsnTOM>
- いきものがかり 「ブルーバード」：
<https://www.youtube.com/watch?v=KpsJWFuVTdI>

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。

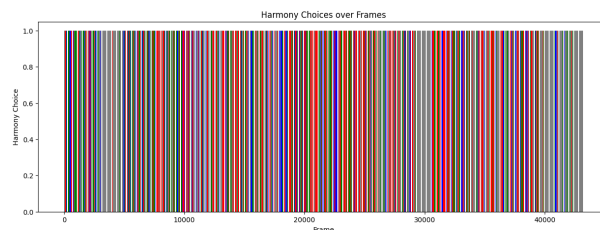
3.2 実験結果



(a) となりのトトロ 「さんぽ」



(b) DISH// (北村匠海) 「猫」



(c) いきものがかり 「ブルーバード」

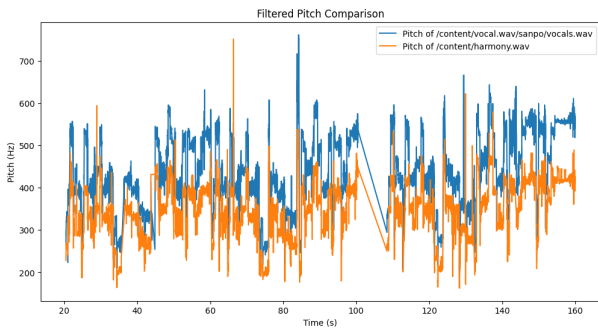
図 1 副旋律の音高として選ばれた副旋律に対する音程の候補（青：長 3 度、緑：短 3 度、赤：完全 4 度、灰色：どれにも該当せず）

表 1 副旋律の音高として選ばれた音名の分布

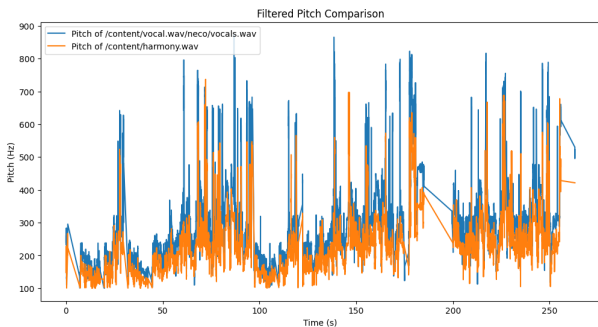
タイトル	長三度	短三度	完全 4 度	総数
となりのトトロ 「さんぽ」	6006	8215	9868	32587
DISH// (北村匠海) 「猫」	7561	14330	14711	55833
いきものがかり 「ブルーバード」	5115	8718	11455	43224

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 1 に示す。集計せずに時間軸上に色で図示したものを図 1 に示す。主旋律と副旋律の基本周波数を時間軸上に色で図示したものを図 2 に示す。全ての曲において、選ばれた音程は完全 4 度が最も多く、それ以外は短 3 度、長 3 度の順に多かった。ただし、長 3 度が 5000 フレーム以上で選ばれており、特定の音程への過度な偏りは見られなかった。

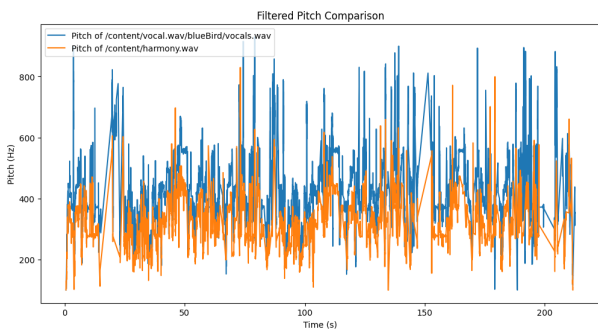
和声学的観点からは、多くのフレームでコード進行を考



(a) となりのトトロ 「さんぽ」



(b) DISH// (北村匠海) 「猫」



(c) いきものがかり 「ブルーバード」

図 2 主旋律と副旋律の周波数 (青：主旋律の周波数，オレンジ：副旋律の周波数)

慮した副旋律になっていたと言える。ただし、いくつかの箇所において和声学的に不自然な音高も見られた。

また、ノイズ成分については、フレームごとに音程を決めているため、主旋律の1つの音符に対して、副旋律の音高が途中で何度も入れ替わることがあった。このような場合には、ノイズ成分のように聞こえることもあった。

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、音響信号に対して音源分離とピッチシフトを用いて副旋律を生成することを試行した。少しノイズ成分が混じりながらも全体的には、多くの箇所において和声的に妥当な副旋律を生成することができた。

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。HamoKara [1] などでも使われている隠れマルコフモデルなどを用いて、調性などを考慮しつつ副旋律の音高を決めていく工夫が求められる。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、科研費 22H03711, 21H03572 の支援を受けた。

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.