

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～

白石 美南^{a)} 小笠原 梢^{a)} 北原 鉄朗^{a)}

^{a)} 日本大学 文理学部 情報科学科

1. はじめに

カラオケの楽しみ方の1つに、主旋律と副旋律を2人で歌ってハーモニーを奏でることがあげられる。しかし、副旋律(いわゆるハモリパート)の歌唱を習得するのは簡単ではない。まず、ハーモニーを奏でたい楽曲の原曲に副旋律が存在しないとき、副旋律を自分で考える必要がある。しかし、副旋律を自分で考えるのは、作編曲の経験がないような一般の歌唱者には困難な作業である。また、副旋律は主旋律に比べて馴染みが薄いため、主旋律に合わせて歌ったときに主旋律につられてしまう可能性がある。

本稿では、これらの問題を解決するハモリパート練習支援システムを提案する。本システムでは、副旋律をカラオケ用のMIDIデータから自動的に生成する。主旋律に対して和声付けを行う研究は、これまで多数行われてきたが、そのほとんどが伴奏なし主旋律にコード進行や副旋律を付与するもの¹⁾か、主旋律に複数の声部を付与して3声部や4声部の和声を生成するもの^{2)~5)}であった。本研究では、主旋律および伴奏がすでにある状態で副旋律を生成する。これを確率ベースおよびルールベースの2通りのアプローチで実現する。また、主旋律につられない練習支援としては、マイクから入力された歌唱者の音声から音高(F0)を推定して画面に表示することで、本来の音高とのずれを歌唱者が自覚できるようにし、正しい音高での歌唱を促す。

2. システム概要

本システムは、ユーザが選択した楽曲に対して副旋律を生成し、音とグラフィクスにより提示することで、副旋律の練習を支援するシステムである。ユーザが楽曲を選択すると、その楽曲の歌詞情報が埋め込まれたMIDIデータが読み込まれる。まず、MIDIデータを分析して副旋律を生成する。(生成手法は3章で述べる)この際、副旋律担当の歌唱者の音域に合わせるため、副旋律を主旋律よりも高くする(いわゆる「上でハモる」)か低くする(いわゆる「下でハモる」)かを選ぶことができる。その後、MIDIデータの再生を開始する。主旋律と副旋律がガイドメロディとしてMIDI音源から演奏され、画面のピアノロール上に表示される。MIDIデータの演奏中にマイクから歌唱を入力すると、その音高を推定してピアノロール上に表示する。これにより、本来の音高と歌唱時の音高がどの程度一致しているかを視覚的に確認できる。その一致度が一定の値を下回ったときは「もっと自分の声を聴きましょう」とリアルタイムに表示される。

3. 副旋律の自動生成手法

本システムでは、MIDIデータが読み込まれると、副旋律が自動的に生成される。副旋律は主旋律と同じ歌詞で歌うことに鑑み、副旋律は主旋律と同じ音符数およびリズムであるものとする。すなわち、主旋律と副旋律は音符単位で1対1で対応する。また、楽曲にメリハリをつけるために、楽曲の一部分(たとえばサビ)のみに副旋律を付けるのが望ましい場合も存在するが、実際のカラオケ場面で提示された副旋律

を歌わないのは難しくないため、システム上では主旋律の最初から最後までに副旋律を付与する。以下、副旋律生成手法として、隠れマルコフモデル(HMM)を用いた手法とルールベースによる手法を述べる。

3.1 HMMによる副旋律自動生成手法

主旋律の音符列を n_1, \dots, n_m 、副旋律の音名列を h_1, \dots, h_m とする。これらは0以上11以下の整数を取り、各々の値が音名(C, C#, ..., B)に対応する。主旋律の各音符 n_i の発音区間における伴奏の特徴を多次元ベクトル b_i で表し、当該楽曲の調を k とする。このとき、求めるべきものは、

$$P(h_1, \dots, h_m | n_1, \dots, n_m, b_1, \dots, b_m, k) \quad (1)$$

が最大になる h_1, \dots, h_m である。副旋律のマルコフ性を仮定し、事前に与えられる主旋律および伴奏における依存関係を省略すれば、式(1)は

$$P(h_1)P(h_1|k)P(n_1|h_1)P(b_1|h_1) \dots P(h_m|h_{m-1})P(h_m|k)P(n_m|h_m)P(b_m|h_m) \quad (2)$$

と書くことができる。これは、副旋律における各音名 h_i を状態とし、主旋律 $\{n_i\}$ および伴奏 $\{b_i\}$ を観測系列としたHMMに相当する。ただし、状態遷移確率は調 k に依存する。このようにモデル化すると、主旋律 $\{n_i\}$ と伴奏 $\{b_i\}$ が与えられたときに副旋律 $\{h_i\}$ を付与する処理は、このモデルに対してViterbiアルゴリズムを適用することで実現できる。

以下、各確率の計算方法について述べる。

- $P(h_1)$ は初期確率であり、等確率とする。
- $P(h_i|k)$ は調 k が与えられたときの各音名の出現確率であり、 h_i が調 k におけるダイアトニックノートで、かつ主旋律の音名に対して一度、三度、五度の関係にあるときには0.15、それ以外のダイアトニックノートの場合は0.1、ダイアトニックノートでないときは0.03とする。
- $P(h_i|h_{i-1})$ は副旋律における音名の遷移確率である。これは、副旋律は主旋律と類似した特徴を持つべきだという観点から、当該楽曲の主旋律から音名の遷移確率を計算し、これを用いる。
- $P(n_i|h_i)$ は主旋律と副旋律の関係を表す。最も典型的な方法は主旋律の三度下に副旋律を付けることである。そこで次式により定義する。

$$P(n_i|h_i) = \begin{cases} 0.014285714 & |n_i - h_i| = 0 \text{ のとき} \\ 0.028571429 & |n_i - h_i| = 1 \text{ のとき} \\ 0.064285714 & |n_i - h_i| = 2 \text{ のとき} \\ 0.214285714 & |n_i - h_i| = 3 \text{ のとき} \\ 0.285714286 & |n_i - h_i| = 4 \text{ のとき} \\ 0.214285714 & |n_i - h_i| = 5 \text{ のとき} \\ 0.071428571 & |n_i - h_i| = 6 \text{ のとき} \\ 0 & |n_i - h_i| \geq 7 \text{ のとき} \end{cases}$$

- $P(b_i|h_i)$ は副旋律と伴奏の関係を表す。伴奏に副旋律と短2度や長2度などの関係の音が含まれると、不協和音の原因となる。そこで、次のように定義する。 b_i を12次元ベクトル $b_i = [b_{i,0}, \dots, b_{i,11}]$ (各要素は互いに独立)とする。 $b_{i,j}$ は、主旋律の音符 n_i の発音時刻から消音時刻までの区間において、対応する音名が発音している時間の割合を表す。たとえば、 n_i が二分音符で、その区間の伴奏にいたなら、 $b_i = [1.0, 0.0, 0.0, 0.0, 0.5, 0.0, 0.0, 0.5, 0.0, 0.0, 0.0, 0.0]$ となる。副旋律に対して短2度や長2度の伴奏音が含まれるの

A Support System for Practicing Harmonized Singing in Karaoke: Automatic Generation of Harmonized Voice Part and Implementation of the Prototype System
by Mina Shiraiishi, Kozue Ogasawara and Tetsuro Kitahara (Nihon University)

表 1 副旋律生成結果の精度

	HMM			ルールベース		
	平均	最大	最小	平均	最大	最小
上	52.3%	64.5%	24.3%	55.5%	75.0%	33.6%
下	53.6%	71.0%	30.5%	64.2%	87.9%	37.6%

を防ぐには、 $|h_i - j| \leq 2$ (ただし $h_i \neq j$) を満たす j に対して $b_{i,j}$ ができるだけ 0 に近づけばよい。そこで、 $|h_i - j| \leq 2$ かつ $h_i \neq j$ のとき、 $P(b_{i,j}|h_i)$ が正規分布 $N(0.0, \sigma_1^2)$ (σ_1 は実験的に定めた定数) に従うとする。それ以外の j に対しては $P(b_{i,j}|h_i)$ は一様分布に従うとする。ただし、実装上の都合により、 σ_2 を十分に大きな定数として正規分布 $N(0.5, \sigma_2^2)$ で一様分布を近似している。

3.2 ルールベースによる副旋律自動生成

次の処理により副旋律を生成する。

- (1) 楽曲を半小節ごとに分割して伴奏パートの分析を行う。分析の結果得られるベクトルの系列 b_1, \dots, b_T の内容は前節と同じであるが、主旋律のリズムと同期しておらず、半小節ごとである点が異なっている。
- (2) 主旋律の各音符 n_i に対して次の処理を行う。
 - (a) 音符 n_i に時刻が対応する要素を $\{b_1, \dots, b_T\}$ から取り出す (b_t とする)。
 - (b) $b_t = (b_{t,0}, \dots, b_{t,11})$ から上位 4 つを取り出し、それらの音名を取得する (取得した 4 つの音名の集合を C_i とする)。
 - (c) C_i に主旋律より長三度下, すなわち $n_i - 4$ が含まれていれば, その音名 $h_i = n_i - 4$ を副旋律の音名とする (「下でハモる」の場合)。
 - (d) C_i に主旋律より短三度下, すなわち $n_i - 3$ が含まれていれば, その音名 $h_i = n_i - 3$ を副旋律の音名とする。
 - (e) C_i に主旋律より完全四度下, すなわち $n_i - 5$ が含まれていれば, その音名 $h_i = n_i - 5$ を副旋律の音名とする。
 - (f) C_i に主旋律より長三度下, 短三度下, 完全四度下のいずれの音名も含まれていない場合, 直前の音名から主旋律と並行に移動するように副旋律の音名を定める。すなわち, $h_i = h_{i-1} + (n_i - n_{i-1})$ とする。「上でハモる」の場合も同様に定義する。

4. 副旋律評価

提案手法によって生成した副旋律がどの程度適切か検証する。実験に使用する楽曲数は 85 曲で、音楽大学生に作成してもらったデータと比較する。生成された副旋律データと正解データがどの程度の割合で一致したかを検証する。

4.1 評価結果

HMM, ルールベースでそれぞれ生成された副旋律の正解との一致率 (以下精度とする) を表 1 に示す。上下ともに、HMM で生成した副旋律よりルールベースで生成した副旋律のほうが平均精度が高い結果となった。

4.2 副旋律評価考察

HMM で生成した副旋律がルールベースよりも精度が下がった理由はいくつか考えられる。

まず第 1 に、HMM で生成した副旋律には、ユニゾンが多く含まれていたことがあげられる。正解データ中に含まれるユニゾンの割合は、全曲を平均して 0.20% であった一方、HMM で生成した副旋律では 4.78% であった。そのうち、正解データと一致するユニゾンを生成できた割合は 1.22% であった。また、正解データ中に含まれるユニゾンの特徴として、これらは休符前後に多く見られる傾向にあったが、HMM で生成した副旋律は休符の有無に関わらずユニゾンを生成していたため、精度が下がってしまったと考えられる。ルールベースで生成する副旋律では、ユニゾンが生成されないが、

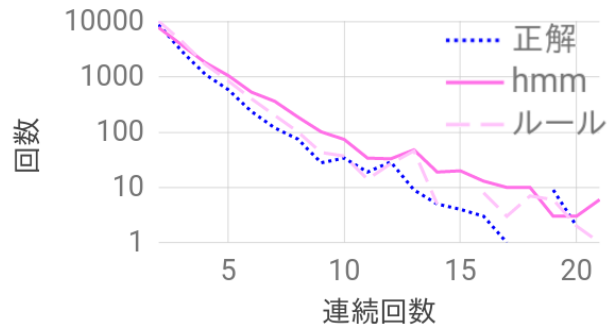


図 1 85 曲中で同じ音名が連続する箇所の出現回数

正解データ中のユニゾンの割合が少ないため、精度への影響が少なかったと考えられる。

第 2 に、HMM で生成した副旋律は、主旋律の音名の前後関係による影響が大きかったことがあげられる。図 1 は正解データ、ルールベース、HMM で生成した副旋律のうち、連続で同じ音に移した回数を表す。ここから、HMM で生成した副旋律は、特に連続で同じ音に移す回数が多く、本来副旋律中に高低差をつけなければいけない部分でも、同じ音が連続で続くことになってしまったため、精度が下がったと考えられる。

HMM で生成した副旋律が、同じ音に移す回数が多くなってしまった理由としては、利用する遷移確率の影響があると考えられる。同じ音が 10 音以上連続で続いた曲 (30 曲) の遷移確率に着目すると、「下でハモる」場合では、 $A\sharp$ を除いた 11 種類の音名において、「上でハモる」場合では、 $C\sharp, F$ を除いた 10 種類の音名において、同じ音名に移す確率が他の音名に移す確率が高かった。一方、それ以外の 55 曲の遷移確率に着目すると、「下でハモる」場合では、 $D\sharp, E, A, A\sharp, B$ の 5 種類の音名において、「上でハモる」場合では、 $C\sharp, D, D\sharp, E, G, G\sharp$ の 6 種類の音名において、同じ音名に移す確率が他の音名に移す確率が高かった。これらのことから、HMM で副旋律を生成するにあたって、主旋律の音名の前後関係から算出した遷移確率の影響が大きかったためだと考えられる。

5. おわりに

本稿では、カラオケでのいわゆる「ハモリ」(副旋律)を練習するためのシステムを提案した。本システムには副旋律を自動生成する機能があり、HMM, ルールベースの両手法において一定の精度が確認できた。今後は、実装した練習システムをユーザに使用してもらい、有効性についても検証していく予定である。

謝辞 本研究は、JSPS 科研費 16K16180, 16H01744, 16KT0136, 17H00749 から支援を受けた。

参考文献

- 1) 菅原 啓太 *et al.*: HMM と音符連鎖確率を用いた旋律への自動和声付け, 情報処理学会研究報告音楽情報科学 (MUS), 2003 巻, 127 号 (2003-MUS-053), pp. 49-54, 2003.
- 2) エバンズ ベンジャミン ルカ *et al.*: 和声学に基づく合唱譜自動作成システム, 和声学に基づく合唱譜自動作成システム "CMY" 転回系の実装と評価, 情報処理学会研究報告, 2013 巻, 37 号, pp. 1-4, 2013.
- 3) 鈴木 峻平 *et al.*: ベイジアンネットワークを用いた四声体和声付け:コードノードの有無による出力結果の比較, 第 76 回情報処大会, 2014 巻, 1 号, pp. 311-312, 2014.
- 4) Yamada, T. *et al.*: Four-part Harmonization: Comparison of a Bayesian Network and a Recurrent Neural Network, *Proc. of CMMR*, pp. 137-148, 2017.
- 5) Hild, H. *et al.*: HARMONET: A Neural Net for Harmonizing Chorales in the Style of J.S.Bach, *Advances in Neural Information Processing 4*, pp. 267-274, 1991.