

## 統計的アプローチに基づく即興演奏補正

石田 克久<sup>†</sup>北原 鉄朗<sup>‡</sup>柳川 貴央<sup>†,\*</sup>武田 正之<sup>†</sup><sup>†</sup>東京理科大学理工学部情報科学科<sup>‡</sup>京都大学大学院情報学研究科知能情報学専攻

\*現在, ヤマハ株式会社

## 1. はじめに

即興演奏は、旋律創作と演奏を同時に行う難しい演奏形態である。そのため、音楽的に自然な旋律を創作しながら演奏するには、高度な技術と経験を要する。そこで我々は、即興演奏において音楽的に不自然な音を自然な音にリアルタイムに補正するシステムについて、研究を進めている<sup>1)</sup>。

このようなシステムの実現は、非常に困難な課題である。なぜなら、演奏された音が「音楽的に不自然」かの判断には、主観が伴うからである。そのため、この課題を扱った研究例は少ない。“Coloring-in Piano”<sup>2)</sup>は、演奏中に間違った音を正しい音に補正するシステムだが、演奏する曲の音高をあらかじめ入力しておく必要がある。“音機能固定マッピング楽器”<sup>3)</sup>は、楽器のインターフェースを従来の音高ではなく機能別に配置した新楽器であり、本研究とは問題設定が異なる。“INSPIRASTION”<sup>4)</sup>は、アヴェイラブルノートスケールから外れる音(アウト音と呼ぶ)をすべて補正する。しかし、アウト音が常に「音楽的に不自然」とは限らないため、アウト音をすべて補正するのは必ずしも適切ではない。

そこで我々はすでに、打鍵されたアウト音を補正すべきかどうかを、統計的に決定する手法を提案した<sup>1)</sup>。この手法では、旋律データベース(既存の楽曲の旋律、コード、調を多数収録したもの)からさまざまな条件下でのアウト音の使用頻度を算出し、使用頻度の低いアウト音のみ補正する。しかし、条件として直前の音しか考慮されていなかった。また、定量的な評価をするには至らなかった。本稿では、N-gramを採用することで、より大局的な旋律の流れを考慮できるよう手法を拡張する。さらに、実際の即興演奏データに補正の必要な箇所を手でラベルづけしたものをを用いて、補正精度を定量的に評価する。

## 2. 即興演奏の補正手法

演奏補正における主要な課題は、補正すべき音(補正対象と呼ぶ)をどのように決定するかである。この課題に対する1つの解決法は、アウト音をすべて補正対象とすることである。なぜなら、アウト音は、伴奏ときれいなハーモニーを形成しにくいとされているからである。しかし、すべてのアウト音が伴奏ときれいなハーモニーを形成しないわけではなく、むしろすべてのアウト音を補正することは、表現の幅を過度に狭めることにつながる。

本手法では、打鍵されたアウト音を補正すべきかどうかを、N-gramによる旋律モデルに基づいて決定する(図1)。まず、旋律を特徴抽出とN-gramによりモデル化する。そし

Statistics-based Performance Correction for Improvisation  
by Katsuhisa Ishida<sup>†</sup>, Tetsuro Kitahara<sup>‡</sup>, Takahiro Yanagawa<sup>†</sup>  
and Masayuki Takeda<sup>†</sup> (<sup>†</sup>Tokyo Univ. of Science, <sup>‡</sup>Kyoto Univ.)  
それぞれのコードに適している音で構成されるスケール<sup>5)</sup>。

表1 特徴ベクトルの各要素(かっこ内は取りうる値)

・対象音の種類(chord, key, else)
・対象音と直前の音との音高差(短2度, 長2度, 短3度以上)
・対象音の発音時刻が8分音符レベルで表か裏か(True, False)
・対象音とその直前の音との間に休符があるか(True, False)

て、演奏された旋律に対応するN-gram確率を、あらかじめ用意された旋律データベースから求め、この確率値に基づいて補正すべきかを決定する。

## 2.1 特徴抽出

旋律の各音の特徴を特徴ベクトルとして表現する。現在の実装で用いている特徴ベクトルを表1に示す。本稿では、特徴ベクトルが $x$ で表される音を「音 $x$ 」と表す。

## 2.2 N-gramによる旋律のモデル化

与えられた旋律の次にどのような音が来やすいかを数量的に表すため、旋律をモデル化する。このモデルは、旋律 $X = x_1 \cdots x_{n-1}$ の次に音 $x_n$ が続く確率 $P(x_n|X)$ を与えるモデルと考えることができる。ここでは、 $x_n$ がその直前の $N-1$ 個の音 $x_{n-N+1} \cdots x_{n-1}$ に依存して決められると考え、

$$P(x_n|X) = P(x_n|x_{n-N+1} \cdots x_{n-1}) \\ = \frac{P(x_{n-N+1} \cdots x_n)}{P(x_{n-N+1} \cdots x_{n-1})}$$

と定義する。これは、さまざまな旋律の出現確率をN-gramを用いてモデル化したことに相当する。

## 2.3 旋律のモデルに基づく補正対象の決定

即興演奏において旋律 $X$ の後にアウト音 $x_n$ が打鍵されたとき、その音が自然かどうかは、旋律データベースから求めたN-gram確率 $P(x_n|X)$ で表される。なぜなら、この値が高いということは、実在する旋律でも $X$ の後に $x_n$ が続くことがよくある、ということを示しているからである。そこで、この値がしきい値より低いとき、 $x_n$ を補正対象とする。

## 2.4 補正後の音高の決定

アウト音 $x_n$ が補正対象となると、アウト音 $x_n$ の音高をさまざまな音高(ただし非アウト音)に変更したときの $P(x|X)$ を求め、この値が最大になる音高を補正後の音高とする。

## 3. 実装・実験

## 3.1 実装

提案手法をWindows上でC言語を用いて実装した。旋律データベースは、スタンダードジャズの楽譜集<sup>6)</sup>の全曲(208曲)のメロディとコード名を入力して作成した。総小節数は6,836小節、総音符数は18,897音である。また、しきい値は0.10とした。

本システムでは、伴奏用データはあらかじめMIDIファイルとして用意する。ユーザは、この伴奏用データ(コード情報含む)の再生に合わせ、MIDIキーボードで単音のみの即

表3 実験結果

	全 体			初 心 者			中 級 者			上 級 者		
	再現率	適合率	F 値									
全 補 正	0.7822	0.3636	0.4964	0.7005	0.4242	0.5307	0.9123	0.5131	0.6568	0.7072	0.2012	0.3133
文献 1) の手法	0.5048	0.5565	0.5294	0.3975	0.5652	0.4667	0.6471	0.7082	0.6763	0.4283	0.3356	0.3763
提案手法 (N = 2)	0.7737	0.4977	0.6057	0.6628	0.5066	0.5743	0.9099	0.6622	0.7665	0.7072	0.2985	0.4198
提案手法 (N = 3)	0.7682	0.4982	0.6044	0.6190	0.5078	0.5579	0.8969	0.6585	0.7594	0.7072	0.3032	0.4244

表2 被験者とラベルの詳細

	人数	小節数	総音符数	要補正音
初心者	10 人	64 小節/人	3,108 音	12.03%
中級者	15 人	64 小節/人	3,177 音	8.22%
上級者	12 人	64 小節/人	2,660 音	3.38%
全 体	37 人	64 小節/人	8,945 音	8.11%

初心者：演奏経験 1 年未満，即興演奏経験なし

中級者：演奏経験 3～5 年程度，即興演奏経験なし

上級者：演奏経験 5 年以上，もしくは即興演奏経験あり

興演奏を行う。そうすると，MIDI 音源からは提案手法に基づいて補正された音が発音される。

### 3.2 実験方法

補正候補の決定が適切かどうかについて実験する。まず，37 人の被験者に即興演奏をしてもらい，その演奏データの補正すべき箇所を手でラベルづけする（「補正すべき」とラベルづけられた音を「要補正音」と呼ぶ）。そして，提案手法を適用して補正し，補正候補決定が適切になされたかを再現率，適合率，F 値の観点から評価する：

$$\text{再現率} = \frac{\text{補正された音のうち要補正音の個数}}{\text{要補正音の総数}}$$

$$\text{適合率} = \frac{\text{補正された音のうち要補正音の個数}}{\text{補正された音の総数}}$$

$$F \text{ 値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}}$$

なお，補正は，提案手法だけでなく，アウト音をすべて補正する手法（全補正と呼ぶ），文献 1) の手法でも行い，比較する。また，被験者とラベルの詳細を表 2 に示す。

### 3.3 実験結果と考察

実験結果を表 3 に示す。全体で，提案手法の F 値が，全補正に比べて  $N = 2$  (bi-gram) のときに 0.1093， $N = 3$  (tri-gram) のときに 0.1080 向上した。また，文献 1) の手法と比べると， $N = 2$  (bi-gram) のときに 0.0763， $N = 3$  (tri-gram) のときに 0.0750 向上した。これにより，提案手法の補正対象決定処理は，全補正，文献 1) の手法より適切と言える。

再現率・適合率で分けて考えると，提案手法は，全補正に比べて再現率が 1～2% 下がり，適合率が 13% 程度向上した（全体の場合）。これは，補正の取りこぼし（要補正音を補正しないこと）を最小限に押さえながら，過補正（補正する必要のない音を補正すること）を考慮できたと考えられる。文献 1) の手法では，高い適合率が得られる代わりに，再現率が大幅に低下した。それに対し，本手法では，再

現率の低下を最小限に押さえることが出来た。

中級者の補正精度は，全手法を通して高かった。中級者は，アヴェイラブルノートスケール内の音を使えば一応自然な旋律ができるということを経験的に知っている人が多い。そのため，打鍵ミスでアウト音を弾く人が少なからずいた（それに対して，上級者は狙って，初心者はわけも分からずアウト音を弾くことがあった）。本手法は，このような打鍵ミスによるアウト音を精度良く検出できたと考えられる。

一方，上級者の補正の精度は，全手法を通してあまり良くなかった。これは，上級者のなかにクラシック音楽の音楽経験者が多かったためと考えられる。すなわち，上級者の演奏がクラシック音楽風の旋律になっており，本システムが持つ旋律データベースとは旋律の傾向が一致しなかったからと考えられる。また，上級者の演奏には，補正すべきか迷うような音もいくつかあった。今後は，同一演奏を複数人でラベルづけし，より詳細に評価していくことも必要である。

本稿で扱った問題は，自然な旋律・不自然な旋律は何かという音楽における根本的な問題を含んでいる。我々は本実験により，この問題に対して統計的アプローチがある程度有効に働くことが確認できたと考えている。

## 4. おわりに

本稿では，補正対象を統計的アプローチに基づいて決定する即興演奏補正システムを提案した。旋律を N-gram でモデル化し，旋律の出現確率に基づいて補正対象を決定することで，適切な補正を可能にした。

ただし，本稿の実験では補正対象決定の評価にとどまり，補正後の旋律のなめらかさやシステムの使いやすさの評価は行っていない。今後は，これらの評価を通じてシステムをより洗練させるとともに，より大規模なデータベースの整備を行う。また，ジャンルごとや演奏者ごとにデータベースを構築することで，ジャンルの特性や演奏者の手癖などを考慮した演奏補正も検討していく。

謝辞 渡辺義大氏をはじめ，実験にご協力頂いた全ての方々に感謝する。

## 参 考 文 献

- 1) 柳川 他：“即興演奏における演奏補正システム”，情報処理学会第 64 回全国大会，1L-5，2002。
- 2) 大島 他：“Coloring-in Piano：表情付けに専念できるピアノの提案”，情処学論，2001-MUS-42，pp.69-74，2001。
- 3) 西本 他：“創造的音楽表現を可能とする音楽演奏支援手法の検討—音機能固定マッピング楽器の提案”，情処学論，39，5，pp.1556-1567，1998。
- 4) 谷井，片寄：“音楽知識と技能を補うピアノ演奏システム INSPIRATION”，情処学論，43，2，pp.256-259，2002。
- 5) 岩田 他：“定番実用音楽事典”，ドレミ楽譜出版社，2001。
- 6) 伊藤：ザ・プロフェッショナルスタンダード・ジャズハンドブック，中央アート出版社，1992。

ただし，文献 1) で導入されていた「前補正」は行わずに評価した。これは，予備実験の結果，この前補正で行われていた音高の推移方向の保存が，必ずしも有効ではなかったからである。