

テクニカルノート

N-gram による即興演奏の旋律補正

石 田 克 久[†] 北 原 鉄 朗^{††} 武 田 正 之[†]

本稿では、即興演奏における不自然な旋律を自動的に補正するシステムについて述べる。旋律補正の一アプローチとしてアヴェイラブルノートスケールから外れる音(アウト音)を補正対象とする方法が考えられるが、アウト音が必ずしも不自然とは限らない。本研究では、アウト音が補正されるべきか否かを、N-gram モデルを用いて決定する手法を提案する。実際の即興演奏に対して補正処理を行ったところアウト音をすべて補正する手法に比べて、適合率を約 13%改善(ただし再現率は約 2%低下)することができた。

N-gram Based Melody Correction for Improvisation

KATSUHISA ISHIDA,[†] TETSURO KITAHARA^{††} and MASAYUKI TAKEDA[†]

This paper describes a system that corrects unnatural melodies in improvisation. It is not suitable to correct all notes out of the available note scale because these notes are not necessarily unnatural. We propose a method for determining notes to be corrected based on the N-gram model. Experimental results show that our method improves the precision rate of melody correction.

1. はじめに

近年、計算機の発達により、ジャムセッションを楽しむ機会が増えつつある。たとえば、「ジャムセッションシステム」¹⁾は、計算機内に仮想ミュージシャンを構築することで、計算機とジャムセッションを楽しむ機会を与える。「Open RemoteGIG」²⁾は、インターネットなどの広域ネットワークを介した遠隔地どうしのジャムセッションを実現する。また、PDA を用いた持ち運びの容易な電子楽器³⁾や、ウェアラブル型の電子楽器⁴⁾も提案されている。このようにハードウェアの整備は進みつつあるにもかかわらず「ジャムセッションが以前より身近になった」とはいえないのが現状である。なぜならジャムセッションを行うのに必要不可欠な即興演奏が旋律創作と演奏を同時に行う難しい演奏形態であり、高度な技術と経験を要するからである。

即興演奏が通常の演奏よりも難しいのは、演奏者が、

どの音を出せば自然な旋律になるか(たとえば伴奏と協和するか)をその場で考えながら演奏しなければならないからである。そのため、即興演奏は、楽器演奏ができる人であっても難しい。そこで、本研究では、「楽器演奏自体はできるが、どの音を出せば自然な旋律になるかが瞬時に判断できない人」を対象に、即興演奏における旋律の不自然な個所を自動的に補正するシステムを提案する(図1)。本システムの使用により、即興演奏の不自然な旋律が軽減されるので、即興演奏にあまり自信のない人でも、ためらわずにジャムセッションに参加できるようになる。

しかしながら、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、主観をとまなうからである。そのため、この課題を扱った研究例は少ない。「Coloring-in Piano」⁵⁾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」⁶⁾は、楽器のインタ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、この楽器の演奏方法を一から覚え直す必要がある。「INSPIRATION」⁷⁾は、アヴェイラブルノートスケールから外れる音(ア

[†] 東京理科大学大学院理工学研究科情報科学専攻

Department of Information Sciences, Graduate School of Sciences and Technology, Tokyo University of Science

^{††} 京都大学大学院情報学研究科知能情報学専攻

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

それぞれのコードに適した音で構成されるスケール。このスケール

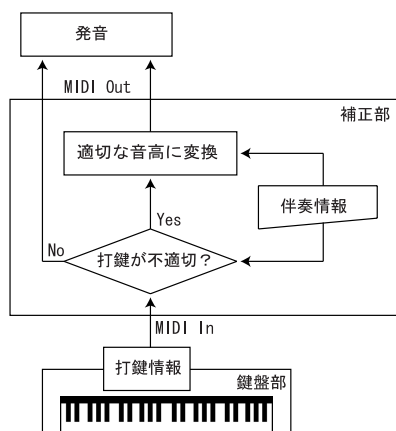


図 1 提案するシステムの概要．演奏された旋律から不自然な音を検出して補正を行う

Fig. 1 The outline of the proposed system. The system detects notes to be corrected in melodies and corrects them.

ウト音と呼ぶ)をすべて補正する．しかし、アウト音がつねに「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．

本稿では、打鍵されたアウト音を補正すべきかどうかを、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)に基づいて、統計的に決定する手法を提案する．「どのような条件でアウト音が使われやすいか」を N-gram でモデル化し、旋律データベースにおける N-gram 確率が低いアウト音のみを補正対象とする．

2. 旋律補正手法

旋律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)をどのように検出するかである．この課題に対する 1 つの解決法として、アウト音をすべて補正する(全補正と呼ぶ)方法が考えられる．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を、N-gram による旋律モデルに基づいて決定する(図 2)．まず、旋律を特徴抽出と N-gram によりモデル化する．そして、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．

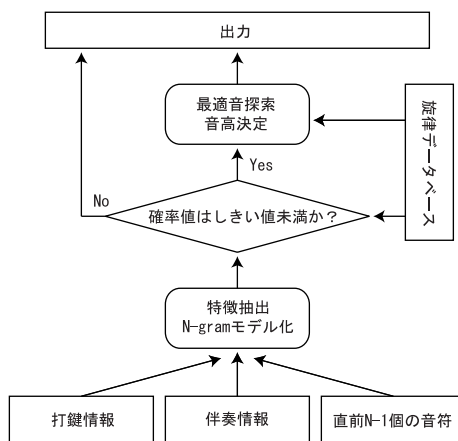


図 2 補正内部処理．打鍵された音と演奏済みの音から N-gram 確率を算出し、確率が小さい場合にのみ補正処理を行う

Fig. 2 The overview of internal processing. The system first calculates N-gram probabilities of played notes, and then corrects only notes of which the N-gram probabilities are low.

表 1 特徴ベクトルの各要素(カッコ内はとりうる値)

Table 1 Elements of feature vector.

対象音の種類(コード構成音, キー構成音, それ以外)
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)

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．以下、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用いられ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

即興演奏の旋律において旋律 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

ル中の音を用いると、コードとよく響きあうとされる．



図 3 不自然な音を補正した例．印のついた音が音楽的に不自然であると判断され，補正された

Fig. 3 An example of melody correction. The marked note was corrected by both methods.

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 補正後の音高決定

アウト音 x_n が補正対象となると，この 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，この値が最大となる音高に補正する．

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⁸⁾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． $N = 2$ (bi-gram)， $N = 3$ (tri-gram) で実装し，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 キーボードを用いて単音のみの即興演奏を行う．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 旋律補正例

補正前と補正後の旋律の一部を図 3，図 4 に示す．これらの楽譜はそれぞれ，1 番上が補正前の旋律，2 番目が提案手法 ($N = 3$)，3 番目が全補正で補正を行った旋律である．図 3 の印のついた音符はアウト音であり，また，実際に不自然な響きを生じるもので，補正されるべき音である．この音に対しては全補正，提案手法とも正しく検出できた．また，図 4 の 2 つの旋律で印の付いた音符はアウト音であるが，実際には不自然な響きを生じるものではなく，補正されるべきではない．これらについて，全補正では補正されたが，提案手法では補正されなかった．このように，提



図 4 アウト音を許容した例．印のついた音はコード，調からは外れており，全補正では補正された．しかし，提案手法によれば不自然ではないと判断され，補正されなかった

Fig. 4 Examples of unchanged melodies. The marked notes are out of the available note scale, but are judged to be appropriate by proposed method.

表 2 被験者とラベルの詳細
Table 2 Details of subjects and labels.

	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

案手法による補正は，演奏者の表現の幅を狭めることのない補正対象の決定がなされている．

3.3 評価実験

補正対象の決定が適切かどうかについて実験する．あらかじめ 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個所を人手でラベル付けする（「補正すべき」とラベル付けされた音を「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，全補正とも比較する．また，被験者とラベルの詳細を表 2 に示す．

表 3 要補正音検出実験結果

Table 3 Experimental results of determining notes to be corrected.

	全 体			初 心 者			中 級 者			上 級 者		
	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	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
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

3.4 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，全補正に比べると bi-gram で 0.1093，tri-gram で 0.1080 向上した．これにより，提案手法の補正対象決定処理は，全補正より適切といえる．

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，提案手法が，要補正音の取りこぼし（要補正音を補正しないこと）を最小限に抑えながら，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ことができる．

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）．あるいは，より高度な演奏を目指してアウト音を積極的に使おうとした結果，逆に不自然な旋律になってしまった人もいた．本手法では，このような明らかに不自然な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，上級者の中にクラシック音楽の演奏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付けし，より詳細に評価していくことも必要である．

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旋律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旋律補正も検討していく．

謝辞 有益なご助言をくださった後藤真孝氏（産業技術総合研究所）に感謝する．また，柳川貴央氏，渡辺義大氏をはじめ，本研究にご協力いただいたすべての方々に感謝する．

参 考 文 献

- 1) 青野裕司：ジャムセッションシステム，コンピュータと音楽の世界，長島洋一ほか（編），pp.283–305，共立出版（1999）．
- 2) 後藤真孝，根本 亮：Open RemoteGIG：遅延を考慮した不特定多数による遠隔セッションシステム，情報処理学会論文誌，Vol.43，No.2，pp.299–309（2002）．
- 3) 寺田 努，塚本昌彦，西尾章治郎：2つの PDA を用いた携帯型エレキベースの設計と実装，情報処理学会論文誌，Vol.44，No.2，pp.266–275（2003）．
- 4) 前川督夫，西本一志，多田幸生，間瀬健二，中津良平：ネットワーク型ウェアラブル音楽創奏システムと日常生活空間演出構想の提案，日本バーチャルリアリティ学会論文誌，Vol.6，No.2，pp.69–78（2001）．
- 5) 大島千佳，宮川洋平，西本一志：Coloring-in Piano：表情付けに専念できるピアノの提案，情報処理学会研究報告，2001-MUS-42，pp.69–74（2001）．
- 6) 西本一志，渡邊 洋，馬田一郎，間瀬健二，中津良平：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，情報処理学会論文誌，Vol.39，No.5，pp.1556–1567（1998）．
- 7) 谷井章夫，片寄晴弘：音楽知識と技能を補うピアノ演奏システム“INSPIRATION”，情報処理学会論文誌，Vol.43，No.2，pp.256–259（2002）．
- 8) 伊藤伸吾：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）．

（平成 15 年 6 月 24 日受付）

（平成 15 年 9 月 5 日採録）