

フレーズに対する心地よさの客観評価法の性能評価

三宅勇紀[†] 井上凌輔[†] 梅村祥之[†]

概要：先に提案した楽曲の心地よさに関する客観評価法の検討において実験参加者は1名であった。そこで、本研究は複数人の主観評価を用いて客観評価法にて機械判定することで、より一般的な結果を得ようとする研究である。そこで、1フレーズおよそ8音符のメロディに対しての客観評価法を3つの方法で性能評価した。1つ目は複数人の主観評価結果を個人ごとに主観評価法によって機械判定する方法である。2つ目は個人ごとの正解率より一般性を高めるため、複数の評価者による判定を正解判定として、客観評価法の正解率を調べる方法である。3つ目は客観評価法が機能することが出来るのかを確認することで、性能評価を行う方法である。以上のことから客観評価法の有用性を確認することが出来た。

Performance evaluation for the objective evaluation method for comfortableness of musical phrases

YUKI MIYAKE[†] RYOSUKE INOUE[†]
YOSHIYUKI UEMURA[†]

Abstract : In our previous study "A proposal for an method to evaluate comfortableness of phrases of music constructed by 8 notes", the number of participant was only 1. In this study, we used some ratings by some participants as the correct data for the machine learning and examined the generalized performance for the objective evaluation method for comfortableness of musical phrases. Firstly, we constructed machine decision method based on the correct data evaluated by individual participant. Secondly, we constructed machine decision method based on the correct data evaluated by some participants. Thirdly, we adopted the objective evaluation method for phrases generated by an automatic composition method.

1. はじめに

自動作曲に関する多くの研究がなされ、曲生成の様々なアルゴリズムが提案されている。[1], [2]. 文献[2]で示される「要件 (B) 徴収になじみのある音楽スタイルを踏襲する」ために、自動作曲アルゴリズムの中に、生成結果が妥当な結果であるかを判定して、生成曲から妥当な取捨選択するモジュールを組み込むものがある。本研究も同様に、自動作曲アルゴリズムによって生成された曲の中から妥当な曲を取捨選択するための客観評価法を構築するための研究である (図1)。

前報[3]では実験参加者1名による273曲の主観評価を正解判定として、客観評価法にて機械判定するといった内容であった。

本研究では複数人の主観評価を用いて客観評価法にて機械判定することで、より一般的な結果を得ようというものである。性能評価の方法として3つの方法を提示する。複数の実験参加者ごとに機械判定して正解率を確認する方法と、実験で得た複数人の客観評価から曖昧な評価以外を抽出して正解率を確認する方法と、コンピュータで自動生成した曲を主観評価して機械判定で正解率を確認する方法である。

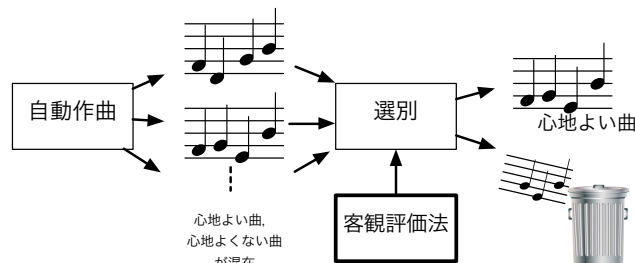


図1 研究の位置づけ

Fig. 1 Research framework.

2. 主観評価される曲の選定

本研究では世界各国の民謡が収録された Essen folksong collection を使用する。このコーパスを使う理由として、楽曲が単旋律で、フレーズ境界や小節の情報があることがあげられる。曲数は8,456曲である。

3. 個人適合

3.1 目的

次の実験を行う目的は、各実験参加者の判断を教師信号として機械学習し、機械判定したときの正解率が実験参加者によってどのような結果となるのか調べるのが目的である。

3.2 主観評価実験

今回の実験では Essen folksong collection からヨーロッパ曲が3,948曲とヨーロッパでない曲の2,254曲と合わせて

[†] 広島工業大学 情報学部 情報工学科
Hiroshima Institute of Technology

6,202 曲を用いる。

それらの楽曲について、次のような主観評価をしてもらい、心地よさの評価を得る。

実験参加者： 大学生 16 名である。

楽曲の提示： MIDI ファイルを Apple 社製コンピュータ Macintosh の QuickTimePlayer で演奏して得られるピアノ音のサウンドファイルを音源とした自作の演奏ソフトを使用する。イヤホンで聴取する。再生操作を実験参加者が行い、実験参加者のペースで評価実験を進める。同じ曲を何度も聴いても良い事とする。

評価方法： 心地よさを、悪い（評価値 1）から良い（評価値 5）までの 5 段階で評価する。

3.3 特徴量の種類と頻度分布

主観評価実験の結果からさらに選択したデータセットを正解判定として客観評価法にて機械判定するべく、特徴量を算出する。

今回用いた特徴量は音程系列、音程予測、音高変化数、音高輪郭である。音程系列においては音程系列の出現頻度に関する特徴量の最大と平均、音程予測においては音程予測の出現頻度に関する特徴量の最大値を用いる。これらの特徴量の説明は文献[3]の 3 章 2 節「特徴量の算出方法および値の分布」を参照されたい。

図 2、図 3、図 4、図 5 にそれぞれの特徴量の頻度分布を示す。良い曲と悪い曲で色を分けて、重ねて表示している。

グラフで色同士が重なって表示されていないほど、強力に判定できるということである。

図 4 の音高変化数のグラフについて、2 色の重なりが少ないことからこの特徴量はとても強力だということが判明した。よって、音高変化数は有用な特徴量である。

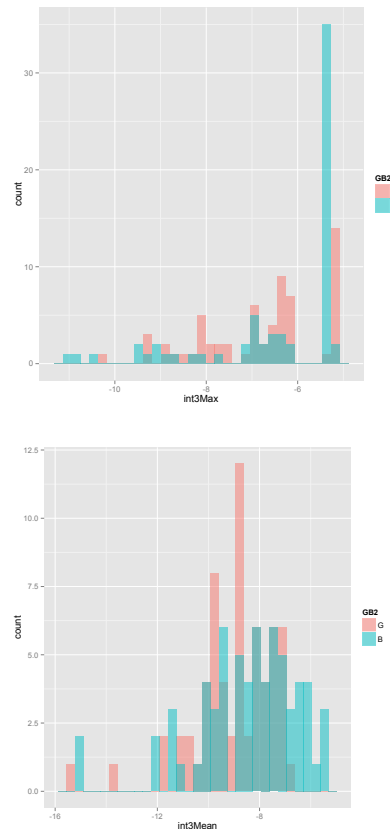


図 2 音程系列の出現頻度に関する特徴量の最大と平均の頻度分布

Fig. 2 Histograms of features about frequency of interval sequence

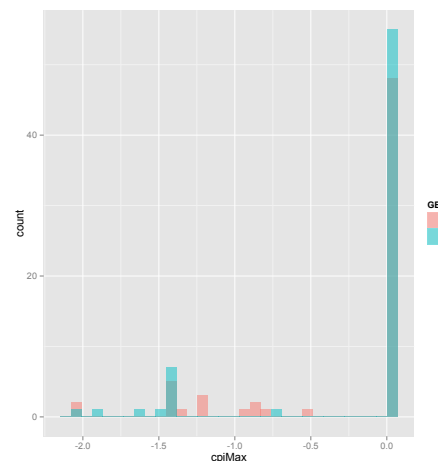


図 3 音程予測の出現頻度に関する特徴量の最大の頻度分布

Fig. 3 Histogram of features about prediction of interval sequence

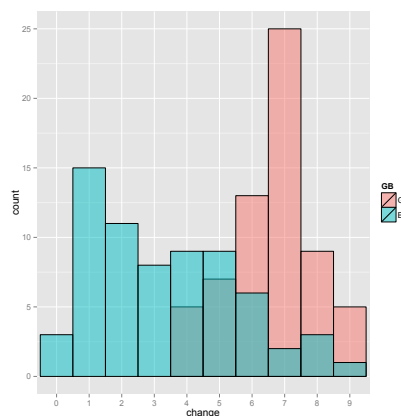


図 4 音高変化数の頻度分布

Fig. 4 Histogram of number of changes

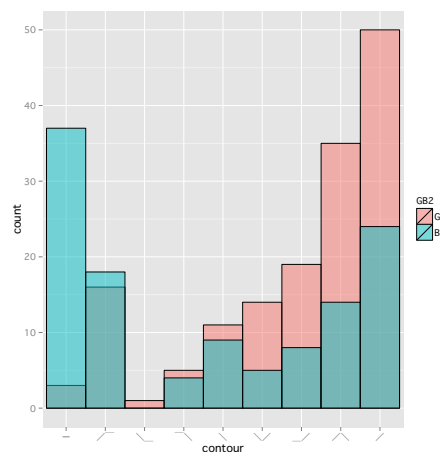


図 5 音高輪郭の頻度分布

Fig. 5 Histogram of pitch contour

3.4 個人結果ごとの機械判定の方法

前の主観評価実験では本実験で評価された Essen folksong collection の 6,202 曲のうち、1 フレーズの音符数が 8 音符に該当する 1,770 曲を用いていた。しかし、16 名に対して行った主観評価実験で主観評価された曲の中にこの 1,770 曲が完全に含まれていない。なので、主観評価された曲のうち、該当した 1,206 曲の主観評価を用いて前報研究の実験を 16 名に対して行う。

機械判定に使用する特徴量の算出方法は前節と同じ方法を使用する。複数の特徴量を用いた機械判定法として、パターン判定の分野で広く使われている SVM を用いる。SVM のソフトウェアとして統計解析用ソフトウェア R で動作するパッケージ kernlab を用いる。性能評価にあたり 10 fold cross validation 法によるオープンテストを行う。

このとき、16 名に対して行った主観評価実験の評価方法が 5 段階から 3 段階に変換する。その理由は、1 フレーズ 8 音符を対象としているため、細かな段階を区別して評価することは困難であろうと考えたからである。5 段階評価から 3 段階評価への変換は図 6 を参照されたい。



図 6 3 段階評価への変換

Fig. 6 Transform from five-point rating scale to three-point rating scale

3.5 判定結果

16 名のうち 1 名に正解率 70% という高い結果が出た。チャンスレベルは 53% なのでこの正解率は十分なものである。他 15 名はチャンスレベルほどの正解率しか出なかった。そこで、前報の主観評価実験で得た評価を用いて再度機械判定した。その結果チャンスレベル 56% に対して正解率 84% という高い結果が出た。

高い正解率を出した実験参加者 2 名のうち、前報の実験で得た主観評価をした人物は以前にも主観評価実験に参加した経験のある人物で、もう 1 人は主観評価実験に参加したことがない人物であった。

正解率の良かった 2 名の判定結果を図 7 に示す。

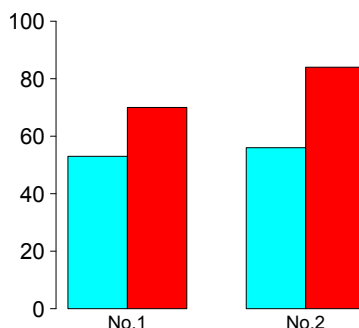


図 7 2 名の正解率

Fig. 7 accuracy rate of 2 participants

3.6 考察

17 名中 2 名が 84% と 70% の正解率であった。他 15 名はチャンスレベルに等しい正解率となった (図 7)。

正解率の高い実験参加者のうち 1 名は主観評価実験の経験者であり、評価する基準が実験参加者の中で安定した主観評価することが出来た為、正解率が高いと考えられる。

上記以外の実験参加者は未経験者である。そのうちの 1 名が高い正解率を得た。その他の実験参加者はチャンスレベルの正解率であった。

安定した評価の出来る実験参加者については高い正解率が期待できる。

4. 主観評価の 2 名による 2 段階選択

4.1 目的

前章で、ある実験参加者が心地よい曲と判定した曲を著者がいくつか聴いてみると心地よくないと判断される曲も含まれていた。前章では各個人の好み为正解判定として

使用されているので、複数の評価者による判定と異なるためと考えられる。そこで、より一般性を高めるため、複数の評価者による判定を正解判定として、客観評価法の正解率を調べるのが目的である。

4.2 評価値の2段階選択の方法

主観評価実験で得られた主観評価結果は6,202曲に16名分の評価が付与している。これをまとめて1つの評価にする。そのために、前報の文献[4]の3章2節「主観評価結果のデータ解析」に記述されている方法を用いて、主観評価実験で得られた6,202曲を心地よい曲のクラスと、心地よくない曲のクラスに分類する。以上により6,202曲から心地よい曲が531曲と、心地よくない曲が563曲の合わせて1,094曲が分類された。

それらの曲を、実験参加者1名が3段階で評価値を付与する。その結果、1,094曲から心地よい、よくない曲合わせて273曲に絞り込んだ。

そして、もう1名の実験参加者によってさらに見直しを行い、評価値に同意できるものとして273曲の中から、241曲が抽出された。

さらに、6音符から10音符までの曲に限定した。この理由は本研究が1フレーズを対象とした研究だからである。1フレーズに相当する基本的な音符数は8個なので、前後幅を持たせて6音符から10音符までの曲を使用する。こうして、抽出された曲は120曲となった。

楽曲を選曲する流れを図8に示す。

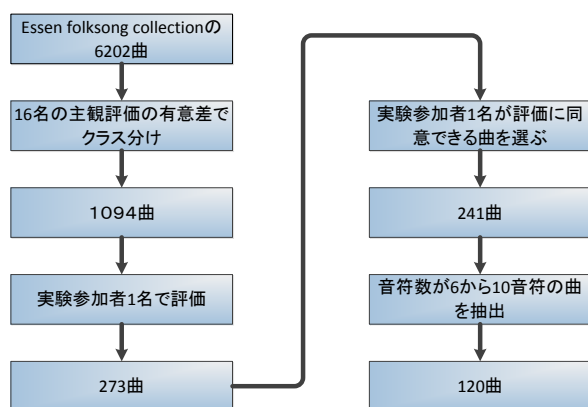


図8 選曲の流れ

Fig. 8 Flow chart of choosing pieces

4.3 デモ曲

最後に本実験において Essen folksong collection の中から心地よいと評価された曲2曲と、心地よくないと評価された曲2曲の楽譜を図9に示す。



(b)心地よくない曲

(c)心地よい曲

(d)心地よくない曲

図9 選曲された曲の例

Fig. 9 Examples of selected pieces

4.4 機械判定法および判定結果

前章と同じ方法で特徴量を算出し、機械判定を行っていく。SVMによる機械判定を行う際の性能評価も同様に、10 fold cross validation 法によるオープンテストを行う。すべての特徴量を用いて判定した結果はチャンスレベル51%に対して正解率80%であった。

4.5 考察

この実験で前章の実験参加者16名の主観評価結果を統合し、2名による段階的な選曲をすることによって、より一般性が高い結果となることが出来た。結果も正解率80%と前章の結果よりわずかに下がってしまったが、依然と高い結果を示している。しかし、わずかながらにも正解率が下がってしまったことから、一般性がより高くなるにつれて、正解率は低くなることが考えられる。

5. 自動生成曲の対する性能評価

5.1 目的

この実験の目的は、図1のような枠組みの1例を試行的に行い、客観評価法が機能することが出来るのかを確認するためである。試行的検討のため、小規模な実験に留める。

1フレーズ8音符のリズムが4分音符で一定なメロディを生成し、客観評価を付与して機械判定を行う。

5.2 自動作曲の方法

今回楽曲を生成するに際して簡易的な以下の方法を用いる。

手順1: 1,024個の正規乱数の累積和の系列を作成する。

手順2: 3点ずつ平均を求めていくことによって、系列の長さが1/3になる。

手順3: 手順2を系列のサンプル数が8個になるまで繰り返す。

手順4: 8個のサンプル値を音高に割り当てる。C4の音が数値の1になるように割り当て、数値1の増減を半音に割り当てる。下限はC1で上限はB6である。

手順4: 各音高を確率1/2でCの和音に対応するド、ミ、ソ、いずれか最寄りの音高に変更する。

図10にこの手順のブロック図を示す。

図11には、実際に手順の通り生成する過程のグラフを示す。

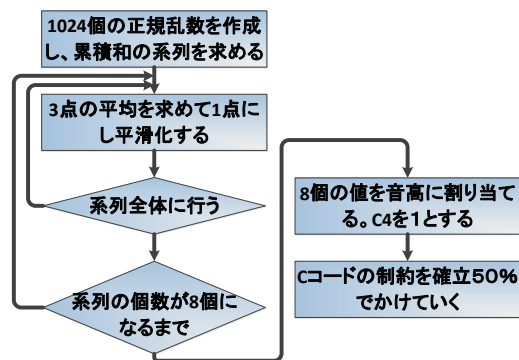


図 10 自動作曲のフローチャート

Fig. 10 Flow chart of automatic composition

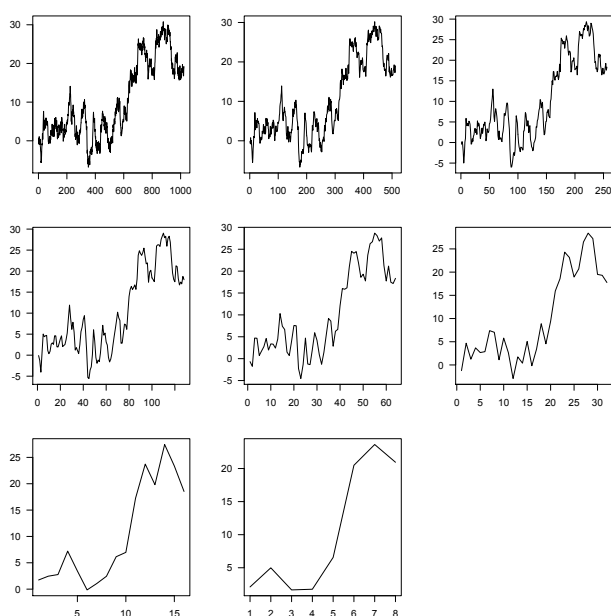


図 11 生成過程のグラフ

Fig. 11 Waveforms generated step by step

5.3 主観評価実験

今回の実験では、前節の方法で自動生成した曲を用いる。その曲を聴いて、心地よいと思った曲のメロディを保存していく。これを 20 曲分保存できるまで続ける。心地よくない曲のメロディデータに関してはランダムで生成した 20 曲のメロディとする。これらの合計 40 曲のメロディデータをもとに客観評価法にて機械判定を行っていく。

実験参加者： 大学生 1 名である。

楽曲の提示： MIDI ファイルを Apple 社製コンピュータ Macintosh の QuickTimePlayer で演奏して得られるピアノ音のサウンドファイルを音源とした自作の演奏ソフトを使用する。イヤホンで聴取る。再生操作を実験参加者が行い、実験参加者のペースで評価実験を進める。同じ曲を何度も聴いても良い事とする。

評価方法： 心地よさを良い、悪い、の 2 段階評価のよって評価を行う。

5.4 機械判定法および判定結果

第 3 章と同じ方法で特徴量を算出し、機械判定を行っていく。SVM による機械判定を行う際の性能評価も同様に、10 fold cross validation 法によるオープンテストを行う。その結果、チャンスレベル 50%に対して正解率は 65%であった。ただし、試行回数が少ないため統計検定による有意差は出ていない。

6. まとめ

1 フレーズに対する心地よさの客観評価法のその性能について、3 つの実験により性能評価を行った。主観評価実験で対象とする曲を Essen folksong collection とし、リズムを除くために音符をすべて 4 分音符とした。主観評価された曲から特徴量を算出し、客観評価法を構成して性能評価した。

1 つ目の実験は、17 名分の主観評価を機械学習し機械判定することによって実験参加者ごとの正解率を求めた。そのうちの 2 名は高い正解率 (70%と 84%) であった。その他はチャンスレベルの正解率であった。

曲の嗜好は人による違いが大きい。しかし多くの人から支持される曲もある。これを踏まえて複数名が一致した評価となる曲に対し客観評価法を適用した。

すなわち、2 つ目の実験として複数名が一致した評価となった曲に対して客観評価法を適用した。その結果、正解率 80%を得た。

客観評価法の最終的な利用形態での活用法を小規模に試みた。すなわち、自動生成した曲を用いた客観評価法の性能評価を行った。簡易的に生成したテンポが一定の楽曲から心地よいと思った楽曲のメロディを 20 曲分収集した。そして、心地よくない曲はランダムに 20 曲収集した。その主観評価結果に対し客観評価法にて機械判定を行った。その結果、正解率 65%となった。

以上の結果、客観評価法の有用性を確認できた。

謝辞 本研究における主観評価実験は、広島工業大学の多くの方々の御協力頂いた、関係各位に深謝する。

参考文献

- [1] G.Nierhaus: Algorithmic Composition, Springer(2009).
- [2] 松原正樹, 深山寛, 奥山健太, 寺村佳子, 大村英史, 橋田光代, 北原鉄朗: 創作過程の分類に基づく自動音楽生成研究のサーベイ, コンピュータ ソフトウェア, Vol30, No.1, pp.101-118(2013).
- [3] 梅村祥之: 楽曲コーパス中の 8 音符からなるフレーズに対する心地よさの客観評価法の提案, 情報処理学会研究報告, Vol. 2014-MUS-105(2014).
- [4] 梅村祥之: 規則的に生成した 4 音符からなる楽曲を用いた楽曲の心地よさに関する客観評価指標, Vol.2014-MUS-104(2014).