Vol.2014-DCC-8 No.1
Vol.2014-MUS-105 No.1
2014/11/20

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

梅村 祥之[1,a)]

概要：我々が先に情報処理学会第 104 回音楽情報科学研究会で発表した「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」は，およそ 1 小節に相当する長さを対象とした楽曲の心地よさに関する客観評価法であった．
今回，先の手法を一部修正し，およそ 1 フレーズに相当する 8 音符を扱えるようにした．評価対象曲を Essen folk song collection の曲とし，実験参加者 1 名による 273 曲の主観評価結果を正解判定として，客観評価法にて機械判定した結果，正解率 84%という高い値を得た．

キーワード：音楽情動，客観評価指標，楽曲コーパス

A proposal for an method to evaluate comfortableness of phrases of music constructed by 8 notes

Abstract: Our previous study “Objective index about musical emotion using pieces of music constructed by regularly generated 4 notes” deals with an objective evaluation method about comfortableness of music constructed by 4 notes that correspond about 1 measure.
We modified the previous method in order to deal with music including 8 notes corresponding to about one phrase. We adopted pieces of music in Essen folksong collection as the targets, and obtained the result of decisions about comfortableness by a participant. As the result of comparison between decisions by the participant and decisions by the objective evaluation method, we achieved an 84% decision accuracy.

Keywords: musical emotion, objective index, musical corpus

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 [2], [3]．文献 [3] で示される「要件 (B) 聴衆に馴染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，生成曲から妥当な曲を取捨選択するモジュールを組み込みむものがある．

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**）．

音楽における情動の研究がなされている [4][8] ．Huron による ITPRA 理論 [11] では，音楽を聴く人は音楽の流れから先の展開を予測し，期待するメカニズムによって情動が生じるとしている．Juslin らによる BRECVEM 理論 [8] では，期待の要素の他，脳幹反応といった生理的なものからエピソード記憶といった常時のメカニズムまで複数の要素を広範囲に扱っている．これらの理論は情動のメカニズ

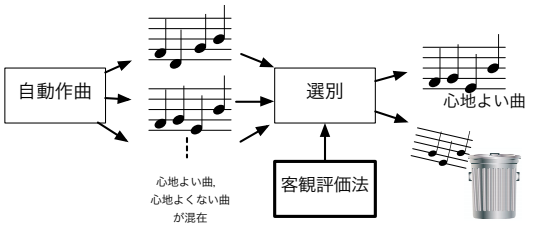

図 1 研究の位置づけ
Fig. 1 Research framework.

[1] 広島工業大学
Hiroshima Institute of Technology, Hiroshima 731–5193, Japan
[a)] y.umemura.im@it-hiroshima.ac.jp

ムを説明するための定性的なモデルの提案と言う面が強く，定量的な計算モデルにはなっていない．

実験心理学的な方法で主観評価実験のデータを解析して，人々の音楽に対する嗜好を扱った研究がある．文献 [5] は，音楽の嗜好 (preference, liking) に関する人の評価構造を調べる音楽心理学的研究である．広範囲の音楽ジャンルを扱い，26 ジャンル 52 曲について，多くの実験参加者（実験 1 が 706 名など）に，嗜好を 9 段階で評価してもらう．その結果を主成分分析して得られた結果から，嗜好を構成する次の 5 つの要因を抽出している．1) Mellow factor（心地よさ，滑らかさ）; 2) Urban factor（都会的）; 3) Sophisticated factor（高級感）; 4) Intense factor（激しさ）; 5) Campestral factor（田舎的）
この研究は，主観評価結果を分析するものであり．楽曲から客観評価指標を抽出したり，機械評価する研究ではない．

また，音楽検索において，楽曲から音響特徴を抽出して，その類似性などに基づいて検索を行う方法が研究されている [6][7]．特徴量の中で情動に関する特徴量は検索にとって有用なため，様々な特徴量が研究されている．その中には，低次の特徴量として，MFCC といった音の周波数成分に基づく特徴量があり，音楽の性質を活かした特徴量として，周波数の替わりに，ピッチクラスを扱った特徴量などが扱われている．

一方，本研究は自動作曲された楽曲から心地よい曲を取捨選択するための客観評価法を開発することを目的としている．そのために，評価対象の曲を極めて単純なものから始め，次第に複雑なものにしてゆく方針をとっている．先の報告 [1] において，評価対象を実際の曲ではなく，コンピュータで規則的に生成した 4 音符からなる楽曲を用いた．一般に音楽は，音高パターン，リズム，和音の 3 つの要素を持つと考えられるが，この中のリズムと和音の要素を除外して，音高パターンのみを扱った．具体的には，4 分音符のみで構成してリズムの要素を除き，短旋律にして和音の要素を除いた．そして，曲を構成する音符の数を極めて少ない 4 音符に限定し，4/4 拍子の 1 小節に相当する長さを扱った．各音符がダイアトニックスケールの 7 種類の音高（ドレミファソラシ）からなる 4 音符の曲の種類は 4 の 7 乗＝ 2,041 曲となる．そこで，これら 2,041 曲に対し，実験参加者に心地よさを 5 段階評価してもらい，主観評価値と得た．その結果から，有意差検定に関する統計処理を行うことによって，心地よい曲（以下，Good と称する）465 曲，心地よくない曲（以下，Bad と称する）465 曲を選別して主観評価結果データセットとした．親近性に関する特徴量等を定義し，機械判定法 SVM を用いて Good, Bad の判定を行った結果，主観評価結果を正解としたときの，機械判定による正解率が 90% という高い値を得た．

その後，実験結果を吟味した結果，楽曲の中に判定の容易なものがいくつか含まれていることが分った．そこで，

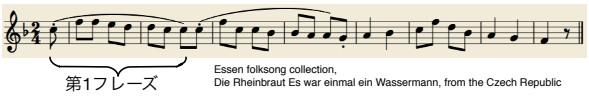

図 2 楽譜例
Fig. 2 An example of scores.

それらの曲を除いた場合の正解率について，本報の付録に記載する．

さて，メロディを知覚する際の最小限のまとまった固まりをフレーズといい，およそ 2 小節程度で構成される（**図 2**）．

Essen folksong collection[9] では，1 フレーズの音符数の最頻値が 8 音符である．そこで，本研究では 1 小節に相当する 4 音符の次の長さとして，8 音符からなる 1 フレーズの曲を対象とする．先の 4 音符を対象とした研究は，実際の曲ではなくコンピュータで規則的に生成した曲を対象としたが，本研究では実際の楽曲を扱う．具体的には Essen folksong collection のヨーロッパ曲を扱う．それを対象に，4 音符の際の特徴量を改良して適用して，心地よさの客観評価を行う．音楽要素の中のリズムと和音に関しては 4 音符の場合と同様に，リズム一定で，かつ，短旋律を対象とする．

規則的に生成した 4 音符からなる曲を主観評価する場合に比べ，1 フレーズからなる実際の曲を主観評価する場合には，人により主観評価結果が大きく異なる．そのため，多人数による主観評価結果に対して客観評価法を検討するのは，次のステップとし，本研究では 1 名の主観評価結果を基に，その判定結果と一致するような客観評価指標を検討する．

2. 主観評価実験および集計

本研究で 1 フレーズ 8 音符からなる曲を対象とするにあたり，Essen folksong collection の中の地域コードがヨーロッパの曲 6,202 曲のうち，第 1 フレーズの音符数が 8 音符の曲 1,770 曲を評価対象とする．これに該当する曲の譜例を先に図 2 に示した．各曲から第 1 フレーズのみを抜き出し，音符を全て 4 分音符に変更し，テンポ 120 で演奏する．なお，Essen folksong collection の曲は全て単旋律である．

次に主観評価実験方法について述べる．

実験参加者: 男子大学生 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．イヤホンで聴取する．再生ボタンを実験参加者が操作し，実験参加者のペースで評価実験を進める．同じ曲を何度聴き直しても良い．

評価方法: 心地よさを，良い (評価値 3)，どちらでもない (評価値 2)，悪い (評価値 1) の 3 段階で評価する

1,770 曲それぞれの主観評価値が「どちらとも言えない」を意味する 2 のデータを除いて主観評価値付きの楽曲データセットを作成する．その結果，主観評価値 1（以下，Good と称する）の曲が 154 曲，主観評価値 3（以下，Bad と称する）の曲が 119 曲得られた．

Good154 曲, Bad119 曲について，同じ実験参加者が，再度，2 段階の主観評価を行い，再現性を調べた結果，一致率 90%, κ 係数 0.81 と高い一致を示す値であった．以下，この主観評価値 Good, Bad を予測する客観評価指標を検討する．

3. 特徴量

3.1 4 音符を対象とする客観評価法からの変更

4 音符を対象とする客観評価法において用いた特徴量のうちのいくつかの特徴量を 8 音符からなる楽曲の客観評価にも用いる．個々の特徴量の説明に入る前に，4 音符を対象とする客観評価法における 9 種の特徴量のどれを採用し，どれを採用しなかったかについて，表 1 にまとめる．

不採用とした特徴量について，その理由を述べる．4 音符の際に用いた特徴量において，音高系列に基づく特徴量と隣接音符間の音程の系列に基づく特徴量の両方が含まれている．しかし，両系列から得られる特徴量は相関が高い．そのため，音程系列に基づく特徴量のみとした．楽曲コーパスから音程系列の出現頻度を求める際に，1 音符ずつ移動させながら頻度を求めるのか，フレーズ先頭の系列のみから頻度を求めるかによって，2 種類の指標を定義できる．それらの指標も，相関が高いため，フレーズ先頭の系列のみから頻度を求める指標は不採用とする．「調性」に関しては，今回のデータに対して特徴量を算出したところ判別力が低かったため不採用とした．

3.2 特徴量の算出方法および値の分布

音符の n-gram に基づく特徴では，系列の出現頻度を楽曲コーパスから算出して用いる．本研究で用いる楽曲コーパスも，4 音符の際と同様の Essen folksong collection から，4 音符の際と同様の算出方法で算出したものを用いる．以下，**図 3**，**図 4** を参照しながら説明する．特徴量「音程 3-gram」も「音程予測」も，本研究で扱う 8 音符の系列に対し，1 音符ずつ移動させながら，4 音符毎に指標を求める．すると，5 ヶ所から指標が得られる．5 ヶ所から得られた 5 個の指標の最小値及び最大値を 8 音符全体の特徴量とする．同図に，本研究のデータセットにおいて Good と主観評価された 154 曲に対する特徴量の頻度分布と Bad と主観評価された 119 曲に対する特徴量の頻度分布を色を変えて重ね合わせ表示したグラフを掲載する．音程予測で，

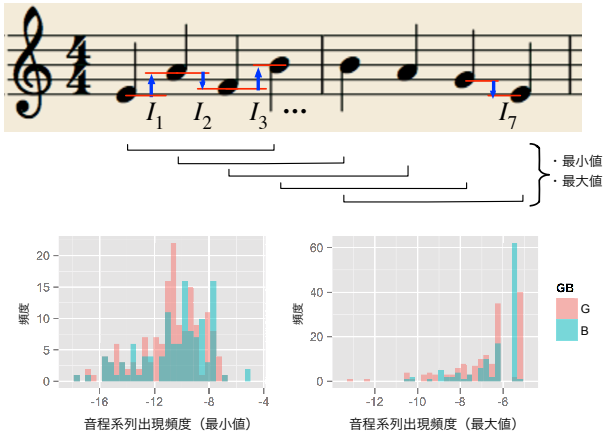

図 3 「音程系列の出現頻度」の算出方法及び頻度分布
Fig. 3 A method of extracting the feature “frequency of interval sequence” and the histogram.

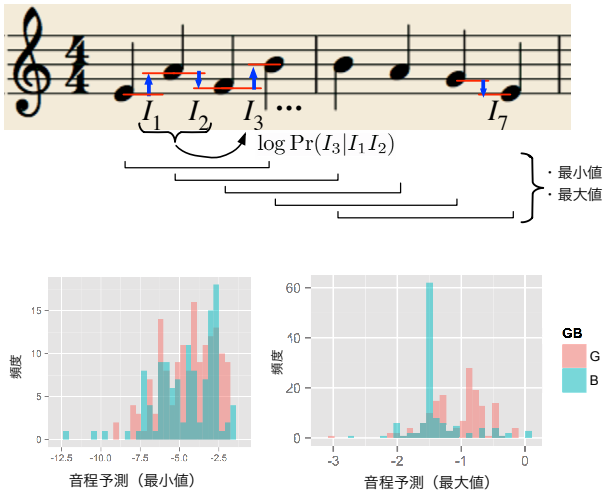

図 4 「音程予測」の算出方法及び頻度分布
Fig. 4 A method of extracting the feature “prediction of interval sequence” and the histogram.

5 ヶ所の最大値を指標とする場合に，Good と Bad をよく分離できることがわかる．

音高輪郭と音高変化数については，4 音符の場合と同じ定義である．**図 5**，**図 6** に算出方法の説明図，及び，本研究のデータセットにおいて Good と主観評価された 154 曲に対する特徴量の頻度分布と Bad と主観評価された 119 曲に対する特徴量の頻度分布を色を変えて重ね合わせ表示したグラフを示す．両者とも，Good と Bad を分離する能力を有することが分る．

4. 機械判定法および判定結果

前章で得られた特徴量を用いて，Good 154 曲，Bad 119 曲を機械判定する問題を扱う．

複数の特徴量を用いた機械判定法として，パターン認識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

表 1 特徴量の選定
Table 1 Selections of the features.

大分類	小分類	採用／不採用	備考
n-gram	音高 4-gram	不採用	音程に基づく指標を使う
	フレーズ先頭音高 4-gram	不採用	4 音符毎の処理
	音程 3-gram	修正	4 音符毎の処理
	フレーズ先頭音程 3-gram	不採用	4 音符毎の処理
予測	音高の予測	不採用	音程に基づく指標を使う
	音程の予測	修正	4 音符毎の処理
調性		不採用	性能
その他	音高輪郭	採用	
	音高変化数	採用	

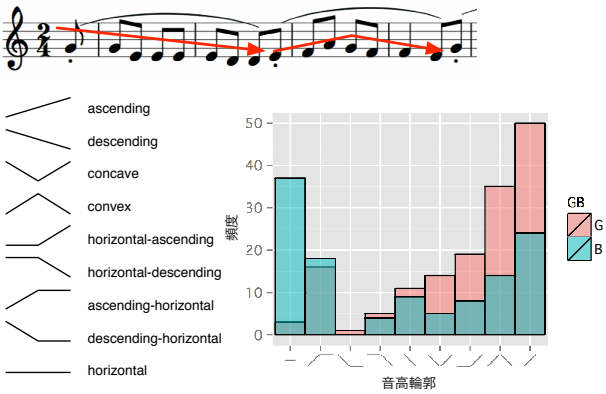

図 5 「音高輪郭」の算出方法及び頻度分布
Fig. 5 A method of extracting the feature “contour” and the histogram.

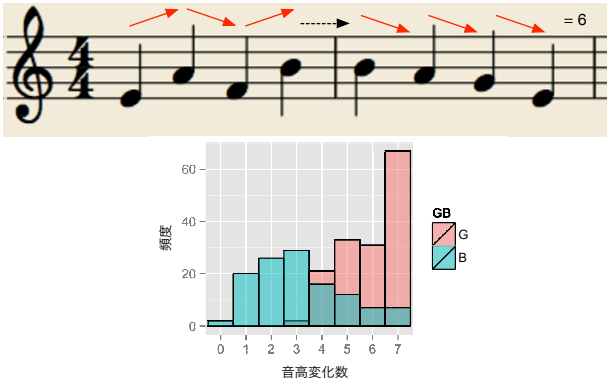

図 6 「音高変化数」の算出方法及び頻度分布
Fig. 6 A method of extracting the feature “number of changes of pitches” and the histogram.

パッケージ kernlab を用いる [10].

それに先立ち，用いる特徴量の選択について検討する．今回，全章で説明した 6 個の特徴量（音程系列の出現頻度の最小値，同最大値，音程予測の最小値，同最大値，音高輪郭，音高変化数）と，それらのペアからなる交互作用項を機械学習，機械判定における説明変数として用いる．6 個の特徴量と 6 個から選び出された特徴量ペアの数は $6 + {}_6C_2 = 21$ 個となる．能力の低い説明変数を採用すると，全体の判定能力が低下するため，21 個全てを使うのではなく特徴選択を行う．R の中のパッケージ MASS で提供される関数 stepAIC は線形モデルにおける特徴選択を自動で行う関数である．そこで，R に付属する線形モデルを扱う関数 lm の結果を stepAIC に渡すことによって特徴選択を行う．その結果，21 個中，次の 21 個の特徴量が選択された．

intMin, intMax, cpiMin, cpiMax, change, contour, intMin:cpiMin, intMin:contour, intMax:change, intMax:contour

ここで，A:B は特徴量 A と特徴量 B の交互作用項を表す．

以上によって選ばれた特徴量を用いて，SVM による機械判定を行う．SVM は，機械学習を行い，その結果を用いて機械判定を行うものである．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．その結果，正解率 0.84 を得た．

本データは 2 クラスを Good:Bad=154:119 の割合で含むため，全て Good と答えれば正解率が 154/(154+119) = 0.56 となる．この値をベースラインとして記号 base を用いる．また，9 個の特徴量をそれぞれ単体で，SVM によって機械判定したときの正解率を求める．その際にも，先と同様 10 fold cross validation を行う．以上の結果をまとめ，ベースライン，9 個の特徴量単独での SVM による機械判定，複数特徴量を用いた SVM による機械判定での正解率およびエラー率をグラフに表す（**図 7**）．

また，SVM による機械判定の過程を可視化表示する目的で，使用したライブラリ kernlab の関数 ksvm から得られる decision value の値の頻度分布を，Good, Bad それぞれについて，重ね合わせ表示する．decision value の値の正負により 2 クラスの判定がなされる．グラフ中に，判定境界を表す decision value=0 の位置を波線の縦線で示す．

5. まとめ

楽曲の心地よさの客観評価法の開発を行っている．前回の報告「規則的に生成した 4 音符からなる楽曲を用いた楽

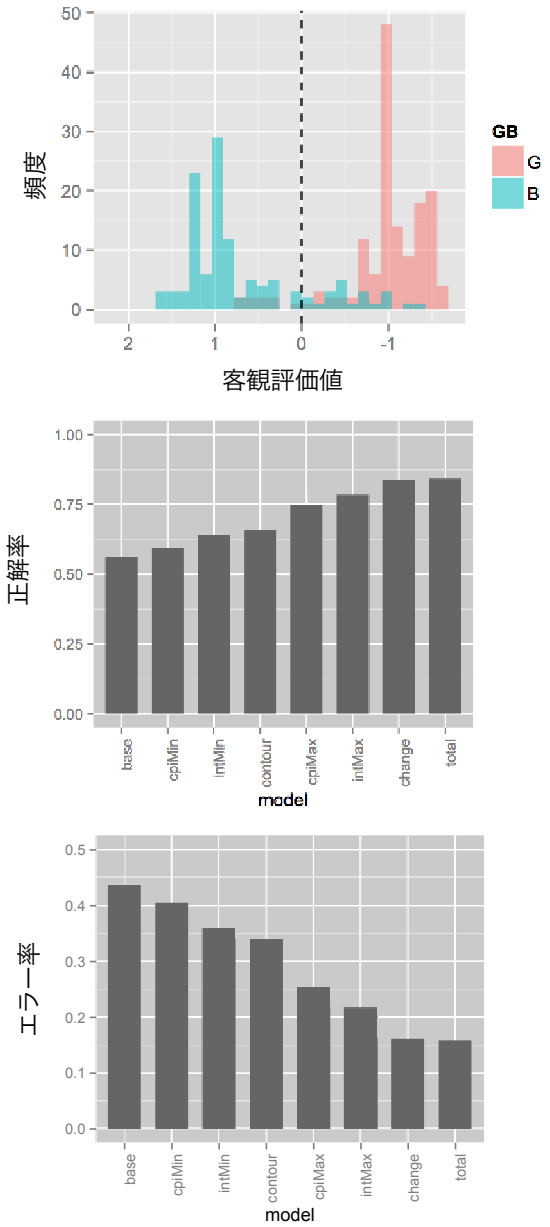

図 7 機械判定の正解率とエラー率
Fig. 7 Accuracy rate and error rate.

曲の心地よさに関する客観評価指標」では，およそ 1 小節に相当する 4 音符の長さを対象とした客観評価法であった．本報告は，およそ 1 フレーズに相当する 8 音符を対象とした．リズムの要素を除くために，音符を全て 4 分音符に変更した．先の研究で提案した特徴量に若干の修正を行い，客観評価法を構成して性能評価した．評価対象曲を Essen folk song collection の曲とし，実験参加者 1 名による 273 曲の主観評価結果を正解判定として，客観評価法にて機械判定した結果，正解率 84%という高い値を得た．

実際の曲の心地よさを主観評価する場合，人により主観評価結果が大きく異なる．そのため，多人数による主観評価結果に対して客観評価法を検討するのは，次のステップとし，本研究では 1 名の主観評価結果を基に，その判定結果と一致するような客観評価指標を検討した．多人数による主観評価結果に対して客観評価法の性能がどうなるかを検討することが今後の課題である．

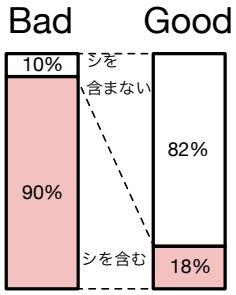

図 A·1 Good と Bad に対する「シ」の含まれる割合 (4 音符)
Fig. A·1 The proportion of songs including notes B in the Good data set and the Bad data set.

謝辞 本研究における実験において，広島工業大学の多くの方々に御協力頂いた．関係各位に深謝する．

参考文献

[1] 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, 情報処理学会研究報告，2014-MUS-104(2014).
[2] G. Nierhaus: Algorithmic Composition, Springer(2009).
[3] 松原正樹, 深山覚, 奥村健太, 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, Vol.30, No.1, pp.101-118(2013).
[4] 大村英史, 柴山拓郎, 寺澤洋子, 星 (柴) 玲子, 川上愛, 吹野美和, 岡ノ谷一夫, 古川聖: 音楽情動研究の動向—歴史・計測・理論の視点から—, 日本音響学会誌，Vol.69, No.9, pp.467-478(2013).
[5] P. J. Rentfrow, L. R. Goldberg, D. J. Levitin: *The Structure of Musical Preferences: A Five-Factor Model*, Journal of personality and social psychology, Vol.100, No.6, pp.1139—1157(2011).
[6] 帆足啓一郎: 音楽情報の検索，映像情報メディア学会誌, Vol. 64, No. 5, pp. 701-707(2010).
[7] Y. Yang, H. Chen: *Machine Recognition of Music Emotion: A Review*, ACM Transactions on Intelligent Systems and Technology, Vol. 3, No. 3, Article 40(2012).
[8] P. N. Juslin and J. A. Sloboda: *Music and Emotion, in The Psychology of Music, Third Edition*, D. Deutsch Eds.(Academic Press, 2013).
[9] *Essen Associative Code and Folksong Database*, http://www.esac-data.org
[10] A. Karatzoglou, A. Smola, K. Hornik and A. Zeileis: *Kernlab-an S4 package for kernel methods in R*, Journal of Statistical Software, Vol.11, No.9 (2004).
[11] D. Huron: *Sweet anticipation*, The MIT Press(2006).

付　　録

A.1　4 音符を対象とする客観評価法の追加実験

先の報告において，規則的に生成した 4 音符からなる楽曲 2,041 曲を対象として客観評価法を検討した．その際，

楽曲の中に判定の容易なものがいくつか含まれていることが分った．そこで，それらの曲を除いた場合の正解率について報告する．

具体的には，「シ」で終わる曲がいくつも含まれており，心地よくない曲との評価になっている．シは導音と呼ばれ主音に進行して解決したいという心理が働くため，シで終了すると違和感を感じると解釈できる．今回のデータの場合，Bad のうちシを含む割合が 90%で Good のうちシを含む割合が 18%と大きく異なった（**図 A·1**）．そのため，4 音符中のいずれかがシであるか否かという論理値を特徴量とすると，その特徴量だけの判定で正解率は 86%に達する．

そこで，Good, Bad のデータセットを次のように作り替える．Good はそのままにする．これまで，Bad として選択された曲を除き，主観評価値の低い曲から順に，シを含むか含まないかを調べ，含まなければ採用し，Good の数と同数になるまで繰り返す．このようにして採用された曲は，「心地よい」，「心地よくない」，「どちらとも言えない」の中の「どちらとも言えない」曲となるため，これらの曲を Nutral と称する．Good, Bad, Neutral の関係を**図 A·2** の (a), (d) に示す．

Good と Bad は平均値の差の検定に基づいて選定しているが，Neutral はそのような処理を経ていない．Neutral の曲に対する実験参加者の主観評価値の頻度分布と，Good の曲に対する実験参加者の主観評価値の頻度分布はかなり重なり合うことになる．この様子を同図 (e) に示す．比較のため Good の曲と Bad の曲に対する主観評価値の頻度分布の重なりを同図 (b) に示す．

以上の準備を基に先に提案した客観評価指標を使って，Good と Neutral のデータセットに対して SVM による機械学習，機械判定を行う．Good, Bad の機械判定の際と同様に 10 fold cross validation 法によるオープンテストによって正解率を測定する．正解率は 68%となった．

Good と Bad の判定における正解率 90%に対して大きく低下している．しかし，実験参加者の判断もばらつきが大きく，人のパフォーマンスも低下している．そこで，次のように人の正解率を定義する．

すなわち，人の主観評価値が 5 段階評価の中の 1 か 2 なら人は Neutral と判定したとし，4 か 5 なら，Good と判定したとする．3 の場合は実験参加者の半数ずつがそれぞれ Neutral, Good と判定したとする．したがって，Neutral に対して，実験参加者が Neutral と判定し，Good に対して，実験参加者が Good と判定した割合から，人の正解率を算出できる．先の報告における Good, Bad の判定タスクと本報における Good, Neutral の判定タスクでの実験参加者の正解率，客観評価法の正解率および，ランダムに判定したときの正解率 50%をまとめて同図 (g) に示す．Good と Neutral を判定するタスクにおいて，人の判定のばらつきを考慮に入れると，客観評価法の性能は人と同等以上となっている．

なお，同図 (c), (f) に客観評価紙票の性能を可視化表示する目的で，2 クラス分類での各クラスの客観評価値の頻度分布を示す．客観評価値には，使用した SVM のライブラリ関数 ksvm が decision value として出力する値を用いた．SVM の関数 ksvm の内部で decision value の正負により 2 クラス判定がなされる．

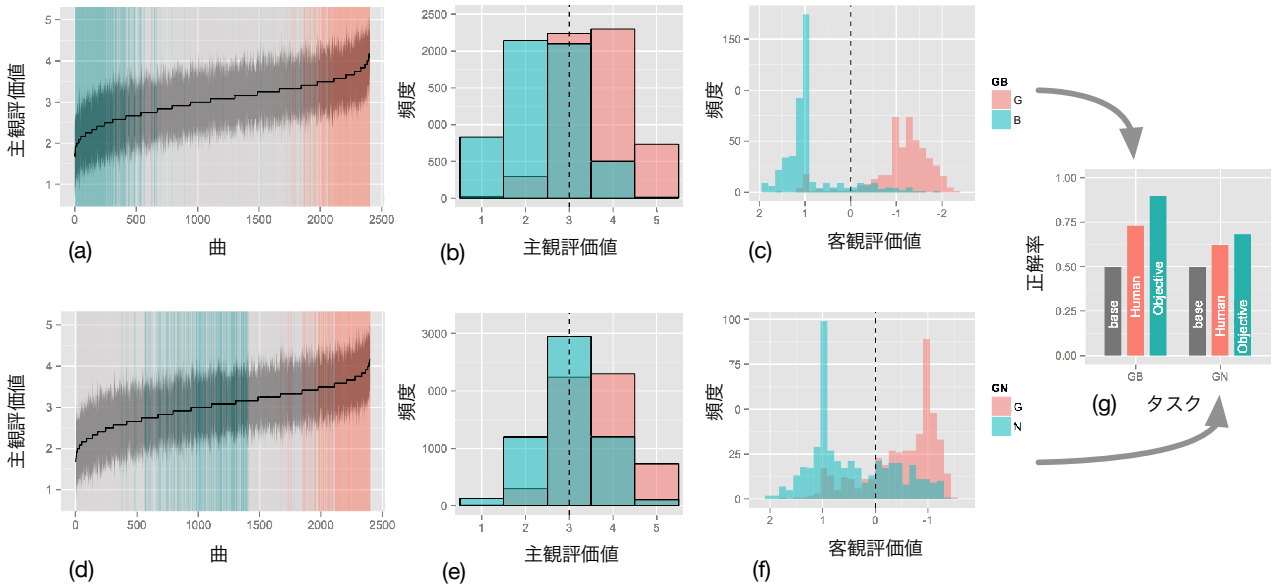

図 A·2 Good と Bad および Good と Neutral に対する実験参加者および客観評価法による判定 (4 音符)

Fig. A·2 The results of decisions by the perticipants and the objective evaluation method on the Good data set and the Bad data set.