

音響心理実験による音楽嗜好計算モデルの構築

Computation Model for Favorite Music Estimation

齊藤 幹
SAITO Kan

富士通株式会社 (〒140-8508 東京都品川区大井 1-20-10 E-mail: cans@jp.fujitsu.com)

ABSTRACT. As expanding speech recognition technology using HMM, we can make computation model for favorite music estimation, and apply to music recommendation system. The HMMs are individual for user, and trained with music characteristic vectors and user's evaluation of music. Since the HMM recognizes unknown music, it can estimate user's evaluation of the music. If the estimator is high, the music may be favorite for the user.

1. 背景

現在、音楽データベースから好みの曲を探すために取りうる手法は少ない。ジャンルや作曲家名といったメタデータを手がかりに検索するか、実際に曲の一部を試聴するなど、手間がかかる手法を取らざるをえない。このような手法では、検索キーなど対象に関する事前知識を必要とする。従って、既知曲の検索は比較的容易であるものの、未知の楽曲を発見することは非常に難しい。現在の時点では、好みの未知楽曲をデータベース中から検索する合理的な手法は存在しない。

一方、ヒトの音楽認知過程を表現しようと試みた、いくつかの音楽知覚モデルが存在する [1]。これらのモデルでは、非常に多くのルールやアドホックな仮定が複雑にからみ合い、コンピュータ上への実装が困難である。自然言語処理など他の分野においては、対象の統計的片寄りに着目した確率モデルが成果をあげている。

以上より、本研究では、メタデータによらない特徴を楽曲から抽出し、それを確率モデルに学習させて音楽嗜好モデルを構築する。音声認識分野での実績やツールキットの充実を考慮し、確率モデルとして Hidden Markov Model (HMM) を利用する。一般に、HMM は時系列データのパターンマッチングに適している。これにより、メタデータや音楽的構文規則を一切用いることなく楽曲の特徴を捕らえ、その曲をどの位好んでいるのか推定が可能になる。この嗜好の推定値によって、好みの未知曲を検索することもできるであろう。

2. 目的

本プロジェクトでは、以下の 4 項目について実現を目指している。

- 音楽嗜好に最も影響を与える特徴量の抽出
- 構文知識を用いずに音楽嗜好をモデル化
- 音楽嗜好における個人性の表現
- 複数の未知曲から好みの曲を検索

以下、各項目について説明する。

音楽嗜好に最も影響を与える特徴量の抽出 通常、楽曲中にはテンポや楽器の種類などさまざまな情報が含まれている。しかし、これらの情報すべてを扱うとなると情報量が多く、処理に多くの時間と記憶空間が必要で

効率が悪い。また、多くの特徴パラメータの中に重要な特徴が埋もれてしまう可能性もある。さまざまな特徴の中から、ヒトの音楽嗜好に最も影響を与える特徴量を抽出できれば、後々の処理において高い信頼性で効率良く楽曲の分析が可能となる。

構文知識を用いずに音楽嗜好をモデル化 ヒトの音楽嗜好は非常に曖昧であるものの、同一人物であれば一定の傾向を持つと考えて良い。この傾向を確率モデルを用いて表現し、音楽嗜好モデルの構築を行う。その際、従来のモデルのような音楽的構文知識を用いず、統計情報のみを使いモデルを構成する。以上の手法により、ユーザが評価した曲が少ない状態においても、「嗜好のすき間」を補間するような効果があると考えている。

音楽嗜好における個人性の表現 一般に、ヒトの音楽嗜好は個人差がある。この差を確率モデルで表現することを目指す。

複数の未知曲から好みの曲を検索 未知曲の特徴パラメータを嗜好モデルへ入力することにより、その楽曲の嗜好スコアを推定できる。この推定値が妥当なものであれば、多くの楽曲の中から推定値が高いものを選択し、好みの曲のみを探し出すことも可能になる。このような検索システムによって、ユーザ自身でも気付いていない音楽嗜好の発見につながることを期待している。

3. 実現手法

以下の手順に従い、音響心理実験の結果を用いて HMM を重み付け学習させ、ヒトの音楽嗜好を表現する (図 1)。

- 1) 嗜好調査
- 2) 楽曲の特徴抽出
- 3) HMM の学習
- 4) 嗜好の推定

(1) 嗜好調査

聴取した楽曲と嗜好との関係を調査するため、評定尺度法を用いた官能検査を行う。評定尺度法は日常用語で表記した数種のカテゴリーを用いて、被験者に感覚量の判断を求める。従って、被験者にとって判断は容易である。また、他の刺激と比較せずに絶対判断を求めるため、継続時間の長い刺激に対する反応を探るには適している。ただし、被験者の回答は等間隔性の保証がない序数尺度になる。そこ

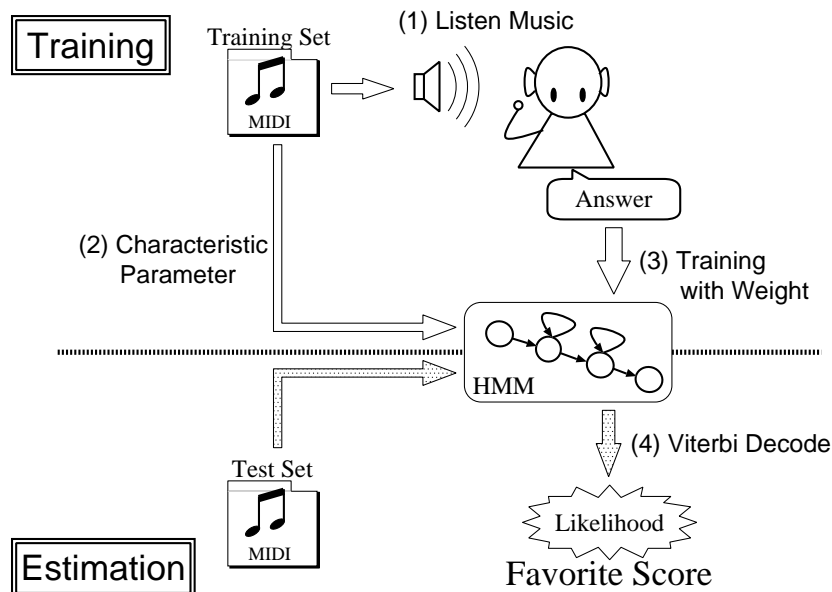


図 1: 音楽嗜好モデルの概要

で、範疇判断の法則 [2] により、評価値が間隔尺度として扱えることをあらかじめ確認する。

まず、対象となる要素を限定するため、被験者に提示するすべての楽曲 MIDI ファイルに関して、音色、音の強さおよび速さを統一する。それを演奏して楽曲を被験者に提示し、どのくらい好みの曲なのかという嗜好スコアを被験者に答えさせる。この楽曲の提示と嗜好スコアの回答を、HMM の学習に充分な量のデータが得られるまで繰り返す。

(2) 楽曲の特徴抽出

被験者に聴かせた楽曲から、ヒトの嗜好に強い影響を与える特徴パラメータを抽出する。

一般に、音楽にはビートとメロディのふたつの要素がある。ヒトが音楽を知覚する際には、まとまったビートパターンおよびメロディパターンを単位とし、両者を統合して解釈しているといわれている [1]。

そこで、ビートパターンに強く影響を与える音の長さ、およびメロディパターンを決定する音の高さを特徴パラメータとして採用し、MIDI ファイルから抽出する。合わせて、音長および音高の経時変化に対して多項式近似を行い、1 次微分、2 次微分を求め、特徴パラメータとする (図 2)。また、和声に関しては、和声構成音のうち最低音がヒトの知覚にもっとも大きな与えることが分かっているため、最低音のみを特徴抽出に用いる。

(3) HMM の学習

被験者に聴かせた楽曲の特徴パラメータを用い、各被験者ごとに別々の HMM を学習させる。その際、心理実験で求めた嗜好スコアにより重み付け学習を行い、個人の音楽嗜好を表現する。これにより、個人の音楽嗜好に対応した HMM が、各人についてひとつずつ構成できる。

通常の HMM 学習と同様に、Baum-Welch 法 [3] を用いて複数の特徴系列から HMM を学習させる。このアルゴリズムでは、各特徴系列に関する状態遷移の頻度を累積し、モデルパラメータの最尤推定を行う。ここで、本研究では、遷移頻度を加算する際に各特徴系列ごとの重みを掛け、モデルパラメータに重みを反映させる。また、学習重みには必要に応じて正の実数を掛け、線形のスケールングを行う。図 3 はその概念を表したものである。重みの高い特徴パラメータ系列に対しては、HMM を構成する確率密度関数の値が高くなり尤度も高くなると考えられ、またそ



Tone	67	67	69	67	72	71
$\frac{d}{dt}$ Tone	0	0	+2	-2	+5	-1
$\frac{d^2}{dt^2}$ Tone	0	0	+2	-4	+7	-6
Duration	3	1	4	4	4	8
$\frac{d}{dt}$ Duration	0	-2	+3	0	0	+4
$\frac{d^2}{dt^2}$ Duration	0	-2	+5	-3	0	+4

図 2: 特徴抽出法

の逆も成り立つと期待できる。

(4) 嗜好の推定

任意の楽曲の特徴パラメータ系列に対して、Viterbi アルゴリズムを用い個人ごとに HMM の尤度を求めることができる。その尤度は嗜好スコアと相関が高いと考えられ、

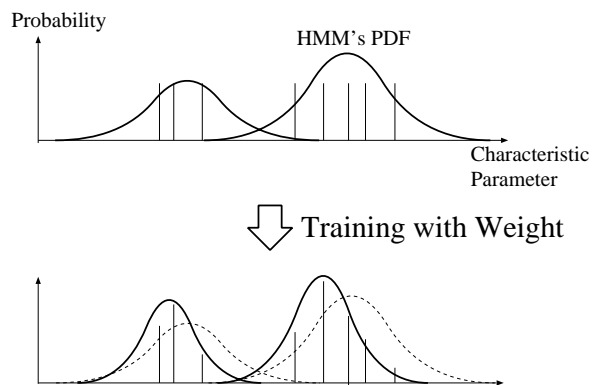


図 3: 重み付け学習の概念

表 1: 楽曲作成条件

ノートナンバ	1(グランドピアノ)
ペロシティ	100
テンポ	120

表 2: 使用機器一覧

MIDI 音源	Roland 社 SD-90
サウンドカード	RME 社 RME 96/8 PST
ヘッドホン	Sennheiser 社 HDA200

嗜好の推定値と見なして良い。

HMM の尤度は状態遷移確率と特徴出力確率の累積確率であり、音の数が増えるとそれに応じて尤度が減少していく。従って、短い楽曲の尤度が高くなり、長い楽曲では尤度は低くなる。そこで、楽曲に含まれる音の数により HMM の尤度を正規化し、一音当たりの尤度を嗜好スコアの推定値として扱う。

4. 実験

HMM がヒトの音楽嗜好を適切に表現できるかどうか、以下に行う実験により確認する。

(1) 嗜好調査

既存の MIDI データベースから複数の楽曲を選択し、すべての被験者に同一のランダムな順序で呈示する。評定尺度法により、被験者にその楽曲がどの程度好みなのかを 7 段階で答えさせる。実験条件を以下に述べる。

被験者 被験者は聴力正常な 26 歳から 28 歳の男性 7 名、女性 5 名の計 12 名で、音楽大学などにおいて音楽の高等教育を受けたものはいない。なお、被験者には本心理実験の参加に対して適正な報酬を支払っている。

使用楽曲データ アイデックス音楽総研社の音楽製作用 MIDI データライブラリ Hyper Groove 2 および 3 から、ピアノ曲を計 200 曲選択した。この楽曲に対して、表 1 の条件で音色（ノートナンバ）、音の強さ（ペロシティ）および速さ（テンポ）を統一する。また、コントロールチェンジやメタイベントなど、ノートオン・オフ以外の命令は削除する。以上の処理により、各呈示曲の間で異なる要素は、音の高さおよび音の長さとなる。1 曲あたりの長さは 10 秒強で、計 200 曲の聴取と回答に約 1 時間を要する。

楽曲呈示システム 使用した機器を表 2 に挙げる。

MIDI 音源がサンプリング周波数 48[kHz] で出力した音響信号をサウンドカードから出力し、ヘッドホンを用いて被験者に呈示する。呈示音量は被験者が快適に聴取できる値に固定した。被験者は PC に接続したテンキーを使って、1 ~ 7 の数値を回答する。この際、被験者が回答を完了するまで次の楽曲を呈示せずに待機し、呈示したすべての楽曲について嗜好スコアを回答させる。実験室は静かなオフィスと同程度の音響環境であり、楽曲の聴取を妨げるような騒音は発生していない。

教示 各被験者には図 4 に示す教示を行う。実験の主旨および作業内容について理解したことを確認した上で実験を実施する。

(2) 特徴抽出

特徴パラメータは音長および音高とし、MIDI ファイルから抽出する。それに加え、それぞれの 1 次微分、2 次微分を用いる。従って、計 6 次元の特徴量となる。特徴抽出条件を表 3 に示す。

図 4: 教示内容

この実験はヒトの音楽の好みを定式化することを目的としています。これから複数の音楽を聴いて頂き、その曲の印象によって、好みの曲であるか否かを指示に従って答えて下さい。答え方は以下の通りです。曲が流れると同時に、以下のような 7 種類の選択肢が表示されます。この 7 つのカテゴリーで判断し、その数字のキーを押して下さい。曲は 10 秒前後続き、変化していますが、必ず全体を聴いた上で、その一部ではなく全体の印象を判断してください。判断は首尾一貫する必要はありません。

- 1) 全く好きな曲ではない
- 2)
- 3)
- 4)
- 5)
- 6)
- 7) 非常に好きな曲だ

なお、答を入力するまで次の曲は始まりません。できるだけ、迷ったりしないでください。各曲について、直ちに浮かんだ印象が第一印象であり、それが我々が望んでいるものなのです。一方、決してごんざいな判断はしないでください。我々はあなたの真の印象を知りたいからです。

表 3: 特徴抽出条件

分析単位	音符
時間分解能	1/16 小節 (16 分音符)
周波数分解能	1/12 オクターブ (半音)
特徴量	音長、音高および 1 次、2 次微分 合計 6 次元
微係数回帰計算長	前後 3 音

(3) モデルの構築

被験者に聴かせた楽曲の特徴パラメータ系列に対して、音響心理実験で得た嗜好スコアにより重み付け学習を行い、嗜好モデルを構築する。既存のツールキットには重み付け学習の機能を持つものがなかったため、HMM ツールキット HTK[4] の改良を行った。

HMM の学習条件を表 4 に示す。各種パラメータは予備実験により適正な値を求めた。

(4) 嗜好推定の結果

構築したモデルが音楽嗜好を正しく表現できているか確認するため、被験者が答えた嗜好スコアと嗜好モデルの対数尤度を比較する。

a) クローズテスト

学習済みの HMM に対して、クローズテストを行う。HMM の学習に用いたすべての楽曲に対して、被験者が回答した嗜好スコアとモデルの対数尤度とを比較する。

まず、心理実験で呈示した 200 曲すべてに関し、嗜好スコアで重み付けして HMM を学習させる。そして、各楽曲の特徴パラメータ系列に対する、一音あたりの対数尤度を計算する。以上の操作を各被験者ごとの HMM につい

表 4: HMM 学習条件

モデル単位	被験者ごと
モデル種	共分散ガウス分布混合型
状態数	各 HMM につき 5 状態
ガウス分布混合数	2 混合
学習重みスケール係数	8.0

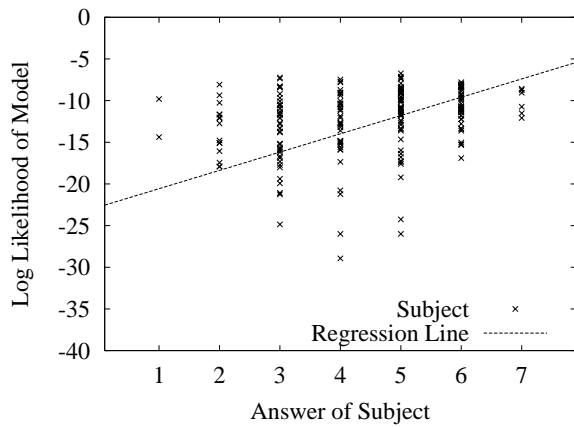


図 5: 嗜好スコアと対数尤度 (クローズテスト)

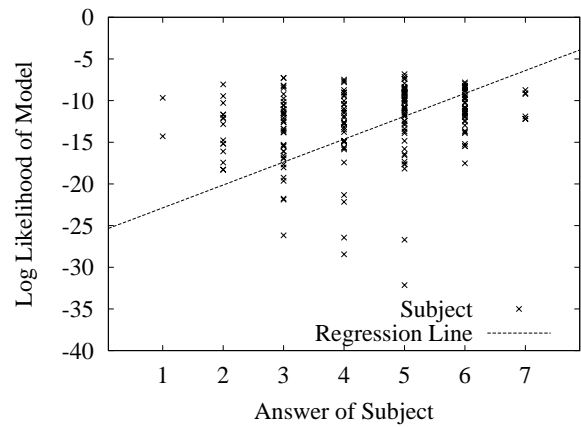


図 7: 嗜好スコアと対数尤度 (オープンテスト)

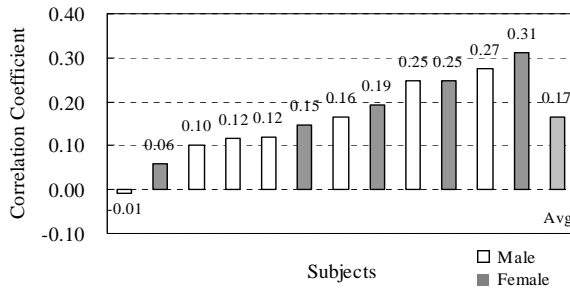


図 6: 嗜好スコアと対数尤度の相関係数 (クローズテスト)

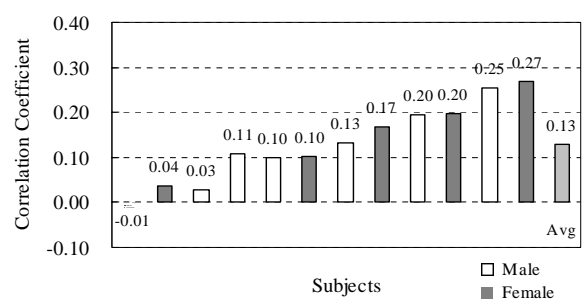


図 8: 嗜好スコアと対数尤度の相関係数 (オープンテスト)

て繰り返す。

最も相関係数が高かった被験者 1 名について、嗜好スコアおよび HMM の対数尤度との関係を図 5 に示す。横軸が被験者が回答した嗜好スコア、縦軸が HMM の対数尤度である。参考として、図中に回帰直線を引いてある。同一の嗜好スコアであっても、HMM の対数尤度は大きくばらついている。しかし、1 次の回帰係数は正の値となっており、全体的に見れば嗜好スコアが高いと HMM の対数尤度も高くなる傾向がある。

すべての被験者について、嗜好スコアと対数尤度の相関係数を図 6 に示す。図中の AVG は全被験者の平均値を表す。双方の間に高い相関があるならば、HMM は音楽嗜好を表現しているといえる。1 名を除き被験者について正の相関があるものの、その値は低い。また、相関係数の値は被験者によってばらついている。従って、学習したモデルが被験者の音楽嗜好を正しく表現できているとはいえない。

b) オープンテスト

学習済みの HMM に対して、オープンテストを行う。HMM の学習に用いていない楽曲に対して、被験者が回答した嗜好スコアとモデルの対数尤度とを比較する。

心理実験で用いた 200 曲からランダムに選択した 190 曲を用い、嗜好スコアで重み付けして HMM を学習させる。学習に用いていない残り 10 曲の特徴パラメータ系列に対して、一音あたりの HMM の対数尤度を求める。この操作を 20 回繰り返し、200 曲すべてに対してオープンテストを行う。これは嗜好モデルが未知曲の嗜好スコアを推定することに相当する。以上の操作を各被験者ごとの HMM について繰り返す。

クローズテストで結果を挙げた被験者に関して、心理実

験で回答した嗜好スコアおよび HMM の対数尤度との関係を図 7 に示す。横軸が被験者が回答した嗜好スコア、縦軸が HMM の対数尤度である。参考として、図中に回帰直線を引いてある。クローズテストの場合と同様の傾向を示している。

また、各被験者について、嗜好スコアと対数尤度の相関係数を図 8 に示す。クローズテストの場合と類似した傾向を示しているものの、相関係数の値は若干低くなっている。

5. 考察

実験結果を見ると、嗜好スコアと HMM の対数尤度との間には、ある程度の相関があるものの相関係数は低い。その原因としては、音楽における長周期の動的特徴を捕らえていない点が挙げられる。100 ミリ秒間程度の特徴により音韻を決定できる音声と比べ、音楽では更に継続時間の長い特徴が重要になる。長周期の特徴を表現できるように HMM の状態数を増加させたものの、相関係数を明確に向上させることはできなかった。よって、そもそも今回用いた特徴抽出手法が不適切であったと考えている。隣接する音のみを比較して特徴パラメータを抽出しており、長周期の変動が特徴量に反映できていない。より長いスパンの変化を特徴量に取り込むために、一定時間離れている音との変化を特徴パラメータとして利用する必要がある。他の研究では、ヒトの音楽嗜好に最も影響を与える特徴量を提案しているものもあり [6]、その成果を踏まえた上で手法を改良していくことが望ましい。

一方、ヒトは音列を群化して知覚しているという知見がある [1]。それに基づき、特定の音列パターンを HMM で表現してそれを連鎖させ、任意の音楽に対応させることが

考えられる。この手法は音楽の多様な構文知識が必要になるうえ、定型的な楽曲にしか対応できない欠点がある。

また、相関係数は被験者によって大きくばらついている。精細な HMM を利用してパターンマッチングを行っても、その傾向に変化はなかった。この事実は、今回用いた特徴抽出法が被験者によっては適しておらず、個人々々によって音楽嗜好に影響を与える要素が異なることを示唆している。被験者の内観報告に和声進行に関する言及があり、特に、和声の特徴抽出に課題があると考えている。和声は最低音のみを特徴量とするのではなく、そのほかの音についても特徴として採り上げる必要があると考えている。

同時に、被験者間の差異を HMM に学習させる重み付き学習が、それほど有効に作用していない点も挙げられる。追加実験を行って重みなし学習と重み付き学習との結果を比較したところ、相関係数の向上は 27% 程度であった。この向上量はやや少ないといわざるをえない。重み付き学習という枠組みは、遷移確率と特徴出力確率を間接的に変更するものであり、必ずしも出力する尤度に直接寄与しない。しかし、学習データが大量にある場合には、本研究での実験よりも重み付き学習が有効に作用すると考えている。

パターンマッチングに HMM のような確率モデルを用いる以上、大量の学習データが必要不可欠である。モデルへ個人性を学習させるためには、個人ごとに大量のデータが必要となる。しかし、大量のデータ収集は被験者の負担が増え、実施は困難である。対策として、まず楽曲の特徴を十分に表現できるようモデルの学習を行い、その後個人性を表現できるようモデルを適応化する手法が考えられる。これにより、楽曲の特徴を学習させる際には被験者の介入がなくなるため、負担を大幅に削減できる。しかし、既存の適応化手法は、HMM が表現する確率分布に対して線形変換を行い、適応化を実現している。このため、非線型な音楽嗜好を十分に表現できる可能性は低い。従って、既存の適応化手法の改良を検討する必要があると考えている。

6. まとめ

本プロジェクトでは 4 つの目的を掲げ、その実現を目指した。各項目について結果をまとめる。

音楽の特徴を表現するパラメータの抽出 楽曲の特徴パラメータとして音長および音高を採用し、和声については最低音のみを特徴抽出の対象とした。しかし、被験者によっては、この特徴量では音楽嗜好を十分に表現できなかったものもある。従って、更に改良を加える必要があり、特に長周期の変動と和声の特徴抽出はその必要性が高いといえる。

構文知識を用いずに音楽嗜好をモデル化 HMM を用いて音楽的な構文知識によらない音楽嗜好モデルの構築を行った。その結果、ある程度は目標を達成できた。しかし、すべての対象について、頑健な結果を得ることができなかった。また、各種のモデルパラメータ調整法を試みたものの、その効果は限定的であった。これは特徴抽出法に問題を含んでいるためだと考えている。

音楽嗜好における個人性の表現 個人ごとに HMM を学習させて音楽嗜好モデルの構築を行い、被験者によって相関係数が大きくばらつく結果を得た。よって、すべての被験者について、音楽嗜好を表現するという目標は達成できてはいない。特徴抽出および重み付き学習の手法が適切であったかどうか、更に検討する必要がある。

複数の未知曲から好みの曲を検索 未知曲から好みの曲を検索するためには、オープンテストの結果が良好で

なければならない。従って、現時点ではこの目的は実現できていない。本目的を実現するためには、先に挙げた 3 つの目的を達成することが先決である。それらが解決できれば本項の達成も可能だと考えている。

7. おわりに

本研究プロジェクトでは、被験者の楽曲評価により HMM を重み付け学習させ、HMM により被験者の音楽嗜好を表現した。しかし、嗜好を十分に表現できておらず、被験者によって結果には大きなばらつきがあった。

今後の課題としては、特徴抽出手法の改良がある。被験者によって、嗜好を判断する際に聴いている部分が異なる可能性があり、それを考慮した特徴抽出法を検討する必要がある。そのためには、音楽認知心理学の専門家とともに、今まであまり研究の対象となつてこなかった認知の個人性について、調査することが望ましい。また、パターンマッチングに確率モデルを用いる以上、大量の学習データが必要不可欠である。音声認識の分野では、複数の HMM 間で状態や確率分布を共有するなど、学習データの不足を補う枠組みが確立している。それに習い、少数の学習データでも効率よく学習できる手法を模索したい。

8. 参加企業及び機関

財団法人 日本産業技術振興協会

参考文献

- [1] 岡田顕宏・阿部純一: コンピュータと音楽の世界, p199-215, 共立出版, 東京 (1999)
- [2] 難波精一郎・桑野園子: 音の評価のための心理学測定法, p72-86, コロナ社, 東京 (1998)
- [3] L. Rabiner and B. H. Juang: Fundamentals of Speech Recognition, Englewood(1993), 古井貞熙 監訳, 音声認識の基礎 (下), p102-187, NTT アドバンステクノロジー, 東京 (1994)
- [4] <http://htk.eng.cam.ac.uk/>
- [5] S. Young, *et al*: The HTK BOOK, Cambridge University, 2000.
- [6] 山根裕明 他: 印象表言語による音楽検索システムの開発, 電子情報通信学会総合全国大会講演論文集, D-14-14, p173, 2002.