

音楽感性空間を用いた感性語による音楽データベース検索システム

池 添 剛[†] 梶 川 嘉 延[†] 野 村 康 雄[†]

本論文では、感性語により音楽を検索するシステムを提案する。提案システムでは、データベース中の曲をマッピングするための検索空間を SD 法ならびに因子分析により生成する。また、データベースに新たな曲を登録する際には、GA とニューラルネットワークにより構成された自動インデクシングシステムにより検索空間へのマッピングを行う。検索の際には、8 つの感性語対の度合い(1-7)をニューラルネットワークに入力することにより、ニューラルネットワークはそれらの入力に対応する感性空間中の座標を出力するので、検索システムはその出力座標値からユークリッド距離の最も近い曲から順番に検索候補としてユーザに提示を行う。提案システムに対して主観評価実験を行ったところ、検索システムに関しては 95% の被験者が満足しており、自動インデクシングシステムに関しては 59% の被験者が満足するという良好な結果が得られた。

Music Database Retrieval System with Sensitivity Words Using Music Sensitivity Space

TAKESHI IKEZOE,[†] YOSHINOBU KAJIKAWA[†] and YASUO NOMURA[†]

In this paper, we propose a music retrieval system with KANSEI words. In the proposed system, retrieval space in which tunes are mapped is generated by semantic differential methods and factor analyses. When a new tune is registered, the tune is mapped in the retrieval space by an automatic indexing system using genetic algorithms and neural networks. In case of retrieval, arbitrary levels of 8 pairs of KANSEI words are input into a neural network, the neural network outputs the corresponding coordinate value in the retrieval space. Then the proposed retrieval systems show a user some candidate tunes in near order of distance from the output value. The subjectivity evaluation experiments to the proposed system demonstrate the good result that 95% of subject was satisfied about the retrieval system and 59% of subject was satisfied about the automatic indexing system.

1. はじめに

近年、インターネットや MIDI (Musical Instrument Digital Interface)、通信カラオケなどの普及にともない、デジタル情報として蓄積される音楽情報が急増している。そのため、音楽データベースから効率良く検索を行うことができるシステムの要求が高まっている。そのような検索システムの 1 つとして、音楽データベースから自分の思い描く曲を探し出す感性語による音楽データベース検索システム^{1)~5)} が提案されている。また、類似研究として絵画データベースを対象とした感性語による検索システム^{6)~14)} も提案されている。本論文では、音楽データベースに対する感性語による検索システムに焦点をあてて議論を進める。また、対象となる音楽データとしては Standard MIDI

File (以下、SMF) を用いることにする。

感性語による音楽検索を考えた場合、「悲しい」や「楽しい」などの感性語を検索キーとして検索を行うことが想定されるが、同じ「悲しい」でもその程度によって検索したい曲も自ずと異なってくると考えられる。よって、検索キーとなる感性語の程度を入力でき、かつそれを反映した検索結果を提示する検索システムの構築は重要であると考え、そのような検索システムの構築を目指す。感性語の程度を表す場合、感性語に対してその程度を表す副詞(非常に、ややなど)を付け加えた検索キーを用意することも考えられるが、本論文では感性語の対を用意し、その割合を検索キーとするようなシステムを構築する。なぜなら、感性語による検索システムを構築する際に重要となる検索空間を提案システムでは SD 法による主観評価実験を通じて生成しているからである。

また、検索方法としては感性語対の程度を検索キーとして入力し、その検索キー値とデータベース中の各

[†] 関西大学
Kansai University

曲に対応する感性語対の値との距離を求め、最も距離の近いものから提示する方法が考えられる。すなわち、検索空間として感性語対の値そのものを利用する方法がある。しかし、我々は検索空間として、SD法による実験結果に対して因子分析を行った結果得られる因子空間（以下、音楽感性空間と呼ぶ^{15),16)}）を利用した。なぜなら、因子空間にマッピングされた曲がどのような印象を持つ曲であるのかを直感的に理解しやすいと考えたからである。また、検索候補を選ぶ際に必要となるユークリッド距離の計算に対して因子空間を用いれば各因子軸は直交しているため、補正などを加える必要がないからである（感性語対をそのまま利用して距離計算を行う場合には補正が必要となる場合が発生し、その補正方法によって検索性能が左右される恐れがある）。そして、本検索システムでは感性語対の検索キー値を検索空間における座標値に変換するためにニューラルネットワークを用いている。

次に感性語による検索システムとして重要となるのは、新たな曲をデータベースに登録する方法（インデクシング）である。先にも述べたように検索空間はSD法による主観評価実験を通じて生成されているため、本質的には新規登録の際には改めて主観評価実験が必要となるが、それは現実的であるとはいえない。よって、登録したい曲の特徴から検索空間中の座標を決定する、いわゆる自動インデクシングシステムは検索システムにおいては必須となる。そこで、提案する検索システムにおいても自動インデクシングシステムを内在している。提案する自動インデクシングシステムでは登録曲のSMFから得られる特徴量をニューラルネットワークに入力することにより検索空間の座標値（予測値）を決定する。また、SMFから得られる特徴量の中から曲の印象に関連が深いものを抽出するためにあらかじめGAによって必要となる特徴量の選定を行っている。その際、曲特徴量と曲の印象との関係は非線形であると仮定し、それらをニューラルネットワークで関連付け、特徴量の選別の際にも2次の非線形性をGAに導入している。

ここで、感性語対の程度を検索キーとして検索を行うシステムとしては野田らのシステム³⁾ならびに佐藤らのシステム⁴⁾がある。野田らのシステムと提案システムとの違いは検索空間と自動インデクシングシステムにおける特徴量の選定法にある。野田らのシステムは検索キー値とデータベース中の各曲に対応する感性語対の値との距離を求め、最も距離の近いものから提示する方法を採用している。よって、すでに述べたように距離を求める際には検索性能を上げるために補正

を必要とする可能性がある。また、曲の特徴量に関しては音高のみに着目しており、その選定も著者自身が行っている点で提案システムとは異なる。一方、佐藤らのシステムとの違いは、自動インデクシングシステムの作成法にある。佐藤らのシステムでは曲の特徴量の選定を先行研究での結果ならびに著者ら自身の考えに基づき行っている。それに対して提案システムではSMFから得られる特徴量の中から曲の印象に関連のあるものをGAによって自動的に選定している。また、佐藤らのシステムでは特徴量と曲印象とを重回帰分析によって関連付けているのに対して、提案システムではニューラルネットワークによって関連付けている点でも違いがある。

提案システムの評価においては、検索性能ならびに自動インデクシング性能を数値評価ならびに主観評価を通じて検証し、また、主観評価の結果に対して客観的解析も行うことで提案システムの有用性を検証している。

本論文の構成は以下のようになっている。2章では提案システムの概要について述べ、3章では検索システム部について詳細に説明する。また、検索システムの有効性を検証するために行った実験の結果を示し、その考察を行う。そして、4章では自動インデクシングシステム部について詳細に説明するとともに実験の結果を示し、その考察を行う。5章では、まとめと今後の展望について述べる。

2. システム概要

本章では、提案システムの概要について説明する。図1が提案システムの構成である。このシステムは、大きく分けて「音楽感性空間生成部」と「検索システム部」の2つの部分から構成される。

「音楽感性空間生成部」では、検索空間となる音楽感性空間を生成する。その方法は、まず、音楽データベースに登録したい曲を用いて、SD法による試聴実験を行う。そして、その結果に因子分析を施すことにより、各因子を軸とした、曲印象の心理的な相関を表す空間を生成する。この空間が音楽感性空間であり、本論文では検索空間として用いる。なお、この方法で生成された音楽感性空間座標値を実際値と呼ぶことにする。

このようにして、検索空間となる音楽感性空間が生成される。しかし、音楽データベースが大規模になると、すべての曲に対してSD法で曲印象を数値化することが非常に困難となる。そこで、音楽データベースに追加したい曲を入力することによりその曲の音楽感

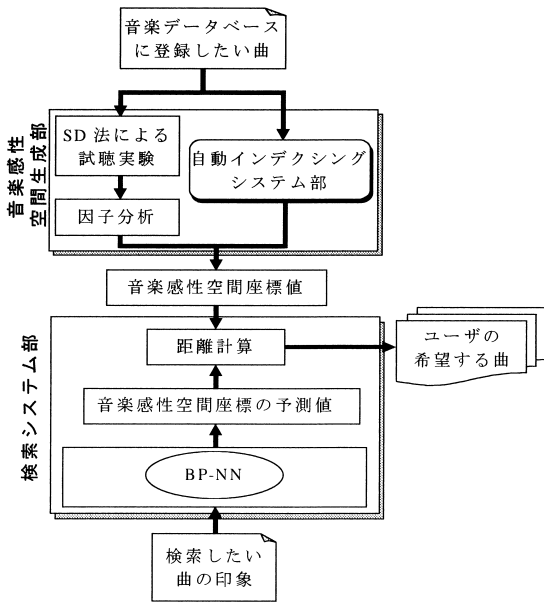


図1 提案システムの概要
Fig. 1 Outline of proposed system.

性空間座標値を自動的に求める、自動インデクシングシステム部を用意しておく。

「検索システム部」は、入力された曲印象に見合う曲を音楽データベースから探し、その曲を出力する部分である。詳細は、次章で説明する。なお、検索システム部に入力する感性語をSD法により数値化されたものとする。つまり、「非常に」や「やや」といった各感性語の細かいニュアンスによる検索を可能としている。

3. 検索システム部

本章では、検索システム部について詳細に説明する。また、検索システムの有効性を検証するために行った実験の結果を示し、その考察を行う。

3.1 検索システム部の説明

図2に検索システム部の構成を示す。このシステム部では、誤差逆伝播ニューラルネットワーク(以下、BP-NN)を用いて入力された印象に対する音楽感性空間座標の予測値(以下、予測値)を求めることにより検索を行い、その候補曲を出力する。その方法は、まず、SD法により数値化された曲印象を入力信号、実際値を教師信号としてBP-NNで学習を行う。そして、学習したBP-NNに検索したい曲印象を入力することにより、入力された曲印象に対する予測値を求める。ここで、音楽感性空間は因子分析の解であるため、音楽感性空間上では曲印象が似た曲は近くに、逆に異なる

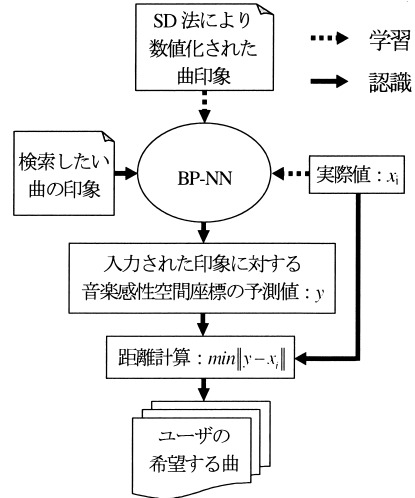


図2 検索システム部の構成
Fig. 2 Organization of retrieval system part.

表1 試聴実験条件
Table 1 Term of experiment.

聴取曲	75曲(ジャンル: Classic, Jazz, Ragtime, etc.)
SD法による評価指標	すべて7段階評価で、用いた感性語は、 明るい—暗い 重い—軽い かたい—やわらかい 安定—不安定 澄んだ—にごった 滑らか—歯切れの良い 激しい—穏やか 厚い—薄い の8種類

る曲は遠くに位置する。そこで、この予測値と各曲の実際値とのユークリッド距離を求め、その距離が小さい実際値の曲から順に候補としてシステムが曲を出力する。

3.2 試聴実験による音楽感性空間の生成

ここでは、検索システム部の有効性に関する検証を行うために必要となる音楽感性空間を生成した結果を示す。

まず、今回行ったSD法による試聴実験の条件について説明する。実験を行うにあたり、被験者、感性語、楽曲を決める必要がある。そこで、聴取サンプルとして音楽経験が十分にある、関西大学の交響楽団部員50人に対して、表1に示す感性語、楽曲を用いて実験を行った。なお、楽曲のジャンルは、クラシック、ジャズ、ラグタイムなどである。ここで、表1における感性語は、心理学の分野における音楽を評価するための尺度に関する研究¹⁷⁾の研究成果に基づいて選出した。具体的には、同論文では音楽を評価する際に適切

表 2 呈示曲
Table 2 Presentation tune.

グループ	呈示曲
1	1 曲目 ~ 30 曲目
2	16 曲目 ~ 45 曲目
3	31 曲目 ~ 60 曲目
4	46 曲目 ~ 75 曲目
5	1 曲目 ~ 15 曲目 61 曲目 ~ 75 曲目

である感性語とその主成分負荷量が示されているので、各主成分から主成分負荷量の大きい上位 2 つの感性語を選び出し、実験に使用した。ただし、今回はより多くの人が満足するシステムの構築を目指しているため「好き-嫌い」といった個人差が激しいと考えられる感性語は省いた。また、今回使用した楽曲は SMF (Standard MIDI File) 形式のものであり、市販されているものやインターネットで入手可能なものである。ただし、音色による効果をなくすために、ピアノ曲のみを用いた。

次に、実験方法について説明する。まず、会議室内で YAMAHA P-300 の内蔵 MIDI 音源、内蔵スピーカを用いて被験者に楽曲を呈示し、1 曲が終了した時点で 30 秒以内に評価を行ってもらう。また、各曲について 20 人のデータが得られるように被験者を 10 人ずつ 5 つのグループに分け、各グループごとに表 2 に示す楽曲を呈示した。また、楽曲および感性語の順序効果をなくすために、各グループに対して楽曲および感性語の呈示順序をランダムに変えて実験を行った。なお、我々はより多くの人が満足するシステムの構築を目指しているため、実験データに異常値が存在する場合はその異常値を除外する必要がある。そのような方法として、異常値を手作業で取り除いてから平均をとることが一般的であるといえるが、本論文では異常値を取り除くという手間を省くために 20 人の中央値をとり実験結果を求めた。以上の条件および方法で実験を行って得られた結果に、主因子法による因子分析¹⁸⁾および基準化バリマックス回転¹⁸⁾を施したものが図 3 である。図 3 を見ると、各因子において各曲が全体的に散らばっていることが確認できる。このことから、今回用いた楽曲はバラエティに富んでいるといえる。また、因子を第 5 因子まで求めた理由は、第 6 因子以降の寄与率が小さいためである。ここで、寄与率とは各因子が元の情報をどれくらい説明しているかを表す指標である。そして、第 5 因子までの累積寄与率が約 95% であることより、この音楽感性空間で曲印象の大半が説明できるといえる。なお、因子の解釈をすると第 1 因子においては「明るい-暗い」の因子負荷量が

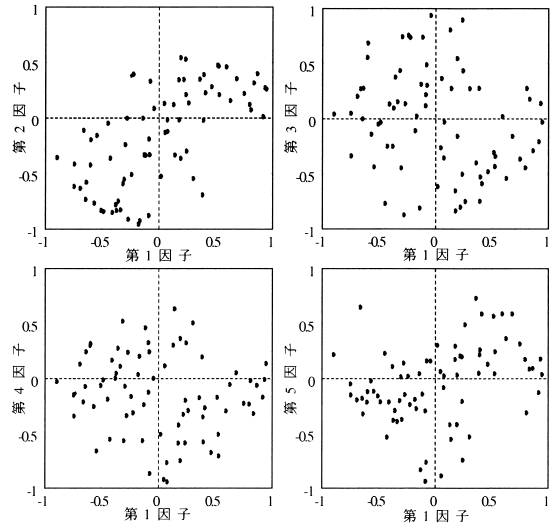


図 3 音楽感性空間座標値

Fig. 3 Music sensitivity space coordinate value.

大きいことから「明度」、第 2 因子においては「激しい-穏やか」の因子負荷量が大きいことから「力量」、第 3 因子では「安定-不安定」の因子負荷量が大きいことから「安定度」、第 4 因子は「重い-軽い」の因子負荷量が大きいことから「質量」、第 5 因子は「滑らか-歯切れの良い」の因子負荷量が大きいことから「躍動性」とそれぞれ解釈した。

3.3 検索システム部の評価

ここでは、前節で生成した音楽感性空間を用いて検索システム部の構築を行い、その有効性に関する検証を行う。

まず、実験条件について説明する。用いた BP-NN は 8 対の感性語を入力する 8 入力、5 因子に対応する 5 出力を持つ 1 つの NN からなる。BP-NN の学習は学習係数 0.1、慣性係数 0.9 で行った。また、試聴実験を行った 75 曲の中からランダムに選び出した 50 曲を学習曲、残りの 25 曲を未知曲とした。

この条件で検索システム部を構築し、試聴実験の結果得られた未知曲の評価値を入力した際の予測値を求めた。その予測精度を表 3 に示す。なお、表中の BP-NN において 20-10 とは、中間層が 2 層で、第 1 層目が 20 ユニット、第 2 層目が 10 ユニットからなるネットワークを表し、相関係数は実際値と予測値の相関係数であり、平均 2 乗誤差は空間内の距離を表している。また、参考のために BP-NN の代わりに重回帰分析を用いた場合の結果も示している。表 3 を見ると、検索システム部が高い精度で予測できていることが確認できる。また、BP-NN をどのような規模に変化させても安定して高い予測精度を得ていることが分

表 3 検索システム部の予測精度

Table 3 Prediction precision of retrieval system part.

		相関係数	平均二乗誤差
重回帰分析		0.954	1.54×10^{-2}
BP-NN	5	0.983	6.16×10^{-3}
	5-5	0.989	3.96×10^{-3}
	7	0.989	4.31×10^{-3}
	7-5	0.991	3.12×10^{-3}
	7-7	0.990	3.66×10^{-3}
	10	0.991	3.46×10^{-3}
	10-5	0.992	2.98×10^{-3}
	10-10	0.994	2.42×10^{-3}
	15	0.989	3.99×10^{-3}
	15-5	0.991	3.41×10^{-3}
	15-10	0.992	2.81×10^{-3}
	15-15	0.993	2.88×10^{-3}
	20	0.988	4.53×10^{-3}
	20-10	0.991	3.21×10^{-3}
	20-20	0.991	3.37×10^{-3}

表 4 検索システム部の検索精度

Table 4 Retrieval precision of retrieval system part.

		第 1 候補		第 2 候補	
		正解数	正解率 [%]	正解数	正解率 [%]
重回帰分析		17/25	68	22/25	88
BP-NN	5	25/25	100	-	-
	5-5	25/25	100	-	-
	7	25/25	100	-	-
	7-5	25/25	100	-	-
	7-7	25/25	100	-	-
	10	25/25	100	-	-
	10-5	25/25	100	-	-
	10-10	25/25	100	-	-
	15	24/25	96	25/25	100
	15-5	25/25	100	-	-
	15-10	25/25	100	-	-
	15-15	25/25	100	-	-
	20	24/25	96	25/25	100
	20-10	25/25	100	-	-
	20-20	25/25	100	-	-

かる。しかし、この予測精度がどこまで高ければ、検索システムとして良いシステムなのかを評価することができない。そこで、その予測値を用いて実際に検索システム部で検索を行う。その結果を表 4 に示す。なお、特定の 1 曲に対して試聴実験で得られた曲印象を検索システム部に入力した際に、その曲が検索されれば正解とした。また、実験に用いたデータベースは、試聴実験を行った 75 曲である。表 4 を見ると、どの規模の BP-NN を用いても高い正解率を得ていることが分かる。また、検索システム部では最低でも第 1 候補で 96%、第 2 候補で 100% の正解率を得ており、安定して高い正解率を得ていることも分かる。なお、学習

曲と未知曲の組合せを変えて同様の実験を 5 回行ったが、結果は同等であった。ここで、曲印象と感性語の関係が線型的であるかを調べるため、図 2 の BP-NN を重回帰分析にかえ同様の実験を行った結果も表 4 に示す。これより重回帰分析では、検索精度が悪いことが分かる。よって、このことから我々が扱うシステムにおいて曲印象と感性語の関係は線型的でなく、非線型的な関係であると考えられる。

このように検索システム部の有効性が確認できる。しかし、検索システム部において第 1 候補での正解率が必ずしも 100% ではない。そこで、検索システム部における不正解曲についての考察を行う。第 1 候補で 100% の正解率を得ていない BP-NN の規模は 15 と 20 であるが、それぞれ 41 曲目 (Escovado) を 36 曲目 (Pastime Rag No.1), 59 曲目 (The Aeroplane) を 21 曲目 (The Thunderer) と誤認識していた。なお、36 曲目は学習曲である。そこで、これらの曲を音楽感性空間上で眺めてみたところ、誤認識を起こしていた曲どうしは他の曲に比べて非常に近い距離関係にあった。このことより、曲印象の非常に近いところで誤認識を起こしているといえる。なお、実際にそれらの曲を聞き比べたところ、同じような印象を与える曲であった。ここで、我々の提案するシステムの目的は、特定の曲を検索することではなく、入力された印象に見合う曲を検索することである。そのため、この誤認識は検索システム部として何ら問題がないといえる。つまり、数値的には検索システム部は非常に有効であるといえる。

しかし、実際のユーザが満足できるかを検証する、実際の設定での評価実験を行う必要がある。そこで、今回試作したシステムを実際のユーザに利用してもらうことにより、検索システム部の主観的評価を行った。図 4 に試作システムのユーザインタフェースを示す。このシステムでは、ユーザが検索したい曲の印象を図 4 のユーザインタフェースを通じて入力するとそれに見合った曲の検索を行い、第 1 候補として検索された曲を再生する。主観的評価実験では、被験者に図 4 に示す検索システムを用いて 10 回検索してもらい、各回に第 1 候補として検索された曲が自分の入力した印象に対してどれだけ満足できたか (曲別評価) を 5 段階で判定してもらい、そして、実験後に、利用してもらったシステムに対して全体的にどれだけ満足できたか (総合評価) を 5 段階で判定してもらい、ここで、満足度により評価した理由は、システムの性質上、実際に利用するユーザが満足できるかどうか最も重要と考えたためである。以上の方法により、検索



図4 検索システム

Fig. 4 The retrieval system part.

表5 検索システム部の主観的評価実験結果

Table 5 Subjective evaluation result of retrieval system part.

満足度	曲別評価			総合評価		
	曲数 [曲]	割合 [%]	評価 割合	人数 [人]	割合 [%]	評価 割合
5	66/200	33.0	多 (0.5)	6/20	30	多 (5.0)
4	81/200	40.5	多 (0.5)	13/20	65	多 (0.5)
3	27/200	13.5	少 (0.5)	1/20	5	少 (2.5)
2	24/200	12.0	少 (0.5)	0/20	0	少 (1.0)
1	2/200	1.0	少 (0.5)	0/20	0	少 (1.0)

システム部の主観的評価を行うことができる。なお、今回は関西大学工学部電子工学科に所属する学生 20 人を被験者とした。また、検索精度が一番高い、10-10 の BP-NN を用いた検索システム部で実験を行った。

この条件で実験を行った結果が表 5 である。なお、満足度において 5 が非常に満足、1 が非常に不満足を表している。表 5 の曲別評価を見ると、曲によってばらつきはあるが、73.5% の曲で満足という評価を得ていることが分かる。また、総合評価を見ると、95% の人で満足という評価を得ており、全体的にはほぼ満足しているといえる。ここで、曲別評価、総合評価それぞれにおける各満足度の評価に対して、評価人数が多い、少ないということを客観的に評価するために、 χ^2 検定¹⁹⁾ を用いて満足度の有意性を調べる。その具体的な方法は、まず満足度の各評価それぞれの割合が一定であるという仮説を立てる。そして、満足度の各評価に対して、上記の仮説が成立するかどうかを χ^2 検定を用いて判断する。その結果、上記の仮説が棄却された場合、その評価割合は全体に対して多い、あるいは少ないと判断される。この操作を満足度のすべての

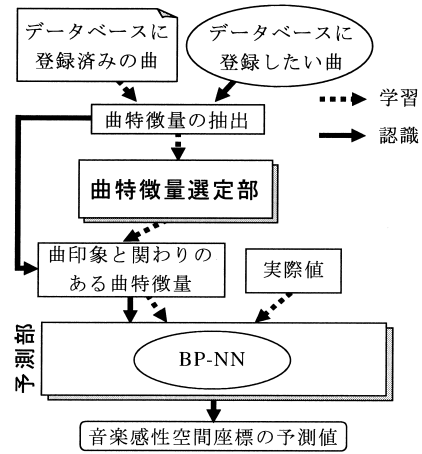


図5 自動インデクシング部の構成

Fig. 5 Organization of automatic indexing system part.

評価に対して行う。その結果が表 5 における評価割合である。ただし、「多」とはその評価をした人の割合が全体に対して多いことを表しており、逆に「少」は少ないことを表している。また、括弧内の数値は検定における有意水準のパーセンテージを表しており、検定力を高めるために有意水準を 5% から順に小さくしていった結果得られた最小の有意水準を示している。つまり、有意水準のパーセンテージが小さいほど有意である度合いが大きいことを表している。この結果を見ると、曲別評価および総合評価において、不満足という評価をした人の割合は少なく、逆に満足と評価した人の割合が多いことが分かる。以上のことより、BP-NN を用いた検索システム部は利用者の立場から有効であるといえる。

4. 自動インデクシングシステム部

本章では、自動インデクシングシステム部について詳細に説明するとともに実験の結果を示し、その考察を行う。

4.1 自動インデクシングシステム部の説明

図 5 に自動インデクシングシステム部の構成を示す。このシステム部は「曲特徴量選定部」と「予測部」の 2 つからなる。「曲特徴量選定部」では、GA と重回帰分析を用いて曲印象にかかわりのある曲特徴量を選定する。「予測部」では、「曲特徴量選定部」で選出された曲特徴量を用いて BP-NN で音楽データベースに登録したい曲の音楽感性空間座標上での予測値を求める。

「曲特徴量選定部」の処理手順を図 6 に示す。まず、GA を用いて曲特徴量を選定するため、曲特徴量の組合せを遺伝子としてコーディングする。そのコーディ

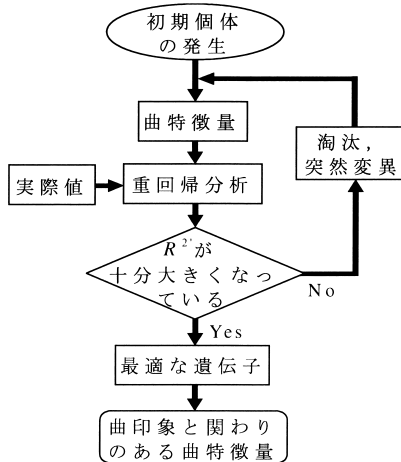
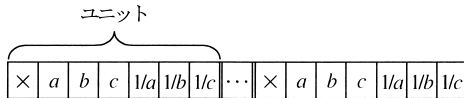


図 6 GA を用いた曲特微量の選定方法

Fig. 6 Selection of music feature parameters using the GA.



- Case1 0 0 0 0 0 0 0 → 入力なし
- Case2 0 0 0 1 0 0 0 → c を入力
- Case3 1 1 0 0 0 1 0 → a×1/b を入力
- Case4 1 0 1 0 0 0 0 → b² を入力

図 7 遺伝子の表記方法

Fig. 7 Notation of gene.

ング方法を図 7 に示す。なお、図 7 は曲特微量が a, b, c の 3 種類の場合の例である。遺伝子は 1 つの曲特微量を表す「ユニット」が複数つながった形で表現される。つまり、ユニットがいくつもつながることにより曲特微量の組合せを表現する。ここで、このユニットで 1 つの曲特微量をどのように表すのかについて説明する。このユニットは、曲特微量の積を用いるかどうかを決める「 x 」セルと、それぞれの曲特微量に対応するセル、それぞれの曲特微量の逆数に対応するセルから構成される。まず、「 x 」セルに 0 あるいは 1 を代入する。もし「 x 」セルに 0 が代入されていれば、残りのセルのうちただか 1 つのセルに 1 を代入し、それ以外のセルには 0 を代入する。その結果すべてのセルが 0 ならば、そのユニットには対応する曲特微量がないことを表す (図 7 Case1)。逆に、1 つのセルに 1 が代入されている場合は、そのセルに対応する曲特微量をそのユニットが表す曲特微量とする (図 7 Case2)。また、「 x 」セルに 1 が代入されている場合は、残りのセルのうち、1 つあるいは 2 つのセ

ルに 1 を代入し、それ以外のセルには 0 を代入する。その結果、2 つのセルに 1 が代入された場合は、それぞれのセルに対応する曲特微量の積をそのユニットが表す曲特微量とする (図 7 Case3)。逆に、1 つのセルに 1 が代入された場合は、そのセルに対応する曲特微量の自乗をそのユニットが表す曲特微量とする (図 7 Case4)。このような方法により、曲特微量の組合せをコーディングすることができる。次に、このように表された遺伝子のうち、どの遺伝子が良い遺伝子なのかを決めるために、各遺伝子に対して適応度の計算を行わなければならない。そこで、遺伝子型を表現型に変換した曲特微量を説明変数、実際値を目的変数として重回帰分析を行い、その結果求めた自由度修正済決定係数²⁰⁾ ($R^{2'}$) を適応度を用いる。なお、自由度修正済決定係数は次式で計算される。

$$R^{2'} = 1 - \frac{S_E / (n - m - 1)}{S_{yy} / (n - 1)} \quad (1)$$

ただし、 S_E, S_{yy} はそれぞれ、実際値と予測値の残差平方和、目的変数の偏差平方和を表し、 n, m はデータの数、説明変数の数を表す。この適応度が大きくなるように世代交代を繰り返すことにより、曲印象にかわりのある曲特微量の組合せが選出される。

「予測部」では「曲特微量選定部」で選び出された曲特微量を用いて、BP-NN で予測値を求める。BP-NN は各因子において個別に NN を用意する。すなわち、5 つの NN を用いている。その方法は、曲印象にかわりのある曲特微量を入力信号、実際値を教師信号として BP-NN で学習を行う。そして、音楽データベースに登録したい曲の曲特微量を学習した BP-NN に入力することで、その曲の音楽感性空間座標上での予測位置が求まる。

このようにして、試聴実験に用いていない曲の音楽感性空間上での予測値を求めることができる。また、図 6 では GA により探索を行う手順を示しており、図 7 から分かるように、遺伝子で表現された曲特微量は最大二次の非線型性を持つことになる。この遺伝子を用いて図 6 に示したように GA で探索を行うので、その結果選出される曲特微量は二次の非線型性までを考慮したものになる。ここで、二次の非線型性までを考慮した曲特微量を選出した理由について説明する。まず、三次以上の高次の非線型性を考慮した曲特微量を選び出すには、二次の場合と同様に、遺伝子の表記方法を工夫すれば可能である。しかし、このよう

各因子ごとに NN を用いた理由は、5 因子に対して 1 つの NN では十分な精度が得られなかったためである。

に改良すると考慮する次数が高くなるにつれて探索領域が指数関数的に増大するため GA で正しく探索が行えない可能性があると考えられる。また、図 5 における「曲特微量選定部」から線型の曲特微量が予測部の BP-NN に入力されると曲特微量と曲印象の非線型性が強すぎるために、学習が十分に行えない可能性がある。このことは、BP-NN の入力自身にあらかじめ非線型性を持たせることで、ニューラルネットで学習できると考えられる。これらのことから、曲特微量の組合せは非線型性を考慮する必要があるが、非線型性の次数はできるだけ低い方が良いといえる。そこで今回は、GA での探索領域との関係上「曲特微量選定部」では二次の非線型性までを考慮した曲特微量を選出することとした。

4.2 自動インデクシングシステム部の評価

ここでは、3.2 節で生成した音楽感性空間を用いて自動インデクシングシステム部の構築を行い、その有効性に関する検証を行う。

まず、実験条件について説明する。システムに入力された SMF から抽出する曲特微量は、文献 21)~24) を参考にして、表 6 に示すものを用いた。なお、表中の低音、中音、高音とはそれぞれ、音高が B2 以下の音、C3 から B4 までの音、C5 以上の音を表す(ただし、中央のドは C3²³⁾)。また、Diatonic, Rhythm, Key, Meter の計算方法は付録に詳述している。さらに、GA で用いるパラメータは、個体数 20, 30, 50 の 3 パターン、交叉確率 1.0, 突然変異率 0.05, 探索終了条件は世代数が 100000 に達したときとし、エリート保存戦略、二点交叉を用いた。そして、選出する曲特微量の組合せの最大数、つまり、遺伝子のユニット数は 21 とした。

この条件で GA による探索を行った際の適応度の変化を図 8 に示す。なお、図 8 は第 1 因子において個体数 50 で探索した場合の結果である。また、平均適応度は初期個体を変えて同様の実験を 10 回行った結果得られたそれぞれの平均適応度を平均したものである。図 8 を見ると、世代数が増えるにつれて最大適応度が 1 に近づいていることが確認できる。また、平均適応度は世代数が大きくなるにつれて大きくなっており、100000 世代ではほぼ収束していることが分かる。これらのことから、GA による探索は十分できていると考えられる。なお、その他の因子においても十分探索できていることが確認できた。そして、その結

表 6 曲特微量

Table 6 Music feature parameters.

曲特微量	意味
Ave(S)-Pitch*	音の高さの平均値 (分散値)
Ave(S)-Velocity*	音の強さの平均値 (分散値)
Ave(S)-Duration	音の長さの平均値 (分散値)
Rest	音の鳴っていない時間の割合
Ave(S)-Density	音数の平均値 (分散値)
Differential	平均音高からの絶対値の平均値
Register	音高の最大値と最小値の差
Diatonic**	音の調和度
Ave(S)-Rhythm**	リズムの平均値 (分散値)
Key**	曲の調性
Meter**	曲の拍子
Tempo**	曲のテンポの初期設定値
Low(Middle)[High]-Number	低音 (中音)[高音] の音数の割合
Low(Middle)[High]-Time	低音 (中音)[高音] が鳴っている時間の割合

*: 音の長さを重みとして掛け合わせて計算される

** : SMF のメタデータからの情報を用いて計算される

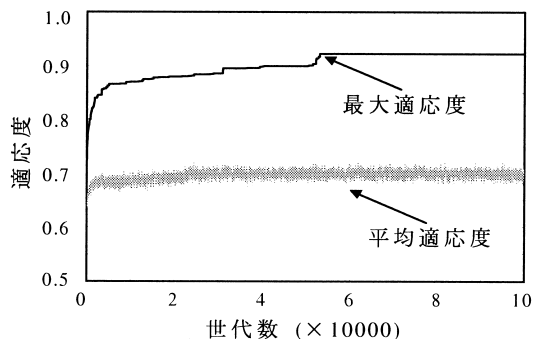


図 8 GA による探索状況

Fig. 8 Searching properties of the GA.

果選出された曲特微量が表 7 である。また、曲特微量に (逆) がついているものは、その曲特微量の逆数であることを表している。さらに、表中の大括弧内の数値は標準偏回帰係数²⁰⁾と呼ばれるものである。つまり、この数値の絶対値の値が大きい曲特微量ほど、その因子に与える影響が大きいことを表している。表 7 では、この標準偏回帰係数が大きいものから順に 4 つのみを示している。

次に、この選出された曲特微量を用いて実際にインデクシングを行う。自動インデクシングシステム部による未知曲の音楽感性空間座標値の予測精度を表 8 に示す。表中の最大平方距離とは 2 乗誤差の最大値のことである。ただし、表 8 は 3.3 節と同じ学習曲、未知曲を用いた場合の結果である。また、各因子において一番学習精度が高い規模の BP-NN を用いた場合の結果である。表 8 を見ると、高い精度でインデクシングできていることが分かる。しかし、この精度が検索システムにどのような影響を与えるのかが分からない

遺伝子のユニット数を 21 より多くして評価してみたが、適応度が求まらない状態になり、GA で探索できなくなった。

表 7 GA で選出された曲特徴量

Table 7 Music feature parameters selected by the GA.

	選出された曲特徴量	
第 1 因子	Ave-Duration×Ave-Rhythm	[2.28]
	Ave-Velocity(逆)×High-Number(逆)	[1.20]
	Ave-Duration(逆)×High-Number(逆)	[-1.17]
	Register の自乗	[-0.93]
	など 20 種類	
第 2 因子	Ave-Density の自乗	[2.22]
	Ave-Density×Harmony	[1.50]
	Ave-Duration の自乗	[-1.18]
	Ave-Velocity×S-Duration	[1.15]
	など 21 種類	
第 3 因子	S-Density×Cadence(逆)	[1.09]
	Ave-Duration×Tempo	[-0.92]
	Key×Ave-Rhythm	[-0.89]
	S-Pitch(逆)×Resister	[0.75]
	など 21 種類	
第 4 因子	Ave-Pitch×Key(逆)	[1.24]
	Rest(逆)×Key	[1.12]
	Rest(逆) の自乗	[-0.91]
	S-Pitch(逆)×Ave-Rhythm(逆)	[0.86]
	など 21 種類	
第 5 因子	Key×Ave-Rhythm(逆)	[-5.69]
	Key の自乗	[4.78]
	Ave-Rhythm の自乗	[-1.96]
	Ave-Velocity×Key(逆)	[1.44]
	など 21 種類	

表 8 自動インデクシングシステム部の予測精度

Table 8 Prediction precision of automatic indexing system part.

	相関係数	平均二乗誤差	予測値と実際値の最大平方距離
重回帰分析	0.9247	3.193×10^{-2}	0.2622(38 曲目)
BP-NN	0.9887	4.026×10^{-3}	0.1298(69 曲目)

表 9 自動インデクシングシステム部の検索精度

Table 9 Retrieval precision of automatic indexing system part.

	第 1 候補		第 2 候補	
	正解数	正解率 [%]	正解数	正解率 [%]
重回帰分析	58/75	77.3	65/75	86.7
BP-NN	72/75	96.0	75/75	100

い。そこで、それぞれのインデクシング結果を用いて実際に検索システム部で検索を行う。その結果が表 9 である。なお、表中で使用している検索システム部における BP-NN の規模は、実際値を用いた検索時に最も精度が高かった 10-10 である。また、表 9 では学習曲と未知曲をあわせた結果を示しているが、その理由はインデクシング精度の悪さは学習曲に対する検索結果に顕著に表れるからである。すなわち、自動インデ

クシングシステムによって曲を新たに登録する場合、検索用ニューラルネットワークの出力座標値とそれに対応する曲の座標値(実際値)の間に自動インデクシングシステムによって追加曲が誤ってマッピングされる場合が想定される。そのような場合、そもそも検索用ニューラルネットワークの出力値に対応していた曲とは異なる曲が検索されることになる。このような状況はインデクシング精度が悪ければ悪いほど頻繁に発生すると予測される。そして、そのことは本システムの検索方法からも理解できるように、学習曲に対する検索結果を評価することによって確認することができる。表 9 を見ると、高い正解率を得ていることが分かる。また、第 1 候補で 96.0%、第 2 候補で 100% の正解率を得ており、十分な正解率を得ているといえる。なお、学習曲と未知曲の組合せを変えて同様の実験を 5 回行ったが、結果は同等であった。ここで、曲特徴量と曲印象の関係が線型的であるかを調べるために、図 5 の BP-NN を重回帰分析に置き換えて同様の実験を行った。その結果が表 9 の重回帰分析である。これより重回帰分析では、検索精度が悪いことが分かる。よって、このことから我々が扱うシステムにおいて、曲特徴量と曲印象の関係が線型的でないといえる。

ここで、自動インデクシングシステム部での第 1 候補における不正解曲について考察する。不正解曲は 21 曲目(The Thunderer)、27 曲目(A Ragtime Nightmare)、41 曲目(Escovado)であり、それぞれ 59 曲目(The Aeroplane)、44 曲目(Feitico)、36 曲目(Pastime Rag No.1)と誤認識していた。なお、36 曲目と 44 曲目は学習曲である。そこで、これらの曲を音楽感性空間上で眺めたところ、検索システム部の場合と同様、非常に近い距離関係にあった。そのため、この誤認識は何ら問題がないといえる。つまり、自動インデクシングシステム部は非常に有効であるといえる。

次に、検索システム部でも述べた理由により、自動インデクシングシステム部の主観的評価を行った。その実験方法は、まず、被験者に 3.2 節で作成した 75 曲の音楽データベース(データベース 1)と自動インデクシングシステム部により新たな 86 曲を追加した 161 曲の音楽データベース(データベース 2)それぞれに対して 3.3 節で作成した検索システム部を用いて 10 回検索してもらい、それぞれのデータベースから第 1 候補として検索された曲の曲別評価を 5 段階で判定してもらおう。なお、3.3 節の実験時と同じ被験者であり、データベース 1 とデータベース 2 の検索キーは同一であり、提示順序は被験者によって変えている。

表 10 自動インデクシングシステム部の主観的評価実験結果
Table 10 Subjective evaluation result of automatic indexing system part.

(a) 曲別評価
(a) Evaluation of each tune.

満足度	データベース 1			データベース 2		
	曲数 [曲]	割合 [%]	評価 割合	曲数 [曲]	割合 [%]	評価 割合
5	66	33.0	多 (0.5)	49	24.5	多 (2.5)
4	81	40.5	多 (0.5)	69	34.5	多 (0.5)
3	27	13.5	少 (0.5)	42	21.0	中 (5.0)
2	24	12.0	少 (0.5)	35	17.5	少 (5.0)
1	2	1.0	少 (0.5)	5	2.5	少 (0.5)

(b) 人別評価
(b) Evaluation of each person.

満足度	データベース 1			データベース 2		
	人数 [人]	割合 [%]	評価 割合	人数 [人]	割合 [%]	評価 割合
5.0-4.5	1	5	少 (5.0)	0	0	少 (2.5)
4.5-4.0	9	45	多 (0.5)	4	20	多 (5.0)
4.0-3.5	9	45	多 (0.5)	8	40	多 (0.5)
3.5-3.0	1	5	少 (5.0)	7	35	多 (0.5)
3.0-2.5	0	0	少 (1.0)	1	5	少 (2.5)
2.5-2.0	0	0	少 (1.0)	0	0	少 (2.5)
2.0-1.5	0	0	少 (1.0)	0	0	少 (2.5)
1.5-1.0	0	0	少 (1.0)	0	0	少 (2.5)

その結果、データベース 2 (161 曲) の方が満足である、あるいは、データベース 1 (75 曲) とデータベース 2 (161 曲) との間に満足の差を感じないという結果が得られたならば、自動インデクシングシステム部は精度良くインデクシングを行っているといえる。つまり、データベース 2 (161 曲) では半数以上が新たに登録された曲であるため、インデクシング精度が十分高くなければデータベース 1 (75 曲) を用いた方が満足であるという結果が得られるためである。このような方法により、間接的に自動インデクシングシステム部の主観的評価を行うことができる。

この方法で実験を行った結果が表 10 である。表 10(a) は曲別評価、(b) は人別評価を示している。人別評価とは曲別評価の評定値を各人について平均したものである。まず、表 10(a) を見ると、データベース 1 (75 曲) では 73.5% (13.0%)、データベース 2 (161 曲) では 59.0% (20.0%) の曲が満足 (不満足) と評価されていることが分かる。また、表 10(b) より全体的にデータベース 2 (161 曲) の方がデータベース 1 (75 曲) に比べ 0.5 だけ満足度が低い傾向にあることが分かる。これらのことから、どちらの評価においてもデータベース 2 (161 曲) に比べデータベース 1 (75 曲) の方が、高い満足度を得ていると考えられる。しかし、表 10 を見ても分かるように、データベース

2 (161 曲) の評価においても満足している傾向がうかがえる。そこで、それぞれのデータベースの評価に対して、この結果を客観的に評価するために検索システム部と同様に χ^2 検定を行い、満足度の有意性を調べた。その結果が表 10 中の評価割合である。ただし、「中」とは有意差がなかったことを表している。この結果を見ると、データベース 2 (161 曲) においても人別評価、曲別評価のどちらに対しても不満足と評価した人の割合は少なく、逆に満足と評価した人の割合が多いことが分かる。このことから、客観的にデータベース 2 (161 曲) においても満足しているといえるが、データベース 1 に比べて評価が低いのも事実である。この理由は自動インデクシングシステムの精度が 100% ではないためである。

しかし、データベース 2 における評価もデータベース 1 の評価に比べて支障をきたすほどの劣化は見られない。このことから、自動インデクシングシステム部は利用者の立場においても有効に機能していると考えられる。

5. ま と め

本論文では、従来の感性語を用いた音楽データベース検索システムがかかえている問題点をふまえ、新たな感性語を用いた検索システムの提案を行った。そして、提案システムを用いて実験を行った結果、相関係数や正解率などを用いた数値的評価およびシステムの利用による主観的評価において良好な結果が得られた。このことから、今回提案したシステムは非常に有効であるといえる。しかし、今回提案したシステムには次の点において改善の余地がある。

- (1) 曲特徴量の組合せにおける遺伝子の表記方法。
- (2) 主観的評価における評価方法。
- (3) 個人差に関する検討。

まず、(1) に関しては、曲特徴量の組合せにおいて非線型要因を考える際、曲特徴量どうしのべき乗関係を用いている。しかし、対数を用いれば、べき乗関係はすべて線型化して扱うことができる。その結果、GA の探索効率が向上すると考えられる。また、(2) に関しては、今回は満足度を用いて評価を行ったが、一般的に、感性語による検索の性質上、検索意図自体が曖昧になるため、検索結果の適合率、ヒット率については甘い評価になると考えられる。そのため、検索システムから検索された曲とデータベース中の他の曲とを比較して、検索された曲が最も意図に即したものであるかを答えるといった対策をとる必要があるといえる。さらに、(3) に関しては、4.2 節での考察でも述べた

ように、提案システムは各個人ごとに対応することができない。そのため、BP-NN などを用いて、各個人の音楽感性空間を今回用いた音楽感性空間に変換することにより、提案システムを各個人ごとに対応させる必要があるといえる。

今後は、これらの課題の解決に取り組んでいく予定である。

謝辞 本研究の遂行にあたり、ご協力いただいた関西大学大学院工学研究科電子工学専攻林享祐氏に感謝いたします。

参 考 文 献

- 辻 康博, 星 守, 大森 匡: 曲の局所パターン特徴量を用いた類似曲検索・感性語による検索, 電子情報通信学会技術研究報告, SP 96-124, pp.17-24 (Nov. 1992).
- 増井誠生, 柿元俊博: 感性語評価を利用した音響情報検索の試み, 情報処理学会研究報告, MUS26-17, pp.115-122 (Aug. 1998).
- 野田達也, 森山 剛, 小沢慎治: 印象語による楽曲検索システム, 電子情報通信学会総合大会論文集, D-12-11, p.184 (Mar. 1999).
- 佐藤 聡, 菊池幸平, 北上 始: 音楽データを対象としたイメージ検索のための感情価の自動生成, 情報処理学会研究報告, DBS 118-8, pp.57-64 (May 1999).
- 門田暁人, 松本健一, 井上克郎: 既存の音楽を手掛かりとする音楽検索システム, 情報処理学会全国大会講演論文集, 5G-3, pp.19-20 (Sep. 1999).
- 栗田多喜男, 加藤俊一, 福田郁美, 坂倉あゆみ: 印象語による絵画データベースの検索, 情報処理学会論文誌, Vol.33, No.11, pp.1373-1383 (1992).
- 藤田健二, 古郡 了, 畑 秀二: 視覚心理空間を利用した画像検索法, 情報処理学会研究報告, CV 51-4, pp.1-8 (Nov. 1987).
- 平林扶佐子, 的場ひろし, 笠原 裕: 感覚的・情緒的情報の蓄積と検索, 電子情報通信学会技術研究報告書, DE 88-34, pp.33-39 (Jan. 1989).
- 戸嶋 郎, 八村広三郎: 絵画からの画面構成の抽出と検索への応用, 情報処理学会論文誌, Vol.40, No.3, pp.912-920 (1999).
- 塩澤秀和, 西山晴彦, 松下 温: 人間のあいまいな感性を反映する絵画検索システム, 情報処理学会研究報告, HI 54-5, pp.33-40 (1994).
- 大坪英昭, 亀井且有: 感性情報による画像検索システムの構築, システム制御情報学会論文誌, Vol.10, No.9, pp.455-462 (1997).
- 田中大典, 前田茂則, 池田克夫: 感性語間の関係に着目した画像検索システム, 電子情報通信学会技術研究報告書, PRMU 98-261, pp.49-56 (Mar. 1999).
- 木本晴夫: 感性語による画像検索とその精度評価, 情報処理学会論文誌, Vol.40, No.3, pp.886-898 (1999).
- 宝珍輝尚, 山田恵一, 都司達夫: 感性の主因子に基づく画像・自然音アクセスに関する一検討, 情報処理学会研究報告, DBS 121-12, pp.77-84 (May 2000).
- 坂本 崇, 梶川嘉延, 野村康雄: 音楽感性空間における非線形判別分析を用いた曲印象別グループの分割, 情報処理学会論文誌, Vol.40, No.4, pp.1901-1909 (1999).
- Saiwaki, N., Katayose, H. and Inokuchi, S.: An Approach to a Computer Assisted Composition System with KANSEI Parameter, *Proc. 2nd. Intl. Workshop in AI and Music* (1989).
- 川原 浩, 野波健彦: 音楽教育研究における実験的研究(II)—享身体験におけるイメージの言語化に関する分析, 広島大学教育学部紀要第4部, 26, pp.75-85 (Sep. 1977).
- 芝 祐順: 因子分析法, 東京大学出版会 (1979).
- 佐藤 信: 官能検査入門, 日科技連出版社 (1978).
- 菅 民郎: 多変量解析の基礎(上), 現代数学社 (1993).
- Rowe, R.: *Interactive Music Systems—Machine Listening and Composing*, The MIT Press (1993).
- 日間賀充寿, 大西 昇, 杉江 昇: 情動に相関のある楽曲中のパラメータについて, 情報処理学会研究報告, MUS 8-2, pp.7-12 (Nov. 1994).
- 枇 薫: コンピュータ&MIDI ②テクニカルブック, 音楽之友社 (1988).
- 石桁真礼生, 丸田昭三, 金光威和雄, 末吉保雄, 飯田 隆, 飯沼信義: 楽典—理論と実習, 音楽之友社 (1975).

付 録

A.1 曲特徴量の計算方法

ここでは、今回用いた曲特徴量の一部について説明する。

A.1.1 Diatonic

楽曲中で使われている音がメジャースケール(マイナースケール)上の音である割合を表し、次式で計算される。ただし、 D_{major} , D_{minor} , D_{all} とはそれぞれ、調性が長調の場合におけるメジャースケール上の音符数, 調性が短調の場合におけるマイナースケール上の音符数, すべての音符数を表す。

$$Diatonic = \frac{D_{major} + D_{minor}}{D_{all}} \quad (2)$$

A.1.2 Rhythm

音楽理論上、拍子記号の分母となる音符長が基本のリズムを構成する音符長になる²⁴⁾。そこで、リズムを

構成する音符の長さ注目し、拍子記号の分母となる音符長を基準にして、各音符長が基準音符長より長いかわりに特徴量を生成する。具体的には、基準音符長の2倍以上の音符長の個数を R_{long} 、2分の1以下の音符長の個数を R_{short} 、それ以外の音符の個数を R_{middle} とすると、次式で計算される。そして、その平均値を Ave-Rhythm、分散値を S-Rhythm とした。

$$Rhythm = \frac{0 \times R_{short} + 1 \times R_{middle} + 2 \times R_{long}}{R_{short} + R_{middle} + R_{long}} \quad (3)$$

A.1.3 Key

曲の調性は、一般的に、長調あるいは短調で表される。そこで、調性の特徴量を次式で計算する。ただし、 K_{major} 、 K_{minor} とはそれぞれ、長調である時間の割合、短調である時間の割合を表す。

$$Key = 0.95 \times K_{major} + 0.05 \times K_{minor} \quad (4)$$

A.1.4 Meter

曲の拍子とは、単純拍子/複合拍子/混合拍子の3種類に分類される²⁴⁾。単純拍子/複合拍子/混合拍子それぞれの演奏時間の総和を M_{sim} 、 M_{com} 、 M_{mix} とすると、次式で計算される。

$$Meter = \frac{1 \times M_{sim} + 2 \times M_{com} + 3 \times M_{mix}}{M_{sim} + M_{com} + M_{mix}} \quad (5)$$

(平成12年10月16日受付)

(平成13年9月12日採録)



池添 剛(正会員)

平成11年関西大学工学部電子工学科卒業。平成13年同大学大学院博士課程前期課程修了。在学中、感性情報処理の研究に従事。



梶川 嘉延(正会員)

平成3年関西大学工学部電子工学科卒業。平成5年同大学大学院博士課程前期課程修了。同年富士通入社。平成6年関西大学助手。平成10年同大学専任講師。平成13年より同大学助教授。主に電気音響変換機のCAD、適応信号処理、非線形信号処理の研究に従事。最近では知能ロボットにも興味を持つ。工学博士。電子情報通信学会、日本音響学会、電気学会、IEEE、EURASIP各会員。



野村 康雄(正会員)

昭和36年大阪大学工学部通信工学科卒業。昭和38年同大学大学院修士課程修了。同年松下電器入社。同社無線通信研究所勤務。昭和45年大阪大学研究生。昭和50年同大学助手。昭和51年関西大学講師。同助教授を経て、現在、工学部電子工学科教授。主に電気音響変換機のCAD、人工知能の研究に従事。工学博士。電子情報通信学会、日本音響学会、IEEE各会員。