

# PCM データに対応した 感性語による音楽データベース検索システムに関する研究

大塚 玲朗<sup>†</sup> 梶川 嘉延<sup>†</sup> 野村 康雄<sup>†</sup>

<sup>†</sup> 関西大学工学部電子工学科

〒 564-8680 大阪府吹田市山手町 3-3-35

E-mail: †{leo,kaji,nomura}@joho.densi.kansai-u.ac.jp

**あらまし** PCM データに対応した感性語による音楽データベース検索システムの構築法を提案する。従来にも感性語による音楽データベース検索システムに関する研究は数多く見られるが、これらの研究において一般に扱われているデータ形式は MIDI 形式であり、現在一般に流通しているデータ形式が PCM 形式であることから、汎用性に乏しいことは否定できない。そこで本研究では PCM データに対応したシステムを構築することで汎用性の向上を目指した。主観評価実験の結果、半数以上のユーザーが満足しており、かつ、MIDI データに対応したシステムと主観的満足度、および感性語評価値予測精度において大きな差はなく、本システムの有効性が実証された。

**キーワード** 感性, PCM, 音楽検索, データベース

## A Study on a Music Retrieval System for PCM Data by Sensitivity Words

Leo OHTSUKA<sup>†</sup>, Yoshinobu KAJIKAWA<sup>†</sup>, and Yasuo NOMURA<sup>†</sup>

<sup>†</sup> Department of Electronics, Faculty of Engineering, Kansai University

3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

E-mail: †{leo,kaji,nomura}@joho.densi.kansai-u.ac.jp

**Abstract** We propose a music retrieval system for PCM data by sensitivity words. Many papers have already proposed various music retrieval systems by sensitivity words, but these systems can handle only MIDI data. These systems consequently lack flexibility because PCM data are generally distributed. We therefore construct a music retrieval system for PCM data in order to solve this problem. Subjective evaluation experiments show the result that the proposed system demonstrate can satisfy more than half of the subjects and accordingly demonstrate the validity of the proposed system.

**Key words** KANSEI, PCM, music retrieval system, database

### 1. はじめに

近年、インターネットの普及に伴い情報量が大規模化するなか、作曲者、曲名等の情報を持たない利用者は探したい曲を検索するのが大変困難になってきている。その解決法の一つとして感性語による検索システムが提案されている。先の研究 [1] [2] では、いわゆる一般的な曲印象の音楽感性検索システムが提案されており、高い精度で感性語評価値を予測し、主観評価実験において高い満足度が得られる楽曲検索システムの構築に成功した。しかし、これらの研究において扱われていた音楽データは MIDI 形式であり、現在一般に流通しているデータ形式が PCM 形式であることから汎用性に乏しいことは否定できない。

そこで本研究では、PCM データに対応した感性語による楽曲検索システムを構築した。

本システムにおいては、楽曲から抽出される曲特徴量からニューラルネットワークを通すことで楽曲の印象値を予測する。PCM データを約 0.1 秒ごとに FFT を施し、特徴量として時間的なスペクトルの変化具合や全体的な音の高さ、音の散らばり具合など 14 種の特徴量を抽出した。

また、楽譜情報が容易に抽出できる MIDI データに対応したシステムも構築し、予測精度の比較、主観評価実験における満足度の比較を行った。さらに、ランダムで楽曲を推薦するシステム、および試聴実験によって得られた感性語評価値をデータベースに保存して検索を行うシステムとの主観評価実験におけ

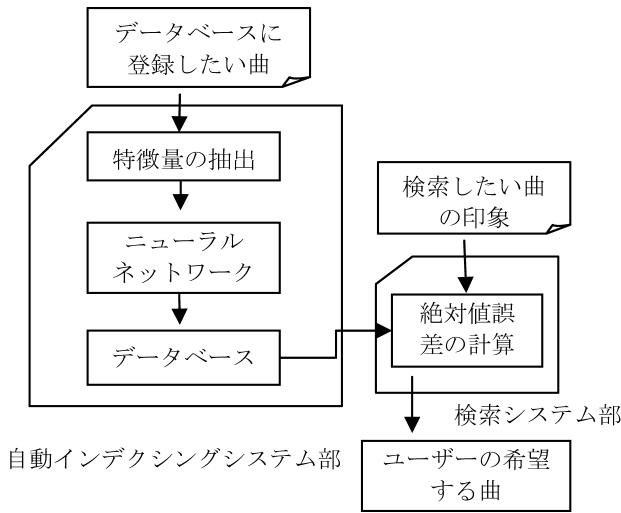


図1 システム概要

る満足度の比較も行った。

## 2. 提案システムの概要

システムの構成を図1に示す。提案システムは自動インデクシングシステム部と検索システム部で構成される。以下に詳細を述べる。

### 2.1 自動インデクシングシステム部

自動インデクシングシステム部では、未知の楽曲に自動的に感性語評価値を付与し、データベースに保存する。まず4096サンプル（約0.1秒）ごとに曲が終了するまでFFTを施す。その際の窓関数にはハミング窓を用いた。その後、特徴量を抽出する。抽出した特徴量は表1の通りである。以下に抽出した特徴量の詳細について述べる。

#### 特徴量

##### $V\_Spe$ （スペクトル変化度）

$i$  回目のFFTにおける周波数点番号  $j$  の振幅値を  $A_{ij}$  とし、FFTをした回数を  $N$  とすると、スペクトル変化度  $V\_Spe$  は、

$$V\_Spe = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{2049} |A_{ij} - A_{(i-1)j}|$$

と定義される。これは、約0.1秒あたりの各周波数点番号における振幅の変化量であり、楽曲のスペクトルの変化具合を表す。

また、人間の感覚的な音の高さを考慮して周波数帯域を200[Hz]以下をlow, 200[Hz]~600[Hz]までをmid, 600[Hz]以上をhighと設定し、それぞれの帯域における特徴量が全体の特徴量を占有している割合を  $Low\_Spe$ ,  $Mid\_Spe$ ,  $High\_Spe$  とし、特徴量として用いる。

##### $Count$ （平均音数）

周波数領域において10[dB]以上の変化のあったサンプリング点の個数を数えたものの時間平均であり、10[dB]以上の変化のあったサンプリング点の個数を  $n$  とすると、

$$Count = \frac{n}{N}$$

と定義される。

表1 抽出した特徴量

特徴量	内容
$V\_Spe$	スペクトル変化度
$High\_Spe$	$V\_Spe$ における高周波帯域成分の割合
$Mid\_Spe$	$V\_Spe$ における中間周波帯域成分の割合
$Low\_Spe$	$V\_Spe$ における低周波帯域成分の割合
$Count$	10[dB]以上変化のあったサンプリング点の時間平均
$High\_Count$	$Count$ における高周波帯域成分の割合
$Mid\_Count$	$Count$ における中間周波帯域成分の割合
$Low\_Count$	$Count$ における低周波帯域成分の割合
$Ave\_Fre$	周波数に音圧値を重みとして掛けたものの平均値
$High\_Fre$	$Ave\_Fre$ における高周波帯域成分の割合
$Mid\_Fre$	$Ave\_Fre$ における中間周波帯域成分の割合
$Low\_Fre$	$Ave\_Fre$ における低周波帯域成分の割合
$Ave\_Amp$	振幅の平均値
$D\_Fre$	周波数強度の分散値

表2 感性語

感性語一覧
明るいー暗い 重いー軽い
硬いーやわらかい 安定ー不安定
澄んだー濁った 滑らかー歯切れよい
激しいーおだやか 厚いー薄い

$Count$  についても  $V\_Spe$  と同様に周波数帯域 low,mid,high における特徴量占有率  $Low\_Count$ ,  $Mid\_Count$ ,  $High\_Count$  も特徴量として用いる。

##### $Ave\_Amp$ （振幅の平均値）

$N$  回FFTした後、周波数点番号  $j$  の振幅値の平均値  $\overline{A}_j$  を、

$$\overline{A}_j = \frac{1}{N} \sum_{i=1}^N A_{ij}$$

とすると、振幅の平均値  $Ave\_Amp$  は、

$$Ave\_Amp = \frac{1}{2049} \sum_{j=1}^{2049} \overline{A}_j$$

と定義される。これは、各周波数点番号における振幅値の平均であり、楽曲における音の大きさを表す。

##### $Ave\_Fre$ （平均周波数強度）

各周波数点番号における振幅値の平均値の総和  $\overline{A}_{all}$  を、

$$\overline{A}_{all} = \sum_{j=1}^{2049} \overline{A}_j$$

とし、周波数点番号  $j$  の周波数を  $F_j$  とすると、平均周波数強度  $Ave\_Fre$  は

$$Ave\_Fre = \frac{1}{\overline{A}_{all}} \sum_{j=1}^{2049} F_j \cdot \overline{A}_j$$

と定義される。これは、各周波数点番号において、振幅値と周波数を掛けたものを周波数強度とし、それらの総和を各周波数点番号における振幅値の平均値の総和で割ったものであり、

表 3 ニューラルネットワーク構築条件

学習法	バックプロパゲーション法
学習係数	0.1
慣性係数	0.9
重みの初期値	±0.3
学習曲	50 曲
未知曲	25 曲
学習回数	50000 回

表 4 予測精度 1

ネットワーク規模	相関係数
14-14-8	0.418
14-20-8	0.332
14-28-8	0.300
14-14-14-8	0.491

楽曲においてどの部分の音の高さに強い音が鳴っているかを表す。

*Ave\_Fre* についても *V\_Spe* と同様に周波数帯域 low, mid, high の特徴量占有率 *Low\_Fre*, *Mid\_Fre*, *High\_Fre* も特徴量として用いる。

### *D\_Fre* (周波数強度の分散値)

周波数強度の分散値 *D\_Fre* は

$$D\_Fre = \frac{1}{A_{all}} \sum_{j=1}^{2049} \bar{A}_j \cdot (F_j - Ave\_Fre)^2$$

と定義される。これは周波数強度の分散を各周波数点番号における振幅値の平均値の総和で割ったものであり、楽曲の音の散らばり具合を表す。

*V\_Spe*, *Count* においてはスペクトルの時間変化を特徴量として抽出し、楽曲の激しさや歯切れよさを予測するのが狙いであり、*Ave\_Amp*, *Ave\_Fre*, *D\_Fre* においては楽曲の全体的な音の高さや強さ、音の散らばり具合を特徴量として抽出することで楽曲の軽さや薄さなどを予測するのが狙いである。

また、high, mid, low 周波数帯域における特徴量の占有率を抽出することで、その楽曲がどの周波数帯域に多くの特徴量を有するかを知ることができる。

これら 14 個の特徴量からニューラルネットワークを通すことにより 8 対の感性語評価値を予測し、データベースに保存する。使用した感性語を表 2 に示す。評価値は 7 段階評価である。

## 2.2 検索システム部

検索システム部ではユーザーが入力した 8 対の感性語評価値と、データベース内の楽曲の感性語評価値との絶対値誤差を計算し、その値の最も小さい楽曲を推薦する。

## 3. 自動インデクシングシステムの構築方法

本節では自動インデクシングシステムの構築方法について述べる。まず、ニューラルネットワークに学習をさせる時に必要となる楽曲に対する理想感性語評価値を得るため、試験実験を行った。以下に試験実験について述べる。

表 5 感性語評価値と特徴量の相関

感性語	相関の強い特徴量			
歯切れよい -滑らか	<i>Count</i> (0.749)	<i>D_Fre</i> (-0.467)	<i>V_Spe</i> (0.426)	<i>Ave_Amp</i> (0.409)
厚い -薄い	<i>Count</i> (0.520)	<i>D_Fre</i> (-0.514)	<i>Ave_Fre</i> (-0.446)	<i>Ave_Amp</i> (-0.430)
やらかい -硬い	<i>Count</i> (-0.616)	<i>D_Fre</i> (0.492)	<i>Ave_Amp</i> (-0.483)	<i>V_Spe</i> (-0.467)
安定 -不安定	<i>Ave_Fre</i> (-0.189)	<i>High_Count</i> (0.186)	<i>High_Fre</i> (-0.181)	<i>D_Fre</i> (-0.176)
激しい -穏やか	<i>Count</i> (0.749)	<i>Ave_Amp</i> (0.687)	<i>V_Spe</i> (0.661)	<i>D_Fre</i> (-0.490)
軽い -重い	<i>High_Fre</i> (0.472)	<i>Mid_Fre</i> (-0.449)	<i>Ave_Fre</i> (0.419)	<i>High_Spe</i> (0.411)
濁った -澄んだ	<i>D_Fre</i> (-0.529)	<i>Ave_Amp</i> (-0.520)	<i>Count</i> (-0.523)	<i>High_Fre</i> (-0.477)
暗い -明るい	<i>V_Spe</i> (-0.506)	<i>Count</i> (-0.482)	<i>Ave_Amp</i> (-0.477)	<i>D_Fre</i> (0.313)

表 6 予測精度 2

感性語	入力数 3	入力数 4
歯切れよい-滑らか	0.428	0.691
厚い-薄い	0.650	0.561
やらかい-硬い	0.618	0.666
激しい-穏やか	0.300	0.653
安定-不安定	0.021	0.011
軽い-重い	0.599	0.402
濁った-澄んだ	0.431	0.537
暗い-明るい	0.710	0.743

### 3.1 試験実験

評価法は SD 法で行った。評価値は 7 段階評価である。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 2 に示す感性語を用いて実験を行った。楽曲は 75 曲であり、ジャンルはクラシック、ジャズ、ラグタイムなどである。なお、本研究ではピアノ曲のみを扱った。

次に実験方法について述べる。被験者に楽曲を提示し、一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、各グループごとに楽曲を提示した。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提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値を取ることが一般的であるといえるが、本研究では異常値を取り除くという手間を省くために 20 人の中央値を取り実験結果を求めた。

### 3.2 ニューラルネットワークの構築法

ニューラルネットワークの構築方法により手法 1 と手法 2 に分け比較検討を行った。なお、ニューラルネットワークの構築条件は手法 1,2 とともに表 3 に示すとおりとする。ただし、入力値、出力値ともに 0.05 から 0.95 の間の値に正規化した。

### i) 手法 1

入力を 14 個の特徴量, 出力を 8 つの感性語評価値とするニューラルネットワークを用いる. 未知曲 25 曲の予測精度を表 4 に示す. 予測精度は理想値との相関係数で表す. 表 4 における相関係数は 8 つの出力の平均値である. なお, ネットワーク規模は一番左側の数字が入力ユニット数, 一番右側の数字が出力ユニット数であり, その間の数字が中間ユニット数である.

### ii) 手法 2

学習データにおいて曲特徴量と感性語評価値との相関を調べ, それぞれの感性語に対していくつかの相関の強い曲特徴量をニューラルネットワークの入力とした. つまり, それぞれの感性語対に対して別個のニューラルネットワークを構築し, 一つの感性語対に対して数個の相関の強い曲特徴量をニューラルネットワークの入力とするのである. 表 5 に感性語評価値と相関の強い曲特徴量を示す. 本研究では入力数を 3 および 4 としたときの未知曲の予測精度を調査し, 予測精度の高い入力数を採用した. 中間層は 1 層で, ユニット数は入力ユニット数と同数とした.

未知曲 25 曲のニューラルネットワークの入力数ごとの予測精度を表 6 に示す. 予測精度は手法 1 と同様に理想値との相関係数で表す. それぞれの感性語において予測精度の高い入力数を採用すると, 8 対の感性語の平均相関係数は 0.570 であった.

以上よりニューラルネットワークの構築方法は手法 2 の方が有効であることが分かった. 以後, 手法 2 のニューラルネットワークを用いた自動インデクシングシステムシステムにおいて有効性を検討していく.

## 4. システムの有効性の検討

システムの有効性を検討するために MIDI データに対応したシステムを構築し, 予測精度の比較を行った. また, 主観評価実験を行い主観的な有効性も調査した.

### 4.1 MIDI データによる自動インデクシングシステム

前節で述べたものと同様の手法を用いて MIDI データに対応したシステムを構築した. MIDI データからは楽曲の正確な楽譜情報が容易に抽出できる. 使用した楽曲は PCM に対応したシステムと同じである. 抽出した特徴量は表 7 の通りである. 以下に抽出した特徴量の詳細について述べる.

#### 特徴量

##### *Rest\_Times* (休符の数)

楽曲中の休符の数を数えたものである.

##### *Ave\_Pitch* (音高の平均値)

音高の平均値 *Ave\_Pitch* は,

$$Ave\_Pitch = \frac{1}{n} \sum_{i=1}^n P_i$$

と定義される. 但し,  $n$  は音符数,  $P_i$  は  $i$  番目の音符のキーナンバー値である.

##### *D\_Pitch* (音高の分散値)

音高の分散値 *D\_Pitch* は,

$$D\_Pitch = \frac{1}{n} \sum_{i=1}^n (P_i - Ave\_Pitch)^2$$

と定義される.

##### *Density* (音符密度)

音符密度 *Density* は,

$$Density = \frac{n}{T}$$

と定義される. 但し,  $T$  は演奏時間である.

##### *Tempo* (テンポ)

楽曲における速度記号の値である.

##### *Ave\_Velocity* (音の強さの平均値)

音の強さの平均値 *Ave\_Velocity* は,

$$Ave\_Velocity = \frac{1}{n} \sum_{i=1}^n V_i$$

と定義される.

##### *D\_Velocity* (音の強さの分散値)

音の強さの分散値 *D\_Velocity* は,

$$D\_Velocity = \frac{1}{n} \sum_{i=1}^n (V_i - Ave\_Velocity)^2$$

と定義される.

##### *Ave\_Gate* (音の長さの平均値)

音の長さの平均値 *Ave\_Gate* は,

$$Ave\_Velocity = \frac{1}{n} \sum_{i=1}^n D_i$$

と定義される. 但し,  $D_i$  は  $i$  番目の音符のゲートタイム値である.

##### *D\_Gate* (音の長さの分散値)

音の長さの分散値 *D\_Gate* は,

$$D\_Gate = \frac{1}{n} \sum_{i=1}^n (D_i - Ave\_Gate)^2$$

と定義される.

##### *Harmony* (ハーモニー)

ハーモニー *Harmony* は,

$$Ave\_Velocity = \frac{s}{A}$$

と定義される. 但し,  $A$  は楽曲全体の音数,  $s$  はスケール上にある音の数とする.

##### *Key* (調性)

メジャーキーの楽曲は 1, マイナーキーの楽曲は 0 と定義する.

これら 11 種の特徴量からニューラルネットワークを用いて感性語評価値を予測する. また, ニューラルネットワークの構築方法は前節と同様に手法 1 と手法 2 を比較検討する.

表 7 抽出した特徴量 (MIDI)

特徴量	内容
<i>Rest_Times</i>	休符の数
<i>Ave_Pitch</i>	音高の平均値
<i>D_Pitch</i>	音高の分散値
<i>Density</i>	音符密度
<i>Tempo</i>	楽曲の速さ
<i>Ave_Velocity</i>	音の強さの平均値
<i>D_Velocity</i>	音の強さの分散値
<i>Ave_Gate</i>	音の長さの平均値
<i>D_Gate</i>	音の長さの分散値
<i>Harmony</i>	ハーモニー
<i>Key</i>	調性

表 8 予測精度 1(MIDI)

ネットワーク規模	相関係数
11-11-8	0.421
11-15-8	0.500
11-22-8	0.532
11-11-11-8	0.290

手法 1 による予測精度を表 8 に示す。また、学習曲において感性語評価値と相関の強い特徴量を表 9 に示し、手法 2 による予測精度を表 10 に示す。手法 2 において予測精度の高い入力数を採用した場合の平均相関係数は 0.571 であった。MIDI データに対応したシステムにおいても PCM データに対応したシステムと同様にニューラルネットワークの構築方法は手法 2 のほうが有効であった。感性語評価値と相関の高い特徴量においては PCM データから抽出したものと MIDI データから抽出したものでは MIDI データから抽出したものが若干高いが、大きな差はなく PCM データからも有効な特徴量が抽出できているといえる。また、未知曲の予測精度においても大きな差はなく、このことから PCM データから有効な特徴量が抽出できたといえる。さらに、MIDI データから抽出した特徴量と感性語評価値との相関があまり高くないことから、PCM データから正確な楽譜情報を抽出できたとしても感性語評価値と相関の高い特徴量にはなりえないことが分かる。

#### 4.2 主観評価実験 1 の概要

主観評価は 5 段階評価で行った。また、比較対照となるシステムの満足度も同時に調査した。各システムの内容を以下に示す。

システム 1: MIDI データに対応した自動インデクシングシステムによって予測された感性語評価値をデータベースに登録し、検索を行う。

システム 2: PCM データに対応した自動インデクシングシステムによって予測された感性語評価値をデータベースに登録し、検索を行う。

被験者が 8 対の感性語評価値を入力すると 2 つのシステムから同時に楽曲が推薦される。順序効果をなくすため推薦される楽曲の試聴順序はランダムに決定した。被験者にはどのシステムから楽曲が推薦されるか分からないように、楽曲の試聴順序

表 9 感性語評価値と特徴量の相関 (MIDI)

感性語	相関の強い特徴量			
歯切れよい ー滑らか	<i>Ave_Vel</i> (0.701)	<i>Density</i> (0.610)	<i>D_Gate</i> (0.605)	<i>Ave_Gate</i> (0.598)
厚い ー薄い	<i>Ave_Vel</i> (0.634)	<i>D_Gate</i> (-0.525)	<i>Density</i> (0.511)	<i>Harmony</i> (-0.436)
やらかい ー硬い	<i>Density</i> (-0.601)	<i>D_Gate</i> (0.568)	<i>Ave_Gate</i> (0.532)	<i>Rest_Times</i> (-0.486)
安定 ー不安定	<i>Key</i> (-0.391)	<i>Tempo</i> (0.280)	<i>Density</i> (-0.208)	<i>Harmony</i> (-0.166)
激しい ー穏やか	<i>Ave_Vel</i> (0.781)	<i>Density</i> (0.751)	<i>D_Gate</i> (-0.728)	<i>Ave_Gate</i> (-0.699)
軽い ー重い	<i>Ave_Pitch</i> (0.474)	<i>Ave_Vel</i> (-0.359)	<i>Tempo</i> (0.215)	<i>Key</i> (0.189)
濁った ー澄んだ	<i>Ave_Vel</i> (0.578)	<i>Ave_Pitch</i> (-0.543)	<i>Harmony</i> (-0.460)	<i>Key</i> (-0.325)
暗い ー明るい	<i>Key</i> (-0.572)	<i>D_Gate</i> (0.562)	<i>Density</i> (-0.547)	<i>Ave_Gate</i> (0.510)

表 10 予測精度 2(MIDI)

感性語	入力数 3	入力数 4
歯切れよいー滑らか	0.716	0.832
厚いー薄い	0.526	0.380
やらかいー硬い	0.459	0.439
激しいー穏やか	0.769	0.894
安定ー不安定	0.035	0.238
軽いー重い	0.441	0.565
濁ったー澄んだ	0.302	0.332
暗いー明るい	0.728	0.686

のみを提示し、提示順序とシステム名は関係がないことを説明した。これは試行回数が増えることによって被験者がシステムの内容を推測するのを防ぐためである。また、データベースに保存されているのは未知曲の 25 曲であり、被験者は 15 名で、1 人当たり 10 回試行した。

#### 4.3 主観評価実験 1 の結果

表 11 及び表 12 に主観評価実験結果を示す。表 11 は全 150 試行におけるそれぞれの評価値の延べ曲数で、表 12 は 15 人の被験者のそれぞれに 10 回の試行の満足度の平均値を求めたものである。

表 11 を見るとシステム 1 では 79.3%(12.6%)、システム 2 では 60.0%(16.7%) の楽曲が満足 (不満足) と評価されていることがわかる。また、表 12 を見ると、10 回の試行の平均の満足度が 3.5 を超えた人数の割合はシステム 1 では 73.4%、システム 2 では 53.3% であった。人別評価において、満足度がシステム 1 の方が大きく高いが、これはシステム 1 では満足度の平均値が 4.1 や 3.6 であった人が多いのに対して、システム 2 では満足度の平均値が 4.0 や 3.5 の人が多かったためである。曲別評価においてシステム 1 とシステム 2 の有意水準を  $\chi^2$  検定によって求めると約 0.1 であり、それほど差はないといえる。全体的に若干システム 1 の方が満足度が高いが、これは比較的印象に残りやすい「激しいー穏やか」や「歯切れよいー滑らか」

表 11 曲別評価 1

	システム 1		システム 2	
	曲数	割合 [%]	曲数	割合 [%]
5	47	31.3	34	22.7
4	57	38.0	56	37.3
3	27	18.0	35	23.3
2	14	9.3	18	12.0
1	5	3.3	7	4.7

表 12 人別評価 1

満足度	システム 1		システム 2	
	人数	割合 [%]	人数	割合 [%]
5.0-4.5	0	0	0	0
4.5-4.0	7	46.7	2	13.3
4.0-3.5	4	26.7	6	40.0
3.5-3.0	3	20.0	6	40.0
3.0-2.5	1	6.7	1	6.7
2.5-2.0	0	0	0	0
2.0-1.5	0	0	0	0
1.5-1.0	0	0	0	0

表 13 曲別評価 2

	システム 1		システム 2		システム 3	
	曲数	割合 [%]	曲数	割合 [%]	曲数	割合 [%]
5	53	35.3	39	26.0	25	16.7
4	63	42.0	50	33.3	33	22.0
3	21	14.0	34	22.7	39	26.0
2	10	6.7	20	13.3	37	24.7
1	3	2.0	7	4.7	16	10.7

表 14 人別評価 2

満足度	システム 1		システム 2		システム 3	
	人数	割合 [%]	人数	割合 [%]	人数	割合 [%]
5.0-4.5	2	13.3	0	0	0	0
4.5-4.0	5	33.3	2	13.3	0	0
4.0-3.5	5	33.3	6	40.0	3	20.0
3.5-3.0	3	20.0	5	33.3	7	46.7
3.0-2.5	0	0	2	13.3	3	20.0
2.5-2.0	0	0	0	0	3	20.0
2.0-1.5	0	0	0	0	0	0
1.5-1.0	0	0	0	0	0	0

などの予測精度がシステム 1 の方が高いためであると考えられる。しかし、有意水準も高く、満足度に大きな差はないことから PCM データに対応したシステムと MIDI データに対応したシステムではほぼ同等の有効性があるといえる。

#### 4.4 主観評価実験 2 の概要

比較対照となるシステムを変更し、満足度を調査した。各システムの内容を以下に示す。

システム 1: 試聴実験によって得られた感性語評価値をデータベースに登録し、検索を行う。

システム 2: 自動インデクシングシステムによって予測された感性語評価値をデータベースに登録し、検索を行う

システム 3: 被験者の入力値に関係なくランダムに楽曲を提示する。

その他の設定は前主観評価実験と同様である。

#### 4.5 主観評価実験 2 の結果

表 13 及び表 14 に主観評価実験結果を示す。表 13 は全 150 試行におけるそれぞれの評価値の延べ曲数で、表 14 は 15 人の被験者のそれぞれに 10 回の試行の満足度の平均値を求めたものである。

表 13 を見るとシステム 1 では 77.3%(8.7%)、システム 2 では 59.3%(18.0%)、システム 3 では 38.7%(35.4%) の楽曲が満足(不満足)と評価されていることがわかる。また、表 14 を見ると、10 回の試行の平均の満足度が 3.5 を超えた人数の割合はシステム 1 では 79.9%、システム 2 では 53.3%、システム 3 では 20.0%であった。

人別評価、曲別評価ともにシステム 1 は非常に満足度が高い。これは本検索システムが有効であることを示している。システム 1 の曲別評価において不満足な曲が 8.7%あるが、これは、データベースに登録されている曲が 25 曲と少なく、被験者

が希望した楽曲がデータベースに存在しなかったためと考えられる。また、曲別評価において、システム 1 と 2、2 と 3 の有意水準を  $\chi^2$  検定によって求めると共に 0.05 を下回り有意差が存在することがわかった。このことから、本研究において構築したシステム 2 はランダムで楽曲を推薦するシステム 3 と有意差があることから有効であるといえるが、試聴実験によって得られた感性語評価値をデータベースに保存して楽曲を推薦するシステム 1 と有意差があることから精度の面で向上の余地があるといえる。しかし、上に述べたようにデータベースに登録されている楽曲が 25 曲と少なく、被験者が希望した楽曲がデータベースに存在しない可能性が高いなかで、システム 2 の満足度が人別評価、曲別評価共に 50%を超えたことは十分評価できると思われる。

以上のことから本研究で構築したシステムは半数以上の被験者が満足しており、かつランダムで楽曲を推薦するシステムと有意差があることから有効であるといえるが、試聴実験によって得られた感性語評価値をデータベースに保存して楽曲を推薦するシステム 1 と有意差があることから精度の面で向上の余地はある。

## 5. ま と め

本稿では PCM データに対応した感性語による楽曲検索システムを構築し、その有効性を検証した。前節で述べたように、半数以上の被験者が満足しており、かつランダムで楽曲を推薦するシステムと有意差が存在しているので、本システムは有効であるといえる。また、MIDI データに対応したシステムと主観評価実験における満足度に大きな差がなかったことから本システムは有効であるといえる。

感性語評価値と相関の高い特徴量は、PCM データから抽出したものと MIDI データから抽出したものでは MIDI データから抽出したものの方が若干高いが、大きな差はなく PCM デー

タからも MIDI データから一般的によく抽出される特徴量と同等に有効性のある特徴量を抽出できたといえる。

ニューラルネットワークの構築方法についても、感性語評価値と相関の強い曲特徴量を入力情報とすることで、精度が向上することが確認できた。

しかし、主観評価実験において試聴実験によって得られた値を使用したシステムと有意差が存在することから、精度の面で向上の余地があるといえる。

**謝辞** 本研究の一部は関西大学重点領域研究ならびに関西大学学術助成基金の援助の下で行われた。

#### 文 献

- [1] 池添剛, 梶川嘉延, 野村康雄, “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”Vol.42, No.12, pp.32101-6, Dec.2001.
- [2] 熊本忠彦, 太田公子 “印象に基づく楽曲検索: システムの実装と評価”, 情処研報, MUS46-6, pp.37-42, Jul.2002.
- [3] 橋本周司, “感性情報処理の諸相”, 映像情報メディア学会誌, Vol. 52, No. 1, pp. 41-45, Jan. 1998.
- [4] 高澤嘉光, 保原信, “計算機と音楽”, 情報処理学会誌, Vol. 29, No. 6, pp. 538-622, Jun. 1988.
- [5] 平田圭二, 鈴木孝, 柏野邦夫, “音楽情報処理”, 情報処理学会誌, Vol. 35, No. 9, pp. 799-835, Sep. 1994.
- [6] 平賀譲, “音楽情報処理 - 芸術, コンピュータ, そして人間の邂逅”, 情報処理学会誌, Vol. 39, No. 5, pp. 468-471, May 1998.
- [7] 佐藤 信, 官能検査入門, 日科技連出版, 東京, 1978.
- [8] 井口征士, 中村敏枝, 小林重順, 感性情報処理, オーム社, 東京, 1994.
- [9] 橋本周司, 大照完, 仮想音楽空間, オーム社, 東京, 1994.
- [10] 平田圭二, 平賀譲, 橋本周司, 長崎洋一, bit 別冊 コンピュータと音楽の世界 - 基礎からフロンティアまで -, 共立出版株式会社, 東京, 1998.
- [11] 豊田英樹, 非線形多変量解析 - ニューラルネットによるアプローチ -, 朝倉書店, 東京, 1996.
- [12] 市川紘, 階層型ニューラルネットワーク - 非線形問題解析への応用 -, 共立出版, 東京, 1993.
- [13] 松岡清利編著, ハイテクシリーズ ニューロコンピューティング - 基礎と応用 -, 朝倉書店, 東京, 1992.