

実環境で受信した楽音をキーとする楽曲探索法

黒住 隆行[†] 柏野 邦夫[†] 村瀬 洋^{†*}

A Robust Music Searching Method Using an Audio Signal Query Received in the Real World

Takayuki KUROZUMI[†], Kunio KASHINO[†], and Hiroshi MURASE^{†*}

あらまし 実環境で収録された楽曲の断片をキーとし、膨大な楽曲データからそれと同一の楽曲を探索する手法を提案する。本手法は、時系列探索、すなわち、ある信号の断片がそれとは別の長時間の信号のどの時点にあるかを探索するという枠組みに基づく。時系列探索についてこれまで、時系列アクティブ探索法など高速性に主眼をおいた手法が提案されているが、実環境で収録された音のように、様々なひずみを含んだ信号を精度良く探索する手法については検討されていなかった。本手法の基本的なアイデアは、時間周波数空間上で局所領域ごとに正規化を行った後、機器特性や環境雑音などによるひずみに対して頑健な部分空間に射影して照合を行うというものである。本手法を用いて、実環境でマイクや PHS で収録した 10 秒の音響信号の断片が、116 時間分の CD 品質の蓄積信号のどの曲のどの時刻に対応するかを調べる課題について実験を行ったところ、単純なパワーによる正規化とスペクトル照合を用いた場合に 12.8% であった探索精度が、71.2% にまで改善し、本手法の有効性が確認された。本手法により、様々なひずみに対して頑健な音響信号探索が可能になると考えられる。

キーワード 音響信号探索、正規化、部分空間、ノイズ、検索

1. まえがき

近年、音や映像などマルチメディア情報の流通が盛んになっており、それらの探索や検索の技術が極めて重要になってきている。そこで我々は、音や映像の探索について研究を進めている。これまで、時系列探索、すなわち、具体的な音や映像（目的信号）を指定して、それと類似する音や映像が膨大なデータベース（蓄積信号）のどこに存在するかを探索する問題について検討してきた。時系列探索とは、図 1 に示すように、目的信号と蓄積信号から特徴抽出を行っておき、蓄積信号中に目的信号と同じ大きさの窓をかけ、目的信号と蓄積信号中の注目箇所との類似度を計算しながら、窓をずらし、目的信号と最も類似した蓄積信号中の箇所を探索結果として出力するという枠組みである。

時系列探索は、高速に、かつ高精度に行わなければ

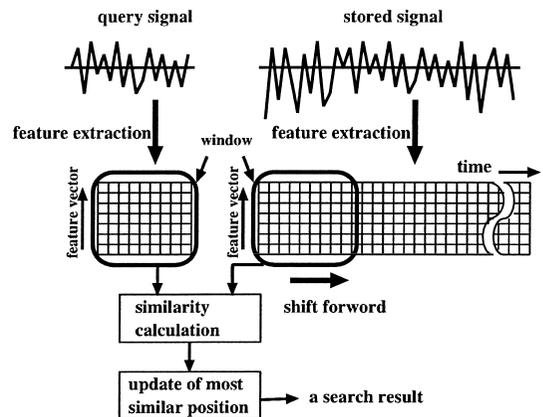


図 1 時系列探索の概要
Fig. 1 Overview of time-series search.

ならない。我々は既に、高速性を主眼とする手法である時系列アクティブ探索法 [1] を提案した。時系列アクティブ探索法は、ヒストグラム照合とその枝刈りによる高速化を特徴とするものであり、テレビ放送のような比較的品質の良い音や映像において有効性の高い

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所、厚木市

NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, 243-0198 Japan

* 現在、名古屋大学大学院 情報科学研究科

手法であった．また，既知の特徴ひずみを事前に学習により求めておき，それを用いて原信号にひずみを付加しながら探索する手法 [2] についても検討した．

さて，時系列探索の応用範囲を考えると，必ずしもテレビ放送のような高い品質が確保されたような場面だけにとどまらず，様々な場面が想定される．特に近年では，ノート型コンピュータ，PDA，携帯電話といった携帯端末の普及が著しく，これら携帯端末を用いて音や映像を扱う機会も増えてきている．もし，携帯端末に入力された実環境の音や映像を探索キーとして情報検索ができれば有用性が高い．例えば，街角で流れている楽曲を携帯電話に入力することによって，その楽曲に関するアーティストや CD，コンサート情報などを検索するといったようなことも可能になるであろう．

そこで本論文では，実環境で収録された音をキーとした音響信号探索を想定し，精度良く探索する手法について検討する．なお，ここでは，マイクロホンアレーを用いた雑音除去法 [3] のような特殊なデバイスを用いた手法については議論しない．また，編曲された楽曲やハミング探索などの類似音楽探索 [4] ではなく，同一の音源を探索する一致探索を対象とする．

本論文の基本的なアイデアは，時間周波数空間上の局所領域ごとに正規化を行った後，機器特性や環境雑音などによるひずみに対して頑健な部分空間に射影して照合を行うというものである．以下，提案手法の概要と，その妥当性を検討するための実験について順に説明する．

2. 問題と解決の方針

信号中のひずみを吸収する手法は，これまで，実環境下での音声認識の研究分野において様々なものが提案されている [5]～[8]．例えば，加法的雑音を除去するため雑音スペクトルを信号スペクトルから差し引くスペクトルサブトラクション (SS) [9]，乗法的ひずみを吸収するためのケプストラム平均正規化 (CMN) [5] など，簡便で有用性が高い方法が提案されている．しかし，本論文で扱うような実環境で流れている楽曲の断片をキーとした時系列探索を対象とするものはなかった．そこで本論文では，時系列探索に適した汎用的な特徴抽出法という立場から，特徴ひずみに頑健な方法を検討することにする．

さて，実環境で収録された音に含まれる特徴ひずみには，様々なものが考えられる．例えば，スピーカな

どの信号発生源の機器の特性によるひずみ，実環境中の雑音，反響，吸収などによるひずみ，入力端末の特性によるひずみなどである．もし，これらのひずみをすべてモデル化し，モデルに基づいてひずみを補正できるならば，探索精度を向上させることができるだろう．しかし現実的には，入力端末側の特性は既知とできる場合もあるが，スピーカ特性や環境雑音等を事前に知ることは難しい．そこで，これらのひずみを吸収し，ひずみの変動の少ない特徴を抽出することが必要である．

本論文で扱う特徴ひずみを具体的に例示したものが図 2 と図 3 である．図 2 は，ある音楽の同一の部分について複数の条件で収録し，それぞれの信号からフーリエ変換 (サンプリング周波数 4000 Hz，フーリエ変換に使用したサンプリング点数 4096) により求めたパワースペクトルを 0～2000 Hz の平均パワーで割って，更に 50 Hz ごとにサンプリングして求めた周波数

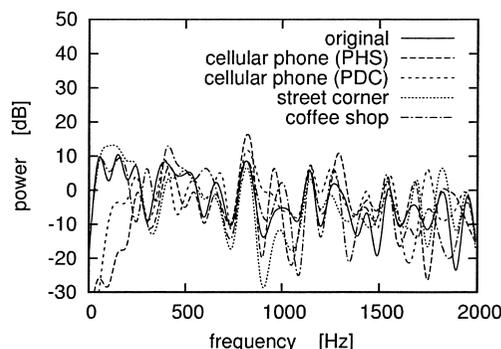


図 2 楽曲の一部分 1

Fig. 2 A segment of a musical piece 1.

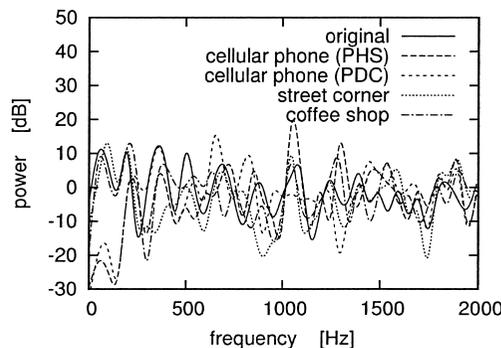


図 3 楽曲の一部分 2

Fig. 3 A segment of a musical piece 2.

分布を示したものである。収録の条件は、原音（CD品質）、実験室で高音質携帯電話（PHS^(注1)）による収録、実験室で中音質携帯電話（PDC^(注2)）による収録、街頭でのマイク収録、及び喫茶店でのマイク収録の5種類である。これらの信号間には大きな差異があるが、探索においては、これらを同一のものとみなして認識しなければならない。一方、図3は、図2に示したものと別の音楽の一部分の周波数分布である。これら図2と図3は別のものとして認識されなければならない。このような信号の探索においては、単純に周波数分布をマッチングしたのでは探索精度が非常に低い。そのため、特徴ひずみに対しては変動が小さくかつ、音響信号の内容に対しては変動を大きくするような特徴抽出法が必要であると考えられる。

本論文の基本的なスタンスは、雑音や周波数特性の違いに対して、できるだけ不変な特徴を選ぶというものである。人のざわめき声や車のエンジン音など加法的雑音やマイクの位置の変化による周波数特性の時間変動は、数秒程度の短時間では変動が十分小さいと考えられる。そこで、数秒程度の短時間では、加法的雑音、周波数特性が定常であるとみなし、あるひずみ要因によって変動が起こった信号の時間周波数分布 $Y(t, i)$ が次式のように表されると仮定する。

$$Y(t, i) = aX(t, i) + b \quad (1)$$

ただし、時刻 t の原信号の周波数 i のパワーを $X(t, i)$ とし、 a は周波数特性の差異によって起こる乗法性ひずみを表す定数、 b は加法的雑音を表す定数とする。ここで、 a や b をキャンセルするような特徴を選ぶことで、加法的雑音や周波数特性の違いを吸収することができると考えられる。このために、数秒程度の短時間の周波数特徴を求め、その時間周波数空間上での局所領域の統計量を用いて正規化することを考える。例えば、各周波数帯域ごと独立に数秒程度の周辺領域の平均を差し引き標準偏差で割ることで、 a や b に対して不変な特徴が得られる。

また、図2、図3を見ると、周波数帯域によっては、特に変動の大きい成分、少ない成分などまちまちである。そこで、探索においても特定の帯域の分解能を高めることが探索精度向上につながると考えられる。そこで、周波数特徴を特徴ひずみに対しては変動が小さくかつ音響信号の内容に対しては変動が大きい部分空間に射影することで、より特徴ひずみに頑健な特徴抽出

を行うことが考えられる。

3. 探索の原理

本手法の基本的なアイデアは、特徴ひずみによる変動を吸収するため、二つの変動の吸収処理を行うことにある。すなわち、周波数特徴を抽出した後、時間周波数空間上の局所領域ごとに正規化し、特徴ひずみに頑健な部分空間への射影を行う。探索の具体的な処理は、(1) 周波数特徴の抽出、(2) 時間周波数空間上の局所領域ごとの正規化、(3) 部分空間への射影、(4) 時系列探索の順で行う（図4）。また、部分空間への射影を行うためには、探索に先立って事前に特徴ひずみに頑健な部分空間を求めておく必要がある。

3.1 周波数特徴の抽出

音響特徴としては、単位時間当りのゼロ交差数、短時間パワースペクトル、LPC ケプストラム、MFCC (Mel frequency cepstral coefficients) など [10] が考えられる。本論文では、その中でも代表的な周波数特徴である、フーリエ変換により求めた短時間パワースペクトルを用いた。ここで、時刻 t の音響信号のフーリエ変換後の特徴 $X(t)$ の k 番目の要素を、

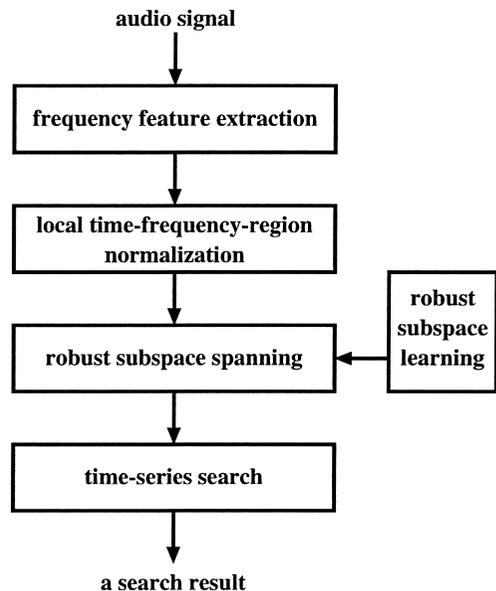


図4 処理の手順

Fig. 4 Procedure of the proposed processing.

(注1): 32 bit/s ADPCM 方式を用いた簡易型携帯電話。

(注2): 6.7 bit/s VSELP 方式を用いた携帯電話。

$$X(t, k) = \sum_{n=0}^{N-1} x(t+n) e^{-j \frac{2\pi kn}{N}} \quad (2)$$

とする。ただし、 $x(t)$ を時刻 t における信号の値、 N は周波数特徴の分析窓の長さ、 k の最大値は標準化定理より $N/2$ である。ここで、パワースペクトル $P(t, k)$ を次式により求める。

$$P(t, k) = |X(t, k)|^2 \quad (3)$$

ここで、このパワースペクトルを s きざみで抽出する。すなわち、時間方向 i 番目の周波数特徴 $Q(i, k)$ は、

$$Q(i, k) = P(s_i, k) \quad (4)$$

である。

3.2 時間周波数空間上の局所領域ごとの正規化

この処理の目的は、数秒程度の短時間で変動の小さい加法性雑音や周波数特性を吸収することである。ここでは、各周波数帯域ごとに、ある時間区間の値から平均と標準偏差を求め、それらを用いて正規化を行う。すなわち、正規化後の周波数特徴 $y(i)$ の k 番目の要素は、

$$y(i, k) = \frac{1}{\sigma(i, k)} (Q(i, k) - m(i, k)) \quad (5)$$

である。ただし、

$$m(i, k) = \frac{1}{2M} \sum_{i=-M}^{M-1} Q(i, k) \quad (6)$$

$$\sigma(i, k)^2 = \frac{1}{2M} \sum_{i=-M}^{M-1} (Q(i, k) - m(i, k))^2, \quad (7)$$

M は局所時間内の周波数特徴の平均、標準偏差を求めるための時間窓の大きさの半分の値である。

なお、本正規化は、実環境中の音声認識でよく使用される CMN (Cepstrum mean normalization) に類似した手法である。CMN は、周波数特性を吸収するため、短時間のケプストラムの平均を差し引くという処理を行う手法である。この手法は、ケプストラムの平均を計算する時間窓の中で周波数特性が定常であるとする、平均を差し引くことにより周波数特性の差による変動を吸収できるという考え方に基づくものである。それに対し、本正規化は、加法性雑音を吸収するために短時間の周波数特徴の平均 $m(i, k)$ を差し引く、周波数特性を吸収するために短時間の周波数特徴

の標準偏差 $\sigma(i, k)$ で割る、という処理を行うものである。またこれは、時間区間内の値をサンプルと見立てたガウシアン正規化と考えることもできる。周波数特徴の平均、標準偏差を計算する時間窓の中で加法性雑音と周波数特性が定常であるとする、この処理により加法性雑音と周波数特性の両方を吸収することができると考えられる。

3.3 部分空間の構成

続いて、正規化後の周波数特徴を部分空間に射影する。この処理の目的は、周波数特徴を特徴ひずみに対しては変動が小さくかつ音響信号の内容に対しては変動が大きい部分空間に射影することで、より特徴ひずみに頑健な特徴抽出を行うことである。

ここで用いる手法は、よく知られた手法である主成分分析 (PCA) と類似した手法であるが、雑音の平均化処理を行うという点で異なる。すなわち、特徴ひずみに対する変動は無視し、音響信号の内容に対する変動のみを考慮した PCA を行う。

具体的には、CD の音楽などの音響信号中のある周波数特徴を L 個用意し、 l 番目の信号部分の正規化後の周波数特徴を y_{l0} とする。また、周波数特徴 y_{l0} それぞれに対して特徴ひずみのある信号を C 種類用意し、 c 番目の種類の周波数特徴を y_{lc} とする。すなわち、クラス数 L 個、各クラス C 個の学習サンプルを用意する。

ここで、特徴の平均を用いて共分散行列 R を計算する。

$$R = \frac{1}{L} \sum_{l=1}^L (\bar{y}_l - \bar{y})(\bar{y}_l - \bar{y})^t \quad (8)$$

ただし、各クラスごとの平均 \bar{y}_l を

$$\bar{y}_l = \frac{1}{C+1} \sum_{c=0}^C y_{lc}, \quad (9)$$

クラス平均の平均 \bar{y} を

$$\bar{y} = \frac{1}{L} \sum_{l=1}^L \bar{y}_l \quad (10)$$

とする。ここで、 R の固有ベクトルを求める。固有ベクトルは、次式の固有値問題により求める。

$$R\phi_u = \lambda_u \phi_u \quad (11)$$

ただし、 ϕ_u は互いに直交する固有ベクトルのうちの

固有値が u 番目に大きいものである。

ここで得られた、固有ベクトル ϕ_u からなる部分空間に射影した特徴、すなわち、

$$z_u = y\phi_u \quad (12)$$

を要素にもつ特徴ベクトル z を用いて探索を行う。

なお、本手法は、従来の PCA に比べて、同じ学習サンプル数でも平均しか使用しないため、比較的少ない計算量で部分空間を求めることができるというメリットもある。本論文では、提案の部分空間以外に、従来の PCA、線形判別分析 (LDA) についても比較実験した。

3.4 時系列探索

最後に、求められた特徴ベクトルを用いて時系列探索を行う。時系列探索は、蓄積信号中に目的信号と同じ大きさの窓をかけ、目的信号と蓄積信号中の注目箇所との類似度を計算しながら、窓をずらし、目的信号と最も類似した蓄積信号中の箇所を探索結果として出力する。目的信号の長さが周波数特徴の分析窓の長さよりも大きい場合は、目的信号から周波数特徴のある時間間隔で抽出し、各々の特徴ベクトルの要素からなる特徴ベクトルを探索用の特徴ベクトルとして用いる。また、本論文では、類似度としてユークリッド距離を用いることにする。すなわち、

$$D = \sqrt{(Z_q - Z_s)^t(Z_q - Z_s)} \quad (13)$$

で定義される距離を用いる。ただし、 Z_q は目的信号の特徴ベクトル、 Z_s は蓄積信号の照合箇所の特徴ベクトルである。本論文では、この距離尺度に基づき探索を行った。

4. 実験

本手法の有効性について検討するため、様々な音質の音響信号を目的信号として、蓄積信号から同一部分を探索し、探索精度を評価する実験を行った (図 5)。ここでは、時間周波数空間上の局所領域ごとの正規化による効果を確認する予備実験 (実験 1)、部分空間による効果を確認する予備実験 (実験 2)、騒音下での精度評価を行う実験 (実験 3) を行った。

4.1 実験データの収集

実験に先立って、楽曲、収録場所、携帯端末を複数選び、実環境下にて楽曲の収録を行った。これらの収録条件を表 1 に示す。収録した音響信号は、表 1 の楽曲、場所、機器すべての組合せ、すなわち合計 1428

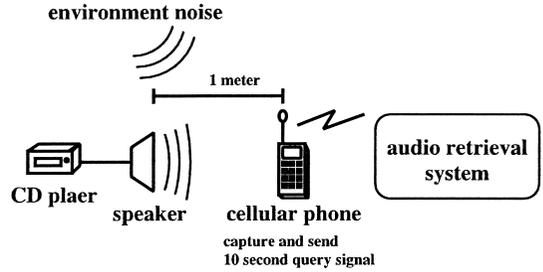


図 5 実験装置

Fig. 5 Experiment setting.

表 1 実環境で収録した音響信号

Table 1 Audio signals captured in the real world.

	内容
楽曲	34 曲 (様々なジャンル) ・学習用データ 17 曲 ・評価用データ 17 曲
収録場所	騒音レベル小 4 箇所 ・静かな部屋 ・アイドリング状態の車 ・コンビニエンスストア ・カラオケ店 騒音レベル大 3 箇所 ・人通りの多い街頭 ・騒がしい喫茶店 ・交通量の多い交差点付近
携帯端末	・マイク ・高音質携帯電話 (PHS, 通信会社 2 社) ・中音質携帯電話 (PDC など, 通信会社 3 社)

表 2 蓄積信号

Table 2 Stored signal.

楽曲	116 時間 (3310 曲)

個の音響信号 (各 42 秒) である。蓄積信号として用いた楽曲は実環境で収録したものの原信号を含む CD 品質の音響信号 116 時間分を用いた (表 2)。

4.2 時間周波数空間上の局所領域ごとの正規化による効果 (実験 1)

まず、時間周波数空間上の局所領域ごとの正規化による効果を調べるため探索実験を行った。実験は、目的信号の長さを 10 秒とし、目的信号の切出し部分を 200 回変えて探索精度の測定を行った (音響信号は 8000 Hz で標準化、特徴抽出のパラメータは $s = 512$, $N = 8192$, $M = 16$)。また、目的信号は、表 3 の条件で収録した音響信号を合計 12 種類用意し、合計 2400 回の探索を行った。探索時に使用した特徴ベクトルは、160 次元の特徴ベクトル (正規化後の周波数特徴の 0 ~ 4000 Hz を 16 分割し、各周波数帯域ごとに

成分を平均化し、1秒ごとに10個の周波数特徴からなる特徴ベクトルを用いた。本実験における探索例を図6に示す。次に、実験結果を表4に示す。ここでは、探索した時刻が正しい時刻の±1秒以内となった場合に正解とした。なお、5位累積分類率とは、5位候補までに正しい時刻が入った率とした。また、比較のため、パワーによる正規化(PN)、SS、CMNについても同様の実験を行った。比較に用いたそれぞれの正規化処理は、PNについては、各成分を全体のパワーで割る処理、すなわち、特徴ベクトルを

$$y(i, k) = \frac{1}{\sum_{n=0}^K Q(i, n)} Q(i, k), \quad (14)$$

とした。ただし、 K はFFTの分析における最大周波数とする。SSについては、得られた周波数特徴を1

秒間について平均し中央の周波数特徴から差し引く処理、すなわち、

$$y(i, k) = Q(i, k) - m(i, k), \quad (15)$$

とした。CMNについては、まず周波数特徴の対数を逆FFTすることによりケプレンシ k' のケプストラム $Q'(i, k')$ を求め、 $Q'(i, k')$ を1秒間について k' ごとに平均し中央のケプストラムから差し引く。すなわち、

$$y'(i, k') = Q'(i, k') - m'(i, k'), \quad (16)$$

を求める。ただし、

$$m'(i, k') = \frac{1}{2M} \sum_{i=-M}^{M-1} Q'(i, k'), \quad (17)$$

とする。最後に、 $y'(i, k')$ をFFTし、最後に指数をとる処理をした。提案の正規化は他の正規化法に比べて探索精度が数十%向上している。これにより、周波数特徴の時間周波数空間上の局所領域ごとの正規化が有効であることが確認された。

4.3 部分空間による効果(実験2)

次に、部分空間に射影することの有効性を調べるため探索実験を行った。ここでは、表1の学習用データの中から目的信号と同じ収録場所、携帯端末で収録した合計12個の音響信号からそれぞれ400箇所無作為に切り出したものを学習サンプルとして部分空間の学習を行った。学習により得られた部分空間を用いて、学習用データとは異なる評価用データ(表1)を用いて探索実験を行った。実験結果を表5に示す。なお、ここでは、比較のため、4.2同様に周波数帯域ごとに平均したもの、全学習サンプルをPCAしたもの、LDAを行ったものについても同様の実験を行った。探索時に使用した特徴ベクトルは、160次元の特徴ベクトル(使用する固有ベクトルの数を16とし、求めた特徴ベクトルを10秒の目的信号から1秒ごとに抽出)を用いた。なお、探索精度は4.2と同様の方法で測定した。また、使用した特徴ベクトルの次元数は、4.2

表3 実験1, 2(予備実験)で使用した音響信号
Table 3 Audio signals used in exp. 1 and exp. 2.

蓄積信号	12分(CDより17曲)
目的信号	2400個の10秒の断片(表1評価用データより)
目的信号の収録条件	収録場所: ・騒音レベル小4箇所 収録端末: ・マイク, PHS 2機種

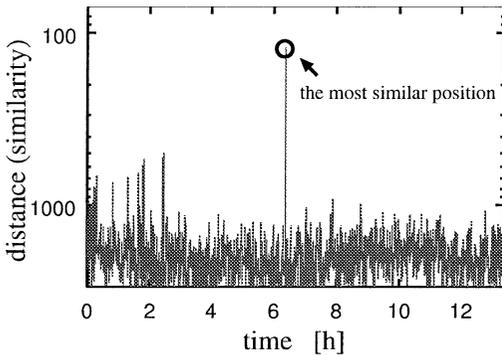


図6 探索の例
Fig. 6 An example of search.

表4 時間周波数空間上の局所領域ごとの正規化による効果(実験1)

Table 4 Effect of local time-frequency-region normalization (exp. 1).

	正解率	5位累積分類率
PN	12.8%	16.5%
SS	2.9%	3.5%
CMN	20.7%	25.6%
提案の正規化	61.5%	65.9%

表5 部分空間による効果(実験2)

Table 5 Effect of robust subspace spanning (exp. 2).

	正解率	5位累積分類率
提案正規化 + PCA	64.0%	68.9%
提案正規化 + LDA	64.8%	69.4%
提案正規化 + 提案部分空間	71.2%	74.8%

では周波数帯域を 16 分割することにより次元削減されていたのに対し、本節の実験で使用した特徴ベクトルの次元数は、上位 16 の固有値をもつ固有ベクトルのみを使用することで次元削減されている。提案の部分空間は他の手法に比べて探索精度が数%向上している。よって、クラス平均のみを使用した PCA により探索精度が向上することが確認された。

4.4 騒音下での精度評価 (実験 3)

次に、更に騒音の大きい環境や音質の異なる携帯電話を用いて探索精度を評価するため、より多種類のひずみのある目的信号 (表 6) を用いた実験を行った。表 7 に探索精度を示す。なお、括弧内には時刻を特定する探索の精度ではなく、蓄積信号の 3310 曲のうちどの楽曲であるかを特定する探索の精度も示した。また、探索時に使用した特徴ベクトルは、160 次元の特徴ベクトル (使用する固有ベクトルの数を 16 とし、求めた特徴ベクトルを 10 秒の目的信号から 1 秒ごとに抽出) を用いた。騒音レベル小では、マイク、高音質携帯電話で収録したもので探索した場合、71.8% の探索精度が得られている。しかし、騒音レベル大では、中音質携帯電話で収録したもので探索した場合、探索精度が 15.2% というように著しく低下しており、今後更に検討が必要である。

最後に参考まで、処理時間の測定を行った。ここでは CPU は Intel 社 Pentium III (1 GHz)、ハードディスクは ULTRA160 SCSI 対応を用いた。表 8 に結果

表 6 実験 3 で使用した音響信号
Table 6 Audio signals used in exp. 3.

	内容
目的信号	それぞれ 2400 個の 10 秒の断片 (表 1 評価用データより)
目的信号の収録条件	収録場所: ・騒音レベル小 4 箇所 ・騒音レベル大 3 箇所 収録端末: ・マイク、高音質携帯電話 2 機種 ・中音質携帯電話 3 機種

表 7 探索精度 (実験 3)
Table 7 Search accuracy (exp. 3).

	騒音レベル小	騒音レベル大
マイク、高音質携帯電話	71.8% (72.0%)	27.6% (27.6%)
中音質携帯電話	39.1% (39.5%)	15.2% (15.4%)

表 8 処理時間
Table 8 Processing time.

	処理時間	
前処理時間	18.2 時間	(18.6 時間)
探索時間	1.82 秒	(1.83.1 秒)

をまとめる。なお、ここに示す時間は、メモリに読み込んだ後から計測した時間を示し、括弧内にはハードディスクからメモリに読み込む時間を含めた時間を示した。ここでは SSDA [11] を用いて探索の高速化を行っている。前処理は蓄積信号の長さの 15% 程度の時間で完了し、探索は 100 時間程度の蓄積信号を 2 秒弱で完了することが確認された。

5. む す び

本論文では、実環境で収録された楽曲の断片をキーとし、膨大な楽曲データからそれと同一の楽曲を探索する手法を提案した。本手法は、部分空間上に特徴を射影することにより、特徴ひずみの少ない周波数帯域を選択することを特徴とし、その前処理として周波数特徴の時間周波数空間上の局所領域ごとに正規化を行う。実験を行ったところ、これらの効果により探索精度が向上することが確かめられた。例えば、蓄積した CD 品質の音楽音響信号をマイクや PHS で収録した 10 秒間の探索キー信号によって探索する実験では、騒音レベルが比較的小さい場所での収録の場合、71.2% の精度 (収録した音響信号の断片が、CD 品質の蓄積信号のどの曲のどの時刻に対応するか特定する) が得られ、本手法を用いない場合の 12.8% に比べてかなりの精度の改善が可能ながわかった。今後は、より大規模なデータを用いて本手法の可能性を詳細に評価し、本手法の高速探索法の適用についても検討する。更に、将来は、より条件の厳しい場合、すなわち、騒がしい場所での収録の場合や、PDC 方式などの中音質携帯電話で収録した探索キーによる探索の場合においても有効性の高い手法について更に検討する予定である。

文 献

- [1] 柏野邦夫, ガビン スミス, 村瀬 洋, “ヒストグラム特徴を用いた音響信号の高速探索法—時系列アクティブ探索法” 信学論 (D-II), vol. J82-D-II, no. 9, pp. 1365–1373, Sept. 1999.
- [2] 黒住隆行, 柏野邦夫, 村瀬 洋, “時系列アクティブ探索法のための特徴変動に頑健な確率ディザボータリング” 信学論 (D-II), vol. J84-D-II, no. 8, pp. 1817–1825, Aug. 2001.
- [3] 小特集—マイクロホンアレー, 音響誌, vol. 51, no. 5, pp. 384–414, 1995.
- [4] K. Lemstrom and S. Perttu, “SEMEX—An efficient Music Retrieval Prototype,” Int. Symp. Music Information Retrieval, 2000.
- [5] S. Furui, “Cepstral analysis technique for automatic speaker verification,” IEEE Trans. Acoust.

Speech Signal Process., vol.ASSP-29, no.2, pp.254-272, 1981.

- [6] 藤本雅清, 有木康雄, “音声の時間変化モデルに基づく音声信号推定法を用いた非定常雑音下での音声認識,” 信学技報, SP2000-78, 2000.
- [7] 庄境 誠, 中村 哲, 鹿野清宏, “ケプストラム平均正規化法と HMM 合成法に基づくモデル適応化法,” 信学論 (D-II), vol.J80-D-II, no.10, pp.2636-2644, Oct. 1997.
- [8] 中川聖一, “音声認識研究の動向,” 信学論 (D-II), vol.J83-D-II, no.2, pp.433-457, Feb. 2000.
- [9] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-27, no.2, pp.113-120, 1979.
- [10] L.R. Rabiner and B.H. Juang, Fundamentals of speech recognition, Prentice Hall, 1993.
- [11] D.I. Barnea and H.F. Silverman, “A class of algorithms for fast digital image registration,” IEEE Trans. Comput., vol.C-21, no.2, pp.179-186, 1972.

(平成 14 年 10 月 11 日受付, 15 年 2 月 24 日再受付)



黒住 隆行 (正員)

平 9 都立大・理・物理卒。平 11 北陸先端科学技術大学院大学情報科学研究科博士前期課程了。同年 NTT に入社。現在, NTT コミュニケーション科学基礎研究所に所属。パターン認識, 画像処理, マルチメディア情報検索に興味をもつ。平 14 本会業績賞

受賞。



柏野 邦夫 (正員)

平 2 東大・工・電子卒。平 7 同大学院電気工学専攻博士課程了。工博。同年 NTT に入社。現在, NTT コミュニケーション科学基礎研究所主任研究員。音響信号の認識・分離・探索, 及び情報統合の研究に従事。平 14 から 1 年間英国ケンブリッジ大

客員研究員。メディア情報を対象とする信号処理及び知識処理に興味をもつ。平 5 情報処理学会奨励賞, 平 9 人工知能学会全国大会優秀論文賞, 平 11 日本音響学会粟屋潔学術奨励賞, 平 13 本会ソサイエティ論文賞, 平 14 本会業績賞各受賞。情報処理学会, 日本音響学会, 人工知能学会, 日本音楽知覚認知学会, IEEE, 各会員。



村瀬 洋 (正員)

昭 53 名大・工・電子卒。昭 55 同大学院修士課程了。同年日本電信電話公社(現 NTT)に入社。以来, 文字・図形認識, コンピュータビジョン, マルチメディア認識の研究に従事。平 4 から 1 年間米国コロ

ンビア大客員研究員。平 13, NTT コミュニケーション科学基礎研究所メディア情報研究部部長。平 15, 名古屋大学大学院情報科学研究科教授, 現在に至る。工博。昭 60 本会学術奨励賞, 平 4 電気通信普及財団テレコムシステム技術賞, 平 6 IEEE-CVPR 国際会議最優秀論文賞, 平 7 情報処理学会山下記念研究賞, 平 8 IEEE-ICRA 国際会議最優秀ビデオ賞, 平 13 高柳記念電子科学技術振興財団高柳記念奨励賞, 平 13 本会ソサイエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞各受賞。情報処理学会, IEEE 各会員。