

LI-18 携帯電話で受信した音をキーとする音響信号探索 A Robust Audio Searching Method Using an Audio Signal Query Received with Portable Phones

黒住 隆行*
Takayuki Kurozumi

柏野 邦夫*
Kunio Kashino

村瀬 洋*
Hiroshi Murase

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である。

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质の良い音や映像において有効性の高い手法であった。また、既知の特徴歪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歪みを付加しながら探索する手法 [2] についても検討した。

ところで、近年では、ノート型 PC, 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が増えてきている。もし、携帯端末でとらえた音や映像を探索キーとして情報検索ができれば、街角で流れている楽曲を携帯端末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。しかし、このような実環境下での探索を考えた場合、携帯端末の特性や環境雑音による信号の歪みにより探索精度が著しく低下するという問題がある。そこで本稿では、携帯端末により実環境で収録した音に基づく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。

実環境下での雑音吸収に関する研究は既に多数行われている。例えば、加法性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) [3], 乗法性歪みを吸収するためのケプストラム平均正規化(CMN) [4] など、簡便で有用性が高い方法が提案されている。また、マイクロホンアレーを用いて雑音除去を行う手法 [5] も多数提案されているが、小型の入力端末への適用には限界があると考えられる。さらに、音楽の探索という観点から見ると、音楽固有の特徴(高調波成分など)からシンボリックな情報(例えば単音など)を抽出し、その情報に基づいて探索を行う方法も考えられ、このような音楽検索の研究も近年盛んになってきている [6]。しかし、一般の音楽に対して精度良く単音抽出を行うことは困難であり、このよう

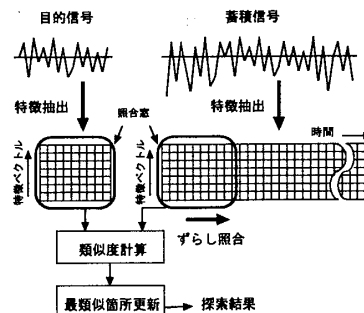


図1: 時系列探索の概要

なアプローチにも限界があると考えられる。

そこで本稿では、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、特徴歪みに頑健な方法を検討することにする。本稿の基本的なアイデアは、機器特性や環境雑音などによる歪みに対して頑健な部分空間を構成して照合を行うというものである [7]。その際、時間周波数空間上の局所領域ごとに正規化を行う。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. 探索の原理

探索は、(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) 時系列探索の手順で行う。本稿では、音響特徴の周波数特徴としてFFTにより求めた短時間パワースペクトルを用いた。時間周波数空間上の局所領域ごとに正規化を行う目的は、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、ある時間区間について平均と標準偏差を求め、それらを用いた正規化、すなわち、時間区間内の値をサンプルとみだたガウシアン正規化を行う。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、周波数特徴を特徴歪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、より特徴歪みに頑健な特徴抽出を行うことである。具体的には、学習用信号としてCDの楽曲などの原信号と同一楽曲について歪みのある信号を複数用意し、それらを同一箇所について平均を求める。平均を複数箇所について求め、それらから計算される共分散行列、すなわち、楽曲の同一箇所を同じ級とする級間共分散行列を求める。そして、その級間共分散行列の固有ベクトルを求め、固有値の大きい固有ベクトルから成る部分空間を構成する。この部分空間に正規化済周波数特徴を射影することで歪みに頑健な特徴を抽出することができると考えられる。なお、本手法は、従来の主成分分析(PCA)や判別分析(LDA)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特徴歪みに頑健な部分空間を求めることができるというメリットがある。

* 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

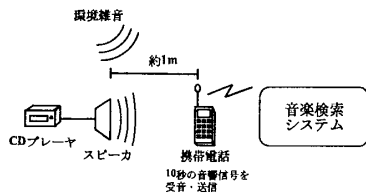


図 2: 探索実験の例 (携帯電話を用いた場合)

3. 実験

本手法の有効性を検討するため、さまざま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 2)。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実験 (実験 1)、部分空間による効果を確認する実験 (実験 2)、および、実環境での精度を確認する実験 (実験 3) を行った。

実環境下での探索の問題を検討するためには、実際に想定される場所や端末について検討することが重要である。そこで、実験に先だって、音楽探索がされうると想定される楽曲、場所、端末を検討し、表 1 に示す全ての組み合わせについて収録を行った。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 116 時間分の音響信号を用いた (表 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を調べるため探索実験を行った (実験 1)。実験は、目的信号の長さを 5 秒とし、目的信号の切り出し部分を 200 回変えて探索精度の測定を行った (音響信号は 8000Hz で標準化、特徴抽出の間隔は 0.06 秒、FFT の分析窓の大きさは 1 秒、平均と標準偏差を求めるための時間区間を 2 秒とした)。また、目的信号は、表 1 の条件で収録した音響信号のうち雑音レベル 1 のマイクと PHS の合計 12 種類を使用し、合計 2400 回の探索を行った。探索には 80 次元の特徴ベクトル (正規化済周波数特徴の 0~2000Hz を 16 分割し、各周波数帯域毎に成分を平均化したものを、1 秒毎に 5 つ使用した) を用いた。また、類似度としてユークリッド距離を用いた。実験結果を表 3 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。また、比較のため、信号のパワーによる正規化 (PN), SS, CMN を行ったものについても同様の実験を行った。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。

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 (実験 2)。ここでは、表 1 の学習用データの中

表 1: 実環境で収録した探索キーとなる音響信号

楽曲	20 分 (ロック, ポップス, ジャズ, クラシックを含む 34 曲) 10 分は学習用, 残りの 10 分はテスト用
場所	4 箇所 (雑音レベル 1) オフィス, アイドリング中の車内, コンビニエンスストア, カラオケ店 3 箇所 (雑音レベル 2) 人通りの多い街頭, 混雑した喫茶店, 交通量の多い交差点
端末	マイク, PHS (通信会社 2 社) 携帯電話 (通信会社 3 社)

表 2: 蓄積信号

楽曲	116 時間 (3310 曲)
----	-----------------

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 300 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テスト用データ (表 1) を用いて探索実験を行った。実験結果を表 4 に示す。なお、ここでは、比較のため、PCA, LDA についても同様の実験を行った。探索には 80 次元の特徴ベクトル (固有値の大きさ上位 16 の固有ベクトルにより変換して得たものを、1 秒毎に 5 つ使用した) を用いた。提案の部分空間は他の手法に比べて探索精度が向上している。

次に、本提案手法を用いて、より多種類の歪みのある目的信号 (表 1 の全ての雑音レベル, 端末を使用) を用いて実験を行った (実験 3)。表 5 に探索精度を示す。雑音レベル 1 で、マイク, PHS で収録したもので探索した場合、探索精度が 69.3% というように高い精度が得られている。しかし、雑音レベル 2 で、携帯電話で収録したもので探索した場合、探索精度が 10.0% というように著しく低下しており、今後さらに検討が必要である。

4. むすび

本稿では、携帯電話を用いて実環境中で流れている音を探索キーとした音響信号探索を、時間周波数空間上の局所領域ごとの正規化、部分空間への射影を用いた時系列探索により行った。提案法では、マイク, PHS 品質の目的信号を用いた場合、良好な探索精度が得られたが、携帯電話を用いると探索精度が低下し、改善の必要があることがわかった。今後、さらに高速かつ高精度な探索法について検討する。

謝辞 日ごろ御指導を頂く NTTCS 研の石井所長, 管村部長及び基識 G の諸氏に感謝致します。

参考文献

- [1] 柏野, スミス, 村瀬. “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, 1999.
- [2] 黒住, 柏野, 村瀬. “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, 2001.
- [3] Boll. “Suppression of acoustic noise in speech using spectral subtraction”. *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-27, No.2, 1979.
- [4] Furui. “Cepstral analysis technique for automatic speaker verification”. *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-29, No.2, 1981.
- [5] “小特集—マイクロホンアレー—”. 音響誌, Vol.51, No.5, 1995.
- [6] Lemstrom, Perttu. “SEMEX - An efficient Music Retrieval Prototype”. *MIR 2000*.
- [7] 黒住, 柏野, 村瀬. “特徴ひずみに頑健な部分空間を用いた携帯端末による音響信号探索”. 信学技法 PRMU2001-43, 2001.

表 3: 提案の正規化による効果 (実験 1)

	PN	SS	CMN	提案正規化
正解率	8.8 %	2.5 %	12.7 %	58.3 %

表 4: 提案の部分空間による効果 (実験 2)

	PCA	LDA	提案部分空間
正解率	68.2 %	61.8 %	68.8 %

表 5: 精度評価 (実験 3)

	雑音レベル 1	雑音レベル 2
マイク, PHS	69.3 %	23.4 %
携帯電話	30.0 %	10.0 %