論　文

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——

柏野　邦夫[†]　　ガビン　スミス[†*]　　村瀬　洋[†]

A Quick Search Algorithm for Acoustic Signals Using Histogram Features
——Time-Series Active Search——

Kunio KASHINO[†], Gavin A. SMITH[†*], and Hiroshi MURASE[†]

あらまし　放送など長時間の音響信号中から，特定のテーマ曲やCMなど，目的音響信号の有無及び時刻を高速に探索する方法を提案する．従来の，スペクトルや波形のずらし照合に基づく探索では，長時間の音響信号を探索対象とした場合，計算量が膨大となるという問題があった．これに対し本論文で提案する方法では，スペクトル特徴のヒストグラムに基づいて探索を行うことにより，大幅に計算時間を短縮することができる．例えば，ワークステーションを用いた実験では，音響信号からあらかじめスペクトル特徴を抽出しておいた場合，6時間の音響信号から目的音響信号（15秒間）を所要約2.3秒で正しく探索できることがわかった．また，白色ガウス雑音の重畳に対しては，SN比20dBまで頑健であることがわかった．

キーワード　音響探索，ヒストグラム，枝刈り，ベクトル量子化，時系列アクティブ探索法

1. まえがき

本論文では，長時間の音響信号（これを入力信号という）と，探したい具体的な音響信号（これを参照信号という）が与えられたとき，入力信号（input signal）中に参照信号（reference signal）が存在するかどうか，また存在するとすればどこにあるかを，高速に探索する方法を提案する．

マルチメディアデータベースの検索などに音響情報を役立てようとする研究は，既にいくつか行われてきている．そのほとんどは，内容検索を目的として，音響情報のインデクシングや分類を試みたものである．このような研究は，時間領域や周波数領域における様々な特徴量に基づくものと，ワードスポッティングに基づくものに大別され，前者の例としては[1]〜[4]など，また後者の例としては[5], [6]などがある．これらのほかに，テレビコマーシャルに特化して種々の経験則を適用し，その検出を試みた報告もある[7].

本論文で扱う問題はこれらとは異なり，参照信号が具体的に与えられていることと，入力信号中に参照信号が（含まれているとすれば）ほぼそのままの形で含まれている（スペクトルの変動が小さい）ことを仮定する．このような問題を，本論文では一致検索と呼ぶ．ただし，必ずしも，文字列探索のように入力信号のある部分が参照信号と全く一致していることは仮定せず，現実のアナログ信号に見られる程度の小さい変動や雑音の重畳はあり得るものとする．

このような音響信号の一致検索は，内容検索に比べて応用範囲が狭いように思われがちであるが，実は，放送におけるコマーシャル（CM）の検出や統計情報の作成，放送やインターネットにおける楽曲等の音データ使用のチェックなど，幅広い応用が考えられる．現在，文字列の高速探索アルゴリズムが広く用いられているのと同様，音響信号の高速探索アルゴリズムも，今後音響情報ハンドリングの基本技術として重要になるものと考えられる．

もちろん，音響信号の一致検索自体は，スペクトルや波形のずらし照合などの従来技術で解決できる問題である．また，既知の信号の検出は，信号検出理論として体系化も行われている[8]．しかし従来法では，長時間の入力信号や，多数の参照信号を対象とする場合

[†] NTTコミュニケーション科学基礎研究所，厚木市
NTT Communication Science Laboratories, 3–1 Morinosato-Wakamiya, Atsugi-shi, 243–0198 Japan.
[*] 現在，ケンブリッジ大学

には，計算量が膨大となるため実用的な処理時間でもれなく検出あるいは探索することは難しい．高速化のために，照合の仕方や時間方向のずらし方を粗くすることも考えられるが，その場合は探索もれの発生など探索精度の低下が避けられないという問題があった．

本論文で提案するのは，十分な精度を維持したまま，探索を大幅に高速化する方法である．我々は，提案法を時系列アクティブ探索法（time-series active search）と呼ぶ．高速化の鍵は，スペクトル特徴のヒストグラムを用いる点にある．すなわち提案法では，ヒストグラムの性質を利用することにより，全探索と同じ精度を保証したままで，実際に行う照合計算の回数を全探索に比べ 1/100 から 1/500 程度に削減することができる．これに加え，照合計算自体も，特徴を直接比較するよりも計算コストが小さい方法を用いることができる．これらの効果により，スペクトルや波形のずらし照合に比べ極めて高速に探索を行うことができる．

以下 **2.** において，提案法の概要を説明する．**3.** において，提案法の処理速度と処理精度に関して評価実験を行う．**4.** をむすびとする．

2. 時系列アクティブ探索法

2.1 方法の概要

図 1 に，時系列アクティブ探索法の概要を示す．本方法は，Vinod らが画像の空間探索手法として提案した「アクティブ探索法」[9], [10] と同様の枝刈り手法を，音響信号に対する時間探索に適用したものである [11].

はじめに，参照信号に対する処理を行っておく．すなわち，まず参照信号に対して特徴抽出を行い，特徴ベクトルの時系列を得る．次に，特徴ベクトルを，予め定めた方法に従って分類し，各分類に属する特徴ベクトルの出現回数を数えることによって，特徴ベクトルのヒストグラム（ヒストグラム特徴）を作る．ヒストグラムは，参照信号のすべての区間に対して一つ作成してもよいが，特徴ベクトル N_{div} 個の部分に時間分割して，それぞれについてヒストグラムを作成してもよい．後者の場合，信号の時間順序を考慮に入れることができる．

次に，入力信号に対する処理を行う．すなわち，参照信号に対する処理と同様に，まず特徴抽出を行って特徴ベクトルの時系列を得る．次に，この時系列に対し，参照信号と同じ長さの時間窓をかけ，特徴ベクトルのヒストグラムを作る．参照信号において時間分割したヒストグラムを作成した場合には，入力信号においても同様に時間分割したヒストグラムを作成する．

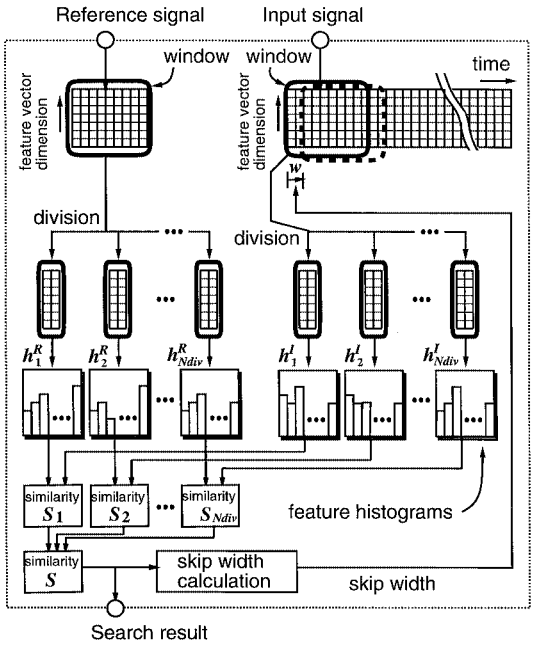

図 1　時系列アクティブ探索法の処理の流れ
Fig. 1　Block diagram of the proposed search algorithm.

続いて，ヒストグラム同士の類似度（similarity）を計算する．類似度があらかじめ設定したしきい値（これを探索しきい値という）以上の場合には，入力信号中に参照信号が発見されたことになる．類似度が探索しきい値に満たない場合には，入力信号に対する時間窓を時間方向にずらして探索を進める．このとき，後述するように，必ずしも一単位（特徴ベクトル一つ分）ずつずらす必要はなく，現在の類似度の値に応じて，探索もれを起こさないことを保証したまま，ずらすことのできる時間幅（スキップ可能幅；skip width）を求めることができる．このようにスキップ可能幅を求めて時間窓をずらすことによって，1 単位ずつ時間窓をずらす場合に比べて，計算量を大幅に削減することができる．

2.2 特徴抽出

特徴ベクトル $\boldsymbol{f}(k)$ は，

$$\boldsymbol{f}(k) = (\ f_1(k),\ f_2(k), \cdots,\ f_V(k)\ )\ , \qquad (1)$$

と書くことができる．ここで k は標本化された時刻，V は特徴の次元数である．$\boldsymbol{f}(k)$ の各要素は，正規化

された短時間パワースペクトルである．すなわち

$$f_j(k) = \alpha(k)\, Y_j(k), \tag{2}$$

$$Y_j(k) = \sum_{t=k-K+1}^{k} y_j^2(t), \tag{3}$$

$$k = lM \quad (l = 1, 2, \cdots), \tag{4}$$

である．ここで $y_j(t)$ は時刻 t における帯域通過フィルタ j の出力波形，M は特徴ベクトルの時間刻み，K は特徴ベクトルの計算に用いる時間区間の長さである．また $\alpha(k)$ は正規化のための係数であり，

$$\alpha(k) = \frac{1}{\max\limits_{j}(\ Y_j(k)\ )} \tag{5}$$

と定義される．

帯域通過フィルタ $y_j(t)$ は，例えば 2 次の IIR フィルタで実装することが可能である．このほか特徴としては，FFT に基づく周波数スペクトルや，MFCC（メル周波数ケプストラム係数）などを用いることも可能であるが，本論文では，十分な精度が得られ，比較的短時間で特徴抽出できる帯域通過フィルタを用いた．

2.3 ヒストグラムによる特徴のモデリング

提案法では，信号の特徴をモデル化する手段としてヒストグラム特徴を用いる．Swain らは，画像認識において，ヒストグラム特徴によって物体認識が可能であることを示している [12].

信号の類似性の判定は，ヒストグラムの照合に基づいて行う．ヒストグラム同士の類似度としては，ユークリッド距離など各種の定義が考えられるが，ここではヒストグラム重なり率を用いる．時間分割された i 番目のヒストグラムにおけるヒストグラム重なり率 S_i は，次のように定義される．

$$S_i(h_i^R, h_i^I) = \frac{1}{D_i} \sum_{l=1}^{L} \min(h_{il}^R,\ h_{il}^I) \tag{6}$$

ここで h_i^R と h_i^I は，それぞれ参照信号と入力信号に対する i 番目の時間分割に対するヒストグラムであり，h_{il}^R，h_{il}^I はそれぞれの l 番目のビンに含まれる度数である．また L はヒストグラムのビンの数，D_i は i 番目のヒストグラムの総度数である．時間窓全体における類似度 S は，S_i を用いて次のように定義できる．

$$S(h^R, h^I) = \min_i(S_i(h_i^R, h_i^I)) \tag{7}$$

ヒストグラム重なり率による類似度の定義を用いた理由は，(1) 計算が簡単であること，(2) 後述のように，時間窓の周辺における類似度の上限値が簡単な計算によって求められること，及び (3) 既に画像の物体認識に適用され，好ましい結果が得られていること [12], の 3 点である．

ヒストグラムの作成においては，特徴ベクトルを適切に分類する必要がある．分類を行う一つの方法は，ベクトル量子化である．通常のベクトル量子化では，学習ベクトルのクラスタリングによって符号帳を作成しておき，各クラスタの重心ベクトルと入力ベクトルの距離に基づいて量子化を行うことが多い．もちろんこのような方法でも差し支えないが，本論文では，より計算量の少ない方法として，特徴ベクトルの各次元を単にいくつかの区間（ビン）に分けることによって分類する方法を考える．この方法も，ベクトルをいくつかのパターンに分類しているという意味で，本論文ではベクトル量子化と呼ぶことにする．

さて，ある特徴ベクトルがビン $(b_1, b_2, \cdots, b_V)$ に属する確率を $P(B_1 = b_1, B_2 = b_2, \cdots, B_V = b_V)$，あるいは単に $P(b_1, b_2, \cdots, b_V)$ と書くと，ヒストグラムのエントロピー H は

$$H = -\sum_{b_1}\sum_{b_2}\cdots\sum_{b_V} P(b_1, b_2, \cdots, b_V) \cdot \log P(b_1, b_2, \cdots, b_V) \tag{8}$$

となる．ヒストグラムのエントロピーが最大となるのは，$P(b_1, b_2, \cdots, b_V)$ が $b_1, b_2, \cdots, b_V$ について等確率となるときである．このとき，ヒストグラムのもつ平均情報量が最大となるため，ヒストグラムの識別性能も最大となることが期待できる．更に，特徴ベクトルの各次元が独立であると仮定すると，

$$P(b_1, b_2, \cdots, b_V) = P(B_1{=}b_1)P(B_2{=}b_2)\cdots P(B_V{=}b_V) \tag{9}$$

が成り立つ．このことから，結局，$P(B_j = b_j)$ がすべての b_j の値について定数（等確率）であるとき $(j = 1, 2, \cdots, V)$，ヒストグラム全体の平均情報量が最大化されることがわかる．$P(B_j = b_j)$ を等確率とするためには，学習段階において，あらかじめ特徴ベクトルをサンプリングして，同数の特徴ベクトルが各ビンに入るようなビンの境界値を定めておけばよい．

ところで，各次元におけるビンの数は，上記の議論とは別に，適切に定める必要がある．ビンの数を増や

し，特徴ベクトルを細かく分類すればするほど，照合に必要な計算量が増えるとともに，ノイズや信号のひずみの影響を受けやすくなる．しかしビンの数が少なすぎると，信号を十分識別することができなくなる．本論文では，**3.** に述べるように，実験的に特徴ベクトルの各要素に対するビン数を定めた．特徴ベクトルの次元を V，各要素に対するビン数を B とすると，ヒストグラム全体でのビン数（ベクトル量子化における符号帳のサイズ）L は，

$$L = B^V \tag{10}$$

で与えられる．

2.4 類似度の上限とスキップ可能幅

ヒストグラムは特徴ベクトルの時系列を分類し累積したものであるから，入力信号の特徴ベクトルに対する時間窓の移動に伴って，式 (6) の類似度が急激に変化することはない．例えば，時間窓が一単位移動したとき，最も急に類似度が増加するのは，今まで類似度（本論文の場合，参照信号のヒストグラムとの重なり）に寄与していなかった特徴ベクトルが一つ時間窓外に出て，新たに時間窓内に入ってきた特徴ベクトルが類似度に寄与する場合である．このことから，式 (6) における類似度 S_i の変化率（時間窓の一単位の移動における S_i の増分）の絶対値は，決して $1/D_i$ を超えないことがわかる．

すなわち，入力信号に対する時間窓の先頭が n_1 番目の特徴ベクトルであるときの，i 番目の時間分割における類似度を $S_i(h_i^R, h_i^I(n_1))$ とすると，時間窓が n_2 番目の特徴ベクトルまで移動したときの類似度の上限値 $S_i^u(h_i^R, h_i^I(n_2))$ は，$n_1 < n_2 < n_1 + D_i$ のとき

$$\begin{aligned} &S_i^u(h_i^R, h_i^I(n_2)) \\ &\quad = S_i(h_i^R, h_i^I(n_1)) + \frac{n_2 - n_1}{D_i} \end{aligned} \tag{11}$$

で与えられる．ただし，式 (6) の定義から S_i は 1 を超えないので，式 (11) で与えられる $S_i^u(h_i^R, h_i^I(n_2))$ が 1 を超えるときは，1 が上限値となる．

今我々は，類似度が探索しきい値 θ を超える箇所を探し出そうとしている．したがって，式 (11) で与えられる上限値が θ 以下となる区間に対しては，照合を行う必要がない．そこで，式 (11) において，上限値を θ で置き換え，$n_2 - n_1$ を w_i とおいて整理することにより，i 番目の時間分割におけるスキップ可能幅 w_i を求めることができる．

$$w_i = \begin{cases} \mathrm{floor}\,(D_i(\theta - S_i)) + 1 & \text{if } S_i < \theta, \\ 1 & \text{otherwise.} \end{cases} \tag{12}$$

ただし，$\mathrm{floor}(x)$ は x を超えない最大の整数を表し，類似度が θ を超える箇所については全探索を行う（時間窓を 1 単位ずつずらす）こととしている．

更に，時間窓全体に対するスキップ可能幅 w について考える．式 (7) より，S_i のうちの一つでも θ 以下であれば，S も必ず θ 以下となるので，w_i のうちの i についての最大幅まで時間窓を移動させても，移動中に S が θ を超えることはない．このことから，w は

$$w = \max_i(w_i) \tag{13}$$

と求めることができる．

時系列アクティブ探索法は，類似度を計算するごとに w を求め，時間窓を特徴ベクトル w 個分だけずらして，また類似度を計算するという操作の繰り返しで構成される．この操作により，類似度が低い（参照信号がありそうもない）時点では大きくスキップし，類似度が高い時点では緻密に類似度を計算するという適応的な探索動作が，探索しきい値を設定するだけで自動的に行われることになる．このとき重要なのは，経験則によるスキップなどとは異なり，上記の議論から明らかなように，類似度が探索しきい値を超える箇所はもれなく探索できることが保証されている点である．

実際の探索では，すべての時間分割について w_i を計算するよりも，順に類似度 S_i を評価していき，探索しきい値に満たない S_i が発見された時点で直ちに時間窓をスキップする方が効率的である．その場合のスキップ可能幅は，それまでに計算されている w_i のうちの最大値とすればよい．また，現在我々が実装している探索システムの最終的な出力としては，類似度が探索しきい値を超えた箇所をすべて出力するのではなく，類似度が探索しきい値を超えた箇所のうちで，類似度が極大となっている時点のみを出力するようにしている．

2.5 探索しきい値の設定

予備実験を行った結果，様々な参照信号や入力信号に対するヒストグラム重なり率は，それぞれ異なった統計的特徴をもつことがわかった．このため，探索しきい値は，固定の値を定めておくのではなく，参照信号と入力信号に応じて定めた方がよい．そこで我々は，探索しきい値 θ を次式で与えることを考えた．

$$\theta = m + c\,\sigma. \tag{14}$$

ここで，m と σ は，それぞれ，与えられた参照信号に対して入力信号をサンプリングし，予備的に類似度の計算を行って収集した類似度値の平均と標準偏差であり，c は経験的に与えられる係数である．式 (14) は，入力信号中で，参照信号との類似度が平均的な値から飛び抜けて高い部分を探索することを意味している．

3. 実　　験

前章に述べた探索法を小型ワークステーションに実装し，探索速度と探索精度に関して実験を行った．実験に用いたワークステーションの仕様を表 1 に示す．

3.1 実験 1：探索速度

提案法の探索速度を評価するため，テレビ放送 6 時間分の音響信号データから，特定の 15 秒の CM を探索するのに要する時間を測定した．

まず，1998 年 1 月 22 日 18 時 21 分から，ある在京民放テレビ局の放送を神奈川県内で受信して家庭用ビデオデッキで 6 時間録画した（VHS HiFi，3 倍モード）．次に，この録画テープを再生して，音響信号を上記ワークステーションに取り込んだ．取込みは，入力信号用として 6 時間分を 1 回取り込んだほか，参照信号用として，同じテープから無作為に 15 秒の異なる CM を 10 本選択して再生し，入力信号用とは別に取り込んだ．いずれの場合も，取込みは標本化周波数 11.0 kHz，量子化精度 8 bit 直線，モノラルで行った．

取り込んだ音響信号は，7 チャネル（$V = 7$）の 2 次 IIR 帯域フィルタバンクで特徴抽出を行った．フィルタの Q は 10 とし，フィルタの中心周波数は 200 Hz から 3.7 kHz の間で対数周波数軸上に等間隔に配置した．特徴の時間区間長 K は 768, 特徴の時間刻み M は 128 とした．すなわち，約 70 ms の区間の平均パワーを約 12 ms ごとに算出したことになる．これらのパラメータはすべて予備実験によって定めた．

探索に要する時間は，(1) 特徴抽出に要する時間（特徴抽出時間），(2) 特徴ベクトルのベクトル量子化に要する時間（ベクトル量子化時間），(3) ベクトル量子化の結果を用いて探索を実行するのに要する時間（探索実行時間，すなわちヒストグラムの作成，類似度の算出，窓の移動の繰返しに要する時間），の三つからなる．これらを順に議論する．なお，時間はいずれも CPU 時間で測定した．CPU 時間は測定ごとに数%程度のばらつきが見られたので，以下では，各値とも 5 回同じ測定を行った平均値を示している．

表 1　実験に用いた計算機の仕様
Table 1　Specification of the workstation used in the experiments.

モデル名	SGI 社 O_2
CPU	R10000 (250 MHz)
メモリ	384 MByte
OS	IRIX Release 6.3
コンパイラ	MIPSPRO C Compiler ver 7.00

3.1.1 特徴抽出時間

特徴抽出時間は，M，V，Kの値によって異なるが，本実験の例では，6 時間分の入力信号と 15 秒の参照信号から特徴を計算するのに要する CPU 時間は約 217 秒であった．すなわち，実時間の約 1%の時間で特徴抽出が可能である．したがって，仮に信号の計算機への取込みと同時に処理を行うとすれば，約 1%の CPU 負荷で特徴抽出が行える．

3.1.2 ベクトル量子化時間

6 時間の入力信号と 15 秒の参照信号に対して，特徴ベクトルは合計 1,861,759 個算出される．ベクトル量子化は，これらを，式 (10) で与えられる L 通りに分類する段階である．本実験では $B = 3$ としたので，$L = 2187$ である．

ベクトル量子化は，既に議論したように特徴の各次元についての大小判定に帰着されているため，特徴抽出に比べて高速な処理が可能である．6 時間 15 秒分（上記個数）の特徴ベクトルのベクトル量子化に要する CPU 時間は，約 1.7 秒であった．これは，すべての特徴ベクトルをメモリにロードしてから，オンメモリで処理する時間を計測したものである(注 1)．なお，この時間には，各ビンに入る特徴ベクトルの数を均等にするようにビンの境界値を決めるための時間は含まれていない．これは，事前に十分な数のサンプリングを行ってビンの境界値を決めておけば，探索のたびごとに境界値を計算し直す必要はないためである．

3.1.3 探索実行時間

探索実行時間の測定結果を表 2 に示す．探索実行時間は，参照信号，入力信号，N_{div}，及び探索しきい値に依存する．表 2 に示した CPU 時間は，10 本の参照信号（CM）について 5 回ずつ測定した平均値である．また本実験では，探索しきい値は式 (14) にかかわらず $\theta = 0.7$ に固定して測定した．また表 2 では，提案法における照合回数が全探索の場合に比べ平均でどれだけ削減されたか（照合回数の比）も併せて記した．

(注 1)：現在の実装では，7 次元特徴ベクトル 1 個を 7 バイトで保持しているので，6 時間分の特徴ベクトルは正味約 13 M バイトとなる．

表 2　探索実行時間
Table 2　Search time.

N_{div}	CPU 時間		速度向上	照合回数の比	参照
	全探索	提案法			
1	35.4 s	0.59 s	60 倍	1/444	図 2
2	35.3 s	0.65 s	54 倍	1/339	図 3
4	35.2 s	0.72 s	49 倍	1/200	図 4
8	35.2 s	0.87 s	40 倍	1/109	図 5

全探索とは，スキップ可能幅 w を 1 に固定した探索のことである．ただし，その場合も $N_{div} > 1$ のときは各分割ごとに順に照合を行い，探索しきい値未満の類似度が発見された時点で直ちに窓を移動させている．

なお表 2 のいずれの場合も，10 本の CM すべてに対して探索結果は正しい（探索もれも，余分な探索もなく，探索結果の時間誤差は 1 秒以内である）ことを確認している．

全探索の場合に比べた提案法における照合回数は，$N_{div} = 1$ の場合が最も削減の効果が大きくなっているが，これはヒストグラム一つ当りの時間幅が大きいほど平均的なスキップ可能幅が大きくとれることによると考えられる．また，いずれの場合も実際の速度向上の比が照合回数の比よりも小さくなっているが，これは，ヒストグラム一つ当りの作成コストが，全探索のように特徴ベクトル一つ分ずつ時間窓がずれていく場合の方が小さいことと，提案法の場合にはスキップ可能量の算出が必要なことの 2 点が理由であると考えられる．

一方，本実験で用いたのと同一の特徴ベクトルを用いて，従来法である特徴ベクトルの相関に基づく探索を行ったところ，すべての特徴ベクトルをメモリにロードした時点から計測して，約 685 秒の CPU 時間を要した．提案法では，例えば $N_{div} = 1$ のとき，量子化ベクトル時間と探索実行時間の和は約 2.3 秒であるので，相関法に比べて約 298 倍探索が高速化されたことになる．

なお，提案法においては，特徴抽出とベクトル量子化は，探索に先立ってあらかじめ行っておくことができ，一度行っておけば，次からは行う必要がないという特徴がある．したがって，多数の参照信号について探索を行う場合などでは，相関法に比べて一層効率的な処理が可能である．例えば，100 本の異なる 15 秒の CM を同一の入力信号に対して探索する場合には，特徴抽出後，上記相関法では約 685 秒 × 100 = 約 68,500 秒（約 19 時間）の CPU 時間を要するのに対し，提案法では，探索実行時間の 100 倍にベクトル量子化時間（6 時間と 1,500 秒分）を加えた時間があれば十分である．本実験の値を用いると，その時間は $N_{div} = 1$ のとき約 61 秒と計算でき，これは相関法の約 1,100 倍の探索速度に相当する．

図 2 ～ 6 において，横軸は時刻，縦軸は類似度（0～1）を示す．

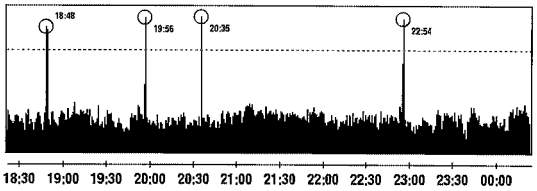

図 2　探索結果（$N_{div} = 1$）
Fig. 2　Search result. ($N_{div} = 1$)

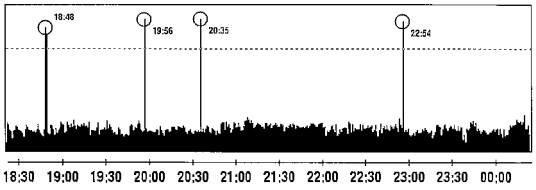

図 3　探索結果（$N_{div} = 2$）
Fig. 3　Search result. ($N_{div} = 2$)

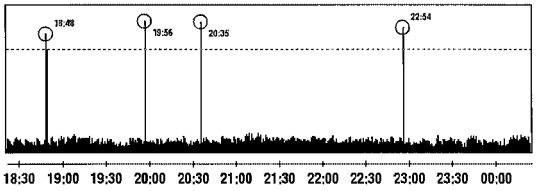

図 4　探索結果（$N_{div} = 4$）
Fig. 4　Search result. ($N_{div} = 4$)

参考のため，図 2 から図 5 に，表 2 の各場合における類似度の時間変化パターン（ある同一の CM を参照信号とした場合）を示す．各図において，○印は探索された時点を，破線は探索しきい値を示す．$N_{div} > 1$ とすることにより時間情報が加味され，しきい値設定に対するマージンが増大していることがわかる．また図 6 に，上記相関法による類似度（参照信号と時間窓内の入力信号における特徴ベクトルの相関値）の時間変化パターンを示す．

3.2　実験 2：探索精度

提案法の探索精度を調べるため，実験 1 とは別のテ

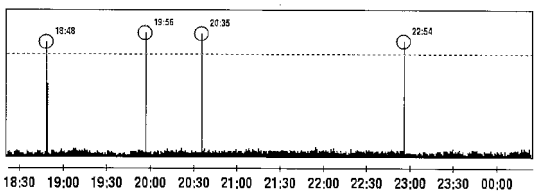

図 5　探索結果（$N_{div} = 8$）
Fig. 5　Search result. ($N_{div} = 8$)

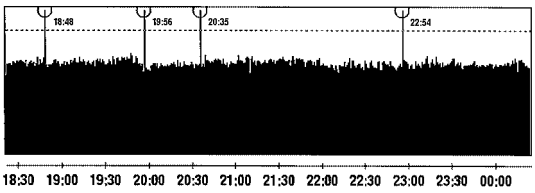

図 6　探索結果（従来法である相関法を用いた場合）
Fig. 6　Search result based on the correlation values.

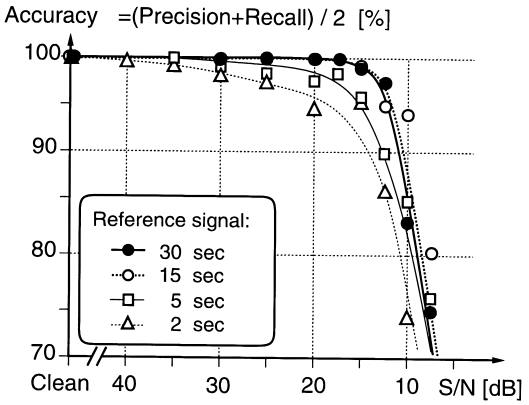

図 7　探索精度（$N_{div} = 1$ のとき）
Fig. 7　Search accuracy. ($N_{div} = 1$)

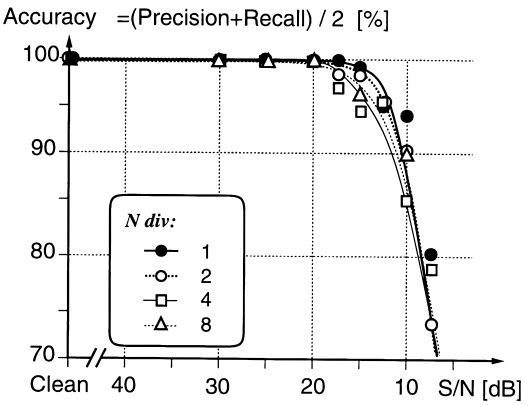

図 8　探索精度（参照信号が 15 秒のとき）
Fig. 8　Search accuracy. (Reference signal: 15 s)

レビ放送の録画を用いて実験を行った．まず，1998 年 6 月 1 日に，実験 1 と同様の方法でテレビ放送を録画し，異なる CM の部分をつなげて 20 分に編集した．これは，繰返しを含まない試料を用いた方が，実験の自動化に都合がよいためである．この 20 分のビデオテープを再生し，音響信号を 2 回に分けてワークステーションに取り込んだ．このうちの一方から，一定の時間区間をランダムな場所から切り出して参照信号とし，他方を入力信号として探索を行った．入力信号に対して白色ガウス雑音を加えた場合についても実験を行った．

本実験では，参照信号の長さ，時間分割数（N_{div}），及び雑音を重畳した際の SN 比をパラメータとした．SN 比は，入力信号 20 分間の平均電力に対して，雑音の平均電力（入力信号に加算する平均 0 の正規乱数の分散）を設定することによって制御した．同一の実験条件において，100 回繰り返して探索を行い，精度を測定した．精度は，適合率（precision rate）と再現率（recall rate）の平均値で評価した．ここで適合率とは，探索結果として出力されたもののうち正しいものの割合であり，再現率とは，探索されるべきもののうち探索結果として出力されたものの割合である．適合率と再現率がともに 100%であれば，探索もれも余分な探索もないことを意味する．適合率や再現率は，探索しきい値の設定によって変化するが，本実験では，式 (14) によって探索しきい値を定めた．すなわち，式 (14) における c の値を 100 回の繰返し中一定とし，その一定値を調節することによって，精度を最大化した値を評価値とした．その他の取込みや探索のパラメータは実験 1 と同様とした．

実験結果を図 7 と図 8 に示す．

図 7 は，時間窓の分割数を $N_{div} = 1$ とし，参照信号の長さをパラメータとして探索精度を測定したものである．図 7 に示されるように，参照時間の長さが 15 秒以上確保できれば，SN 比 20 dB まで探索もれも余分な探索も生じていない．また，SN 比が 30 dB 以上であれば，参照信号が 2 秒であっても，98%程度以上の探索精度が得られることがわかる．これらのことから，提案法の精度は，テレビ放送からの特定の CM の探索などの応用に対しては，十分実用的な水準にあ

ると判断できる．

図 8 は，参照信号の長さを 15 秒とし，時間窓の分割数（N_{div}）をパラメータとして探索精度を測定したものである．図 2 から図 5 においては，N_{div} が大きいほど，探索されるべき箇所と探索されるべきでない箇所とにおける類似度値の差（探索しきい値のマージン）が大きいことが示されていたが，図 8 によれば，探索精度の値は N_{div} の値にあまり依存していない．すなわち，N_{div} を大きくしても，雑音が重畳されたときの探索精度を向上させる効果は明らかでない．これは，探索しきい値のマージンの拡大の効果と，雑音の重畳によって入力信号中の正解場所における類似度が低下する効果とが拮抗しているためであると考えられる．

4. むすび

本論文では，参照信号として具体的な音響信号が与えられたとき，それが長時間の入力信号中のどこに存在するかを高速に探索する手法を提案した．提案法は，各信号の特徴ベクトルをベクトル量子化し，量子化した各符号に対するヒストグラムを作って，ヒストグラム同士の照合によって参照信号と入力信号との類似度を判定するというものである．この過程で，ヒストグラムの代数的性質を利用することによって不要な照合計算を省略できる．実際のテレビ放送の音響信号を用いて実験を行ったところ，あらかじめ特徴抽出を行っておいた場合，6 時間の入力信号から 15 秒の参照信号をもれなく探索するのに要する時間は，小型ワークステーションで約 2.3 秒であった．また白色ガウス雑音の重畳に対しても，SN 比 20 dB までは頑健であることがわかった．一方，ベクトル量子化までをあらかじめ行っておいた場合には，上記探索の所要時間は約 0.59 秒（時間窓の分割を行わない場合）であることがわかった．したがって実用上は，入力信号や参照信号に対して，特徴ベクトルのベクトル量子化までを行った情報を蓄積しておくことにより，効率的な探索が可能である．

提案法は，音響信号ばかりではなく，特徴の変更により映像など他のメディアの探索への応用も期待できる．今後は，他のメディアを併用した探索について検討を進めたいと考えている．

謝辞　日ごろ御指導を頂く NTT コミュニケーション科学基礎研究所の東倉洋一所長，石井健一郎部長，及び萩田紀博部長に感謝する．また日ごろ御協力を頂く同研究所メディア認識研究グループの諸氏に感謝する．

文　献

[1] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio," IEEE Multimedia, vol.3, no.3, pp.27–36, 1996.

[2] S. Pfeiffer, S. Fischer, and W. Effelsberg, "Automatic audio content analysis," Proc. ACM Multimedia, pp.21–30, 1996.

[3] J. Saunders, "Real-time discrimination of broadcast speech/music," Proc. of ICASSP-96, vol.2, pp.993–996, 1996.

[4] S.R. Subramanya, R. Simha, B. Narahari, and A. Youssef, "Transform-based indexing of audio data for multimedia Databases," Proc. IEEE Conf. on Multimedia Computing and Systems, pp.211–218, 1997.

[5] S.J. Young, M.G. Brown, J.T. Foote, G.J.F. Jones, and K.S. Jones, "Acoustic indexing for multimedia retrieval and browsing," Proc. of ICASSP-97, vol.1, pp.199–202, 1997.

[6] C.J. Wellekens and P. Gelin, "Remap for video soundtrack indexing," Proc. of ICASSP-97, vol.2, pp.1423–1426, 1997.

[7] R. Lienhart, C. Kuhmunch, and W. Effelsberg, "On the detection and recognition of television commercials," Proc. IEEE Conf. on Multimedia Computing and Systems, pp.509–516, 1997.

[8] J.C. Hancock and P.A. Wintz, "Signal Detection Theory," McGraw-Hill, 1966.

[9] V.V. Vinod and H. Murase, "Focused color intersection with efficient searching for object extraction," Pattern Recognition, vol.30, no.10, pp.1787–1797, 1997.

[10] 村瀬　洋，V.V. Vinod，"局所色情報を用いた高速物体検索 – アクティブ探索法，" 信学論（D-II），vol.J81-D-II, no.9, pp.2035–2042, Sept. 1998.

[11] G. Smith, H. Murase, and K. Kashino, "Quick audio retrieval using active search," Proc. of ICASSP-98, vol.6, pp.3777–3780, 1998.

[12] M.J. Swain and D.H. Ballard, "Color indexing," Int. J. Computer Vision, vol.7, no.1, pp.11–32, 1991.

（平成 10 年 12 月 17 日受付，11 年 4 月 2 日再受付）

柏野　邦夫　（正員）

平2東大・工・電子卒．平7同大大学院電気工学専攻博士課程了．工博．同年NTTに入社．以来，音響認識・分離・探索，及び情報統合の研究に従事．現在，NTTコミュニケーション科学基礎研究所研究主任．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．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．

ガビン　スミス

平9英国ケンブリッジ大卒．同年9月から7か月間，研修生としてNTT基礎研究所（当時）に滞在．現在，ケンブリッジ大PhDコースに在学中．音響信号処理に興味をもつ．

村瀬　　洋　（正員）

昭53名大・工・電子卒．昭55同大大学院修士課程了．同年日本電信電話公社（現NTT）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．現在，NTTコミュニケーション科学基礎研究所メディア認識研究グループリーダ．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受賞．情報処理学会，IEEE各会員．