音楽情報科学 41−12
（２００１．８．５）

類似度に基づく曖昧文字列照合法と音楽検索への適用

永野 秀尚　　柏野 邦夫　　村瀬 洋

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
〒 243-0198 厚木市森の里若宮 3-1
Tel: 046-240-3667　　Fax: 046-240-4708
{nagano, kunio, murase}@eye.brl.ntt.co.jp

あらまし　本稿では、音や映像のメディア探索のための曖昧文字列照合法を提案し, その類似音楽検索への適用を検討する. ここで類似音楽検索とは多重奏音楽の音響信号を検索キーとして, 長時間の音楽から検索キーに類似する部分を探索することである. この探索においては, 信号間の類似度と信号の伸縮を考慮しなければならないことと, 探索に時間がかかることが問題である. そこで, 符号間の類似度を表現する類似度行列を導入し, これに基づき符号系列化された信号間で, 伸縮を考慮した探索を, 類似度行列のスパース性により高速に行う曖昧文字列照合法を提案し, この類似音楽検索に適用した. 30 曲の類似音楽サンプル用いた予備実験では, 曖昧文字列照合法により, 従来の DP マッチングを用いたずらし照合法と比べ, 約 4 倍高速な探索が行えた.

Similarity-Based String Matching for Media-Information Retrieval and Its Application to Similar-Music Retrieval

Hidehisa Nagano　　Kunio Kashino　　Hiroshi Murase

NTT Communication Science Laboratories, NTT Corporation
3-1, Morinosato Wakamiya, Atsugi-shi, Kanagawa 243-0198 Japan
Tel: +81-46-240-3667　　Fax: +81-46-240-4708
{nagano, kunio, murase}@eye.brl.ntt.co.jp

ABSTRACT　We propose a Similarity-Based String Matching method for media information retrieval and its application to similar-music retrieval. The media information retrieval is here defined as detecting all the segments that are similar to a specified audio or video segments on a long audio or video stream. In such a task, we must consider similarities between features, deal with temporal stretching or shrinking, and also realize quick searching. Thus, the proposed method introduces a similarity matrix with a similarity enhancement technique and the DP matching method with a newly developed acceleration technique. Experiments using 30 similar-music pieces show that the proposed method can retrieve similar music fragments approximately four times faster than the conventional DP matching method, maintaining the same accuracy.

1　はじめに

音や映像のメディア情報の増加, 多様化により様々なメディア情報の探索技術が必要とされている. 我々は特に, メディア探索, すなわち, 長時間の音や映像の信号 (入力信号) と, 探したい音や映像の信号 (参照信号) が与えられたとき, 入力信号中の参照信号に類似する部分を探し出す探索技術について研究を進めてきた. そして, 今回, 高速かつ柔軟な類似探索を目的とした曖昧文字列照合法を提案し, それを類似音楽検索に適用する. ここで提案する曖昧文字列照合法は, 信号の特徴ベクトル間の類似度に基づきながら, 信号の時間伸縮に対応した探索を高速に行う手法である.

例えば, メディア探索において重要なものとして, 音楽検索がある. 音楽検索は, 音楽 CD 等と信号レベルでほぼ同一の音楽の検索をねらうもの [1] と, 信号レベルでは必ずしも一致していないが, 演奏者が異なる同一の曲や, 主旋律が同じ曲や, アレンジの異なる曲等のように, 何らかの意味で類似しているものの検索をねらうものとに大別される. 本稿では, 後

者のような類似音楽検索を対象とする．このような音楽検索に関する研究としては, MIDI 等の記号列を対象としたもの [2] や, 鼻歌, 歌声による音楽検索の研究 [3, 4, 5] 等がある．

我々は, アンサンブル音楽において, 編曲や演奏者が異なるような類似音楽の, 多重奏対多重奏の音響信号による類似音楽検索を検討している．この類似音楽検索では, 編曲の違い等による音響の変動という問題, 音符の挿入や演奏速度の違いによる信号の伸縮という問題, そして, 探索に時間がかかるという 3 つの問題がある．そこで, 音響の変動に対しては, 和音に着目した特徴抽出で対処する．また, 信号の伸縮, 探索時間という問題に対しては, 本稿で提案する曖昧文字列照合法が適していると考えられるので, 適用する．

以下, 本稿では, 曖昧文字列照合法とその類似音楽検索への適用結果を述べる．

2 曖昧文字列照合法

本章では, メディア探索とその手法について述べ, 柔軟な高速類似探索を目的としたメディア探索の手法である曖昧文字列照合法について提案, 説明する．

2.1 曖昧文字列照合法が目的とするメディア探索

曖昧文字列照合法で目的とするメディア探索は, 入力信号の特徴ベクトル系列と, 参照信号の特徴ベクトル系列が与えられたとき, 入力信号中の参照信号に類似する部分を信号の伸縮も考慮して探し出すことである．

曖昧文字列照合法では上記のような探索を高速に行うため, 各特徴ベクトルを符号化し, そして, 入力信号, 参照信号を入力符号系列, 参照符号系列と呼ぶ符号系列にする．そして, この符号系列において探索を行う．ここで, 特徴ベクトル間の類似度に基づいた探索を行うため, 特徴ベクトルの類似度に基づき, 各符号間に類似度を導入する．そして, この符号間の類似度に基づき, 時間伸縮を考慮した探索を行う．

ここでは, 特に以下のような場合を考える．入力符号系列 $\mathbf{T}$ と参照符号系列 $\mathbf{P}$ を

$$\mathbf{T} = [c_T(1), c_T(2), \ldots, c_T(n)], \quad (1)$$

$$\mathbf{P} = [c_P(1), c_P(2), \ldots, c_P(m)] \quad (2)$$

と表す．ここで, $c_T(k), c_P(k)$ は特徴ベクトルを符号化したもので, k は時系列信号において対応する特徴ベクトルが抽出された時間 (符号系列中の順番) を表す．このような $\mathbf{T}$, $\mathbf{P}$ について, 符号間の類似度に基づき, 符号系列の伸縮 (符号の脱落, 挿入) を考慮し, $\mathbf{P}$ と類似する $\mathbf{T}$ 上の全ての部分符号系列の位置を求める．ここでは, 伸縮を考慮したマッチング手法として従来から知られる DP マッチングを用い, DP マッチングの結果, $\mathbf{P}$ と類似する $\mathbf{T}$ 上の全ての部分符号系列の位置を求めることとする (図 1)．

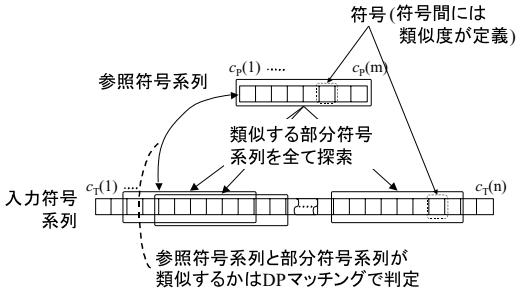

図 1: 符号化後の探索

2.2 曖昧文字列照合法のねらい

上記 2.1節で述べたメディア探索に関連する従来技術としては, テンプレートマッチングを用いたずらし照合法やその高速化手法である SSDA(sequential similarity detecting algorithm) がある．しかし, 本稿で目的とするような時間伸縮を考慮した探索を考えた場合, テンプレートマッチングは, 伸縮への対応の点で問題がある．特に音楽の類似探索では, 音符の挿入や, リズム, 演奏速度の変化による信号および符号系列の伸縮が起こる．

そのため, 符号系列間 (または特徴ベクトル系列間) の照合は DP マッチングで行うが, 全ての類似部分系列を見つけるには DP マッチングを用いたずらし照合 (図 2) を行う必要があり, 大きな探索時間がかかる．なお, 図 2では, 符号系列間での照合の場合を示したが, 特徴ベクトル間の照合の場合も同様である．この場合, 類似度は特徴ベクトル間の類似度を用いる．

また, 従来, 符号系列間の照合に関しては, 文字列照合において BM 法 [6] 等の非常に高速な照合法が知られている．しかし, 文字列照合では, 符号間を一致するかしないかのみで判定し, かつ, 参照符号系列と完全に一致する部分系列を探索する．それゆえ, 符号系列の伸縮等の変動に対応できない点や, 符号間の類似度に基づいた探索が行えないという問題点がある．

そこで, 2.3節で詳述する曖昧文字列照合法を提案する．曖昧文字列照合法は特徴ベクトルを符号化し, 符号間の類似度に基づき, 伸縮に対応した探索を行う．そして, 以下で述べるように, 符号間の類似度を表す類似度行列をスパースにすることで高速に探索する手法である．図 3に従来技術との関係を示す．

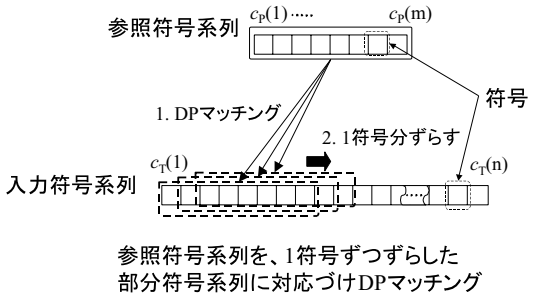

図 2: DP マッチングを用いたずらし照合法

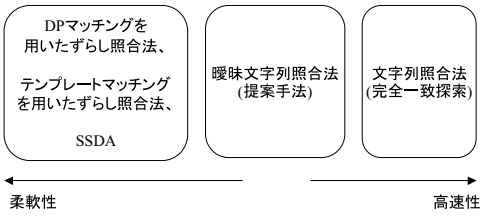

図 3: 従来技術と曖昧文字列照合法の位置づけ

2.3 曖昧文字列照合法の提案

本稿で提案する曖昧文字列照合法は

(1) 特徴ベクトルの符号化

(2) 符号間に類似度を導入し, 類似度行列を設定

(3) 類似度行列のスパース化

(4) スパース類似度行列を用いることによる, 符号系列の伸縮を考慮した高速照合

の 4 つの過程から成り立つメディア探索法である. その概要を図 4に示す.

まず, 第 1 のステップで特徴ベクトルの符号化を行う.

そして, 第 2 のステップで類似度行列の導入を行う. 類似度行列 M は (i, j) 成分が符号 i と符号 j の類似度 (0 から 1 の実数) を表している行列である. そして, 類似度行列に値を設定することで, 符号間の類似度を定義する.

次に, 第 3 のステップでは, 類似度行列において, できるだけ 0 の成分が多くなるようスパース化する. これは符号間の類似度を考慮して, 柔軟でロバストな類似探索を行うとともに, より高速な照合を行うためである.

そして, 第 4 のステップで照合を行う. ここでは符号系列の伸縮に対応するため, DP マッチングを用いたずらし照合法と同じく, DP マッチングを適用するが, 類似度行列のスパース性を用いて, より高速な照

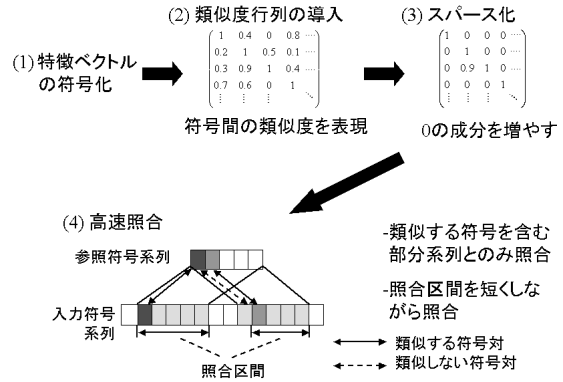

図 4: 曖昧文字列照合法の概要

合を行う. この照合では, DP マッチングを用いたずらし照合法のように, 入力符号系列中のあらゆる部分で照合を行うのではなく, 参照符号系列中で順に選んだ符号と類似する符号をもつ部分系列とのみ照合することで, 高速な探索を行う. また, 各照合において照合区間を短くしながら照合することでさらに高速化する. なお, ここで類似する符号とは類似度が 0 より大きい符号である.

なお, この第 4 のステップの照合法では, 符号間の類似度に同じ類似度を用いた場合, DP マッチングを用いたずらし照合法と同じ探索結果が得られる.

2.4 照合アルゴリズムの詳細

以下, 照合アルゴリズムの詳細について説明する. ここで m' は探索したい類似部分系列の長さ, θ は探索閾値である. そして, 本アルゴリズムは長さ m' で参照符号系列との DP マッチングによる類似度が θ 以上の入力符号系列の部分系列を全て探索する. なお, ここで, 符号間の類似度は (3) でスパース化した類似度行列の値を用いる.

照合アルゴリズム

(1) 各符号について $\mathbf{T}$ での出現位置を調べる.
(2) d = 2
(3) $j = d - i$ として, i を 1 から $\min(m, d-1)$ まで 1 ずつ増やしながら, $c_P(i)$ と類似する符号を j 番目に持つような $\mathbf{T}$ の部分系列 $\mathbf{P}' = [c_{P'}(1), c_{P'}(2), \ldots, c_{P'}(m')]$ を選び, $\mathbf{P}'$ が未照合なら, $\mathbf{P}$ と $\mathbf{P}'$ で図 5のようにして DP マッチングを行い, $\mathbf{P}'$ を照合済とする.
DP マッチングの結果, 類似度が θ 以上のときは, $\mathbf{P}'$ の場所に類似部分があったとする.
(4) d を $(1-\theta)(m+m')+1$ まで 1 増やしながら (3) を繰り返す. □□

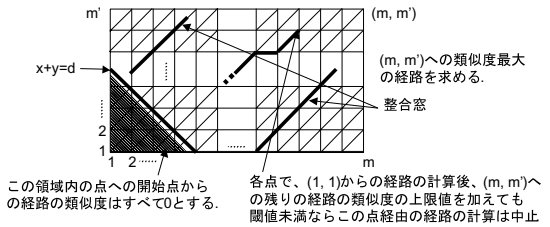

図 5: 高速な照合法

なお, 図 5で横軸と縦軸は $\mathbf{P}$ と $\mathbf{P}'$ に対応する. 図 5の照合では未照合の $\mathbf{P}'$ が選ばれるため, $c_P(x)$ と $c_{P'}(y)$ は, $x + y < d$ なら類似しないため, 斜線部では類似度の計算は行わず, 各点までの経路の類似度は 0 とし, 残りの部分でのみ類似度し, 類似度最大の経路を求める. また, 各点でその点を経由する経路による類似度の上限を求め, それが θ 以上にならない場合はその点経由の経路の計算は中止する.

2.5 従来技術との関連

曖昧文字列照合法の特殊な場合として, 次のような条件を課すことを考える.

- (条件 1) 類似度行列が単位行列
- (条件 2) 伸縮を許さない,
- (条件 3) $\theta = 1$.

(条件 1) から (条件 3) の全てが課された場合, 問題は, 従来の文字列照合に帰着する. そして, 曖昧文字列照合法で, (条件 1) と (条件 2) を満たす場合, $\mathbf{P}$ の部分符号系列 $\mathbf{P}_\theta = [c_P(1), c_P(2) \cdots c_P(\lceil \theta m \rceil)]$ を高速な通常の文字列照合で前処理として探索し, $\mathbf{P}_\theta$ が見つかった部分は類似部分符号系列とし, 照合済とするさらなる高速化法も考えられる.

3 類似音楽検索への適用

本章では, 曖昧文字列照合法の類似音楽検索への適用について述べる. ここで対象とする類似音楽検索は, アンサンブル音楽において, 編曲や演奏者が異なるような類似音楽を多重奏対多重奏の音響信号において探索する類似音楽検索である. このような類似音楽検索においては, 編曲の違いや演奏速度の違いを吸収して探索する必要がある.

そこで, 編曲の違い, 例えば, 楽器の追加や削除による変動の吸収のため, 多重奏において, 和音を特徴ベクトルとして抽出し, この特徴に応じた類似度を用いる. また, 音符の挿入, 削除や演奏速度の違いの吸収のためには, 信号の伸縮に対応した探索を行う. そこで, このような探索を高速に行うため, 曖昧文字列照合法を用い, 上記の特徴ベクトルとその類似度に基づいた符号化を行い, そして, 符号系列の伸縮に対応した探索を行う.

以下, 今回用いた特徴ベクトルについて, 音響信号からの特徴抽出過程について述べる. そして, 特徴ベクトルとその類似度について述べ, 曖昧文字列照合法における符号化について説明する.

3.1 音響特徴抽出

まず, 帯域通過フィルタを用い音響波形信号から各周波数におけるパワースペクトルを得る. 帯域通過フィルタは 75Hz から 9600Hz まで 1 オクターブごとに 48 ずつ 7 オクターブ分, 計 336 個のフィルタを周波数の対数軸上で等間隔に配置して用いた. そして, 各フィルタで, 44ms の時間区間の分析を, 11ms ごとに行い, 各時間, 各周波数におけるパワーを抽出した.

3.2 拍位置抽出

上記で得た時間 - 周波数 - パワー空間のスペクトルを用い, 拍位置を抽出する. この拍位置抽出の導入は, 上記で 11ms ごとに抽出した周波数 - パワースペクトルを各時刻で符号化すると, 符号系列長が非常に大きくなるためである. そこで, 音の変化に着目して, 拍単位で符号化することで, 符号系列長を短くする. また, 拍の長さの違い等による信号の伸縮を吸収することも期待できる.

拍位置抽出は, 文献 [7] に基づき拍位置を抽出する. ただし, 基本拍の推定は行わず, 単に, 拍位置になる確率の高いスペクトル上の点 (拍候補点) がある時刻を拍のある予測位置とし, そして, 近傍の時刻内の拍候補点について, 拍である確率を計算し, 確率が最大でかつ閾値以上の点の時刻を拍位置とした.

このようにして得られた各拍位置を用いて, 時間 - 周波数 - パワー空間のスペクトルを拍 - 周波数 - パワーのスペクトルにする. ここでは, 各拍で, 各周波数において拍内でのパワーの最大値をその周波数でのパワーとした.

3.3 倍音除去

次に, 上記で得られた各拍において, 倍音除去を行い [7], 各音 (音高, 音名) の有無を調べる. この倍音除去は, 多重奏においては複数の楽器が同時に演奏されるため, その倍音により, 音の抽出において, 本来楽器が演奏している音以外の音が抽出されることを防ぐために行う.

まず, 各拍で周波数方向のパワーのローカルピークについて, 最も周波数の近い音に割り当て, そ

の音が存在したとする． このようにして，この拍でローカルピークであった音を $a_1, a_2, \ldots, a_N$ とし，そして，$A_1, A_2, \ldots, A_J$ を基本音とする． そして，$h_{i,1}, h_{i,2}, \ldots, h_{i,\delta}$ を a_i の 1 倍音から δ 倍音とし，この拍内で存在する a_i の倍音の個数 c_i を

$$c_i = \sum_{k=i}^{N} c_i(k) \tag{3}$$

とする．ここで，

$$c_i(k) = \begin{cases} 1 & \text{ある } l \text{ で } a_k = h_{i,l} \\ 0 & \text{上記以外} \end{cases} \tag{4}$$

である．そして，$c_i \geq \varepsilon$ のとき，a_i の対応する音 A_j が存在したとする． ここで ε は倍音除去の閾値である． すわわち，a_i の δ 倍音のうち ε 個以上が存在すれば，この拍内で a_i の対応する音が存在したとする．

3.4 特徴ベクトルと類似度

上記で得られた基本音の有無に着目し，各拍の特徴ベクトルを 12 ビットの特徴ベクトルを構成する．この特徴ベクトルは，多重奏における和音に着目した特徴である．

特徴ベクトルの各ビットは下位から順に A，A#，B，... ，G# の各音に対応しており，この拍内で対応する音が存在したとき 1 にそうでないとき 0 とする． 例えば，A(220Hz)，E(330Hz)，A(440Hz)，C(523Hz)，A(880Hz) が拍内に存在した場合，この拍の特徴ベクトルは 2 進表現で "000010001001" とする．

そして，多重奏音楽における類似探索のため，特徴ベクトル間の類似度は，特徴ベクトルを a と b としたとき，

$$\frac{w(a \otimes b)}{12} \tag{5}$$

とする．ここで $a \otimes b$ は a と b のビットごとの排他的論理和で，$w(a)$ は a のハミング重みである．

3.5 符号化と符号間類似度

今回，曖昧文字列照合法の特徴ベクトルの符号では，上記の 4096 種類の特徴ベクトルをそのまま符号とした．これは上記特徴が 12 ビットでコンパクトに符号化されていることと，この参照符号系列 (入力符号系列) で符号をシフトすることや，ビット位置を入れ換えて探索することで移調，変調された音楽の検索も可能となるからである．

そして，この符号化では，符号間類似度も，特徴ベクトルと同様に式 (5) により定義した．

4 実験

3章で述べたようにして，音楽を符号系列化し，類似音楽検索を，(a) DP マッチングを用いたずらし照合法，(b) 曖昧文字列照合法と同じスパース類似度行列による符号間類似度を用いての DP マッチングを用いたずらし照合法，(c) 曖昧文字列照合法，により行った． なお，(a) では 3.5節で述べた符号間類似度を用いた．そして，各照合法で，探索時間，探索精度を評価した． これは，類似度行列のスパース化による探索精度の変化と，曖昧文字列照合法による高速化の効果を確認するためである．なお，(a) と (b) においては探索時間は同じであり，(b) と (c) では同じ探索結果が得られるため，探索精度が同じになる．

実験には多重奏のアンサンブル音楽を録音したものを用いた．楽曲は 5 種類の曲を用い，各曲について 2 種類の編曲を用意した．そして，各編曲毎に，奏者 A による演奏を 2 回，他奏者 B の演奏を 1 回，計 3 回の演奏を録音した．このようにして，各曲ごとに 6 個の演奏を録音し，5 曲で合計 30 個の録音をまとめた計約 30 分の音響信号 (入力信号) から探索することとした．検索キー (参照符号系列) には各編曲毎に奏者 A の一方の演奏から約 15 秒の断片を 3 つ選び，計 30 の検索キーを選んだ．このとき，拍に基づく符号化をした結果，入力符号系列は長さ 5656，参照符号系列は平均 52 の長さの符号系列となった．

各断片の探索では，移調を考慮して符号を 1 ビットずつシフトして得られる 12 種類の参照符号系列で系列長 m の $\pm 6\%$ 内の全ての m'(平均 4.3 種類) について探索し，この全ての探索にかかる時間を探索時間とした．このとき，探索における整合窓については，経路上の点 (x, y) は $|x - y| < 0.06m$ を満たすようにした．

探索精度については，探索結果中の正しい探索結果 (正答) の割合と，探索すべきもの (要検出区間) のうち探索結果に含まれた割合が等しい探索閾値での両割合の平均とした．なお，各検索キーについて，検索キーと同じ曲の該当区間を要検出区間とすることとした．すなわち，各検索キーについて，検索キー自体を含む入力信号の 6 箇所を要検出区間とした． これは，単に曲名を検索結果とするよりも厳しい設定である．

ある探索結果が正答と判断する基準については，その探索結果の区間がある要検出区間と重複率

$$\text{重複率} = \frac{\text{要検出区間} \cap \text{探索結果の区間}}{\text{要検出区間} \cup \text{探索結果の区間}} \tag{6}$$

が 0.5 以上のとき，この探索結果を正答とするとした [5]． また，ある要検出区間が探索されたかの判断もある探索結果との重複率が 0.5 以上の時，探索された

表 1: 探索結果の比較

	(a) DP マッチングを用いたずらし照合法	(b) DP マッチングを用いたずらし照合法 ((c) と同じ符号間類似度)	(c) 曖昧文字列照合法 (提案手法)
探索精度	73%	73%	73%
探索速度 (秒)	8.7	8.7	2.1

(探索結果に含まれた) とする.

実験で用いたアンサンブル音楽では 1 曲のうち, 繰り返しにより, 要検出区間以外の箇所にも検索キーに類似する部分が含まれることがある. 実験ではこのような箇所を探索結果とした場合は不正答としてある. このような探索結果も正答とした場合は, 探索精度の値はより大きくなると考えられる.

表 4に実験結果を示す. なお, 探索時間の計測には PC(Linux, Pentium4 1.5GHz) を用いた. このとき, 類似度行列のスパース化は 0.85 以上の成分を 1 に, 他は 0 にすることで行った.

表 4から, 類似度行列のスパース化による探索精度の低下はないことがわかる. そして, スパース化した類似度行列を用いた曖昧文字列照合法 (提案手法) では約 4 倍の高速化が達成できていることがわかる.

5 まとめ

本稿では, 入力信号と参照信号の特徴ベクトル系列が与えられたとき, 各特徴ベクトル系列を符号化し, その符号間の類似度行列のスパース性を利用し, 類似度に基づきながら伸縮を考慮した探索を高速に行う曖昧文字列照合法を提案した. そして, 本提案手法を類似音楽検索に適用した. そして, 実験では曖昧文字列照合法を用いることで, DP マッチングを用いたずらし照合法に比べ, 約 4 倍高速に探索することができた.

今回提案した曖昧文字列照合法では, DP マッチングを用いたずらし照合法と同じ探索を行う場合について検討, 評価したが, 今後, より大規模なデータベースに対して検証を行うとともに, 連続 DP(端点フリー DP) を用いる場合についても, 検討, 評価を行う. また, 類似度行列の構成法の工夫による, 類似音楽検索における高精度化, 高速化を検討する.

謝辞 日頃御指導を頂く NTT コミュニケーション科学基礎研究所の石井健一郎所長, 管村昇部長, 萩田紀博部長に感謝する. また日頃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.

参考文献

[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 –. 電子情報通信学会論文誌 D-II, Vol. J82-D-II, No. 9, pp. 1365–1373, September 1999.

[2] 西原祐一, 小杉尚子, 紺谷精一, 山室雅司. 時間正規化を用いたハミング検索システム. 情報処理学会研究報告, 99-MUS-30, pp. 27–32, May 1999.

[3] 蔭山哲也, 高島洋典. ハミング歌唱を手掛かりとするメロディ検索. 電子情報通信学会論文誌 D-II, Vol. J77-D-II, No. 8, pp. 1543–1551, August 1994.

[4] 園田智也, 後藤真孝, 村岡洋一. WWW 上での歌声による曲検索システム. 電子情報通信学会論文誌 D-II, Vol. J82-D-II, No. 4, pp. 721–731, April 1999.

[5] 橋口博樹, 西村拓一, 赤坂貴志, 岡隆一. 鼻歌の旋律と歌詞をクエリーとする楽曲信号のスポッティング検索. In *Technical Report of IEICE*, PRMU2000-118, pp. 79–86, November 2000.

[6] 野下浩平, 高岡忠雄, 町田元. 基本的算法. 岩波講座 情報科学 10. 岩波書店, 1983.

[7] 柏野邦夫. 音楽音響信号を対象とする聴覚的情景分析に関する研究. PhD thesis, 東京大学, 1995.