2005－SLP－57（23）
2005／7／16

# 歌波形データのフレームワイズな音素識別に基づく検索

矢口勇一[†]　岡　隆一[†]

本稿の目的は,「フレームワイズな音素識別ラベルを用いた音声検索」の方式を歌声データ検索に適用し，その有用性を調べることである．すなわち，フレームワイズに音素ラベル化された歌声データを検索対照データとし，クエリーとしての歌声データを検索対象データと同様に処理した後，クエリー列を用いてデータベースから連続 DP によってスポッティング認識することで類似した部分区間の検索を行い，抽出された区間を含む楽曲を検索するものである．楽曲データベース全体から歌声クエリーを用いた楽曲検索率は，4 小節分のクエリーで 78%であった．また、音声クエリーとの検索率の違いも本稿では述べる．

# Song Wave Retrieval Based on Frame-wize Phoneme Recognition

Yuuichi Yaguchi[†] and Ryuichi Oka[†]

A song wave retrieval method is proposed. Both song wave data and a query song wave data are transformed into phoneme sequences by frame-wize labeling of each frame feature. Applying a spotting algorithm called Continuous Dynamic Programming to these phoneme sequences, we can detect a set of similar parts in the song database each of which is similar to a query song wave. Song retrieval rate hits 78% in 4 clauses from whole databases. Additionaly, difference of each query from song wave data and speech wave data remarked in this paper.

## 1. まえがき

近年，音声認識・検索・合成技術は発展を遂げている[1]．ニュース番組の音声認識，カーナビゲーションシステムにおけるコマンド音声認識など，リアルタイム認識の実用化も試みられている[2)3)]．また，女声・男声区別無く自然発話ならできることから，コールセンターでの音声入力や音声翻訳も制約付ではあるが実用化も視野に入ってきた．さらに，音声認識とはシステム概念を異にする音声データベースを検索する技術も実用化されてきている[4]．

しかし，人間が発するもうひとつの音声である歌声に関する検索については研究の試みは音声認識と比べると極めて少ない．歌声を含む音楽データの検索には，これまでメロディーを抽出し、そのメロディー区間と類似する音楽データを検出する試みがある[5]．しかし，歌声音声を対象として，その音韻特徴に注目して歌声音声データを検索する試みはこれまでほとんど行われていない．この理由は，歌声音声の音声認識が，従来の音声認識手法の適用対象とはみなされていないことにある．

歌声には，自然発話の音声と様々に異なるピッチの変動があり，それはメロディーとして捉えられる一方，これら音韻特徴へ影響も与え，さらに歌声での母音の持続長が比較的長いということも相まって，通常の音声認識とは異なる認識アルゴリズムの開発を必要としている．ここでは，歌声音声の単語や文などのカテゴリを認識するという問題を扱うのではなく，歌声音声の音韻特徴に基づいた検索問題を扱うとする．すなわち，音韻特徴に基づくものであるので，メロディ情報は利用されず，歌声のもつ音韻情報に基づく検索問題となっている．

一般に，検索のための音声クエリーが認識され，また検索対象の音声データベースも認識されているという状況では，音声検索問題は単なる記号列の検索問題となる．しかし，先に述べたように歌声音声の単語や文カテゴリを認識する技術がまったく開発されていない状況では別の方法をとらねばならない．そこで本稿では，語彙に依存しない音素の識別による自然発話音声の検索手法[6] を歌声に適用し，その有効性を検証することとする．この方法論をとる理由は，単語や文などの語彙に依存しない音素という音韻情報を利用できるという文献[6] の検索方法が一種の汎用性を持つ音韻

† 会津大学
The University of Aizu

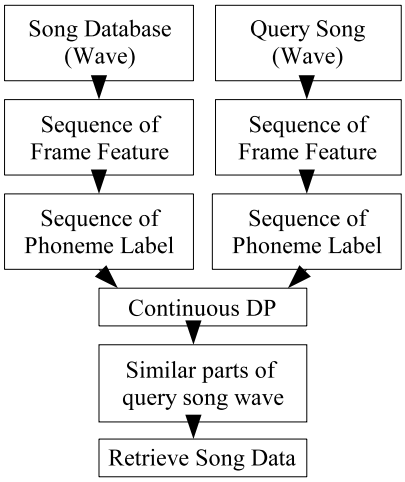


**図 1** 歌声検索システムの構成

コーディングであると考えられ，この汎用性が歌声音声の音韻情報に基づく検索にも適用が可能ではないかという仮説に基づいている．

本論文ではまず，**2.** で語彙に依存しない歌声検索システムの構成とその手法について述べ，**3.** で本研究で使用する音素識別関数とフレーム毎のラベル作成について述べる．さらに，**4.** では連続 DP[7] による歌声検索アルゴリズム[8] を示し，**5.** で入力クエリー長の比較による実験結果を示して **6.** でまとめる．

## 2. 語彙に依存しない歌声検索システムの構成

情報検索の分野では，きわめて大規模な，切れ目の無いデータベースを想定してシステムを作成するが，その大規模なデータベースでいかに Segmentation と Identification を行うかが議論される．Discrete Hidden Markov Model を用いたニュースデータ検索などの，HMM を使った手法[9] などが近年では開発されているが，特に無音区間を含むクエリ列に対して大規模データベースからのスポッティング認識では機能しないことが知られている．

この問題に対して，筆者の一人は，先にデータベース・クエリー双方の音声波形をそれぞれ各フレームごとに音素記号でラベリングし，双方の音素記号列の間で連続 DP マッチングに基づく音声検索を行う方式を提案している[6]．これは，データベースの音声波形とクエリーの音声波形を一度フレームワイズな音素識別ラベルを生成するフィルタに通し，コーディングすることと同義である．コーディングされたデータは，語彙や文などによらない，単位時間毎に作られた記号列のマッチングとして処理されることから，完全に語彙に依存していないことになる．このことは **1.** でも触れた通り，歌の中に含まれるリズムによって認識することの難しい文節の区切りを無視して，各フレーム毎での連続 DP マッチングを行うことができるので，歌声検索を実現するのには非常に都合のいいものである．

| | | | | | | |
|---|---|---|---|---|---|---|
| /a/, | /i/, | /u/, | /e/, | /o/, | /y/, | /w/, |
| /r/, | /n/, | /m/, | /nn/, | /g/, | /j/, | /z/, |
| /f/, | /h/, | /s/, | /sh/, | /b/, | /d/, | /k/, |
| /p/, | /t/, | /ts/, | /ch/, | /sp/ | | |

**表 1** The phoneme label list

このフレームワイズな音素識別ラベルを生成するために，まず，音声のフレーム毎の特徴ベクトルを作成する．ここで作成される特徴ベクトルは音声の方向性パターンと呼ばれているもので，音声認識においてはケプストラムより優れた性能を示すとされ用いられている[10]．まず音声波形より 8ms をフレーム間隔とする 20 チャネルのスペクトルから作られるスペクトル場を作成し，次にこのスペクトル場から上下左右の 4 方向性パターンを 18 次元について抽出し，$(4 \times 18 = 72)$，それを前後 7 フレームについて時間軸方向のみに平滑化した後，その中の 3 フレーム分 $(\{[t-2], [t], [t+2]\})$ の特徴を用いて時刻 t における特徴ベクトルと定めている．この 7 フレーム間の平滑化は，話者毎のイントネーションの幅を吸収する働きを持つので，歌声におけるピッチの差によるエラーを緩和することができると考える．この特徴ベクトルからベイズ識別法を用いたフレームの音素記号を表現することによって，リズム・ピッチ等の歌声と音声における違いを極力抑えることができる．

## 3. フレーム間音素ラベリング

### 3.1 ベイズ識別関数によるフレーム音素認識

ここで用いる識別関数は，音素の各フレームについて音素ラベルが付けられている音声データがあるときに，同一の音素ラベルを持つフレームの集合を用いて以下のベイズ識別関数 $((g^l(x))$ を作成するということである．

$$g^l(x) = \sum_{i=1}^{k} \frac{\{\phi_{l,i}^T (x - \mu_l)\}^2}{\lambda_{l,i}} + \ln \prod_{i=1}^{k} \lambda_{l,i} - 2\ln p(\omega_l) \tag{1}$$

ここで，

$x$：識別したい入力の特徴ベクトル

$\mu_l$：音素 $l$ の学習済み特徴ベクトルの平均

$\lambda_{l,i}$：音素 $l$ の学習済み特徴ベクトルのサンプルによる $i$ 番目の固有値

$\phi_{l,i}$：音素 $l$ の $i$ 番目の固有ベクトル

$k$：用いる固有ベクトルの総数

| | System Phoneme | ATR Phoneme | Song DB Phoneme |
|---|---|---|---|
| Vowel | a | a | a |
| | i | i | i |
| | u | u | u |
| | e | e | e |
| | o | o | o |
| Demi-Vowel | w | w | w |
| | y | y | y |
| | r | r | r |
| | ly | y | ya/yu/yo |
| Fricative | s | s | s |
| | h | h | h |
| | sh | sh | sh |
| | f | f | f |
| | z | z | z |
| | j | j | j |
| Plosive | ts | ts | ts |
| | ch | ch | ch |
| | b | b | b |
| | d | d | d |
| | g | g | g |
| | v | - | - |
| | p | p | p |
| | t | t | t |
| | k | k | k |
| nasal | n | n | n |
| | m | m | m |
| | nn | nn | nn |
| | N | sp | q |
| youon | by | b,y | b,ya/yu/yo |
| | gy | g,y | g,ya/yu/yo |
| | hy | h,y | h,ya/yu/yo |
| | ky | k,y | k,ya/yu/yo |
| | my | m,y | m,ya/yu/yo |
| | ny | n,y | n,ya/yu/yo |
| | py | p,y | p,ya/yu/yo |
| | ry | r,y | r,ya/yu/yo |
| | dy | d,y | d,ya/yu/yo |
| sokuon | pp | sp,p | q,p |
| | tt | sp,t | q,t |
| | kk | sp,k | q,k |
| | cch | sp,ch | q,ch |
| | ssh | sp,sh | q,sh |
| | ss | sp,s | q,s |
| | dd | sp,d | q,d |
| | tts | sp,ts | q,ts |
| | ff | sp,f | q,f |
| yousokuon | kky | sp,k,y | q,k,ya/yu/yo |
| | ppy | sp,p,y | q,p,ya/yu/yo |
| blank | pau | sp | sil |
| other | he | e | e |
| | wo | o | o |

**表 2** 音素対応表

$p(\omega_l)$：音素 $l$ の事前出現確率

とする．これらの識別関数を音素の種類だけ作成（26 個）し，各フレームの特徴に適用して，最大の値を与える識別関数を持つ音素ラベルが，そのフレームの持つ音素ラベルとして決定される．

### 3.2 データベース・クエリ列のフレーム間音素ラベリング

**2.** で説明したフレームの特徴ベクトルによる表現から，**3.** で説明したベイズ識別関数を用いて，各フレームを音素記号によってラベリングしていく．今回の実験で扱う音素ラベルは，ATR SPEECH DATABASE[11] の 503 文・話者 10 名分の計 5030 文から作成したもので，表 1 に示される 26 種類とする．これは，ATR SPEECH DATABASE で使用されている約 50 種の音素ラベル群[11] と，歌声データベースを作る際に使用した 29 種の音素ラベル群から，共通する音素 26 種を選択したものである．なお，音素の対応は表 2 に挙げる．

| | Level 1 | Level 2 | Level 3 |
|---|---|---|---|
| Song DB | 45.71 | 57.44 | 63.85 |

**表 3** オープン話者全体のフレーム単位の音素ラベル変換精度（3 位まで考慮）． Song DB 78 曲のデータの評価.

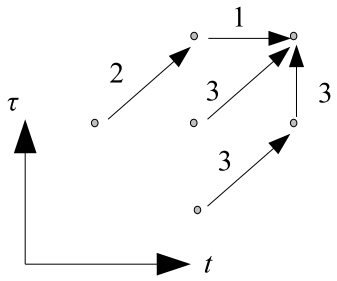


**図 2** 連続 DP における局所パスとそれにつく非対称の重み

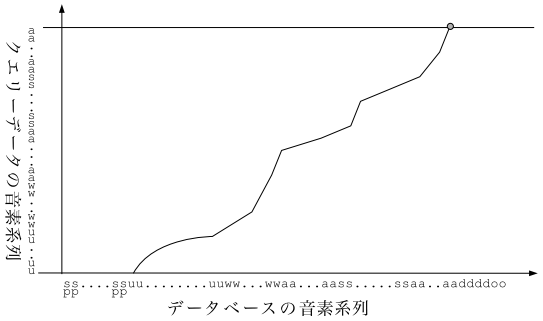


**図 3** 連続 DP におけるマッチングの経路

音声のデータベースから作成した識別関数を用いたことから，歌声データベースに対して人的に付けた音素ラベルと比較するとかなりのエラーがある（表 3）が，ここでは，音素ラベル化に関しては特徴量を抽出して別のデータ列に置き換えるというフィルタの役割があると考え，データベースのデータも，クエリのデータも同様にエラーが発生するコーディングを施している（言い換えれば、音声と歌声の発音は多少なり違う可能性がある）[12] として，実験を進めていく．

## 4. 連続ＤＰによる歌声検索アルゴリズム

### 4.1 歌声波形のフレーム表現に対する連続 DP の適用

クエリーによるデータベースの検索アルゴリズムと

して連続 DP[7] を用いる．これは，**3.** までの処理によって，歌声波形がラベルの系列に変換されたことで，クエリーとデータベースとの間で非線形の連続 DP を使ったスポッティング認識が可能となったからである．まず音声のフレーム表現として，各フレームを音素記号でラベリングして，その 3 位までの候補を用いて，それを

$$F(t) = (f_1(t), f_2(t), f_3(t)) \tag{2}$$

と表現する．ここで，$f_{1/2/3}(t)$ は，$t$ フレームにおける第 1/2/3 候補の音素ラベルとする．式 (2) で第 3 位まで扱うのは，表 3 で示されているように，歌声データベース内における音素ラベルの変換精度が第 3 位までの累積をとると 63.85%の累積的な変換精度の結果を使えることが出来るためである．

さて，連続 DP では，入力の時系列パターンと参照の時系列パターンの要素との間に局所距離というものを定義するが，入力系列は各時刻で上位 3 位までの音素ラベル候補が考えられるため，その系列を

$$F(t), t = 1, 2, 3, ... \tag{3}$$

として，参照パターンを

$$G(\tau) = (g_1(\tau), g_2(\tau), g_3(\tau)) \tag{4}$$

$$(1 \leqq \tau \leqq T) \tag{5}$$

とする．

連続 DP のアルゴリズムでは，参照パターンに対して，入力パターンは無限に入力可能である[8] ことから，参照パターンにクエリーの音素列を，入力パターンにデータベースの音素列を扱うことにする (図 3)．本稿での実験では，参照パターンについて，各フレームで上位 3 位までの音素ラベルを候補とし，入力パターンの 1 位の音素ラベルを用いて連続 DP マッチングを行う．その時の局所距離を，

$$d(t,\tau) = \begin{cases} 0.0 & (if) f_1(t) = g_1(\tau) \\ 0.1 & (if) f_1(t) = g_2(\tau) \\ 0.2 & (if) f_1(t) = g_3(\tau) \\ 1.0 & (otherwise) \end{cases} \tag{6}$$

と定義する．このことから，局所距離を用いる連続 DP の漸化式は，次のように表すことが出来る．**Initial Condition:**

$$P(-1,\tau) = P(0,\tau) = \infty \tag{7}$$

**Iteration** $(t = 1, 2, ...)$**:**

For $\tau = 1$

$$P(t,1) = 3\dot{d}(t,1) \tag{8}$$

For $\tau = 2$

$$P(t,2) = \min \begin{cases} P(t-2,1) + 2 \cdot d(t-1,2) \\ \qquad\qquad +d(t,2) \\ P(t-1,1) + 3 \cdot d(t,2) \\ P(t,1) + 3 \cdot d(t,2) \end{cases} \tag{9}$$

For $\tau = 3$

$$P(t,\tau) = \min \begin{cases} P(t-2,\tau-1) + 2 \cdot d(t-1,\tau) \\ \qquad\qquad +d(t,\tau) \\ P(t-1,\tau-1) + 3 \cdot d(t,\tau) \\ P(t-1,\tau-2) + 3 \cdot (t,\tau-1) \\ \qquad\qquad +3 \cdot D(t-\tau) \end{cases} \tag{10}$$

となり，出力は，

$$A(t) = \frac{1}{3 \cdot T} P(t,T) \tag{11}$$

となる．連続 DP で用いられる局所パスとその重みは図 2 で示されるものとし，図 2 で示される重みは非対称となっている．そのため，累積距離 $P(t,T)$ を計算する最適パスがどのようなものであっても，重みの累積は $3T$ と一定になる．結果として $(p,T)$ の値が標準パターンの長さに依存しないようにする正規化演算が式 (11) に示すように簡単になる．

### 4.2 検索区間の抽出

この連続 DP を用いた語彙に依存しない歌声検索は，歌声データベースも歌声クエリーも単語や文などの記号列へと変換しない．つまり，音声認識の場合のように入力音声についての認識の結果を単語や文の記号表現で表現するものではなく，歌声のクエリーを入力として，それと類似するデータベース中の歌声区間を出力とする．

この場合では, 連続 DP の適用によって上述のように各時刻 $t$ においての連続 DP 値 $A(t)$ が出力される (図 4). この時, その出力値を与える最適パスが決定されるので，最適パスの入力時間軸上の始点も定まることになる．なお，始点の決定にはバックトレースによる方法とフィードフォワードによる方法とがあるが，今回は終点を固定とするフォードフォワード法を用いて始点を決定するようにする．

今， $A(t) \leq \alpha$ を与える入力時間軸上の始点を $S(t)$ とすると，各時刻 $t$ における入力時間上の区間 $N$ を，

$$N(t : A, \alpha) \stackrel{\text{def}}{=} [S(t), t] \tag{12}$$

として定めることが出来る．また，この時 $N(t : A, \alpha)$ と $N(t' : A, \alpha)$ が共有区間を持つ場合，$A(t)$ と $A(t')$ を比較して小さい方に対応する N を選択するようにする．これにより，検索される区間の間では必ず排他的な時間区間を持つようになる．また，この検索は $\alpha$

【歌詞 1】女声
…
夜中の 3 時　目がさめて
携帯にもメールが来ない
朝までずっと　待っている私
結局は睡眠不足
あなたに逢うと文句が言えないまるで飛べない鳥だね
…
【歌詞 2】男声
…
線路は続くよどこまでも　夢と希望のせどこまでも
誰もがみんな手にしてる　切符を握りしめ
汽笛は響くよどこまでも　僕らの心でいつまでも
明日と言う名の未来へと　全力　走ってく
…

**表 4**　歌データベースの例（歌詞表示）. 下線の部分が音声クエリーに含まれているスクリプト部分を示す.

| | 1st clause | /2nd clause | /3rd & 4th clauses |
|---|---|---|---|
| 1. | 夜中の 3 時 | /目が覚めて | /携帯にはメールが来ない |
| 2. | 今でも | /君が好きだと | /言っていいのかな |
| 3. | ちょっと待って | /不思議な | /ときめきあふれ出していく |
| 4. | まだ知らない明日へ | /踏み出して行こう | /勇気を出して さぁ |
| 5. | だから君を | /離さない | /ずっとこの僕見ててくれ |
| 6. | 君の | /心 | /ゆらゆら |
| 7. | 線路は続くよ | /どこまでも | /夢と希望のせどこまでも |
| 8. | 遠く | /こだまのように | /心に響く |
| 9. | やっぱり伝えて | /みたいよ | /わたしの言葉で |
| 10. | きっと新しい | /気持ちを | /連れてくるから |

**表 5**　実験で用いた音声クエリーのデータ（スクリプト表示）

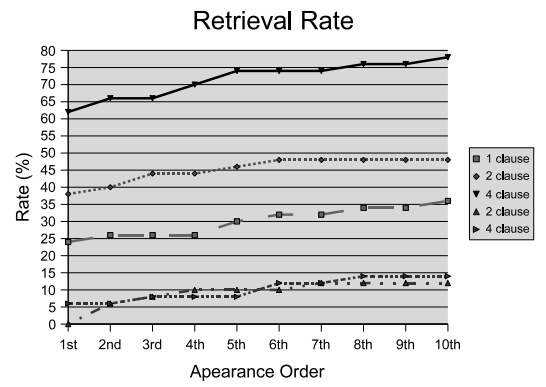


**図 5**　歌声検索実験の結果

の値を大きくすると，一般にはデータベース中で検索される区間の占める割合が大きくるため，検索漏れが少なくなる反面，検索されるべきでない部分も多くなる．この $\alpha$ の適正な値や，検索条件に関する問題は今後の課題とする．また，この検索出力はクエリーと類似しているか否かのみが評価されるので，歌声の音素ラベルへの変換精度の評価とは異なるものとなる．

## 5. 歌声検索の実験

ここでは、1 小節分の歌声，2 小節分の歌声，4 小節分の歌声，2 章節分の音声，4 小節分の音声の，計 5 種類のクエリーを用いて検索の比較実験を行う．データベースは，RWC 研究用音楽データベース：ポピュラー音楽著作権切れ音楽の歌のみのもので，日本語・ソロ歌手のもの計 78 曲を使用する．クエリーでの歌声は，被験者は男性 4 名，女性 1 名で，表 5 に挙げたフレーズをできるだけ音程・リズム共に正確にしようとしたものを使用する．クエリーでの音声は、ゆっくり目に発音した通常音声を使用する．例示として，データベースのサンプル歌詞の抜き出しを表 4 に，クエリーのサンプル歌詞を表 5 にそれぞれ挙げる．なお、データベース中の楽曲の平均長は 3 分 55 秒，全長は 5 時間 5 分 32 秒である．

評価方法として，第 10 位までの各順位までに，各種のクエリーを用いてデータベースから検索した場合に，目的とするスポッティング区間が 50 %以上取り出せている場合を検索されたこととして検索された率を表 5 に示す.

図 4 は，表 5-4. のスクリプトを歌声クエリーとして入力した場合に得られた全データベースからの局所距離のグラフである．

検索実験を行った結果，歌声データベースから 10 位までを検索されたとして，1 小節分の歌声クエリーでは 36%，2 小節分の歌声クエリーでは 48%，4 小節分の歌声クエリーでは 78%となった (表 5)．検索できなかった楽曲の代表として，表 5-10. が挙げられるが，これにはヴォーカルラインにエフェクトとしてエコーディレイがかけられていたことと，女声特有の伸びのある声が逆に男性にとって似たように発声し難かった可能性がある．表 5-2. も同様にエフェクトとしてエコーがかかっていたが，特有の発声によって検索が難しかったと考えられる．なお、セレステを狙った細かいディレイのかかった表 5-9. や軽いエコーのかかった表 5-8. など，エフェクトがかかったものは他にもあったものの、2. と 10. 以外は類似区間の検索に成功している．また、2 小節分，4 小節分の音声データではほとんど検索できなかった．これは、リズムとの差が DP の範囲外であったことと，歌声と音声では音のバリエーションが違う可能性があることの 2 つの側面があると考えられることから，その対応についても今後の議論としたい.

## 6. まとめと考察

本論文では，語彙に依存しない音声検索の方式を歌声に適用して，歌声データベースから歌声・音声のクエリーを用いた検索の場合の比較実験を示した．この比較実験では，ほぼ第 1 位までに検出されないもの

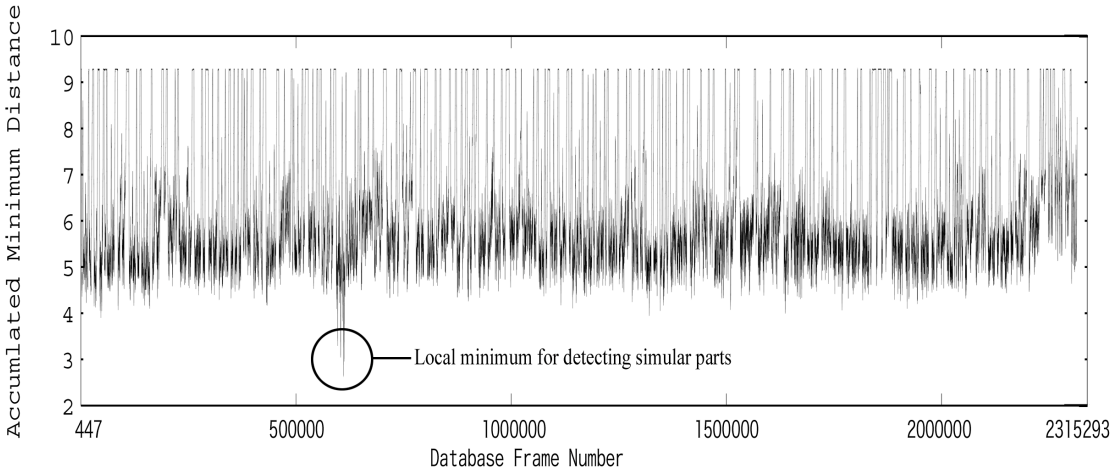


**図 4** CDP による歌声検索の実験：表 5-4. のスクリプトで示される歌声を入力し検索すると，入力と整合する複数箇所で局所距離の累積値がが他と比べて小さくなる．

は，第 5 位以降にはほぼ検出されていないことがわかった．また，歌声では扱うクエリの長さによって検索率は上昇したが，逆に音声検索からは検索率は下降している．これは，音声と歌声のリズムの差が蓄積されてしまったために，DP の範囲では対応できていないことがわかる．歌声クエリーだけに限ってみると，明らかにエコーディレイの強くかかった楽曲に対する検索率が悪かった。これは，エコーディレイがエラーとして認識され、音素列に変換されているからだが，歌声の中の多様性として捉える必要があることから，音素記号列への変換の精度を上げることで対処をしたいと考える．

今後の展望としては，現在使用している DP マッチングでのラベル間の距離はものすごく曖昧であるので，ラベル間の距離をあらかじめ用意しておいて，パスの重みを使ったより精度の高い DP マッチングを使うことで検索率を上げたいと考えている．また、加工される際に加えられるヴォーカルに対するエフェクトがどのくらい検索率に影響を与えているかを調べると共に，BGM をノイズとして捉えた，CD 音源の検索を視野に入れたノイズ除去やノイズフリーな手法について検討したいと考えている．

**謝辞** 本研究を進める上でご支援頂いた，張　建新 ((株) メディアドライブ) 氏，伊原正典 ((株) シャープ) 氏，に深謝いたします．

## 参 考 文 献

1) 特許庁総務部技術調査課, “音声認識技術に関する特許出願技術動向調査報告”, 特許庁, http://www.jpo.go.jp/shiryou/pdf/gidou-houkoku/voice_recognition.pdf, May 2003.
2) 安藤彰男, “ リアルタイム音声認識 ”, 電子情報通信学会, September 2003.
3) 中田和男, “ 日本音響学会編　音響工学講座 7　改訂　音声 ”, コロナ社 , February 1995.
4) (株) メディアドライブ, “CrossMediator”, http://adv.mediadrive.jp/product/.
5) 橋口博樹, 西村拓一, 張　建新, 滝田順子, 岡　隆一, “ モデル依存傾斜制限型の連続 DP を用いた鼻歌入力による楽曲信号のスポッティング検索 ”, 信学論 (D-II), vol.J84-D-II, no.12, pp.2479-2488, December 2001.
6) 岡　隆一, 西村拓一, 張　建新, 伊原正典, “ フレーム特徴の音素記号化に基づく語彙に依存しない音声検索 ”, 信学論 (D-II), vol.J86-D-II, no.6, pp.764-775, June 2003.
7) 古井貞熙, “ ディジタル音声処理 ”, 東海大学出版会, September 1985.
8) Ryuichi Oka, “Spotting Method for Classification of Real World Data”, The Computer Journal, Vol.41, No.8, pp.559-565, 1998.
9) H. Wang, “ Experiments in syllable-based retrieval of broadcast news speech in Mandrin Chinese ”, Speech Communication, vol.32, nos.1-2, pp.49-60, September 2000.
10) 松村　博, 岡　隆一, 小暮一也, 小島有里江, “ スペクトルベクトル場の方向性パターンを用いた不特定話者の単語音声認識 ”, 信学論 (D-II), vol.J72-D-II, no.4, pp487-498, April 1989.
11) ATR, ATR SPEECH DATABASE, Phonetically balanced 503 sentences, 1992.
12) 矢口勇一, 岡　隆一, “歌波形データのフレームワイズな音素識別に基づく検索”, 信学技報, SP2004-50, pp.19-24, August 2004.
13) 後藤真孝, 橋口博樹, 西村拓一, 岡　隆一, “RWC 研究用音楽データベース：ポピュラー音楽データベースと著作権切れ音楽データベース”, 情報処理学会 音楽情報科学研究会 研究報告, 2003-MUS-42-6, Vol.2001, No.103, pp.35-42, October 2001.