社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－MUS－76 (10)
2008／8／7

Hyperlinking Lyrics: 複数の楽曲の歌詞中に共通して登場するフレーズ間へのリンク作成手法

藤原弘将 後藤真孝 緒方淳

産業技術総合研究所

本稿では，複数の楽曲で共通に出現する歌詞のフレーズ間に，ハイパーリンクを作成する手法について述べる．歌詞が既知の楽曲と歌詞が未知の楽曲を両方含む楽曲データベースに対して，提案法は，歌詞が既知の楽曲同士のハイパーリンクと歌詞が既知の楽曲と未知の楽曲の間のハイパーリンクの2種類を考える．前者は，既知の歌詞テキストからキーワードを抽出し，各キーワードが音響信号中に登場する時刻を推定することで，実現する．後者は，キーワードスポッティング技術により，歌詞が未知の楽曲中にキーワードが出現するかどうかを検出することで実現する．提案法の有効性を確かめるために評価実験を行った結果，実験結果には向上の余地が未だ残されているものの，このアプローチの可能性が確認できた．

Hyperlinking Lyrics: A method for creating hyperlinks between phrases common in multiple song lyrics

HIROMASA FUJIHARA, MASATAKA GOTO and JUN OGATA

National Institute of Advanced Industrial Science and Technology (AIST)

We describe a novel method for creating a hyperlink from a phrase in the lyrics of a song to the same phrase in the lyrics of another song. Given a song database consisting of songs with their text lyrics and songs without their text lyrics, our method creates two different types of hyperlinks: a hyperlink between songs with text lyrics and a hyperlink from a song with text lyrics to a song without them. The former hyperlink can be created by extracting potential keywords from all the text lyrics and finding sections temporally locating them in audio signals. For the latter hyperlink, our method looks for sections including voices that sing the keywords by using a keyword spotting technique. Our experiments show the potential of this new approach, although the performance obtained has room for improvement.

1. はじめに

本研究の最終的な目的は，楽曲が歌詞の内容に基づいて相互にリンクされたMusic Web（**図1**）を実現することである．Web上ではハイパーテキスト同士が互いにハイパーリンクされているように，Music Web上では歌詞中のフレーズ同士が互いにハイパーリンクされている．楽曲同士を関連づけるやり方は歌詞以外にもいくつも考えることができるが，本研究では，その中でも楽曲の表現する内容を如実に表す重要な要素として歌詞に焦点を当てた．このように，楽曲の集合をMusic Web上のハイパーリンク構造として表現することで，様々な応用が可能である．例えば，ハイパーリンク構造を分析することで歌詞の意味を元に楽曲をクラスタリングしたり，ハイパーリンクされた歌詞を分析することで，楽曲の再生中にその歌詞に関連した情報を表示したりできる．また，楽曲の再生中に歌詞中のハイパーリンクされたフレーズをクリックすることで，同じフレーズが歌詞に含まれた別の楽曲にジャンプできる新しい音楽鑑賞インタフェースにも応用できる．

従来の楽曲同士の関係を扱った研究として，楽曲間類似度の計算とそれを応用したインタフェースに関するものがある．[1)~5)]．また，楽曲内部の関係を扱った研究例として，サビなどの楽曲構造の自動解析や[6),7)]や，歌詞のと音楽の時間的対応付け[8)~10)]．などがある．一方，歌詞のハイパーリンクは，歌詞の内容に基づき楽曲同士の関係と楽曲内部の関係を同時に扱うものであり，従来は提案されていなかった．

本稿では，複数の楽曲に共通に登場するキーワード（フレーズ）間にハイパーリンクを張る手法について述べる．従来から，歌詞と音響信号の時間的対応付けに取り組んだ研究[8),9)]はあり，歌詞が既知の楽曲に対しては有効に機能するが，歌詞が未知の楽曲に対しては適用できない．また，単独歌唱の音響信号中の歌詞の認識に取り組んだ研究もある[11),12)]．伴奏を含む歌唱から歌詞を高精度に認識できればハイパーリンクの作成に応用できるが，実際は伴奏を含む歌唱中の歌詞の認識は非常に困難で，単独歌唱を対象とした技術は適

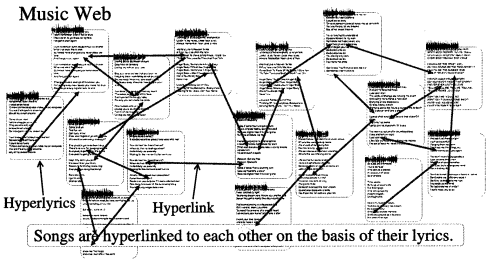

図 1　Music Web 上の歌詞のハイパーリンク．

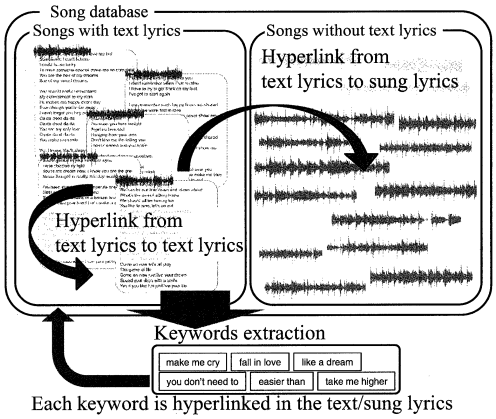

図 2　2 種類のハイパーリンク作成戦略：歌詞が既知の楽曲同士のハイパーリンクと，歌詞が既知の楽曲から未知の楽曲へのハイパーリンク．

用できない．また，音響信号ではなく，テキストとしての歌詞を分析した研究もある[13),14)]が，本研究は音響信号と歌詞の両方を対象としているのでそのままでは適用できない．

歌詞をハイパーリンクするための戦略を図 **2** に示す．本研究では，ユーザが楽曲データベースの全ての楽曲について音響信号（ MP3 ファイルなど）を，そして一部の楽曲について歌詞を保持していることを仮定する．すなわち，楽曲データベースは歌詞が既知の曲と歌詞が未知の曲の 2 種類からなると仮定する．これにより，2 種類のハイパーリンク作成戦略を考えることができる．一つは，歌詞が既知の楽曲同士のハイパーリンクであり，もう一つは，歌詞が既知の楽曲から歌詞が未知の楽曲へのハイパーリンクである．歌詞が未知の楽曲同士のハイパーリンクは，歌詞が既知の楽曲のあるフレーズから二つ以上の歌詞が未知の楽曲にハイパーリンクされている場合に，それらの歌詞が未知の楽曲同士にハイパーリンクを張ることで実現できるので，本稿では明示的には扱わない．

2. 歌詞のハイパーリンク作成手法

本手法は，複数の楽曲に共通して登場する同一のフレーズ同士にハイパーリンクを作成する．つまり，同じフレーズ（本稿ではキーワードと呼ぶ☆）を共有する複数の楽曲について，音響信号中のそのキーワードが発声されている区間同士をリンクによって関連付ける．対象とする楽曲は，歌声だけでなくそれ以外の伴奏楽器の音も含んだ音響信号である．本研究では，対象の楽曲データベースが，歌詞が既知と未知の 2 種類の楽曲から構成されていると仮定し，歌詞が既知の楽曲同士のハイパーリンクと，歌詞が既知の楽曲と未知の楽曲の間のハイパーリンクの 2 種類のハイパーリンクを考える．前者のハイパーリンクは，データベース中の既知の歌詞テキスト全体から，キーワードの候補を抽出し，そのキーワードが楽曲中のどの時刻で発声されているかを推定することで実現する．キーワード発声時刻の推定には，歌詞と音響信号の時間的対応付け手法[9)]を使用する．後者については，データベース中の歌詞が未知の楽曲に対して，抽出されたキーワードのいずれかが含まれているかどうかをキーワードスポッティング技術により推定することで実現する．

2.1 歌詞が既知の楽曲同士のハイパーリンク

本節では，歌詞が既知の楽曲同士のハイパーリンク作成手法について述べる．まず，楽曲データベース中の全ての既知の歌詞を用いて，複数の楽曲に共通して登場する意味のある単語の集まりをキーワードとして抽出する．次に，楽曲の音響信号中から，それらのキーワードが登場する区間（開始時刻と終了時刻）を推定する．最後に，それぞれのキーワードが登場する楽曲の該当区間同士にハイパーリンクを作成する．

2.1.1 歌詞テキストからのキーワード抽出

まず，歌詞をよく表現するキーワードの集合を，楽曲データベースの歌詞テキストから抽出する．楽曲間にハイパーリンクを作成するという観点からは，それぞれのキーワードは少なくとも 2 つの楽曲に登場する必要があり，登場回数は多ければ多いほど良い．一方で，キーワードは長い方が重要な意味を表すことが多いという観点から，なるべく長いキーワードが望ましい．さらに，キーワードは長い方が，2.2 節で述べるキーワードスポッティングの精度が向上するというメリットもある．逆に，助詞などのような短い単語，それだけで意味をなさない短いフレーズは，キーワードとしては不適切である．

そこで，適切なキーワードを抽出するための用件を以下のように整理する．

(a) キーワードを含む楽曲数は多いほうが良い．
(b) キーワードが含む音素数は多い方が良い．

なお，この二つの用件の間にはトレードオフの関係がある．なぜなら一般にキーワードが長くなればなるほど登場楽曲数が減るからである．そこで，本研究では，少なくとも 2 つの楽曲に登場するという条件下で，音素数が最大となるキーワードを抽出する．

上述の用件に基づいて，下記のようなキーワード抽出法を提案する．このキーワード抽出法のアプローチは，まずは大量の短いキーワードを用意し，それらを前後に出現する単語と繋げていくことで，できるだけ長いキーワードを抽出するというものである．

(1) キーワード辞書を空に初期化する．このキーワード辞書に登録されるそれぞれのキーワードには，探索終了フラグを設定することができる．このフラグは，このキーワードに別の語を付け加え

☆ 本稿では，キーワードとは一つ以上の単語からなる歌詞中のフレーズのことを指し，必ずしも 1 単語ではない．

ても，これ以上長いキーワードは生成できないということを表す．

(2) まず，楽曲に登場する全ての語をキーワード辞書に登録する．

(3) キーワード辞書中のそれぞれのキーワードについて，そのキーワードが登場する楽曲数を数える．

(4) 登場曲数が M 曲より少ないキーワードを，キーワード辞書から削除する．

(5) 探索終了フラグが設定されていないキーワードの中で，最も登場曲数が多いキーワードを選択する．全てのキーワードに探索終了フラグが設定されていた場合，アルゴリズムを終了する．

(6) 選択されたキーワードが登場する全ての曲の歌詞で，そのキーワードの前または後に登場する語とそのキーワードを連結させ，より長いキーワード候補を作る．

(7) 作成されたキーワード候補の中で，最も登場回数が多いキーワード候補を選び，その登場回数が M 回より多いければ，キーワード辞書に追加する．なお，新たなキーワードが追加された場合でも，元のキーワードも削除せずに残しておく．

(8) (7) で，最も登場回数が多いキーワード候補の登場回数が M 回に満たない場合，そのキーワード候補の元となったキーワードの音素数が N 個以上であればそのキーワードに探索終了フラグを設定し，N 個より少なければそのキーワードを削除する．

(9) 5. に戻る．

なお，このアルゴリズムを実行する前に，歌詞中の全ての単語の発音を発音辞書から検索しておく．また，歌詞が日本語の場合，前もって形態素解析により歌詞を形態素に分割し各形態素の発音を推定しておく．そして，各形態素を単語として解釈する．

パラメータ M と N は，それぞれ上記の二つの用件 (a) と (b) に対応する．これらのパラメータの適切な値は，歌詞の全体量によって変わるので，これらの値の設定は 3.2 節で述べる．

2.1.2 ハイパーリンクの作成

抽出されたキーワードを使って，双方向のハイパーリンクを作成する．まず，歌詞が既知の曲で，歌詞テキスト中におけるキーワード出現場所を検索する．次に，各楽曲に出現した各キーワードについて，その楽曲の音響信号中のどの時刻に出現するかを推定する．これは，歌詞と音響信号の時間的対応付け手法[9]によって実現できる．最後に，同じキーワードが出現する楽曲の，キーワードが出現する時刻同士にハイパーリンクを作成する．

2.2 歌詞が既知の楽曲から未知の楽曲へのハイパーリンク

この節では，歌詞が既知の楽曲から歌詞が未知の楽曲へのハイパーリンクについて述べる．ここでも，2.1.1 節で抽出されたキーワードを使用する．歌詞が未知の楽曲に対してはテキスト検索によってキーワードを検索することができないので，キーワードが含まれているかどうかを伴奏を含む歌唱の音響信号から自動的に推定する必要がある．

本研究では，音声認識で用いられるキーワードスポッティング技術[15]を応用して，これを実現する．この手法では，キーワードの音響的特徴を表す音響モデルモデル（キーワードモデル）とキーワード以外の音の音響的特徴を表す音響モデル（ガベージモデル）の 2 種類を使用する．そして，分離された歌声の音響信号に対して，キーワードモデルとガベージモデルの最尤経路を計算し，キーワードが登場する区間を検出する．具体的には，一旦キーワードの出現候補を多めに検出しておいて，それらのキーワードを後から絞り込むという戦略をとる．

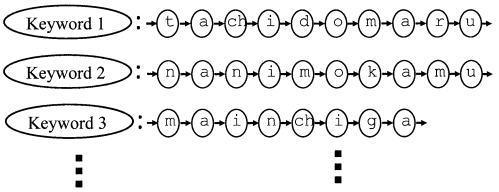

図 3 キーワードモデル（HMM）．

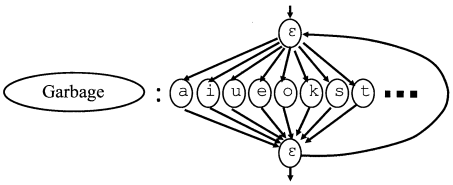

図 4 ガベージモデル（HMM）．

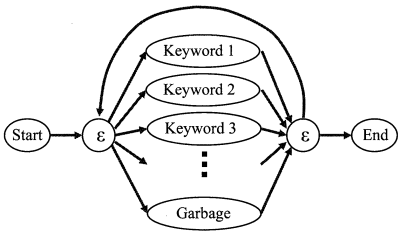

図 5 キーワードモデルとガベージモデルの統合．

2.2.1 多重奏の音響信号からの歌声の特徴抽出

音響モデルを使用した音声認識の手法を適用するためには，歌詞が未知の楽曲から歌声の特徴を表現する特徴ベクトル列を抽出する必要がある．本研究では，文献 9) で使われる特徴量抽出法により特徴ベクトル列を抽出する．この手法は，楽曲の音響信号に混在する歌声以外の伴奏音の影響を低減させることができる．

この手法では，以下の 3 つの処理により，混合音中の歌声を分離再合成する．

(1) PreFEst[16] を用いて，最も優勢な F0 をメロディ（歌声）として推定する．

(2) 推定された F0 の高調波構造を抽出する．

(3) 抽出された高調波構造を再合成し，メロディの音響信号を得る．

その後，再合成された音響信号から，音声認識で一般に使われる特徴量である MFCC，ΔMFCC，Δ パワーを計算する．

2.2.2 音響モデルの学習

キーワードモデルとガベージモデルは，隠れマルコフモデル（HMM）によりモデル化する．キーワード HMM は，**図 3** のようにキーワードを構成する音素を

直列に並べた形で構成される．ガベージ HMM は，**図 4** のようにあらゆる音素が任意の順番で登場できる音素タイプライターによって構成される．キーワード HMM とガベージ HMM は，**図 5** のように並列に並べた形で統合する．

2.2.3 音響モデルの学習

各音素の音響的特徴を表現する音響モデルは，キーワードスポッティングの性能に大きな影響を与える．通常の音声認識に使用される話し声用の音響モデルを，今回のように分離によって歪んだ歌声に適用させるのは難しいので，本研究では歌声専用の音響モデルを一から学習することで作成した．

まず，RWC 研究用音楽データベース：ポピュラー音楽（RWC-MDB-P-2001）[17] 中の日本語楽曲 79 曲に対して，詳細な音素レベルのアノテーションを付与した．そして，それらの音素アノテーションを使用し，3 状態 left-to-right HMM からなるモノフォンの音響モデルを学習した．

2.2.4 キーワード候補の検出とスコア計算

歌詞が未知の楽曲から抽出された特徴ベクトル列に対して，図 5 の HMM の最尤経路を計算する．すると，キーワード HMM はキーワードが歌われている区間の候補に，ガベージ HMM はそれ以外の区間に登場することが期待される．しかし，ガベージ HMM はあらゆる音素が任意の順で登場することを許容するので，原理的に全てのキーワード HMM を表現できてしまい，このままではキーワード HMM が登場しなくなってしまう．そこで，単語挿入ペナルティを導入し，この問題を解決する．単語挿入ペナルティは，経路上の単語の個数に応じて仮説にペナルティを導入するものである．キーワード HMM では 1 つのキーワードが 1 単語と，ガベージ HMM では 1 つの音素を 1 単語と考えることで，キーワード HMM で表現できる区間がガベージ HMM によって表現されてしまうことを防ぐことができる．このようなキーワードスポッティングのフレームワークは，言語モデル（図 5）が，ワードスポッティング用の特別なものである点を除くと，通常の音声認識の基本的な枠組みと同様である．

キーワードの候補区間が検出された後，各候補区間のスコアを計算し，候補区間の絞り込みを行う．まず，各候補区間にガベージ HMM のみを適用した時の尤度を計算する．各候補区間に対するキーワード HMM の尤度は，キーワード候補を検出する際に計算されているのでその値を利用する．そして，キーワード HMM に対する尤度とガベージ HMM に対する尤度の差を，候補区間のスコアとして定義する．最後に，スコアが閾値以上の候補区間を，キーワード出現区間として検出する．

2.2.5 ハイパーリンクの作成

各キーワードについて，キーワードが出現する楽曲を選択する．そして，歌詞が既知の楽曲とそれらの楽曲のキーワード出現区間の間に双方向のハイパーリンクを作成する．

3. 評価実験

本章では，提案法を評価するために行った予備的な実験について述べる．実験には，RWC 研究用音楽データベース：ポピュラー音楽（RWC-MDB-P-2001）中の歌詞が日本語の 79 曲を使用した．

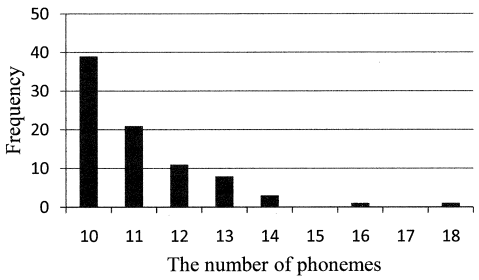

図 6 抽出されたキーワードの音素数分布．

表 1 抽出されたキーワードの例．

Keyword	Phoneme sequence	Number of occurances
が教えてくれたこと	g a o sh i e t e k u r e t a k o t o	2 songs
どこまでも続く	d o k o m a d e m o t s u z u k u	2 songs
心の中	k o k o r o n o n a k a	3 songs
素敵な笑顔	s u t e k i n a e g a o	2 songs
世界中に	s e k a i j y u: n i	2 songs

3.1 歌詞のハイパーリンク作成の評価

まず，2.1.1 節で述べた歌詞テキストからのキーワード抽出を実行した．本実験では，パラメータ M を 2（曲）に，N を 10（音素）に設定した．実験に使用した 79 曲全ての歌詞からキーワードを抽出したところ，84 のキーワードが抽出され，その平均音素数は 11.1 だった．**図 6** は抽出されたキーワードの音素数の分布を表し，表 1 に抽出されたキーワードの例を示す．抽出されたキーワードは，意味のある適切な長さであることが確認できた．なお，本実験では日本語歌詞の形態素解析のために，MeCab[18] を使用した．

次に，歌詞が既知の楽曲から歌詞が既知の楽曲に対するハイパーリンクを，リンク成功率という新しい評価基準を用いて，評価する☆．これは，作成されたハイパーリンク全体の中で，他の楽曲で同じフレーズが歌われている区間に正しくリンクされたものの割合を表す．まず，それぞれの楽曲について，その楽曲の歌詞のみが既知であると仮定し，2.2 節で述べた手法により，残りの 78 曲（歌詞は未知と仮定）にハイパーリンクを作成した．本実験では，各キーワードについて最もスコアの高い候補区間にのみハイパーリンクを作成した．そして，それらのハイパーリンクに対してリンク成功率を計算し，79 曲の平均リンク成功率を求めた．リンク成功率 r は，

$$r = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} s(w(k,i))}{\sum_{k=1}^{K} I_k}, \tag{1}$$

$$s(w(k,i)) = \begin{cases} 1 & \text{if } w(k,i) \text{ is appropriate} \\ 0 & \text{if } w(k,i) \text{ is not appropriate} \end{cases}, \tag{2}$$

として表される．ここで，K は抽出されたキーワードの総数を表し，k は各キーワードを表す．また，I_k は楽

☆ 本稿では，歌詞が既知の楽曲同士のハイパーリンクの評価は行わなかった．これは，このハイパーリンクの性能は，歌詞と音響信号の時間的対応付けの精度のみに依存するからである[9]．

表 2　下限音素数とリンク成功率及び抽出されたキーワード数の関係.

# of phonemes	8	9	10	11	12	13
Link success rate (%)	23.2	27.5	30.1	24.3	35.9	40.0
# of keywords	271	144	84	45	24	13

曲中に k 番目のキーワードが登場する回数を表し（ただし，同じキーワードが一つの楽曲に複数回登場することがある），$w(k,i)$ は k 番目のキーワードに関する i 番目のハイパーリンクを表す．ハイパーリンク $w(k,i)$ は，ハイパーリンク先の区間と，正解データ中の k 番目のキーワードが登場する区間が半分以上が重なった場合，正解と判断した．実験の結果，リンク成功率は30.1%であった．

3.2 キーワードの個数と音素数

次に，歌詞テキストからのキーワード抽出のパラメータ M（2 曲と設定）と N（10 音素と設定）の妥当性を確かめるための評価を行った．2.1.1 節で述べたように，キーワードは少なくとも M 曲以上の楽曲に登場し，各キーワードは少なくとも N 音素以上からなるように，キーワードが抽出される．本研究では，なるべく多くのキーワードが抽出したいので，M の値は 2 に固定した．そして，N の値を変化させながら，リンク成功率を計算した．

表 2 は，N（キーワード中の下限音素数）を変化させた時の，リンク成功率と抽出されたキーワードの個数の変化を表す．リンク成功率とキーワード数のトレードオフを考慮にいれて，3.1 節の実験では N の値を 10 と定めた．

3.3 実験条件の検証

実験条件の妥当性を確かめるための実験を行った．3.1 節と 3.2 節の実験では，実験に使用した 79 曲と，音響モデルの学習に使用した 79 曲は同じ楽曲である．これは，音響モデルの学習データと，評価に用いる正解データの両方に音素レベルのアノテーションが必要であるが，現時点で使用できる音素アノテーションの数が限られているからである．さらに，音響モデルの尤度を楽曲間で比較するためには，全ての楽曲で同じ音響モデルを使う必要があるため，クロスバリデーション法などを利用することもできなかった．とはいえ，音響モデルの学習には 79 曲を使用しているので，それぞれの楽曲の寄与分は大きくないと期待できる．そこで，本節ではこのことを確かめるための実験を行った．

ここでは，歌詞が未知の楽曲に対して，言語モデルとして音素タイプライタを使用した場合の音素認識率を評価した．音素認識率は，正解ラベルとして音響モデルの学習のための音素ラベルを使用し，楽曲全体の長さに対する音素が正しく認識できた区間の割合とした．下記の 2 つの条件で実験を行った．

i. 音響モデルの学習に 79 曲全てを用いた.（オープン）

ii. 10-fold cross validation 法により，ある楽曲の評価に使用する音響モデルの学習データにはその楽曲が含まれないようにした.（クローズ）

なお，本節での実験条件や目的，評価基準は，3.1 節と 3.2 節のものとは全く異なるものである．表 3 に実験結果を示す．表より，実験条件による音素認識率の違いは小さく，オープンに相当する条件の実験でも一般性は大きく失うことはないことを確認した．

表 3　オープンとクローズの 2 種類の音素認識結果.

Condition	i. closed	ii. open
Accuracy	50.9%	49.8%

表 4　3 種類の音響モデルの音素認識結果.

	i. small adapt.	ii. large adapt.	iii. train.
Accuracy	27.1%	32.7%	50.9%

3.4 音響モデルの評価

最後に，本研究で用いた歌声用に学習した音響モデルと話し声用の音響モデルを歌声に適応させたものと比較する．文献 9) では，音響モデルは話し声用の音響モデルを少量（10 曲）の歌声データに適応させたものを用いていた．その音響モデルは，歌詞と音楽の時間的対応付けの問題に対しては満足な精度で動いていたが，今回扱うキーワード検出の問題は，より難易度が高く，高精度な音響モデルを必要とする．そこで，本研究では，79 曲の楽曲に対して詳細な音素ラベルを用意し，音響モデルを一から学習して作成した．本節では，この新しい音響モデルの性能を評価するため，下記の三つの音響モデルの性能を，音素認識の実験によって比較する．

i. (small adaptation) 話し声用の音響モデルを 10 曲のデータに適応させて得られた音響モデル.

ii. (large adaptation) 話し声用の音響モデルを 79 曲のデータに適応させて得られた音響モデル.

iii. (training) 79 曲のデータを用いて，一から学習して得られた音響モデル.

その他の実験条件は，3.3 節の実験の条件 (ii.) と同様で，10-fold cross validation 法により音素認識率を評価した．実験結果を表 4 に示す．表より，条件 iii. では，平均音素認識率が劇的に向上していることがわかり，79 曲のデータで学習した音響モデルの有効性が確認できた．

4. 考　　察

本稿では，歌詞のハイパーリンクを作成する手法，つまり複数の曲に共通で登場する歌詞のフレーズ間に双方向リンクを作成する手法について述べた．これにより，楽曲データベースの楽曲が相互にハイパーリンクされたネットワーク構造（Music Web）を作成することができる．本手法は，楽曲データベース中の全ての楽曲の歌詞が既知である必要がないという利点がある．本研究が，歌詞の内容に基づく音楽検索を，歌詞が未知の楽曲に対しても適用できるようになるための第一歩であると考えている．

また，楽曲の再生に同期して歌詞を表示する音楽再生システム LyricSynchronizer[9),19)] と本手法を統合することも可能である．例えば，楽曲を再生中に歌詞中の下線の付いたキーワードをクリックすることで，同じキーワードが登場する別の曲にジャンプしたりすることができる．

本研究は新しい研究テーマの初期段階であり，また，多重奏の音響信号中の歌詞を扱うことは難易度の高い問題である．そのため，3.1 節の実験で示した本手法の性能は，改善が必要である．そのためには，音響モデルの改善が有効であると考えられる．3.4 節で示した通り，多くの楽曲を使用して学習することで音響モデルの性能が向上する．特に，ある程度の数の詳細な

ラベルを持つ学習データがあると，あとは，詳細なラベルの付いていない学習データを大量に用意することで，音響モデルの性能をさらに高めることができる．

tf-idf（term frequency-inverse document frequency）法などの既存手法と比較して，2.1.1 節で述べたキーワード抽出手法は，本研究の目的，つまり 2 曲以上の楽曲に登場するなるべく長いキーワードを抽出したい場合には適している．tf-idf 法は，キーワードの重要度はある文章中のキーワードの登場回数に比例し，全ての文章中の登場回数に反比例すると仮定している．しかし，今回の目的の場合キーワードの登場回数は重要ではなく，キーワードの長さや登場楽曲数が重要になる．そのため，本研究では独自のキーワード抽出法を開発した．

5. ま と め

本稿では，異なる楽曲に共通して登場する歌詞にハイパーリンクを作成する手法について述べた．本研究では 2 種類のハイパーリンクを扱った．すなわち，歌詞が既知の楽曲同士のハイパーリンクと，歌詞が既知の楽曲から歌詞が未知の楽曲へのハイパーリンクである．本手法では，全ての既知の歌詞からキーワードを抽出した後，HMM に基づくキーワードスポッティングにより多重奏の音響信号からキーワードを抽出する．今後は，全体の性能向上を図ると同時に，実験を大規模なデータベースと現実に近い条件で行う予定である．また，歌詞に基づく楽曲のクラスタリングや，ハイパーリンクされた歌詞により楽曲データベースをブラウジングできるインタフェースなど，本手法を応用したアプリケーションの開発に取り組む予定である．

謝辞 本研究の一部は，科学技術振興機構 CrestMuse プロジェクトの支援を受けた．本研究では音響モデルの学習には，HTK（Hidden Markov Model Toolkit）[20] を使用した．また，RWC 研究用音楽データベースに対する音素ラベルの作成について助言を頂いた伊藤克亘氏（法政大学）に感謝する．

参 考 文 献

1) Tzanetakis, G. and Cook, P.: MARSYAS: A Framework for Audio Analysis, *Organised Sound*, Vol. 4, No. 30, pp. 169–175 (1999).
2) Neumayer, R., Dittenbach, M. and Rauber, A.: PlaySOM and PocketSOMPlayer: Alternative Interfaces to Large Music Collections, *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 618–213 (2005).
3) Goto, T. and Goto, M.: Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces, *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 404–411 (2005).
4) Pampalk, E. and Goto, M.: MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling, *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pp. 367–370 (2006).
5) Lamere, P. and Eck, D.: Using 3D Visualizations to Explore and Discover Music, *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, pp. 173–174 (2007).
6) Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 14, No. 5, pp. 1783–1794 (2006).
7) Müller, M. and Kurth, F.: Enhancing Similarity Matrices for Music Audio Analysis, *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006)*, pp. V–9–12 (2006).
8) Wang, Y., Kan, M.-Y., Nwe, T. L., Shenoy, A. and Yin, J.: LyricAlly: Automatic Synchronization of Acoustic Musical Signals and Textual Lyrics, *Proceedings of the 12th ACM International Conference on Multimedia*, pp.212–219 (2004).
9) Fujihara, H., Goto, M., Jun, O., Komatani, K., Ogata, T. and Okuno, H. G.: Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals, *Proceedings of the IEEE International Symposium on Multimedia (ISM 2006)*, pp. 257–264 (2006).
10) Müller, M., Kurth, F., Damm, D., Fremerey, C. and Clausen, M.: Lyrics-based Audio Retrieval and Multimodal Navigation in Music Collections, *Proceedings of the 11th European Conference on Digital Libraries (ECDL 2007)* (2007).
11) Wang, C.-K., Lyu, R.-Y. and Chiang, Y.-C.: An Automatic Singing Transcription System with Multilingual Singing Lyric Recognizer and Robust Melody Tracker, *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech2003)*, pp. 1197–1200 (2003).
12) Suzuki, M., Hosoya, T., Ito, A., and Makino, S.: Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information, *EURASIP Journal on Advances in Signal Processing*, Vol. 2007 (2007).
13) Knees, P., Schedl, M. and Widmer, G.: Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics, *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 564–569 (2005).
14) Wei, B., Zhang, C. and Ogihara, M.: Keyword Generation for Lyrics, *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, pp. 121–122 (2007).
15) Knill, K. and Young, S.: Speaker Dependent Keyword Spotting for Accessing Stored Speech, Technical Report CUED/F-INFENG/TR 193, Cambridge University (1994).
16) Goto, M.: A Real-Time Music-Scene-Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals, *Speech Communication*, Vol. 43, No. 4, pp. 311–329 (2004).
17) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, pp. 287–288 (2002).
18) MeCab: Yet Another Part-of-Speech and Morphological Analyzer: http://mecab.sourceforge.net/.
19) Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, pp. IV–1441–1444 (2007).
20) HTK: The Hidden Markov Model Toolkit: http://htk.eng.cam.ac.uk/.