音声検索システムのための時間整合を考慮したサブワードモデル構築手法の検討

岩田　耕平†　伊藤　慶明†　小嶋　和徳†　石亀　昌明†　田中　和世††
李　時旭†††

† 岩手県立大学ソフトウェア情報学研究科　〒020–0193 岩手県滝沢村滝沢字巣子 152–52
†† 筑波大学大学院図書館情報メディア研究科　〒305–8550　茨城県つくば市春日 1–2
††† 産業技術総合研究所情報技術研究部門　〒305–8568　茨城県つくば市梅園 1-1-1
E-mail: †g231d002@edu.soft.iwate-pu.ac.jp

あらまし　パソコン・ハードディスクレコーダーの普及に伴い，ユーザが見たい場面を容易に検索できる機能が必要不可欠である．我々は収録されている音声をテキストおよび音声で検索するシステムを提案した [1]．提案手法の特長として単語認識ではなく，音素モデルである monophone や triphone モデルといったのサブワードモデルを用いた認識・検索を行うこと及びサブワードモデル間の音響的な距離を利用することが挙げられる．本稿では語彙フリー検索における適切な局所距離およびサブワードモデルに関する検討を行い，結果を報告する．また，サブワード継続時間に着目して状態数を割り当てたモデルについての検討結果についても報告を行う．

キーワード　音声検索システム，サブワードモデル，音響距離，時間整合モデル

A Study of Subword Models Taking Time Cosistency Consideration for Vocabulary-free Spoken Document Retrieval System

Kohei IWATA†, Yoshiaki ITOH†, Kazunori KOJIMA†, Masaaki ISHIGAME†, Kazuyo TANAKA††, and Shi-wook LEE†††

† Iwate Prefectural University　Sugo 152–52, Takizawa, Iwate, 020–0193 Japan
†† Tsukuba University　Kasuga 1–2, Tsukuba, Ibaragi 305–8550 Japan
††† AIST, Umezno 1–1–1, Tsukuba, Ibaragi 305–8568 Japan
E-mail: †g231d002@edu.soft.iwate-pu.ac.jp

Abstract　According to the recent spread of personal computers and video hard-disc recorders, a new function is needed such that it is easy for users to identify the scene that a user wants to watch in a video data. For this purpose, we proposed a speech retrieval system by a text and speech query [1]. The proposed system has two charactaristics. One is that the system uses subword models such as phone models. The other is that the system retrieves similar section by using acoustic similarity between subword models. In this paper, we discuss the appropriate subword models and the appropriate local distance. We construct other subword model which taking into time cosistency consideration. Through the paper, we discuss what kind of subword model is suitable for proposed SDR system.

Key words　SDR system, Subword models, phonetic distance, time consistency model

1. はじめに

近年，パソコンにおけるマルチメディア環境の整備やハードディスクレコーダーの普及に伴い，ビデオデータは容易かつ頻繁に扱われるようになった．媒体の大容量化により，ビデオデータの長時間保存が可能となり，将来的には全ての番組を録画しておき，録画データの中で見たい番組やカットのみを鑑賞するというビデオスタイルに変わっていくことが想定される．このようなビデオスタイルにおいては，長時間保存したデータの中から自分の見たい場面を検索できる機能が必要不可欠となる．実際の番組の中でどのような内容が話されたかといった詳細情報に関しては一般的に入手することが困難である．また，キーワードが話された部分を検索・特定することは現状では極めて難しい．

ユーザが聞きたい音声の特定区間を検索するための最も簡便な手法の一つとして，テキストまたは音声をクエリとする音声検索が考えられる．音声検索の方法としては，音声認識の結果を利用する方法が代表的である [2] [3] が，辞書に登録されていない語句 (OOV: *Out of Vocabulary*) が検索クエリとして与えられた場合，適切な認識は行われず検索が困難となる．人名や地名，専門的技術用語といった特殊な単語は OOV となる可能性が高いが，それらの語句がビデオデータを特徴付けることが多い．付録に示すように，実際に web で検索される語句は OOV 率が高く，検索システムにおける OOV 対応は非常に重要であるといえる．

そこで，認識の単位を単語とするのではなく，サブワードとすることによって，あらゆる語句の検索を可能とする語彙フリーな検索を提案する．

本稿では，語彙フリー検索における音響距離，および適切なサブワードモデルの検討結果について報告を行う．また，サブワード継続時間に着目して状態数を割り当てたモデルについての検討結果についても報告を行う．

本検索システムでは，音声クエリと音声データベース（ビデオ音声）それぞれをサブワード単位で認識を行い，サブワード系列の照合を行う．サブワードモデルを利用した研究には，単純に記号が一致するか否かで判断する (以下 edit distance と表記) 方式を用いた研究 [8] ，サブワードの間違いやすさを利用した研究 [9] 等がある．本研究では認識誤りに対して頑健な検索を実現するため，サブワードモデルの統計量を用いてサブワードモデル間の音響的な近さ（音響距離）を定義し，それをサブワードモデル間の距離に利用する [6] [7]．本稿内で照合時の局所距離による検索性能の比較を行う．

次にサブワードモデルとして音素より時間的に精緻なモデル化を行ったモデルを導入する．連続する音声データの検索では時間的な整合が重要になるため，この時間軸上の精緻化により検索性能の向上を期待するものである．環境依存の 1/2 音素モデルと 1/3 音素，音素片モデル (SPS: *Sub-Phonetic Segment*) [4] を利用する．環境依存の 1/2 音素および 1/3 音素とは triphone をベースとして前後の音素を考慮した上で，各音素を時間軸方向にそれぞれ 2 つ，3 つのモデルとしたものである．それぞれのモデルは HMM で構成する．

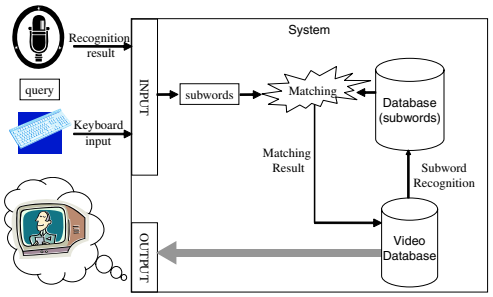

図 1　システム概念図
Fig. 1　System image of proposed SDR system

本稿ではサブワードの継続時間に着目したサブワードモデルの構築を行い，評価を行う．従来の我々の実験で用いたサブワードモデルはすべて同一状態数で構成されていたが，継続時間に合わせて状態数を割り当てることにより，時間整合性が向上し，検索性能が向上することを期待する．

各章の章立ては，2. 章で音声検索システムの概要および検討するサブワードモデルや局所距離について述べる．3. 章で各サブワードモデルにおける検索性能について実験および考察を行う．最後にまとめを述べる．

2. 語彙フリー音声検索方式

2.1 システム概要

本システムはサブワード単位の認識・検索を行うことによって，通常の音声認識システムでは OOV となりやすい人名や地名といった固有名詞や，新語や専門用語といった非頻出単語でも検索を可能とするもので，語彙フリーの音声検索システムである．図 1 にシステムの概念図を示す．

録画したビデオ等の音声データは予めサブワード認識を行いサブワード列のデータベース (DB) として保持する．クエリはキーボード等によるテキストあるいは音声を入力して与えることを想定する．テキストの場合は規則に従いサブワード列に変換し，音声は音声 DB 同様のサブワード認識を行う．クエリのサブワード列と DB 中の全サブワード列間でサブワード間音響距離を局所距離として連続動的計画法（連続 DP）による照合を行う．

本システムでは音声 DB と検索語のサブワード系列間の照合方式に連続動的計画法（連続 DP）を用いる．

連続 DP における検索語の τ 番目のサブワードとデータベースの t 番目のサブワードまでの累積距離 $G(t,\tau)$ は以下の式 (1) で与えられる．式中の $D(p_t,p_\tau)$ は検索語の τ 番目のサブワード p_τ とデータベースの t 番目のサブワード p_t との局所距離を意味している．連続 DP 中，局所距離は計算せずメモリー上のサブワード距離マトリックスを参照するだけである．なお検索語毎にサブワード数が異なるため，累積距離を検索語のサブワード数で正規化した結果を検索語とデータベース要素の距離とする．

$$G(t,\tau)=argmin\begin{cases} G(t-1,\tau-1)+ \\ \quad D(p_t,p_\tau)\cdot 3 \\ G(t-2,\tau-1)+ \\ \quad D(p_{t-1},p_\tau)\cdot 2+D(p_t,p_\tau) \\ G(t-1,\tau-2)+ \\ \quad D(p_t,p_{\tau-1})\cdot 3+D(p_t,p_\tau)\cdot 3 \end{cases} \tag{1}$$

2.2 局所距離

本稿では局所距離として edit distance，サブワードの間違いやすさ (confusion matrix)，サブワード間音響距離 (phonetic distance) を用いる．以下にそれぞれの局所距離の説明および参照サブワード ref と入力サブワード inp の局所距離 $d(ref,inp)$ の算出式を記述する．

edit distance　参照サブワードと入力サブワードが一致するか否かを判断する．

$$d(ref,inp)=\begin{cases} 0 & if \quad ref=inp \\ 1 & otherwise \end{cases} \tag{2}$$

confusion matrix　サブワードの認識誤りを求め，間違いやすさを算出する．同一サブワード間の距離は edit distance と同様 0 とする．

$$d(ref,inp)=0 \tag{3}$$

それ以外の場合については次式を用いて距離を算出する．式において $P(inp|ref)$ は評価用データベース以外の音声データベースを用いて ref に属する音声に対して認識処理を行った結果，inp と認識された確率を示している．少数の認識誤りをした場合と認識誤りのない場合の距離の差を大きくするために累乗を行うこととし，予備実験によって比較を行った結果最も性能の高い 3 乗を用いる．

$$d(ref,inp)=(1.0-P(inp|ref))^3 \tag{4}$$

サブワード間音響距離　我々の提案する局所距離であり，学習した HMM 音響モデルに含まれるガウス分布に対して Bhattacharya 距離を算出する．サブワード間距離は式 (5) の通り，サブワードを構成する N 個の状態間距離の平均を取る．

$$d(ref,inp)=\frac{1}{N}\sum_{i=1}^{N}d_s(s_i^{ref},s_i^{inp}) \tag{5}$$

表 1　各サブワードにおける「イワテ」の表記
Table 1　Each subword expression for the word "Iwate"

Subword	Expression
monophone	i w a t e
triphone	#-i+w i-w+a w-a+t a-t+e t-e+#
1/2 音素	#1i i2w i1w w2a w1a a2t a1t t2e t1e e2#
1/3 音素	#1i ii i2w i1w ww w2a w1a aa a2t tt a1t t2e t1e ee e2#
SPS	#i ii iw ww wa aa at tcl tt te ee e#

状態間距離 $d_s(s_i^{ref},s_i^{inp})$ の算出には式 (6) を用いる．2 つの状態に含まれるあらゆる分布間の距離を算出し，類似度を表現するために最小値を状態間距離として選択する．

$$d_s(s_i^p,s_i^q)=\min_{1\leqq j,k\leqq M} d_B(c_{i,j}^p(x),c_{i,k}^q(x)) \tag{6}$$

各状態に含まれる分布間の距離は式 (7) で求める．

$$\begin{aligned} &d_B(c_{i,j}^p(x),c_{i,k}^q(x)) \\ &=-\log\int\sqrt{g_j(x)g_k(x)}dx \\ &=\tfrac{1}{4}\sum_{\ell=1}^{L}\left\{\frac{(\mu_{j\ell}-\mu_{k\ell})^2}{\sigma_{j\ell}^2+\sigma_{k\ell}^2}+\log\frac{(\sigma_{j\ell}^2+\sigma_{k\ell}^2)^2}{4\sigma_{j\ell}^2\sigma_{k\ell}^2}\right\} \end{aligned} \tag{7}$$

2.3 サブワードモデルの検討

サブワードとは単語よりも細かい単位であり，音素や音節がこれに該当する．本稿ではサブワードモデルとして一般的なモデルである monophone や triphone に加え，時間的に精緻なモデル化を行った 3 つのモデルを構築し，比較する．

我々が考案した 1/2 音素モデル，1/3 音素モデルは 1 つの音素 (triphone) を時間軸上でそれぞれ 2 つ，3 つに分割し，それぞれを別々のモデルとしてモデル化を行う．以降の節で詳述するが，時間的に精緻なモデルとなるとともにサブワード数が減少し，サブワードバイグラム，トライグラムから成る言語モデルの perplexity が低下するため，検索性能の向上を期待できる．さらに SPS モデルを構築する．SPS モデルは国際音声記号 (IPA) に準拠した ASCII コードである XSAMPA をベースにして，この音声記号に対して音響物理的特性を考慮して分割したサブ音声セグメント符号系である [4]．表 1 に「イワテ」という単語を表現する際の各モデルの音素表記を記す．monophone および triphone は 5 音素から構成されるのに対し，SPS は 12，1/2 音素は 10，1/3 音素は 15 となっている．表中#は前後に隣接する単語の最初あるいは最後の音素を表す．サブワードモデル比較実験においては同一状態数の HMM で構成されるサブワードモデルを用いているため，同一語句を表現するためのサブワード数が増えるモデルはそれだけ時間的に精緻なモデル化がなされているといえる．

2.3.1 サブワード音響モデル

各サブワードモデルは学習用音声データベースを利用

表 2　各サブワードモデルの比較
Table 2　Comparison between each Subword Model

	モデル数	精緻度	perplexity	サブワード認識率
monophone	43	1.0	7.91	73.52
triphone	7,956	1.0	4.73	55.89
1/2 音素	1,333	2.0	2.97	65.08
SPS	423	2.2	2.65	77.84
1/3 音素	1,374	3.0	2.02	70.30

して HMM 音響モデルを構築する．音響モデルには自己ループを含む left-to-right 型のモデルを用いる．各モデルはすべて monophone モデルから初期モデルを作成し，学習を繰り返した後に混合数を増加させる．一般に音響モデルは状態共有を行い再学習を行う．本稿においては triphone モデルは状態共有を行うが，状態共有を行わないモデルでの評価とした．

2.3.2　サブワード言語モデル

サブワード言語モデルにはサブワードバイグラムおよびサブワードトライグラムをテキストコーパスを利用して作成する．辞書サイズはサブワード数と一致する．

2.3.3　サブワードモデルの予備的な比較

日本音響学会の新聞記事読み上げ音声コーパス (JNAS: *Japanese Newspaper Article Sentences*) [11] を利用し，各サブワードモデルについて音響モデルおよびサブワードバイグラムとサブワードトライグラムから成る言語モデルの構築を行った．Table 2 に各サブワードモデルの比較を示す．表には JNAS に出現する各サブワードのモデル数とモデルの時間的な精緻度，サブワード perplexity，および ETLDB [12] を用いた認識実験により得られたサブワード認識率（サブワード正解率）を示す．モデルの時間的な精緻度とは，JNAS に含まれる全ラベルファイルを各サブワード系列に変換した際の総サブワード数を表現しており，音素モデルである monophone を 1.0 とした場合の比率である．認識率の向上にはサブワード数の削減および perplexity を低減したモデルが有効であると考える．

2.4　継続時間を考慮したモデルの構築

本稿では音素を単純に分割したモデルを構築しているが，更なる検索性能向上のためには時間整合性を考慮したモデルの学習が重要であると考えた．本稿では時間整合性の高いモデルを構築するために，サブワード継続時間に着目して状態数の割り当てを行う．

図 2 に SPS モデルでの認識を行った際の各サブワード継続時間の分布の模式図を示す．SPS モデルは音素を中心部（母音，長母音，子音）と過渡部に分割してモデル化する．また，破裂音や破擦音を発声する際に生じる無音部（特殊）をモデル化することにより音響特性を考慮したモデル化を行っている．図より，母音や子音，特殊音

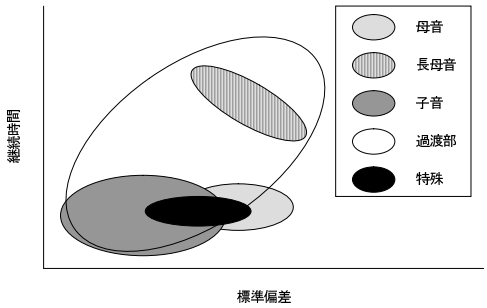

図 2　SPS モデルの継続時間の分布
Fig. 2　Distribution of each SPS model duration time

の継続時間は短く，長母音の継続時間は長いことが分かる．そこで，全てのモデルに対して同一状態を割り当てるのではなく，継続時間に合わせて状態数を変化させることにより性能が向上すると考えた．母音と子音の状態数を少なく，長母音の状態数を多くしてモデルを再学習し，評価を行う．過渡部に関しては分布が大きく，さらに細かい分類を行って状態数を割り当てる必要があると考えられるが，本稿においては継続時間の長いモデルをより精密に学習させるため，長母音と同じ状態数とする．

3.　評 価 実 験

3.1　学習用・評価用音声データ

各サブワードモデルの検索性能を評価するため比較実験を行う．

継続時間を考慮したモデル以外のサブワードモデルは状態数 3 の HMM で構成し，JNAS データを利用して音響モデル，言語モデルを作成した．検索実験に用いるモデルの混合数は局所距離にサブワード間音響距離を用いる際の計算コストを考慮し，全ての実験において 16 混合とした．

言語モデルは，単語ではなくサブワードのバイグラム，トライグラムを用い，JNAS の音素表記から学習した．なお，サブワード認識の辞書は，全てのサブワードから構成される．例えば SPS の場合，辞書の語彙は SPS のモデル 423 個となる．

評価用データは学習用の音声コーパスとは異なる電子技術総合研究所の単語データベース (ETL-DB) [12] を利用する．DB 中の 1,542 語 10 人分の単語セットを用いる．各単語の音素 (monophone) 数は 4 から 12 であった．予め全ての単語について音声認識デコーダ Julius を用いてサブワード認識しておく．

本 DB に含まれる音声データの音響分析条件等を表 3 に示す．分析時のフレームシフトは 10msec/5msec とした．これは予備実験において 1/2 音素および 1/3 音素，SPS モデルは 5msec の性能が高く，monophone および triphone においては 10msec における性能が高い性能を

表 3　音響分析条件

Table 3　Conditions of feature extraction

Sampling	16KHz, 16bits
Feature vector	12 dimentional MFCC_E_D_N_Z
Window	Hamming window
Window length	256 points (16 msec)
Frame interval	10 / 5 msec

示したためである．この原因としては時間的に精緻なモデルの場合，10msec の特徴量では各状態に対して十分な学習データが確保できないためであると考えられる．

3.2　実験条件

本実験ではモデル間の性能比較を目的としているため，連続音声中の検索語の検索ではなく，3.1 で示した単語 DB の全 1,542 単語，10 人分の音声を検索対象データとして検索単語との照合実験とした．

検索語はテキストで入力する場合（テキストクエリ）と音声で入力する場合（音声クエリ）との 2 通りを想定した．テキストクエリの場合，DB に付属している音素ラベルをルールに従いサブワードの系列に変換する．検索のターゲットとなる音声単語はそれぞれ 10 人が発声しているので各単語につき 10 個の正解検索単語が存在する．音声クエリの場合，10 人中 1 人の話者のデータを検索語とし，残りの 9 人のデータを検索対象 DB としたため，各話者についての性能が得られる．

各検索語に対しては類似度の高い順（連続 DP の正規化距離が小さい順）に候補として出力され，Precision-Recall グラフおよび F 値で評価を行う．

3.3　結果および考察

3.3.1　局所距離による検索性能の比較

図 3 に SPS モデルを用いて局所距離を変化させた場合の結果を示す．従来手法である edit distance や confusion matrix の結果と比較して，我々の提案手法であるサブワード間音響距離を用いることの優位性が確認できた．confusion matrix 算出のために用いた認識文の数を今回の実験においては約 5000 としたが，認識文数を増加させることによって精度向上は期待できる．しかし，認識に時間を要することや性能が頭打ちになる箇所の見極めが困難である．この実験によって性能面，コスト面ともに音響距離の優位性が確認できた．

3.3.2　サブワードによる検索性能の比較

図 4 に monophone，triphone，1/2 音素，1/3 音素，SPS の各サブワードの検索性能を示す．各モデルともテキストクエリ，サブワード間音響距離を用いたものである．

検索性能において最も優れているのは，SPS モデルであった．また，時間的に精緻なモデルを用いることによって音声認識に一般的に用いられる triphone よりも高い検索性能が得られ，時間的に精緻なモデル化が有効であることが確認できる．

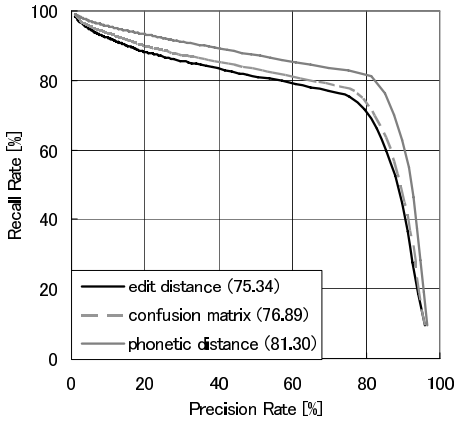

図 3　サブワード間距離による性能比較

Fig. 3　Performance comparison between subword distances

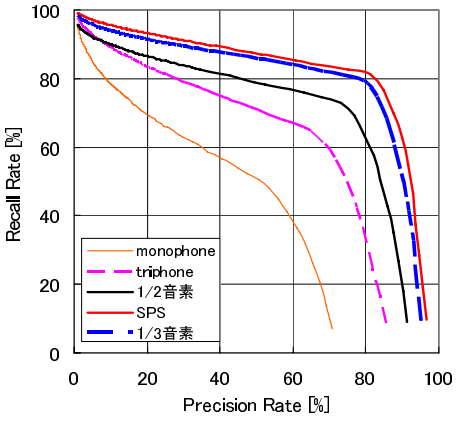

図 4　サブワードモデルによる検索性能の比較

Fig. 4　Performance comparison between subword models

1/2 音素モデルおよび 1/3 音素モデルはそれぞれ 6 状態，9 状態の triphone と類似する．それらのモデルについても構築・評価を行ったが，3 状態の triphone よりも性能が低下した．これは triphone の状態数を 6 状態に増やすと総状態数は $8000 \times 6 = 48000$ となり，学習が十分に行えなかったためと考える．クラスタリングして 2000 状態，4000 状態に状態共有を行った場合でも性能は向上は見られなかった．

3.3.3　音響特性を考慮したモデル

図 5 に 3 状態で固定したモデルおよびサブワード継続時間に応じて 3 状態と 4 状態を割り当てたモデルを使用した際の検索性能を示す．状態数の異なるサブワード間の局所距離算出についてはさまざまな方法が考えられるが，今回の実験では状態数を固定しないモデルの有効性について検証を行うことを目的とするため，局所距離には edit distance を用いることとした．実験の結果，3 状

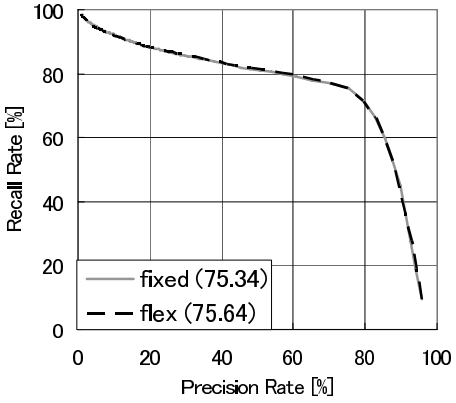

図 5　音響特性考慮モデルとの検索性能の比較

Fig. 5　Performance comparison between acoustic property based subword model

態固定モデルよりも F 値は 75.34 から 75.64 へと若干向上したものの，性能の優位性を確認することはできなかった．これは過渡部のサブワードの継続時間が異なるにも関わらず，全てに 4 状態を割り当てたことが原因であると考えられる．今後はさらに詳細な継続時間の分析を行い，継続時間に合わせた状態数の割り当てを行うことで性能向上を図りたい．

4. おわりに

本稿では，語彙フリー検索における音響距離，適切なサブワードモデル，サブワード継続時間に着目したモデルについての検討を行った．実験を通じて，局所距離にサブワードモデル間の音響距離を導入することによる検索性能向上を得られた．また，時間的に精緻なモデル化によってより詳細なモデル化が可能となり，性能向上につながることが確認できた．音響特性を考慮した SPS モデルが高い性能を示したことなどからモデル化に際して音響特性を考慮する意義はあると予想されるが，本稿で行ったサブワードの継続時間長に着目した状態数の割り当てでは従来手法に対して優位な結果が得られなかった．

今後はさらに音響特性を考慮したモデル化を行い，検索性能向上を目指していきたい．また，本稿内における評価は全て単語 DB の照合実験であったため，今後は連続音声中の候補区間の検出や実データに対する評価を行いたい．

付　　録

1. 検索語の辞書登録語率

表 A·1 に検索サイト yahoo [13] における 2004 年，2005 年の検索キーワードランキング BEST 50 が音声認識ソフトウェア Julius に付属の辞書に登録されている割合を示す．表からも分かるとおり，検索語が辞書に登録されていない可能性は高い．さらに，年が新しくなるたびに新たな言葉が流行するため，単語単位での認識を行う場合には辞書を定期的に更新する必要がある．我々の提案するサブワード単位での認識においてはサブワードの組み合わせであらゆる言葉を再現することが可能であるため，辞書の更新コストをかけずに使用し続けることができる．

表 A·1　検索語の辞書登録語率

Table A·1　Registration rate of popular query word

辞書の種類	2004 年	2005 年
Web 辞書	62%	52%
新聞記事辞書	60%	50%

文　　献

[1] 岩田他：”語彙フリー音声検索における時間精緻化サブワードモデルの検討”，日本音響学会講演論文集，1-1-11, pp. 21-22, Mar. 2006

[2] 桐山他：”話題知識を導入した文献検索音声対話システム”，IEICE Vol.J85-D2 No.5, pp.863-876, 2002

[3] 松下他：”音声入力による Web 検索のためのキーワード認識・抽出法の改善”，電子情報通信学会　研究報告「音声言語情報処理」，2003.

[4] 田中，児島，藤村，伊藤：”汎用音声符号系への符号化と音声処理システムの構築”，日本音響学会講演論文集，2-5-14, pp. 101-102, Mar. 2002

[5] 鹿野清宏他：IT Text 音声認識システム，オーム社，2001

[6] Shi-wook Lee，Tanaka K.，Fujimura N. and Itoh Y.，“Evaluation of speech data retrieval system using subphonetic sequence”，日本音響学会講演論文集，3-Q-3, pp.159-160, Sep. 2002

[7] 小川，山口，高橋，”混合重み係数を考慮した分布間距離尺度による音響モデルの分布数削減”，日本音響学会講演論文集，2-1-23，Sep. 2004

[8] N. Moreau，H.-G. Kim，T. Sikora，“Phonetic Confusion Based Document Expansion for Spoken Document Retrieval”，ICSLP，Vol. 2，pp.1593-1596，Oct 2004.

[9] S. Savitha，P. Dragutin，“Phonetic Confusion Matrix Based Spoken Document Retrieval，”Phonetic Confusionn Matrix Based Spoken Document Retrieval，pp. 81-87，2000.

[10] K.Tanaka，Y.Itoh，H.Kojima，N.Fujimura: Pro.IEEE ASRU2001，Paper A01kt080，pp. 1-4，2001

[11] Katsunobu Itou，“JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research，” J. Acoust. Soc. Japan. (E)，Vol. 20-3，pp. 199-2006，1999.

[12] 速水悟他：”研究用音声データベースのための VCV/CVC バランス単語セットの作成”，電子技術総合研究所彙報，第 49 巻，第 10 号，1985

[13] Yahoo! JAPAN，*http://www.yahoo.co.jp*