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

2004－ＳＬＰ－51　（８）
2004／5／28

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

松下　雅彦†　　中川　聖一†　　西崎　博光††　　宇津呂武仁†††

† 豊橋技術科学大学 情報工学系，〒441-8580 愛知県豊橋市天伯町雲雀ケ丘 1-1
†† 山梨大学 大学院医学工学総合研究部，〒400-8511 山梨県甲府市武田 4-3-11
††† 京都大学 大学院情報学研究科，〒606-8501 京都市左京区吉田本町

E-mail: †{masahiko,nakagawa}@slp.ics.tut.ac.jp, ††hnishi@yamanashi.ac.jp,
†††utsuro@pine.kuee.kyoto-u.ac.jp

あらまし　近年，ウェブ検索の分野では，NTCIR ワークショップ等で競争型のコンテストが行なわれるなど，研究が盛んにおこなわれている．我々は，以前，Web 検索の分野の中でも音声入力による Web 検索の有用性に着目し，NTCIR-3 音声入力 Web 検索タスクにおいて，クエリーの音声認識率を改善することで検索精度を改善する方法を提案した．本稿では，さらにクエリーの音声認識率の改善を図るために，音声認識で用いる認識辞書の語彙サイズを 2 万語から 6 万語へ拡大することを試み，その検索実験について報告する．実験結果より，Web 文書の検索性能を増加させるには，音声クエリーを認識する際の辞書の語彙サイズを増加させ，未知語を減少させることで認識率を改善することが大変有効であることが分かった．また，以前提案した SVM による複数の認識モデルの出力の混合も語彙サイズに関わらず有効であることを示せた．
キーワード　NTCIR3，音声 WEB 検索，複数認識モデル，SVM

Improvement of Keyword Recognition and Extraction for Speech-driven Web Retrieval Task

Masahiko MATSUSHITA†,　Seiichi NAKAGAWA†,　Hiromitsu NISHIZAKI††, and Takehito UTSURO†††

† Department of Information and Computer Sciences, Toyohashi University of Technology
1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, 441-8580 Japan
†† Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi
4-3-11 Takeda, Kofu, Yamanashi 400-8511 Japan
††† Graduate School of Informatics, Kyoto University,
Yoshidahonmachi, Sakyo-ku, Kyoto, 606-8501 Japan

E-mail: †{masahiko,nakagawa}@slp.ics.tut.ac.jp, ††hnishi@yamanashi.ac.jp,
†††utsuro@pine.kuee.kyoto-u.ac.jp

Abstract　This paper descibes speech driven Web retrieval model which accepts spoken search topics (queries) in the NTCIR-3 Web retrieval task. The major focus of this paper is on improving speech recognition accuracy of spoken queries by using multiple LVCSR models with larger vocabulary size and then improving retrieval accuracy in speech-driven Web retrieval. We experimentally evaluate the technique of combining outputs of multiple LVCSR models with a 60,000 vocabulary size in recognition of spoken queries. As a model combination technique, we use the SVM learning. We show that the technique of multiple LVCSR model combination can achieve improvement both in speech recognition and retrieval accuracies in speech-driven text retrieval. Comparing with the retrieval accuracies when a LM with a 20,000/60,000 vocabulary size is used in LVCSRs, the LM with a larger size of the vocabulary also improves retrieval accuracies.
Key words　NTCIR3, Speech driven Web retrieval task, Multiple LVCSR models, SVM

1. はじめに

近年の音声認識技術は非常に進展してきており，その重要性も増してきている．実際に，一般のパソコン上でも動作する音声認識ソフトウェアも増えつつあり，音声入力を用いたアプリケーションも徐々に出はじめている．一方，情報検索の研究も非常に盛んに行なわれており，この2つを組み合わせた研究も最近になり増えつつある．

音声データを用いた検索には，検索対象が音声データである場合と，質問を音声で入力する場合 (音声クエリー) が考えられる．音声データの検索としては，TREC-6 の Spoken Document Retrieval(SDR) トラック [1] で音声データを対象にしたテストコレクションが整備されたことにより盛んに研究が行なわれ始めた．音声クエリー入力による検索としては，Barnett ら [2] が既存の音声認識システムを用いてテキスト検索を行なっている．Crestani [3] も音声入力による検索を行ない，適合性フィードバックによって検索精度が向上することを示している．上記の2つの手法は検索精度の改善についてのみ焦点を当てており，認識精度の改善については触れていない．

音声クエリーの音声認識を改善することにより，検索性能を向上させようと試みる研究も行なわれている．伊藤，藤井ら [4] [5] は NTCIR-3 の Web 検索タスクで，音声入力による検索というサブタスクを提案し，音声認識と検索精度の両方の改善を行なった．彼らは，検索対象から言語モデルを作成し，それにより未知語を減少させ，認識精度の改善を行っている．また我々は音声キーワードと音声データベースを用いて検索実験を行っている．ここでは，グルーピングと呼ばれる手法と N ベスト出力の結果を用いて，音声認識誤りに対して頑健に対処している．また，音声クエリーとして，キーワード入力法と自然文入力法を比較し，後者の方が高い音声認識精度が得られることを示している [6].

我々も同様に NTCIR-3 の音声入力 Web 検索タスクにおいて，音声クエリーの音声認識率を改善する方法として，複数の音声認識モデル出力の混合 [7] を用いる方法を提案した [8]．混合手法として機械学習法である SVM [9] を利用することにより，クエリーの音声認識率の改善が得られ，その結果検索精度を大幅に改善することができた．しかし，音声認識で用いる辞書の語彙サイズが2万語彙に限定されていたということもあり，クエリー中に含まれる検索キーワードの未知語率が高く，伊藤，藤井ら [4] [5] の結果には及ばなかった．本稿では，クエリーを認識する際の語彙サイズを6万語彙に増加し再度検索実験を行ったので，その結果について報告する．

2. NTCIR-3 音声入力 Web 検索タスク

2.1 NTCIR ワークショップ

NTCIR ワークショップ [10] は，情報検索とテキスト要約・情報抽出などのテキスト処理技術の研究をより発展させることを目的とした評価会議である．実験用のデータセットと実験結果を評価するための統一された手順が用意されており，参加グループは用意されたデータを用い，様々なアプローチで研究と実験を行なっている．また，NTCIR では，「テストコレクション」と呼ばれる実験用データセットを整備・公開している．テストコレクションとは，情報検索分野で情報検索システムの性能評価に用いるデータである．

2.2 音声入力 Web 検索タスク

NTCIR-3 の音声入力 Web 検索タスク [5] は筑波大の藤井らがオーガナイザーとなり始められたタスクである．NTCIR 側で用意された文書集合から，音声の検索質問 (query) を用いて検索実験を行なう．実際の実験に用いられているテストコレクションは NTCIR の Web 検索タスクに用いられているものと同じである．検索文書集合は参加者が扱いやすいサイズ (10GB 程度) と，ある程度現実に近いサイズ (100GB 程度) が設定されている．各検索課題ごとに適合度順の検索結果の上位 1000 ページを順位付きで提出し，それを人手で判定するプーリングと呼ばれる手法を用いて正解候補を作成する．ただし，このような便宜的な方法の場合は，正解が一部に偏よる場合があり，実際には正解であっても不正解であると判定する可能性もある．この時，正解候補は4段階 (高度に適合，適合，部分適合，不適合) で判定が行なわれている．

3. 複数音声認識モデル出力の混合

今回の実験も音声入力 Web 検索タスクの音声認識に着目して実験を行なった．音声認識率を向上させることにより検索精度も同時に向上させることを目的としている．そこで，音声認識率を向上させるために，本稿では複数の大語彙連続音声認識モデルの出力を混合する手法 [7] を適用する．前回の実験により，複数モデルの混合法には SVM（サポートベクターマシン）が有効であることが判ったので，今回も同様に機械学習法である SVM を用いることにした．SVM は，一つ一つの属性を次元と見なし，単語の持つ属性の組み合わせをベクトルと考えることにより，2クラスの分類を行なうための座標変換式を学習する手法である．混合手法を用いた場合の検索手順は図1の通りである．まず，複数の音声認識システムを用いてユーザーが発話した検索課題音声の認識を行なう．次に，出力された複数の認識結果を用いて混合処理を行なう．その結果より，不要語を除去し，キーワードリストを作成する．最終的に，作成されたキーワードリストを用いて文書集合より検索を行なう．

3.1 大語彙連続音声認識モデル

3.1.1 デコーダ

大語彙連続音声認識モデルのデコーダとしては次の2種類を用いる．1つ目としては，我々の研究室 (豊橋技術科学大学 中川研究室) で開発した SPOJUS [11] であり，もう1つは CSRC「日本ディクテーション基本ソフトウェアの開発」プロジェクト [12] から提供された Julius(ver.3.2) である．Julius，SPOJUS のどちらのデコーダにおいても2パス探索により認識を行ない，1パス目では単語バイグラム，2パス目では単語トライグラムを，それぞれ使用した．

(a)Julius

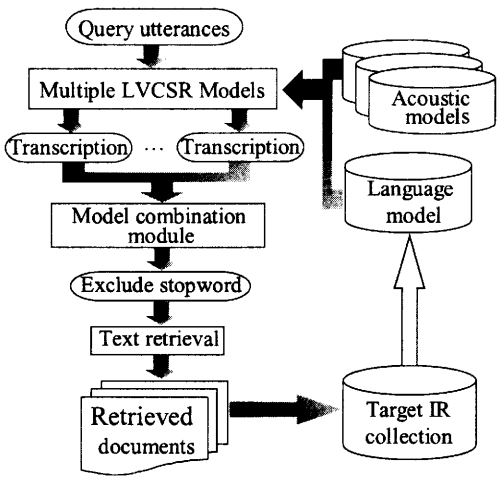


図1　複数音声認識モデル混合を用いたWeb検索

Juliusは，1パス目ではバイグラムを用いた1best近似を行ない，1パス目で得られた単語トレリスを用いて，2パス目で単語間にわたるクロストライフォンを用いている．

(b)SPOJUS

SPOJUSは，1パス目ではバイグラムを用いた1best近似(改良版は1-bestと線形辞書の併用(注1))を行ない200ベストを求める．本実験においては，2パス目でコンテキスト依存モデルは用いていない．

3.1.2　音響モデル

音響モデルの学習には，日本音響学会読み上げ音声JNASを用いた．

(a)Julius

Juliusでは音素を基本単位とするHMMモデル，および，音節を基本単位とするHMMモデルを用いた．16KHzサンプリング，フレーム周期10ms，特徴ベクトルはMFCC(12次元) + ΔMFCC + ΔPOW (計25次元) の条件で行ない，HMMとしては次の4種類を用いて実験を行なう．

- モノフォンモデル
- トライフォンモデ
- 音素内タイドミクスチャ(PTM) モデル
- 音節モデル[14]

(b)SPOJUS

SPOJUSでは音節を基本単位とするHMMを用いており，デコーダと同様に，我々の研究室で開発されたものである．16kHzサンプリング，フレーム周期10ms，25msハミング窓，特徴ベクトルはMFCC(セグメント単位): MFCC(10次元 × 4フレームをKL展開で20次元に圧縮) + ΔCEP + ΔΔCEP + ΔPOW + ΔΔPOW (計50次元)，およびMFCC(フレーム単位): MFCC(12次元) + ΔMFCC + ΔΔMFCC + ΔPOW + ΔΔPOW (計38次元) の2種類，さらに，継続時間制御/自己遷移ループの2種類，合計以下の4種類のモデルで認識を行なう．

(注1)：北岡ら[13]によるSPOJUSの改良版

- MFCC-seg + 継続時間制御
- MFCC-frm + 継続時間制御
- MFCC-seg + 自己遷移ループ
- MFCC-frm + 自己遷移ループ

3.1.3　言語モデル

言語モデルは藤井ら[5]と同じモデルを利用した．この言語モデルは，検索対象の文書集合である100GBのコレクションを用いて作成された．このコレクション中から頻度上位6万語に制限した単語トライグラムおよびバイグラムを作成した．言語モデルの作成方法は，大語彙連続音声認識ツールキットでの作成方法[15]に準じているが，Web文書の場合，英語の文書なども混ざっているので，ASCII文字だけからなる段落などを排除する前処理を追加した．平滑化には，バックオフスムージング (Witten-Bell ディスカウント) を用いた．カットオフはバイグラム，トライグラム共々20とした．バイグラムは前向きのもの，トライグラムは前向きのものと逆向きのものの両方をARPA形式で作成した．

3.2　評価データ

検索対象はNTCIR-3[10]で用意された100GB(1000万ページ) 程度のWeb文書である．Web検索タスクでの検索課題105文を合計10人の話者 (男性5人，女性5人) がそれぞれ発話したものが音声クエリーとして準備されている．ただし，今回の我々の実験は男性5人の音声クエリーだけを用いて行なっている．

評価データとしては検索課題105文中から53文を用いる．ただし，実際に評価として用いているのは正解が公表されている47文のみである．機械学習で用いる学習データは，評価として用いられない検索課題 (52文) を，評価者以外の4人分 (合計208文) を合わせたものを用いる．検索課題の例を以下に示す．

0008　サルサを踊れるようになる方法が知りたい。
0016　ゲノム創薬の最近の動向について述べられている文書を探したい。
0146　ドメスティックヴァイオレンス、DVの現状について調べたい。

先頭の数字は，検索課題番号を示している．認識結果を評価するための正解データとしては，検索課題を書き起こしたテキストを用いている．キーワードリストの評価用の正解データは，認識結果の出力と混合処理結果の出力からキーワードリストを作成する手順と同じ手順を用いて，検索課題の書き起こし結果から作成する．不要語の処理などでヒューリスティックスを用いているため，正解としているキーワードが必ず検索において適切であるとは限らない．

3.3　複数モデルの混合手法

本稿では，先に述べた8種類の大語彙音声認識モデルの出力を混合し，単語認識率を向上させる．

3.3.1　SVMを用いた混合

SVMは，一つ一つの属性を次元とみなし，単語の持つ属性の組み合わせをベクトルと考えることにより，2クラスの分類を行うための座標変換式を学習する．複数の大語彙連続

音声認識モデルの出力を混合する手法 [7] では，クラス分類時に計算される 2 クラスの境界からの距離を信頼度とすることで，音声認識性能の向上を図っている．ベクトルの素性としては次の 3 つを用い，各単語の正誤を判別対象のクラスとする．

- 単語の品詞
- 音節数
- 単語を認識・出力したモデルの情報

SVM では，個々の素性を次元とする多次元空間中の点として記述された事例の集合に対し，全事例を 2 クラスに分類する境界面を学習し，適用時は評価事例と境界面との間の距離 (信頼度) に閾値を設け，単語ラティス上の競合する単語中で，これらの閾値を越える距離を持つ単語が存在すれば，その中で境界面からの距離が最も大きい単語を 1 つ混合結果として出力する．もし，閾値を越える距離を持つ単語が存在しない場合は，そのラティス上では単語は出力されない．

3.3.2 SVM 冗長を用いた混合

SVM 冗長は，正解精度を犠牲にして正解率を上昇させることを目的とした手法である．学習は SVM と同じ方法で行う．適用を行なう場合は SVM とは異なり，単語ラティス上の競合する単語中で閾値を越える距離を持つ単語が存在する場合は，閾値を越える距離を持つ単語はすべて出力する．この手法では，単語ラティス上で競合する単語があったときに複数個以上の単語を出力できる可能性があり，SVM の場合より正解率を向上させることができる．

4. Web 検索

4.1 キーワードリストの作成

キーワードリストを作成するために，混合処理などにより得られた結果から内容語 (主に自立語) を抽出し，さらにそこから不要語を取り除いた．不要語は，ヒューリスティックスにより選択し，検索時によく用いられる単語 (探す，知り，など) と，検索には必要ないと思われる単語 (平仮名 1 文字) を削除した．この処理によって出力された単語を検索のためのキーワードとする．

4.2 検索エンジン

キーワードとして抽出された単語を検索エンジンに入力し，検索を行なう．今回実験に用いた検索エンジンは，藤井ら [5] の作成した統計的手法を用いたものである．クエリー Q とテキスト D_i の類似度 $sim(Q, D_i)$ は次式で計算される．

$$sim(Q, D_i) = \sum_t \left(\frac{TF_{t,i}}{\frac{DL_i}{avglen} + TF_{t,i}} \cdot log \frac{N}{DF_t} \right)$$

ただし，t は検索要求に含まれる索引語，$TF_{t,i}$ はテキスト i 中の索引語 t の出現頻度，DF_t は索引語 t を含む検索対象テキストの数であり，N は検索対象のテキスト総数，DL_i はテキスト i の文書長 (バイト数)，$avglen$ は検索対象中の全テキストに関する平均長である．

検索エンジンにキーワードを入力すると，検索対象である文書集合 (100GB) の各ページに対して上式を用いて類似度を計算し，スコアが高いページから順次列挙する．このエンジンは，キーワードの挿入誤りには比較的頑健であることから，単語正解率 (correct) が検索性能に直接影響を及ぼすと考えられる (検索エンジンよっては，挿入誤りは致命的になる場合もある)．

4.3 検索精度の評価方法

正解判定は 4 段階 (高度に適合，適合，部分適合，不適合) で行なう．実際に評価を行なう場合は，高度に適合と適合の判定は混合して考えている．その他に，ハイパーリンク情報の利用の有無も考慮するため，最終的には以下の 4 つで評価を行なう．

- RC : (高) 適合，ハイパーリンク情報を用いない．
- RL : (高) 適合，ハイパーリンク情報を用いる．
- PC : 部分適合，ハイパーリンク情報を用いない．
- PL : 部分適合，ハイパーリンク情報を用いる．

検索精度は平均適合率を用いて計算を行なっており，各クエリーの検索結果上位 1000 件から計算する．平均適合率は，各再現率レベルでの適合率の平均値 (適合文書が検索された時点での適合率の平均) である．もし，正解文書が検索されなかった場合は適合率は 0.0%である．最終的に，話者 5 人の平均を求めて評価する．

5. 実験結果

5.1 音声クエリーの認識精度

表 1 は評価用検索課題 47 文を，前述の 8 種類の認識モデルで認識を行った際の認識率を示している．各デコーダ毎，および言語モデルの語彙サイズ毎に，認識率が最大／最小であった結果のみを示す．2 万語の認識辞書を用いるよりも，6 万語の辞書を用いた方が全体的に 2〜6%程度の改善がみられている．これは，表 2 に示すように 47 の評価文に対する未知語率の改善が大きく寄与している．語彙サイズを 2 万語から 6 万語にすることにより，ストップワードを含むすべての単語での未知語率で 3.5% の改善が，また，検索キーワードに限れば 9.9% の改善が得られている．助詞，助動詞などの機能語はどの文章にも比較的現われやすいため，語彙サイズの増加の影響はさほど大きくないが，検索のキーワードになりうる名詞などの単語は，語彙サイズを増加することでそれらが未知語となってしまう数を大幅に減らすことができる．

図 2，図 3 は，提案した複数モデルの出力を SVM を用いて混合した場合の認識結果である．男性 5 名が発話した 47 文の音声クエリーの単語認識率およびキーワード認識率の平均を示している．括弧内の数値は言語モデルの語彙サイズを表している．“Julius”，“SPOJUS” は単独のデコーダでの最大の単語正解率を持つ認識モデルの結果である (Julius ではトライフォンモデル，SPOJUS では MFCC-seg + 継続時間制御)．“SVM” は最も正解精度が高くなるように複数の認識モデルの出力を混合したもの，また”SVM 冗長” は，複数の認識モデルから出力された単語候補を一つだけ選ぶのではなく複数個選んだ場合の結果である．当然，正解精度は下がる．図 3 で示すキーワード認識率は音声認識結果から不要語を除去した出力の認識率である．

これらの図より，語彙サイズの増加により大幅な認識率の

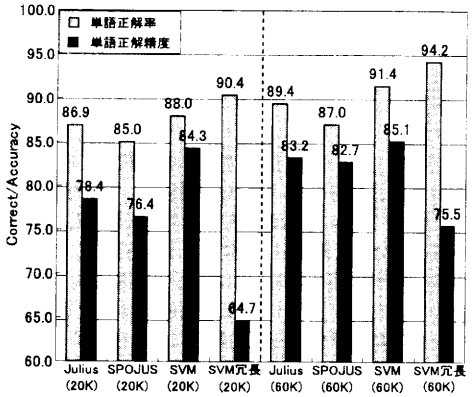


図 2　評価用検索課題 47 文の単語の正解率・認識精度

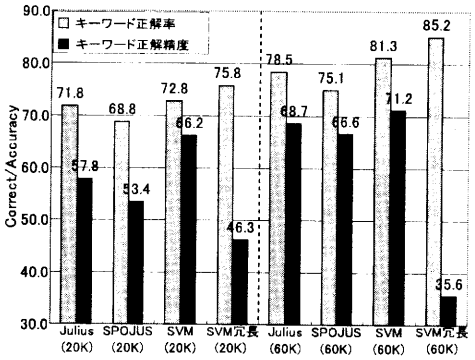


図 3　評価用検索課題 47 文のキーワードの正解率・認識精度

改善が得られた．特に検索キーワードの認識率は，未知語率の改善と同様に大幅に良くなっており，SVM を用いた混合手法で 10%近く改善していることがわかる．また，語彙サイズの大きさに関わらず混合処理を行なった場合には，単独の認識システムの場合と比較して正解率と正解精度の改善が達成されていることがわかる．特に，SVM を用いた場合の認識精度の改善は非常に大きいといえる．SVM 冗長と SVM を用いた場合を比較してみると，正解精度を犠牲にして，正解率の改善を達成できていることがわかる．

表 1　認識モデルの音声認識率括．弧内数値は語彙のサイズ

デコーダ	単語正解率 [%]	単語正解精度 [%]
Julius(20K)	86.9(max)～73.1(min)	78.4(max)～66.9(min)
SPOJUS(20K)	85.0(max)～81.8(min)	76.5(max)～75.0(min)
Julius(60K)	89.4(max)～76.7(min)	83.2(max)～72.6(min)
SPOJUS(60K)	87.5(max)～86.2(min)	82.7(max)～81.5(min)

表 2　語彙サイズの違い（20K，60K）による未知語率の変化

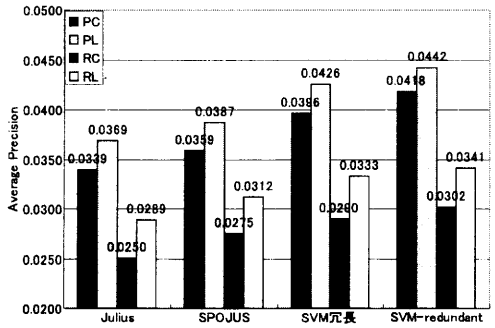
	語彙サイズ	未知語率 [%]
全単語	20,000	4.5
	60,000	1.0
キーワードのみ	20,000	12.7
	60,000	2.8

5.2 検索精度

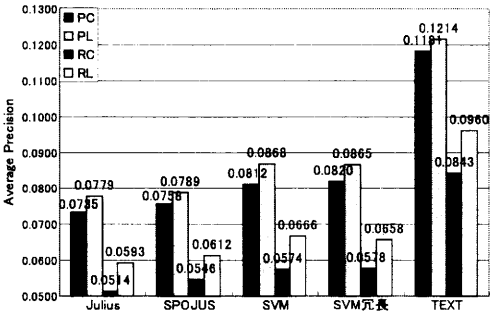
図 4 に音声入力による Web 検索実験の結果を示す．(a) のグラフが語彙サイズが 2 万語，(b) のグラフが語彙サイズ 6 万語で音声クエリーを認識した時の結果である．Julius と SPOJUS はそれぞれのデコーダ中で，最も単語正解率が良かった場合の結果を示している．語彙サイズに関わらずキーワードの認識率が高かった Julius より，低い SPOJUS の方がより高い検索精度を示した．これは，検索エンジンはキーワードの重みを考慮しているため，キーワードの認識率と検索精度が線形な関係でないことによる．したがって，SPOJUS では重みが大きいキーワードがより多く認識されていたと考えられる．SVM を用いた場合の結果は，単独の認識モデルを用いた場合に比べて非常に高い精度を示している．SVM と SVM 冗長を比較した場合，(a) の方では正解率を重視した SVM 冗長の場合のほうがより良い結果を得ることができたが，(b) の方では一概にはそうは言えない．2 万語の場合は，SVM と SVM 冗長とを比べて，キーワードに関して 3%の正解率の改善，挿入誤り 19.9%の増加であるが，6 万語の 3%場合は，3.9%の正解率改善に比べて挿入誤りは 35.6%も増加している．これは，音声入力 Web 検索においては認識精度より正解率が重要な要素であることを示している一方で，本来のクエリーと関係のないキーワードがあまりにも多く挿入されると，若干検索性能を悪くする可能性があると言える．つまり，単語正解率と単語正解精度にはトレードオフが存在する．

次に，音声クエリーを認識する際に用いた辞書の語彙サイズの違い（2 万語と 6 万語）について比較してみると（(a) と (b)），明らかに 6 万語の辞書を用いて書き起こしたクエリーを利用した方が検索結果が良いことがわかる．図 3 に示すように，2 万と 6 万の語彙の差により，キーワードの認識率に約 6～10%の違いが生じているが，検索の精度は，それ以上の改善が得られており，約 2 倍の改善が見られている．複数のキーワードの集合で一つのクエリーを構成するわけであるが，そのクエリーを構成しているキーワードが一つでも欠けてしまったり，他の単語に置き換わってしまうと，本当に必要な文書とは異なる文書を検索してくることになりかねない．この比較からも，クエリー中に含まれる単語を出来るだけ認識辞書でカバーすることは重要であることが言える．図 3 (b) には，テキストクエリー（つまり音声認識が 100%正しく出来た場合）を用いた検索実験結果も示している．6 万語の SVM の検索性能よりもさらに 50%以上の高い改善率が得られている．これに対して音声クエリー（SVM 冗長）は，必要なキーワードの 15%が脱落し，本来のクエリーと無関係なキーワードがクエリーの半分を占めているため，テキストクエリーに比べて検索性能が悪くなっている．

最後に，この実験結果と，同じタスクの実験を行っている藤井ら [4] の実験結果と比較してみる．藤井らは，Julius 単体の認識結果（我々と同じ音響モデル・言語モデル（60K）），我々と同じ検索エンジンを用いて実験を行っているが，若干



(a) 語彙サイズ (20K)



(b) 語彙サイズ (60K)

図 4 検索精度評価

検索性能が異なっている．これは，我々の実験は 5 名の男性話者が発話した合計 235 発話 (47 クエリー ×5) を評価に用いているのに対し，藤井らは男女 10 名が発話した音声クエリー計 470 発話を利用しているためである．我々と藤井らの実験結果で一番ベストな結果を比較してみると(注2)，女性の音声クエリーの認識率の方が男性のものよりも良いという事実にも関わらず，我々の複数の認識モデルを利用した混合を利用した結果の方が良くなっており，SVM による混合法の有効性が示せた．

6. おわりに

本稿では，音声入力 Web 検索タスクにおいて，クエリーの音声認識率 (特に単語正解率) を改善し Web 検索性能を向上させるため，語彙サイズを増加することに焦点を当てた実験を行った．Web 文書の検索性能を増加させるには，音声クエリーを認識する際の辞書の語彙サイズを増加させることにより，クエリーを構成するキーワードの未知語率を大幅に改善し，それに伴って認識率を改善することが大変有効であることが分かった．また，以前提案した SVM による複数の認識モデルの出力の混合も語彙サイズに関わらず有効であることが示せた．

今後は，混合時に挿入単語が多すぎる場合に，検索精度を低下させる原因になるため，距離閾値や競合単語の制限などで，キーワードとなる単語の数のバランスを取るなどの対策を講じた実験を行う予定である．

(注2)：藤井らの実験結果[5] は，PC=0.0676，PL=0.717，RC=0.474，RL=0.552 となっている．

謝　辞

この研究では，NTCIR-3 のデータを数多く利用させていただいた．これらのデータベース提供してくださった NTCIR ワークショップの関係諸氏に深く感謝致します．また，本研究を進めるにあたって適切な助言を頂いた，筑波大学 図書館情報学科の藤井敦助教授に深く感謝致します．また，6 万語の順方向きトライグラムを始め言語モデルを提供して頂いた名古屋大学伊藤克亘助教授に深く感謝致します．

文　献

[1] J. S. Garofolo, E. M. voorhees, V. M. Stanford, and K. S. Jones. Trec-6 1997 spoken document retrieval track overview and results. In *In Proceedings of the 6th Text Retrieval Conference*, pp. 83–91, 1997.

[2] J. Barnett, S. Anderson, J. Broglio, M. Singh, R. Hudson, and S. W. Kuo. Experiments in spoken queries for document retrieval. In *Eurospeech97*, pp. 1323–1326, 1997.

[3] F. Crestani. Word recognition errors and relevance feedback in spoken query processing. In *Fourht International Conference on Flexible Query Answering Systems*, pp. 267–281, 2000.

[4] Atsushi Fujii and Katunobu itou. Building a test collection for speech-driven web retrieval. In *Eurospeech2003*, pp. 1153–1156, 2003.

[5] 伊藤, 藤井. NTCIR-3 ワークショップにおける音声入力型ウェブ検索タスク. 情報処理学会研究報告, Vol. 2002, No. (2002-SLP-43), pp. 25–32, 2002.

[6] 西崎, 中川. 音声キーワードによるニュース音声データベースの検索手法. 情報処理学会論文誌, Vol. 42, No. 12, pp. 3173–3184, 2001.

[7] 小玉康広, 渡邊友裕, 宇津呂武仁, 西崎博光, 中川聖一. 機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合. 情報処理学会研究報告, Vol. 2003, No. (2003-SLP-45), pp. 95–100, 2003.

[8] 松下雅彦, 西崎博光, 宇津呂武仁, 中川聖一. 音声入力による Web 検索のためのキーワード認識・抽出法の検討. 情報処理学会研究報告, Vol. 2003, No. (2003-SLP-48), pp. 21–28, 2003.

[9] T. Joachims. Making large-scale svm learining practical. In *Advances in Kernel Methods Support Vector Learning*. MIT Press, 1999.

[10] K. Eguchi, K. Oyama, E. Ishida, and K. Kuriyama. Overview of Web retrieval task at the third NTCIR workshop. In *Working Notes of the 3rd NTCIR Workshop Meeting*, pp. 1–24, 2002.

[11] 赤松, 花井, 甲斐, 峯松, 中川. 新聞・ニュース文をタスクとした大語彙連続音声認識システムの評価. 情処第 57 回全大講論集, pp. 35–36, 1998.

[12] 河原, ほか. 日本語ディクテーション基本ソフトウェア (99 年度版). 日本音響学会誌 (技術報告), Vol. 57, No. 3, pp. 210–214, 2001.

[13] 北岡, 高橋, 中川. N-best 線形辞書探索と 1-best 近似木構造辞書探索の併用による大語彙音声認識. 電子情報通信学会技術報告, SP2003-26, 2003.

[14] 山本, 池田, 松本, 西谷, 宮澤. コンパクトで高精度な音節モデルの検討. 日本音響学会秋期講演集, Vol. 2002, No. (1-9-22), 2002.

[15] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄 (編). 音声認識システム. IT Text. オーム社, May 2001.