

孤立単語認識と連続基本単語認識の併用に基づく 組織名の音声入力インタフェース

北岡 教英[†] 押川 洋徳[†] 中川 聖一[†]

[†] 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ヶ丘 1-1
E-mail: †{kitaoka,oshikawa,nakagawa}@slp.ics.tut.ac.jp

あらまし 本稿では、高頻度単語と短い単語(基本単語)を併用した音声認識を用いた組織名入力インタフェースを提案する。これは、まず音声で組織名を入力し、音声認識の結果得られた複数の単語・基本単語候補から、ペンタッチで選択、入力するマルチモーダルインタフェースである。組織名といった語彙サイズが大きく、常に新しい組織名が生み出されるため、すべてを登録することが難しく、また音声認識が難しいタスクに対し、単語認識と連続基本単語認識を併用するもので、認識対象のカバー率と認識性能、入力効率の両方の向上を図った。そして、その認識結果から単語・基本単語系列候補、基本単語候補をタッチパネルに表示し、ペンタッチにより簡単に選択して入力が可能な組織名入力インタフェースを考案した。この高頻度単語と基本単語を併用した音声認識をオフラインの認識実験により評価したところ、それぞれ単独での音声認識結果より良い結果が得られた。さらにこの認識結果に基づいてインタフェースを用いた場合の入力可能な割合をシミュレーションすると、約92%で入力が可能となることが分かった。また、このインタフェースを実装してオンラインで被験者実験を行ったところ、音声認識性能の低下で1回の発声では83.3%の入力可能率となったが、2回まで音声入力を許すことで93.3%となった。

キーワード 組織名入力, 音声認識, 孤立単語認識, 連続基本単語認識

Speech interface for organization name input based on combination of isolated word recognition and continuous base-word recognition

Norihide KITAOKA[†], Hironori OSHIKAWA[†], and Seiichi NAKAGAWA[†]

[†] Department of Information and Computer Sciences, Toyohashi University of Technology 1-1 Hibarigaoka,
Tempaku-cho, Toyohashi, Aichi, 441-8580 Japan
E-mail: †{kitaoka,oshikawa,nakagawa}@slp.ics.tut.ac.jp

Abstract In this report, we propose a multimodal interface for organization name input based on combinational usage of isolated word recognition for high frequency names and base-word recognition for low frequency names. Using this interface, the user first utters a organization name and then the system displays multiple word hypotheses and base-word lattice. Then the user selects a word or base-words to construct the name he/she intends to input. There are a huge number of organization names and new names are continuously generated, so the system cannot prepare all the names *a priori*. We addressed this problem to obtain both high coverage and high recognition performance by combining isolated word recognition and continuous base-word recognition. We proposed a multimodal interface to select a whole word or base-words to construct the word the user wanted to input by touch pen. We evaluated this combinational recognition method by off-line experiment and obtained the result better than both isolated word recognition only and continuous base-word recognition only. The simulation result of the system was about 92%. We also evaluated the performance of on-line system and revealed that the user can input a word by 93.3% by at most two trials of the utterances despite the recognition performance degradation.

Key words organization name input, speech recognition, isolated word recognition, continuous base-word recognition

1. はじめに

近年、情報アクセスの手段として、ウェブブラウザがよく用いられる。特に検索エンジンや各種 CGI を用いた情報検索や登録によく用いられるようになってきている。一般に、このようなウェブブラウザの操作は、マウスやキーボードによって操作可能な Graphical User Interface (GUI) が用いられる。これらは主に PC 上で使用することを前提に発達してきた。近年では、インターネットの利用が拡大し、携帯や PDA (携帯情報端末機器) などからもアクセス可能となっているが [12], [13], これらの環境ではマウスやキーボードは不向きである。そこで、キーボードなどの入力手段に代わるインタフェースが必要となる。

このことからウェブブラウザへの音声インタフェースの導入がいくつか検討されてきた [1]~[5], [7], [9], [10]。また、モバイル機器への応用として PDA 上で動作する音声入力インタフェースに関する研究も行われてきた [11]~[13]。

Johnston ら [11] は、PDA 上に表示された地図に対してペンタッチによる位置指定と音声による条件指定の組み合わせなどにより、地図上の情報検索を行うマルチモーダルインタフェースを提供している。

Wang [12] は、PDA への 7 つの入力項目に対し、項目選択はペンで、内容入力は音声で行うインタフェースを提案している。その際、選択された項目に対応した語彙と文法を認識に用いている。

WWW(World Wide Web) において任意の語句を入力できるフォーム入力は、検索や登録などに多く使用され重要であり、これを音声で行うための研究も行われている [18]。中でも、姓名や住所、組織名などを入力する場面は多い。しかし任意文字列や姓名を入力タスクとした場合の音声認識においては、その種類の多さや計算量等により全ての単語や姓、名を登録し単語認識させることは現実的でない [14]。

谷垣ら [14] は、日本人の姓・名データの分析により、未登録語を含む音声認識のための新たな言語モデルを提案している。言語モデルは単語のクラス N-gram と、未登録語認識用のサブワードモデルから成り、クラス N-gram は入力文中の固有名詞 (日本人姓・名) 部分の検出に使用され、サブワードモデルは固有名詞の認識に使用される。この言語モデルを用いることにより、未登録語 (事前に予測できない希有な姓・単語) についてもサブワードモデルの認識結果 (クラスラベル付きの読み) により正しく認識できる可能性がある。これにより旅行会話文中の姓・名の認識実験では、単語登録方式で 73% の認識率に対して、未登録語認識方式で 70% の認識率が得られた。

大森ら [15] は、16 万種の姓入力インタフェースとして、思い込み応答戦略を用いる方法を提案している。これは、頻出上位の姓の聞き誤り先は、やはり頻出上位の姓になりやすいという被験者実験の結果を利用したもので、最初は頻出上位 1 万種の姓で認識し、誤認識であれば、残りの 15 万種の姓で認識する手法である。しかし、認識の困難さを解消しようとしたものではない。そこで、誤認識の場合、属性 (例えば「かとう (加藤)」の場合、「3 文字」や、「か」、「くわえる」など) の入力を要求す

る方法を提案している [16]。

英語において、スペル入力を用いることにより、姓名の認識性能を向上させる方法が試みられている [17]。スペル入力は通常の発声よりも認識率は向上する。しかし、スペル入力のみへの入力は直感的に不自然である。そこで通常の発声とスペル入力を組み合わせる方法が検討されている。[18]~[22]

スペル入力は、日本語の場合においては、ローマ字表現のアルファベット読みによる入力や音節毎に区切って発声する音節単位入力に相当する。しかし、この方法は、ユーザの負担が大きくなり入力方法として好ましくない。

我々は、これまで、単語と音節 N-gram を併用した音声認識により、姓、名の認識を行い、単語候補、音節系列候補、音節ラティス候補をタッチパネルに表示しペンタッチなどにより選択入力するインタフェースによって、任意文字列が入力可能となる割合を向上させる方法をこれまで検討してきた。その結果、姓で約 93%、名で約 95% で入力が可能となった [23]。

そこで、本論文では姓名と同様に語彙数も多く、常に新しい組織が増えるため、新出語もよく発声する組織名の音声入力法を検討した。しかし、組織名は、姓名と違い非常に長い音節列で構成される場合が多い。そのため音節系列による入力では、音節の選択が多くなり難しくなる。一般に長い組織名は複数の単語から構成される場合がある。そこで、姓名の時に使用した音節列を基本単語列に変えることによって組織名の入力を可能とする。これにより、登録されていない単語についても基本単語列によって認識できる。図 1 に、登録組織名の数、基本単語数の数による 13 万種類の日本語組織名のカバー率を示す。組織名の頻度は新聞記事から求めた。基本単語列により多くの組織名がカバーされる。たとえば登録組織名の数を 5000、10000 とした場合にはカバー率は 92.2% および 96.0% であるが、基本単語数 5000 では 95.8% となる。さらに、例えば基本単語 5000 語と登録組織名 10000 語を組み合わせると 98.4% となるなど、高頻度組織名と基本単語を併用することにより、それぞれ単独で動かしただけよりも良い結果が得られることが期待できる (詳細は表 1 参照)。加えて、基本単語ラティス結果が得られることから、この結果を利用したインタフェースも考える。

2. 組織名の認識方法

2.1 基本単語による音声認識

組織名は姓名同様、種類が非常に多く新出語もあるため、全てを単語として登録し孤立単語認識を行うのは、計算量の上からも、網羅性からも適切でない。また、組織名においては複数の単語によって構成されることも多く、登録した組織名自体がほかの単語の一部となっている場合がある。そこで組織名を構成する短い単語 (基本単語と呼ぶ) を認識単位とした One-pass Viterbi 法により連続基本単語認識を用いることを考える。連続基本単語認識を用いる際、何の制約もなく自由に基本単語の接続を許す必要はなく、組織名の言語的特徴を導入することができる。組織名には基本単語の接続に特徴があるため、基本単語単位の統計的言語モデル (N-gram) を適用する。

今回使用した言語モデルは、13 万の機関名に新聞記事データ

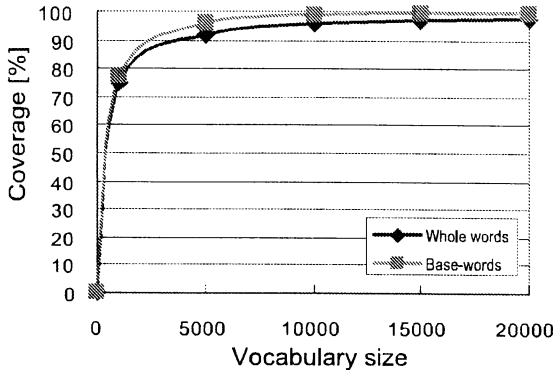


図1 上位N頻度単語もしくは基本単語を登録した場合の13万組織名のカバー率

Fig. 1 Coverage of 130,000 organization names by a lexicon of N-frequent whole words and concatenation of N-frequent base-words.

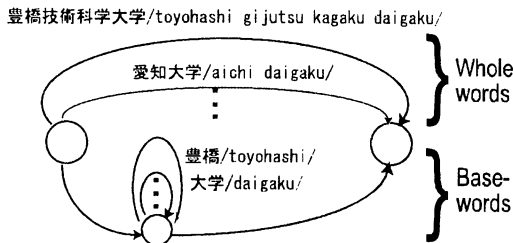


図2 言語モデル例
Fig. 2 Language model

を用いて作成されている。基本単語の連結に関する言語モデルは漢字の読みの違いを考慮した約13万種類の組織名データ[26]に対して、読売新聞記事データより頻度を求めた組織名データを用いてbi-gramを作成した。

2.2 高頻度組織名と基本単語併用時の音声認識

提案する組織名認識では、高頻度組織名の孤立単語認識を基本とする。これに、前節の連続基本単語認識を併用する。その概念を図2に示す。これにより登録した高頻度組織名に加え、基本単語による連続基本単語認識も同時に行える。

2.3 基本単語の選定と単語の分割方法

基本単語の選定には、読売新聞記事データ(1987~2001年)に記載されている日本語キーワードを用いて行った。キーワードは、漢字と読みの組で構成されており、また複数の単語からなるキーワードについては、そのキーワードの前後に短く区切ったキーワードが記載されている。例えば、豊橋技術科学大学に対して/豊橋/や/技術/などの一般的な単語がキーワードとされている。

今回は、短い単語への分割を目的としているので、漢字で2~3文字(2~5音節程度に相当)のものをピックアップした。それらを単位として分割を行う。

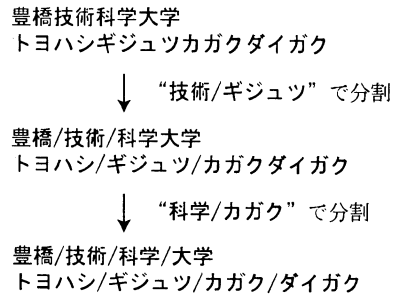


図3 組織名分割例

Fig. 3 Splitting an organization name into base-words

登録する頻度上位N組織名が別の組織名の一部として現れる場合がある。そこで、このN組織名もキーワードに加えた。

分割は、組織名データの一部が、キーワードと、漢字と読みの両方が一致する場合に行った。組織名データ中に、日本語キーワードの漢字と読みが含まれる場合に分割を行う。例を図3に示す。

こうして分割された個々の断片(キーワードに対応しなかった部分も含む)を基本単語と呼ぶ。認識では、これらのうち頻度上位のM単語の連続単語認識を行う。

なお、今回の分割方法はgreedyアルゴリズムであり、最適な分割が得られるとは限らない。また、言語的な繋がりを考慮した分割ではないため、意味的に正しい分割を行えるとは限らない。例えば、“北大西洋:キタタイセイヨウ”の場合、“北/大西洋:キタ/タイセイヨウ”と分割するのが正しいが、“北大/西洋:キタタイ/セイヨウ”のような分割も生じ得る。しかし、このような分割になっても“北大/キタタイ”という基本単語が登録されるので認識には影響しない。組織名の解析方式としては、自然言語処理を用いた方法[24]などがあり、それらを利用することも可能であるが、今回の組織名の基本単語への分割は、短い基本単語への分割を目的としているため、ここでは議論しない。

3. 組織名入力インタフェース

前節の方法で得られる認識結果を用いた単語・基本単語系列候補と基本単語ラティス候補からなる組織名インタフェースを考える。本インタフェースは、組織名が入力されるWebのフォームがクライアントとなって、サーバとして起動される。

ユーザはまず、組織名を音声で入力し、システムはこれを認識する。それで得られた認識結果から、図4のような表示を作成する。単語・基本単語系列候補(図4の漢字候補5best)では、単語と基本単語の連結によって認識された結果の上位候補を表示する。基本単語ラティス候補では、HMMトレリスにおいて、単語・基本単語系列候補を構成する単語と同じ位置で終端となっている基本単語の上位候補を表示する。例えば、“豊橋・技術・科学・大学”と認識した場合、“科学”などと同じ位置で終端となっている単語の上位候補を表示する。基本単語ラティス候補の作成例を図5に示す。こうして得られた候補から、



Japanese character input button
Speech input button
Organization's name candidates
Base word lattice
OK button to transfer the result to client

図4 組織名入力インタフェース例

Fig. 4 The multimodal interface server for organization name input

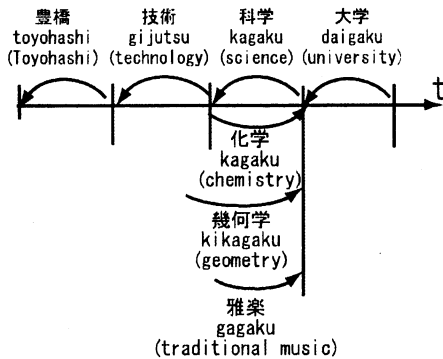


図5 基本単語ラティス作成例

Fig. 5 Creation of base-word lattice.

所望の組織名をペンタッチにより選択して入力する。また、日本語では、漢字入力だけでなく、その発音を表現したかな入力が求められる場合が多い。図4中の“かな入力”ボタンにより、表示をかなに切替えることができる。最後にOKボタンによりクライアントに結果を転送する。

4. 組織名認識実験

今回作成した組織名入力インタフェースの性能を調べるため、連結基本単語認識と単語辞書を併用した音声認識を用いて組織名の認識/入力実験を行った。

4.1 実験条件

組織名用言語モデルを用いて連結基本単語認識実験を行った。実験に使用されたデータは、男性話者5名により発話された100種類の組織名(計500サンプル)である。実験データは人為的に登録数10000語でカバー率が90%、15000語で92%、20000語で95%になるように選んだ(注:13万語の場合(図1)よりカバー率が低い)。登録単語数に対するカバー率を表1に

表1 実験試料の単語カバー率 [%] (括弧内は、組織名データ(13万種)に対するカバー率)

Table 1 Coverage of the test set [%]. (Coverage of 130,000 organization names are also shown in parentheses.)

基本単語数	単語登録数					
	0	1000	5000	10000	15000	20000
0	—	71 (75.0)	88 (92.2)	90 (96.0)	92 (97.4)	95 (98.0)
1000	69 (77.1)	90 (85.3)	95 (95.7)	96 (97.2)	97 (98.4)	95 (98.6)
5000	93 (95.8)	99 (96.0)	99 (97.6)	99 (98.4)	99 (98.8)	100 (99.1)
10000	94 (98.6)	99 (98.4)	99 (98.8)	100 (99.1)	100 (99.2)	100 (99.3)
15000	95 (99.4)	100 (99.1)	100 (99.2)	100 (99.3)	100 (99.4)	100 (99.5)
20000	96 (99.6)	100 (99.3)	100 (99.4)	100 (99.5)	100 (99.5)	100 (99.6)

示す。また表中の括弧内は13万種類の組織名に対する頻度を考慮したカバー率を示している。

音響分析は16kHzサンプリング、25msハミング窓、10msフレームシフトで行い、特徴量は12次元MFCC、 Δ MFCC、 $\Delta\Delta$ MFCC、 Δ パワー、 $\Delta\Delta$ パワーの38次元を用いた。

音響モデルはセグメント単位入力音節HMMで、5状態4出力分布、全共分散4混合のleft-to-right型の環境独立HMMであり音節数は114個である[25]。音響モデルの学習に使用したデータは、ATR連続音声テキスト503文A~Jセット(各50文)の6名の男性話者による文発声データ(3018文)と、サンプルが不足するカテゴリーに対して216単語の音声データのそれぞれから切り出された音節データを用いたML学習により初期モデルを作成し、更に、日本音響学会データベースATR503文A~Jセットの30名の男性話者(ASJ)(合計4518文)と、新聞記事読み上げ音声コーパスの125名の男性話者(JNAS)(合計12703文)を用いてMAP推定により追加学習を行ったものを用いた。

組織名用言語モデルは、組織名データから作成した組織名用の基本単語単位言語モデル(パイグラム)を用いて実験を行なった。基本単語の作成には、新聞記事データに記載されている日本語キーワードを用いて行った。13万の組織名について、新聞記事データ中で使われた回数を頻度として与えた。

表1の結果より、単語のみ、基本単語のみを登録した場合よりも単語・基本単語の組合せた場合の方がカバー率が高くなっており、認識性能が向上することが期待できる。例えば、単語のみを15000語登録した場合よりも、単語10000語+基本単語5000語の方がカバー率が良くなっている。

4.2 高頻度組織名と連結基本単語認識を併用した認識手法の認識性能評価

高頻度組織名と連結基本単語認識を併用した認識を行った結果を表2.3に示す。表2.3より、単語と基本単語を併用した結果、それぞれ単独での認識結果より良くなっていることが分かる。例えば、高頻度組織名を15000個登録した場合より、組

表 2 組織名 1best 認識率 [%]

Table 2 1-best recognition rates of organization names. [%]

基本 単語数	単語登録数					
	0	1000	5000	10000	15000	20000
0		68.6	82.4	83.2	79.4	81.2
1000	70.4	84.6	87.2	86.8	81.8	84.4
5000	81.8	88.8	88.0	86.6	82.2	87.2
10000	80.0	86.6	87.6	87.0	89.8	88.8
15000	80.0	87.4	88.8	89.2	89.8	88.8
20000	79.4	87.8	88.8	89.0	89.4	88.8

表 3 組織名 5best 認識率 [%]

Table 3 5-best recognition rates of organization names. [%]

基本 単語数	単語登録数					
	0	1000	5000	10000	15000	20000
0	—	68.6	83.0	83.6	79.4	81.2
1000	76.6	85.0	89.0	88.6	88.6	89.0
5000	83.0	90.8	89.8	88.8	90.6	90.6
10000	81.6	89.0	90.2	88.8	91.2	90.8
15000	81.4	90.4	91.6	91.2	91.4	89.8
20000	81.0	90.8	91.6	90.8	90.4	90.4

表 4 インタフェースによる入力可能割合 [%](登録単語数は 10000)

Table 4 The rate that users could input the desired name using the interface [%] (Whole word lexicon size was set to 10000)

基本単語 登録数	漢字 系列	読み 系列	基本 単語	システム
0	83.6	83.6	—	83.6
1000	88.6	88.6	89.0	89.8
5000	88.8	89.0	88.6	90.0
10000	88.8	89.0	88.8	90.4
15000	91.2	91.4	91.6	92.6
20000	90.8	91.0	90.6	92.2

組織登録数を 10000 個、基本単語数を 5000 個とした場合の方が認識率が良くなっていることが分かる。この結果、高頻度組織名と連続基本単語認識を併用した認識手法が効果的であるといえる。

4.3 組織名入力インタフェース利用による組織名入力率による評価

組織名入力インタフェースを利用した場合の入力可能割合を表 4 に示す。表 4 中の「漢字系列」は単語・基本単語系列の漢字列 5 候補、「読み系列」は単語・基本単語系列の読み列 5 候補、「基本単語」は単語・基本単語系列を構成する基本単語ラティスからの選択により入力可能な割合を示している。「システム」はいずれかの方法により入力可能である割合を示している。なお、「読み系列」は、図 4 中の“かな入力”ボタンにより、漢字の読みの入力を行った場合の入力可能割合である。

表 2 および表 4 の結果より高頻度単語辞書と基本単語 bigram を併用した音声認識が有効であることが言えるが、基本単語ラティスを利用した場合の入力可能率の向上は若干にとどまった。

表 5 オンライン被験者実験によるインタフェースシステム性能 [%](登録単語数、基本単語数はともに 10000)

Table 5 Interface system performance obtained by online subjective evaluation. (Both whole word lexicon size and basic-word lexicon size were set to 10000.)

被験者	漢字系列		基本単語		システム	
	1 回目	2 回目	1 回目	2 回目	1 回目	2 回目
	計		計		計	
A	68	14	69	14	76	14
	82		83		90	
B	81	9	66	8	84	9
	90		74		93	
C	78	12	70	11	81	13
	90		81		94	
D	89	4	75	4	92	4
	93		79		96	
平均	79.0	9.8	70.0	9.25	83.3	10.0
	88.8		79.3		93.3	

5. オンライン被験者実験

図 4 のインタフェースを実装し、被験者実験を行った。実験は音声対話システムの使用経験のない 20 代男子大学生 4 名にペンタッチ入力可能なノート PC 上でインタフェースを使用してもらった。発話内容は、4.1 節で使用した 100 種類の組織名と同じものを使用し、各話者 1 回ずつ入力を行った。また 1 回目の発話で入力できないと判断した場合、2 回目の入力を行うように指示した。使用した言語モデルには、応答速度も考慮して、高頻度組織名から上位 10000 語、基本単語数は上位 10000 語を登録したものをを用いた。認識実験結果を表 5 に示す。

まず、1 回目の漢字系列の結果をオフラインの音声認識実験に基づいた 5-best の結果 (表 3) と比較すると、大きく認識性能が低下している (88.8% から 79.0%)。これは、オンラインでの実験では音声区間の自動検出など音声認識率の低下要因があり、今回の実験では大きくその影響を受けたものと考えられる。これはシステムを実用的なレベルで運用するための課題である。個人差も大きいことが分かる。

2 回目の結果を見ると、1 回目で入力できなかった場合にもよく回復していると言える。その結果、漢字系列だけでも 90% 近い認識性能を得た。また、漢字系列では入力できなかった場合にも基本単語ラティスによる入力ができる場合が多く見られ、システム全体としては全員に対して 90% 以上の結果となり、このインタフェースの効果が見られた。

被験者実験の終了時に

- Q1. 本システムを利用したいか (したい 5 - 1 したくない)
 Q2. 本システムの性能が向上したら使いたい (したい 5 - 1 したくない)

というアンケート調査を実施した。その結果、Q1 に対しては平均 3.8、Q2 に対しては平均 4.3 という結果となった。認識性能、応答速度が向上すれば十分に便利に利用可能であると期待されていることが分かる。

6. ま と め

高頻度組織名と基本単語を併用した音声認識を提案し、組織名入力インタフェースについて検討を行った。

組織名入力インタフェースでは、一般に組織名は「国際電気通信基礎研究所」や「豊橋技術科学大学」などのように基本単語の連続で表現されるものが多いため、連続基本単語認識と高頻度の単語辞書を併用した言語モデルを使用した認識方法を提案し、評価実験を行った。録音した組織名発声によるオフライン音声認識実験の結果、高頻度の組織名と基本単語を併用することにより、高頻度組織名 10000 語 + 基本単語 10000 語で約 87% で入力することが可能となり、高頻度の組織名のみ 20000 語で約 81%、基本単語のみ 20000 語で約 79% と、それぞれ単独の言語モデルを使用した場合より良い結果が得られた。また、未登録語に関しても認識が行えるため、新出語などにも対応できることがわかった。

この音声認識結果に基づいて組織名入力インタフェースをシミュレーションした実験では、漢字系列候補により約 89%、漢字ラティス候補により約 89% という結果が得られたが、インタフェースシステム全体においては、約 90% と若干の入力率向上となった。

さらに、このインタフェースを実装してオンラインで被験者実験を行ったところ、音声区間自動検出などの精度の影響を受けて音声認識性能が低下し、1 回では 83.3% の入力可能率となったが、2 回まで音声入力を許すことで 93.3% が可能となった。

今後の課題としては、オンラインにおける音声認識率の向上が挙げられる。また、2、3 回の入力では正しく入力できない場合の対処法も考慮していく必要がある。

謝 辞

本稿を執筆するにあたり多くのシステム評価を行ってくれた豊橋技術科学大学情報工学系の清田和誠君に感謝致します。

文 献

- [1] 甲斐充彦, 中野崇広, 中川聖一, “音声認識サーバ-SPOJUS-を利用した WWW ブラウザの音声操作システム”, 情報処理学会研究会資料, SLP20-14, 1998.
- [2] 中野崇広, 甲斐充彦, 中川聖一, “WWW 上のフォーム型音声検索のための音声インタフェースの検討”, 情報処理学会研究会資料, SLP25-1, 1999.
- [3] 中野崇広, 甲斐充彦, 中川聖一, “WWW 上のテキスト入力フォームのための任意文字列入力の音声インタフェース”, 情報処理学会第 62 回全国大会講演論文集, 2-130, 2001.
- [4] 甲斐充彦, 盛 浩和, 中野崇広, 中川聖一, “フォーム型 Web 情報検索サービスのための音声ユーザーインタフェースシステムと操作性の評価”, 情報処理学会論文誌, Vol. 46, No.5, pp. 1318-1328, 2005.
- [5] 押川洋徳, 北岡教英, 中川聖一, “ウェブブラウザにおける任意文字列入力を目的とした音声入力インタフェース”, 日本音響学会講演論文集, 2-P-26, 2002.
- [6] 押川洋徳, 北岡教英, 中川聖一, “音声による姓名入力インタフェース”, FIT2003, 2003.
- [7] 押川洋徳, 北岡教英, 中川聖一, “音節 N-gram と単語辞書を併用した音声認識”, 日本音響学会講演論文集 2-6-11, pp.81-82, 2003.
- [8] 甲斐充彦, 廣瀬良文, 中川聖一, “単語 N-gram 言語モデルを用いた音声認識システムにおける未知語・冗長語の処理”, 情報処

- 理学会論文誌, Vo.40, No.4, pp.1383-1394, 1999.
- [9] R. C. Rose, I. Ariztendi, S.Parthasarathy, “An efficient framework for robust mobile speech recognition services,” Proc. of ICASSP2003, pp.316-319, 2003.
- [10] D. Jouvet, K. Bartkova, L. Delphin-Poulat, A. Ferricux, X. Lamming, J. Monnne, C. Raix, “About improving recognition of spontaneously uttered French city-names,” Proc. of ICASSP2003, pp.544-547, 2003.
- [11] M. Johnston et al., “Multimodal Language processing for mobile information access,” Proc. of ICSLP2002, pp.2237-2240, 2002.
- [12] Y. Y. Wang, “Robust language understanding in MiPad,” Proc. of EUROSPEECH2001, pp. 1555-1558, 2001.
- [13] L.Cornford et al., “The IBM personal speech assistant,” Proc. of ICASSP2001, pp. 1-4, 2001.
- [14] 谷垣宏一, 山本博史, 匂坂芳典, “語彙クラスに依存した未登録語モデルに基づく統計的言語モデル”, 日本音響学会講演論文集, 1-Q-23, pp.177-178, 2000.
- [15] 大森久美子, 斎藤博昭, “大語彙を対象とした音声対話インタフェースにおける自然な応答生成”, 自然言語処理, Vo.10, No.5, pp.23-40, 2003.
- [16] 大森久美子, 斎藤博昭, “語の属性を利用した大語彙確定のための音声対話インタフェース”, 自然言語処理, Vo.11, No.3, pp.123-146, 2004.
- [17] C. A. Kamm, C. R. Shamich, S. Singhal, “Speech recognition issues for directory assistance applications,” *Speech Communication*, 17, pp.303-311, 1995.
- [18] R. Cordoba, R. San-Segundo, J. M. Montero, J. Colas, J. Ferroiros, J. Macias-Guarasa, J. M. Pardor, “An interactive directory assistance service for Spanish with large-Vocabulary recognition,” Proc. of EUROSPEECH2001, pp.1279-1282, 2001.
- [19] F. Neubert, G. Gravier, F. Yvon, G. Chollet, “Directory name retrieval over the telephone in the Picasso project,” Proc. IEEE IVT'98 Workshop, 1998.
- [20] M. Meyer, H. Hild, “Recognition of spoken and spelled proper names,” Proc. EUROSPEECH1997, Vol.3, pp.1579-1582, 1997.
- [21] H. Schramm, B. Rucber, A. Kellner, “Strategies for name recognition in automatic directory assistance systems,” *Speech Communication*, 31, pp.329-338, 2000.
- [22] N. Davidson, F. McInnes, M. A. Jack, “Usability of dialogue design strategies for automated surname capture,” *Speech Communication*, 43, pp.55-70, 2004.
- [23] H. Oshikawa, N. Kitaoka and S. Nakagawa, “Speech interface for name input based on combination of recognition methods using syllable-based N-gram and word dictionary”, ICSLP-2004, pp. 177-180, 2004.
- [24] 岩瀬成人, “自然言語処理を用いた企業名解析方式”, 電子情報通信学会論文誌 (D-II), Vol. J82-D II, No.8, pp. 1305-1314, 1999.
- [25] S. Nakagawa, K. Hanai, K. Yamamoto, N. Minematsu, “Comparison of syllable-based HMMs and triphone-based HMMs in Japanese speech recognition,” in proceedings of International Workshop on Automatic Speech Recognition and Understanding, pp.393-396, 1999.
- [26] 日外アソシエーツ: デジタルコンテンツ「人名辞書」.