

## 単語の重要度を考慮したベイズリスク最小化音声認識と それに基づく情報検索

南條 浩輝<sup>†</sup> 翠 輝久<sup>††</sup> 河原 達也<sup>††</sup>

<sup>†</sup> 龍谷大学 工学部 情報メディア学科  
〒520-2194 大津市瀬田大江町横谷1-5

<sup>††</sup> 京都大学 情報学研究科 知能情報学専攻  
〒606-8501 京都市左京区吉田二本松町

E-mail: [†nanjo@ryukoku-u.jp](mailto:†nanjo@ryukoku-u.jp)

あらまし 情報検索システムのフロントエンドとしての音声認識を目的とした音声認識の新しい評価尺度とそれに基づく認識戦略について述べる。これまで、音声認識の一般的な評価尺度として、全ての単語を一様に扱う「単語誤り率 (word error rate: WER)」が用いられてきた。これに対して、情報検索の観点から各単語の重要度を考慮した「重みつき単語誤り率 (weighted word error rate: WWER)」を提案する。その上で、ベイズリスク最小化 (Minimum Bayes-Risk: MBR) の枠組みに基づいて、重みつき単語誤り率の最小化を行う音声認識を実現する。文書検索システム「音声版ダイアログナビ」を用いて、提案する認識手法が重みつき単語誤り率及び情報検索精度の改善に効果があることを示す。

キーワード 音声認識, ベイズリスク最小化デコーディング, 情報検索, 文書検索

## Information Retrieval based on Minimum Bayes-Risk Decoding considering Word Significance

Hiroaki NANJO<sup>†</sup>, Teruhisa MISU<sup>††</sup>, and Tatsuya KAWAHARA<sup>††</sup>

<sup>†</sup> Faculty of Science and Technology, Ryukoku University  
Seta, Otsu 520-2194 Japan

<sup>††</sup> Graduate School of Informatics, Kyoto University,  
Sakyo-ku, Kyoto 606-8501 Japan

E-mail: [†nanjo@ryukoku-u.jp](mailto:†nanjo@ryukoku-u.jp)

**Abstract** The paper addresses a new evaluation measure of automatic speech recognition (ASR) and a decoding strategy oriented for speech-based information retrieval (IR). Although word error rate (WER), which treats all words in a uniform manner, has been widely used as an evaluation measure of ASR, significance of words are different in speech understanding or IR. In this paper, we define a new ASR evaluation measure, namely, weighted word error rate (WWER) that gives a weight on errors from a viewpoint of IR. Then, we formulate a decoding method to minimize WWER based on Minimum Bayes-Risk (MBR) framework, and show that the decoding method improves WWER and IR accuracy.

**Key words** speech recognition, Minimum Bayes-Risk decoding, information retrieval, document retrieval

## 1. はじめに

話し言葉の音声認識技術の進展に伴い [1] [2], 近年, 自然な音声による情報検索システムの研究・開発が盛んに行われるようになった。音声による情報検索システムでは, 入力音声を認識し, そこから検索要求を解釈することが必要である。従来のデータベース検索型の情報検索システム [3] では, 関係データベースのフィールド名などが検索に必要なキーワード集合となるため, 音声認識はそのようなキーワードを認識すれば検索要求を解釈することができた。これに対して, 近年, マニュアル [4] や Web ページなど, 自然言語テキストで記述された大規模知識ベース (文書) を検索する高度な情報検索システムが研究対象となってきた [5] [6]。このような対象を検索するためには, 入力を自然言語文として受けとり, それを解釈する必要性が指摘されている [7]。これは, 検索対象が構造化されておらず, 検索を達成するためのキーワードを明確に定義できないためである。したがって, 音声認識では発話全体を自然言語文として認識する必要がある。

自然言語文でテキスト文書を検索する最も一般的な手法は, 入力文と検索対象の文書の距離を統計的に算出し, マッチングを行う手法である。距離の算出には, 入力文や各文書における各単語の出現頻度をベクトルの要素として表現して (bag-of-words), そのベクトル間の距離を求める手法が有名である。ベクトル要素の値は, 各単語がもつ情報量を示す *tfidf* 尺度を用いて求められることも多い。このように自然言語文を用いた文書検索システムにおいては, キーワードは明確に定義されていないものの, 検索 (マッチング) に決定的に影響を与える単語 (重要度の高い単語) とそうでない単語の区別が存在する。したがってフロントエンドとしての音声認識では, 発話全体を自然言語文として認識する際に, 重要な単語を優先的に認識する戦略が必要となる。すなわち, 重要でない単語の認識誤り数を減らすのと引替えに検索に影響を与える単語が認識誤りになるという事態を避けるような戦略を必要とする。このような音声認識の戦略を考えた場合, これまでの音声認識精度の一般的な評価尺度である単語誤り率 (WER) は必ずしも最適なものではない。単語誤り率は, 正解テキストと認識結果テキスト間の最小編集距離 (Levenshtein 距離) で定義される尺度であり, この評価尺度においては全ての単語の認識誤りが同等に扱われるためである。

このような背景に基づいて, 本稿では情報検索を目的とした音声認識の新しい評価尺度と認識手法について述べる。具体的には, 情報検索の観点から各単語の認識誤りに異なる重みを与える「重みつき単語誤り率 (WWER)」を評価尺度として定義し, 次に, ベイズリスク最小化 (Minimum Bayes-Risk: MBR) デコーディングの枠組 [8] に基づいて, 重みつき単語誤り率の最小化を目的とした音声認識を行う。我々は, キーワードを利用した学会講演からの重要文抽出 (情報検索) において, このベイズリスク最小化デコーディングが有効であることを示した [9] [10]。本稿は, 提案手法をより一般的な情報検索タスクに適用し, その有効性を報告するものである。

**[HOWTO] Windows XP で音声認識を使用する方法**  
この資料は以下の製品について記述したものです。

- Microsoft Windows XP Professional
- Microsoft Windows XP Home Edition

**概要** この資料では, Windows XP で音声認識を使用する方法について説明しています。Microsoft Office XP の音声認識をインストールしているか, または, Office XP がインストールされたコンピュータを新たに購入した場合は, すべての Office アプリケーションや, 音声認識が利用可能なその他のアプリケーションで音声認識を使用できます。

**詳細** 音声認識は, 音声をテキストに変換するオペレーティングシステムの機能です。音声認識エンジンと呼ばれる内部ドライバによって, 単語が認識され, テキストに変換されます。音声認識エンジンは, ..

図 1 ソフトウェアサポート技術情報の例  
Fig. 1 Example of software support article

表 1 ソフトウェアサポート用知識ベース  
Table 1 Software support knowledge base

Text collection	# documents	size
glossary	4.707	1.4M byte
FAQ	11.306	12M byte
DB of support articles	23.323	44M byte

## 2. ベースラインシステム

### 2.1 情報検索システム

本研究では, 情報検索システムとして「音声版ダイアログナビ」を用いる。「ダイアログナビ」とは, マイクロソフト社のソフトウェアサポート用知識ベースを自然言語文により検索するシステムであり, 東京大学で開発されたものである [11] [12]。検索対象の知識ベースは以下の 3 種類であり, これらは全て自然言語で記述されたテキストである。

- (1) 用例集 (glossary)
- (2) ヘルプ集 (FAQ)
- (3) サポート技術情報

図 1 に, 知識ベースの例を示す。また, 各知識ベースの文書サイズおよびデータサイズは, 表 1 に示されている。

我々は, 「ダイアログナビ」の入力インターフェースとして音声を採用した「音声版ダイアログナビ」の研究・開発をこれまでにやってきた [13] [14]。本稿は, この「音声版ダイアログナビ」の音声認識戦略について述べるものである。

### 2.2 音声認識システム

本研究では, 以下に示す音響モデルおよび言語モデルと認識エンジン Julius rev.3.4 [15] で構成した音声認識システムを用いる。音響モデルには, 新聞記事読み上げ音声 (JNAS コーパス: 40 時間) を用いて学習した性別非依存 PTM triphone モデル [16] を用いる。言語モデルには, 以下に示す 3 種類のテキストコーパス (合計約 690 万単語) を用いて学習した 3-gram

表 2 音声認識結果を用いた情報検索成功率  
Table 2 Success rate of retrieval using ASR result

	IR success rate (%)			ASR rate (WER (%))
	set-1	set-2	total	
Manual transcript	69.91	63.89	67.74	0
ASR result	62.07	57.78	60.52	19.27

モデル（語彙サイズ約 18K 単語）を用いる。

- (1) ソフトウェアサポート用知識ベース（表 1）
- (2) ダイアログナビに寄せられた質問集
- (3) ソフトウェアサポートの疑似対話を書き起こしたもの

なお、本研究での単語の単位は形態素解析システム ChaSen ver2.02 [17] に基づいている。

### 2.3 ベースラインシステムによる情報検索結果

評価データは、音声対話システムを利用したことのない 30 名の被験者による 499 発話である。本実験では、種々のパラメータの推定にディベロブメントセットを用いるため、評価データを 2 つ (set-1: 319 発話, set-2: 180 発話) に分割した。各セットの実験に用いるパラメータは、他のセット（ディベロブメントセット）を用いて推定する。

表 2 に、ベースラインシステムによる検索成功率を示す。ここでは、検索システムが最終的に出力した候補（10-best 候補）の中に、検索対象の文書が存在した場合に検索成功としている。表 2 には、書き起こしテキストを利用した場合の検索成功率も示されている。この結果は音声認識精度が 100% の場合の検索率を示しており、音声認識の改善により達成できる検索成功率の上限を表している。音声認識結果（単語誤り率 19.27%）を用いた場合の検索成功率は、set-1 に対して 62.07%、set-2 に対して 57.78%、平均で 60.52% であり、音声認識誤りにより精度が約 6~8% 低下していることがわかる。

## 3. 情報検索のための音声認識の評価尺度

### 3.1 単語誤り率

音声認識の評価尺度で最も一般的なものは単語誤り率 (WER) であり、これは式 (1) で定義されるものである。

$$\text{WER} = \frac{I + D + S}{N} * 100 \quad (1)$$

ここで、 $N$  は正解文における単語の数、 $S$  は置換誤り単語の数、 $D$  は削除誤り単語の数、 $I$  は挿入誤り単語の数である。

式 (1) から明らかのように、この評価尺度（単語誤り率）は、全ての単語の音声認識誤りを同一に扱っている。すなわち、単語誤り率は情報検索に決定的に影響を及ぼす重要な単語（キーワードのような単語）とそうでない単語（問投詞やフィルターなど）を区別せず、等しく扱う評価尺度である。

### 3.2 重みつき単語誤り率 - 単語誤り率の一般化 -

重要単語の認識誤りは情報検索に影響が大きいので、そのような誤りに重みをつけて評価する必要がある。本節では、各単語の音声認識誤りに異なる重みを与えるために、単語誤り率 (WER)

ASR result	: a b c d e f
Correct transcript	: a c d' f g
DP result	: C I C S C D

$C$ : correct word,  $I$ : inserted word,

$D$ : deleted word,  $S$ : substituted segment

$$\text{WWER} = (V_I + V_D + V_S) / V_N * 100$$

$$V_N = v_a + v_c + v_{d'} + v_f + v_g$$

$$V_I = v_b$$

$$V_D = v_g$$

$$V_S = \max(v_d + v_c, v_{d'})$$

$v_i$ : weight of word  $i$ .

図 2 重みつき単語誤り率 (WWER) の計算例

Fig. 2 Example of weighted word error rate (WWER) calculation

の一般化を行なう。具体的には、以下の式で定義する「重みつき単語誤り率 (Weighted Word Error Rate: WWER)」について述べる。

$$\text{WWER} = \frac{V_I + V_D + V_S}{V_N} * 100 \quad (2)$$

$$V_N = \sum_{w_i} v_{w_i} \quad (3)$$

$$V_I = \sum_{\hat{w}_i \in I} v_{\hat{w}_i} \quad (4)$$

$$V_D = \sum_{w_i \in D} v_{w_i} \quad (5)$$

$$V_S = \sum_{seg_j \in S} v_{seg_j} \quad (6)$$

$$v_{seg_j} = \max(\sum_{\hat{w}_i \in seg_j} v_{\hat{w}_i}, \sum_{w_i \in seg_j} v_{w_i}) \quad (7)$$

ここで、 $v_{w_i}$  は正解文における  $i$  番目の単語 ( $w_i$ ) の重みを表し、 $v_{\hat{w}_i}$  は音声認識結果における  $i$  番目の単語 ( $\hat{w}_i$ ) の重みを表している。また、 $seg_j$  は  $j$  番目の置換誤り区間を表し、 $v_{seg_j}$  は誤り区間  $seg_j$  の重みを表している。この誤り区間の重み  $v_{seg_j}$  は、当該区間  $seg_j$  に含まれる正解系列の単語の重みの合計値と認識結果の単語の重みの合計値のうち、大きいほうとする。参考のため、図 2 に重みつき単語誤り率の計算例を示す。

なお本研究では、誤り単語を同定する際に、単語誤り率を求める際に用いた DP マッチングの結果を用いている。したがって、全ての単語の重みを等しく設定した場合、重みつき単語誤り率は単語誤り率と等しくなる。

## 4. ベイズリスク最小化デコーディング

前章では、情報検索のための音声認識の評価尺度である重みつき単語誤り率 (WWER) の定式化を行なった。本章では、この重みつき単語誤り率を最小化する音声認識手法（デコーディング手法）について述べる。本手法は、ベイズリスク最小化 (Minimum Bayes-Risk: MBR) 枠組み [8] に基づく音声認識手法である。

統計的な音声認識は一般的に、与えられた入力音声信号  $X$  を最もよく説明する単語列  $\hat{W}$  を求めるプロセスとして定式化

される。これを式 (8) に示す。

$$\hat{W} = \operatorname{argmax}_{W'} P(W'|X) \quad (8)$$

ベイズ決定理論に基づくと、音声認識は決定規則  $(\delta(X) : X \rightarrow \hat{W})$  と記述できる。ここで、損失関数を  $l(W, \delta(X)) = l(W, W')$  とすると、音声認識は以下のベイズリスク最小化の枠組みで記述できる [8]。

$$\delta(X) = \operatorname{argmin}_W \sum_{W'} l(W, W') \cdot P(W'|X) \quad (9)$$

なお、式 (8) で示されている一般的な音声認識のプロセスは、式 (9) において 0/1 損失関数を用いる場合と等価であり、文誤り率 (Sentence Error Rate) を最小化するプロセスであるといえる。本研究でのベースライン音声認識システムではこのデコーディング手法を用いている。

次に単語誤り率 (WER) 最小化音声認識について述べる。単語誤り率はテキスト編集における最小編集距離 (Levenshtein distance) から導かれるものであり、式 (1) で与えられるものである。単語誤り率最小化を目的とした場合は、損失関数  $l(W, W')$  としてこの単語誤り率を用いればよいことが知られており、実際にこのような研究が行われてきている [8] [18]。

これに対し本研究では、情報検索システムのための音声認識を行うため、提案する音声認識の評価尺度である「重みつき単語誤り率 (WWER)」の最小化を考える。すなわち、重みつき単語誤り率に基づく損失関数を用いてベイズリスク最小化音声認識を行う。これを定式化したものが、式 (10) である。

$$\delta(X) = \operatorname{argmin}_W \sum_{W'} \text{WWER}(W, W') \cdot P(W'|X) \quad (10)$$

ここで、式 (10) の右辺の事後確率  $P(W'|X)$  はベイズ則により  $P(W', X)/P(X)$  と記述できる。この時、分母の  $P(X)$  は式全体の最小化に影響を及ぼさないため、式 (10) は次のように変形できる。

$$\delta(X) = \operatorname{argmin}_W \sum_{W'} \text{WWER}(W, W') \cdot P(W', X) \quad (11)$$

さらに文献 [8] で有効性が主張されているように、スコア正規化のためのパラメータ  $\lambda$  を導入することにより、決定規則は最終的に次のように記述できる。

$$\delta(X) = \operatorname{argmin}_W \sum_{W'} \text{WWER}(W, W')^{\lambda_1} \cdot P(W', X)^{\frac{1}{\lambda_2}} \quad (12)$$

なお、デコーディング時に最適な単語列  $W$  を直接求めるのは困難であるため、本研究では、まずベースラインの音声認識手法により N-best リストを生成し、次に生成された N-best リストのリスコアリングを行うことで、ベイズリスク最小化デコーディングを行う。

## 5. 情報検索のための単語重み

本章では、検索結果に決定的に影響を及ぼす単語とそうでない単語を定義する方法 (重みの決定方法) について述べる。

### 5.1 tf-idf 尺度に基づく重み

まず、情報検索で一般的に利用されている tf-idf 尺度を用いて単語重みを定義する方法について考える。文書  $d$  における単語  $w$  の tf-idf 値 ( $tf \cdot idf(w, d)$ ) は以下のように定義される。

$$tf \cdot idf(w, d) = tf(w, d) \times \log\left(\frac{N}{df(w)}\right) \quad (13)$$

$$= tf(w, d) \times \log(idf(w)) \quad (14)$$

ここで、 $tf(w, d)$  は文書  $d$  中の単語  $w$  の出現回数であり、 $df(w)$  は単語  $w$  を含む文書数である。  $N$  は全文書数であり、 $idf(w)$  の全文書数  $N$  を  $df(w)$  で除したものである。

このように tf-idf の値は、一般的に各文書ごとに求めるものであるが、本研究では、検索対象の文書全体に対して単語の重要度を求める必要がある。単純に tf-idf 値を全文書で平均する方法は、もともと tf-idf 尺度が表している「ある単語の特定の文書での出現傾向」という情報を打ち消してしまう。すなわち、このように tf-idf 値を平均化する行為は、重みとしての適切さを損なうという問題がある。したがって、本研究では tf-idf 値に基づく単語重みを以下の手続きで求める。

- 1) 検索対象の各文書  $d$  の各単語  $w$  に対して  $tf \cdot idf(w, d)$  値を求める
- 2) 各文書  $d$  に対して、 $tf \cdot idf(w, d)$  値が高い上位 5 単語をその文書の代表単語とする。
- 3) 各単語  $w$  に対して、それが代表単語となった文書の数をカウントし、そのカウントを重みとする。単語の重みが 1 以下の場合は重みを 1 とする。

本稿では、この重みを用いた「重みつき単語誤り率」を  $\text{WWER}_{\text{tf-idf}}$  と表す。

### 5.2 検索対象文書から学習した言語モデルとの類似度に基づく重み

次に、検索対象の文書集合と入力文の一致度を評価する尺度に基づいて重みを定義する方法について述べる。本研究では一致度の尺度として、検索対象である知識ベースのみから学習した単語 N-gram モデルによる単語パープレキシティ [14] を採用する。ある語句のパープレキシティが大きいことは、その語句の知識ベース内での出現頻度が低いことを示し、検索に直接関係ないことを示唆している。対照的に、パープレキシティが小さいことは、その語句の知識ベース内での出現頻度が高く (ターゲットの文書集合と近く)、当該語句が情報検索に重要であることを示唆している。このように、パープレキシティに基づく重みは情報検索に対する語句の重要度を表す指標となっている。本研究では、パープレキシティに基づく重みを以下のよう求める。

文節	PP
<S> O s I M E だ け ず ( 不 要 )	2121.62
( が ) 不 要 に な っ た ( I M E )	21241.56
( な っ た ) I M E 2 0 0 0 を ( 削 除 )	349.24
( を ) 削 除 し た い の で す が ( ど う )	10.56
( が ) ど う し た い で し ょ ( か )	2459.01
( で し ょ ) か ( </S> )	3044.04

<S>, </S>はそれぞれ始端記号, 終端記号

図3 パープレキシティの計算例  
Fig.3 Example of word perplexity

1) 文節ごとに検索対象文書から学習した言語モデルを用いてパープレキシティを算出する。その際、直前文節の最終単語を当該文節の先頭に、後続文節の先頭単語を当該文節の末尾に加える。この様子は図3に示されている。

2) パープレキシティをシグモイド関数を用いて変換し、文節の重みとする。

ここで定義される重みと前述の *tf-idf* に基づいた重みのとの主な相違点は、重みを定義する単位が単語ではなく文節であるという点である。すなわち、この重みは単語列としての特徴(重要度)をとらえた重みとなっている。

本稿では、この重みを用いた「重みつき単語誤り率」を“*WWER<sub>pp</sub>*”と表す。

### 5.3 重みの統合

最後に、*tf-idf* 尺度に基づく単語重みとパープレキシティに基づく単語重みの両方を用いる場合について説明を行う。本研究では、式(15)に示すように、それぞれから求まる重みつき単語誤り率の幾何平均を「重みつき誤り率」「*WWER*」として用いる。

$$WWER = WWER_{tf-idf}^{\phi} * WWER_{pp}^{(1-\phi)} \quad (15)$$

なお、 $\phi$  は統合重みであり、この値はディベロップメントセットを用いて推定する。

## 6. 評価実験

### 6.1 種々の単語重みを用いた音声認識と検索成功率

5章で述べた3種類の重みを用いて、重みつき単語誤り率最小化音声認識を行い、その結果を用いて情報検索を行った。音声認識におけるN-best リスコアリングは発話ごとに行った。なお、N-best リストのNのサイズは100とした。式(12)のリスコアリングパラメータ( $\lambda_1, \lambda_2$ )および式(15)の統合重み $\phi$ は、ディベロップメントセットを利用して求めた。具体的には、set-1をset-2のパラメータを推定するためのディベロップメントセットとして用い、set-2をset-1のパラメータを推定するためのディベロップメントセットとして用いた。パラメータ推定は、ディベロップメントセットの重みつき単語誤り率を最小とるように決定した。

表3 音声認識の評価尺度最小化デコーディングと検索成功率  
Table 3 Comparison of ASR evaluation measure minimization decoding for IR

	IR Success Rate (%)
baseline	60.52
MBR ( <i>WWER<sub>tf-idf</sub></i> )	60.32
MBR ( <i>WWER<sub>pp</sub></i> )	61.12
MBR ( <i>WWER</i> )	61.72

表4 音声認識評価尺度(WERおよびWWER)の改善率  
Table 4 Improvement of WER and WWER

	WER (%)	WWER (%)
baseline	19.27	25.88
MBR ( <i>WWER</i> )	20.31	25.35

表5 認識戦略の比較  
Table 5 Comparison of decoding strategies

ASR method (loss function)	IR Success Rate (%)		
	set-1	set-2	total
baseline (0/1)	62.07	57.78	60.52
MBR (WER)	60.56	59.44	60.12
MBR ( <i>WWER</i> )	<b>62.38</b>	<b>60.56</b>	<b>61.72</b>

種々の音声認識の評価尺度(重みつき単語誤り率)の最小化音声認識とその結果を用いた検索成功率を表3に示す。パープレキシティに基づく重み(*WWER<sub>pp</sub>*)を利用した場合に、検索成功率の精度向上を確認した。*tf-idf* 尺度に基づく重みを単体で使用した場合(*WWER<sub>tf-idf</sub>*)には、精度向上はみられなかったが、パープレキシティに基づく重み(*WWER<sub>pp</sub>*)と統合する効果がみられ、本実験における最も高い検索成功率(61.72%)を得た。提案する重みつき単語誤り率最小化音声認識が、情報検索に有効であることを確認した。

### 6.2 重みつき単語誤り率最小化音声認識の効果

表4は、損失関数に重みつき単語誤り率を定義し、ベイズリスク最小化音声認識を行った場合の「単語誤り率(WER)」と「重みつき単語誤り率(WWER)」を従来の音声認識(0/1損失関数+ベイズリスク最小化音声認識)と比較して示したものである。重みつき単語誤り率の最小化を目的とした音声認識により、単語誤り率が低下したものの、ターゲットとする音声認識評価尺度である「重みつき単語誤り率」は25.88%から25.35%に改善された。この結果は、提案手法の妥当性を示している。

表5は、提案手法(重みつき単語誤り率最小化音声認識)による情報検索成功率を、従来のベイズリスク最小化音声認識(単語誤り率最小化音声認識)を用いた場合と比較したものである。提案手法では、重みつき単語誤り率が改善されたのと同時に検索成功率も61.72%に改善されたことがわかる。逆に、単語誤り率最小化音声認識(従来手法)を行った場合は、単語誤り率の改善はみられたが(19.27% → 18.68%)、検索成功率の

改善はみられなかった (0.40%の精度低下)。

これらのことは、重みつき単語誤り率最小化音声認識の効果を示している。

## 7. おわりに

情報検索を目的とした新しい音声認識の評価尺度である「重みつき単語誤り率 (WWER)」を提案した。情報検索の観点から単語に重みを設定し、重みつき単語誤り率を定式化した。その上で、この新しい評価尺度を最小化するデコーディング手法を実装し、評価を行った。提案した重みつき単語誤り率最小化デコーディングの情報検索における有効性を示した。

謝辞 本研究の一部は、龍谷大学理工学術研究助成基金によるものである。

## 文 献

- [1] S.Furui, K.Maekawa, and H.Isahara. Toward the realization of spontaneous speech recognition - introduction of a Japanese priority program and preliminary results -. In *Proc. ICSLP*, Vol. 3, pp. 518-521, 2000.
- [2] S.Furui. Recent advances in spontaneous speech recognition and understanding. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 1-6, 2003.
- [3] E. Levin, S. Narayanan, R. Pieraccini, K. Biatov, E. Bocchieri, G. Di Fabbriozio, W. Eckert, S. Lee, A. Pokrovsky, M. Rahim, P. Ruscitti, and M. Walker. The AT&T-DARPA communicator mixed-initiative spoken dialogue system. In *Proc. ICSLP*, 2000.
- [4] 伊藤亮介, 駒谷和範, 河原達也. 機器操作マニュアルの知識と構造を利用した音声対話ヘルプシステム. *情処学論*, Vol. 43, No. 7, pp. 2147-2154, 2002.
- [5] S. Harabagiu, D. Moldovan, and J. Picone. Open-domain voice-activated question answering. In *Proc. COLING*, pp. 502-508, 2002.
- [6] C. Hori, T. Hori, H. Isozaki, E. Maeda, S. Katagiri, and S. Furui. Deriving disambiguous queries in a spoken interactive ODQA system. In *Proc. IEEE-ICASSP*, 2003.
- [7] 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, Pascale Fung. 柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム. *情処学研報*, 2001-SLP-39-30, 2001.
- [8] V. Goel, W. Byrne, and S. Khudanpur. LVCSR rescoring with modified loss functions: A decision theoretic perspective. In *Proc. IEEE-ICASSP*, Vol. 1, pp. 425-428, 1998.
- [9] 南條浩輝, 河原達也. 音声理解のための音声認識評価尺度の提案. *日本音響学会研究発表会講演論文集*, 2-1-5, 秋季 2004.
- [10] 南條浩輝, 河原達也. 音声理解のための音声認識評価尺度とベイズリスク最小化デコーディング. *信学技報*, SP2004-122, NLC2004-82 (SLP-54-42), 2004.
- [11] Y. Kiyota, S. Kurohashi, and F. Kido. Dialog Navigator: A question answering system based on large text knowledge base. In *Proc. COLING*, pp. 460-466, 2002.
- [12] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答 - ダイアログナビ - . *自然言語処理*, Vol. 10, No. 4, pp. 145-175, 2003.
- [13] T. Misu, K. Komatani, and T. Kawahara. Confirmation strategy for document retrieval systems with spoken dialog interface. In *Proc. ICSLP*, pp. 45-48, 2004.
- [14] 翠輝久, 駒谷和範, 清田陽司, 河原達也. 音声対話によるソフトウェアサポートタスクのための効率的な確認戦略. *電子情報通信学会論文誌*, Vol. J88-DII, No. 3, pp. 499-508, 2005.
- [15] A. Lee, T. Kawahara, and K. Shikano. Julius - an open source real-time large vocabulary recognition engine. In *Proc. EUROSPEECH*, pp. 1691-1694, 2001.
- [16] 李晃伸, 河原達也, 武田一哉, 鹿野清宏. Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識. *信学論*, Vol. J83-DII, No. 12, pp. 2517-2525, 2000.
- [17] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸. 日本語形態素解析システム 茶筌 version 2.0, Dec. 1999.
- [18] A. Stolcke, Y. Konig, and M. Weintraub. Explicit word error minimization in N-best list rescoring. In *Proc. EUROSPEECH*, pp. 163-165, 1997.