

## 関連度計算補正により認識率向上を図った会話文音声理解システム

岡崎 雄亮<sup>†</sup> 渡部 広一<sup>‡</sup> 河岡 司<sup>‡</sup>

†‡同志社大学大学院工学研究科 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: †dth0738@mail4.doshisha.ac.jp, ‡{hwatabe, tkawaoka}@mail.doshisha.ac.jp

あらまし 本研究では、コンピュータと人間との自然な会話の実現を目指している。そのためには音声認識の技術が欠かせない。しかしながら、現在の音声認識では高い認識率を得るのが困難であり、コンピュータと人間との自然な会話は実現できない。そこで、音声認識装置から得られた認識結果に対して、更に概念ベースと関連度計算を用いて言語処理的な補正を行うことで音声認識精度の向上を図る会話文音声理解システムの構築を目的とし、その手法を提案する。

キーワード 音声認識, 概念ベース, 関連度計算, 会話文音声理解システム

## The Conversation Speech Understanding System that Planned Recognition Improvement by Calculation The Degree of Association

Yusuke OKAZAKI<sup>†</sup> Hirokazu WATABE<sup>‡</sup> and Tsukasa KAWAOKA<sup>‡</sup>

†‡Graduate School of Engineering, Doshisha University

1-3 Miyakodani Tatara Kyotanabe-shi, Kyoto, 610-0394 Japan

E-mail: †dth0738@mail4.doshisha.ac.jp, ‡{hwatabe, tkawaoka}@mail.doshisha.ac.jp

**Abstract** In this research, it is an aim to realize the natural conversation with a computer and the human being. Therefore, the technology of the speech recognition is indispensable. However, it is difficult to get high recognition in the current speech recognition, and it can't realize the natural conversation with a computer and the human being. Then, it is purpose to construct the Conversation Speech Understanding System to plan improvement of the speech recognition precision by revision of the language processing with Concept-Base and Degree of Association for a recognition result provided from speech recognition system, it is suggested the technique of the Conversation Speech Understanding System.

**Keyword** Speech Recognition, Concept-Base, Degree of Association, Conversation Speech Understanding System

### 1. はじめに

近年コンピュータは急速に発展しており、様々な場面で人間の手助けとなっている。コンピュータが人間の手助けとなるためには、人間とのコミュニケーションが必要であり、現在では、人間との会話によるコミュニケーションが可能な知能ロボットが注目されている。このような知能ロボットを実現するためには、知能ロボットが話者の言葉を正確に聞き取り、内容を理解する手段、つまり音声認識が必要不可欠となる。

音声認識の分野では、これまで様々な研究が行われ、多くの製品が提供されている。その音声認識装置のほとんどは、音声認識精度の向上を主たる目的として開発が進められてきたものである。しかし、現在の音声認識においては、話者や環境、語彙などの変化によって高い認識率を得ることが困難であり、音声による人間との自然な会話が実現できていないと言われている。

そこで、人間とコンピュータとの自然な会話を実現するために、音声認識装置から得られた認識結果に対

してさらに言語処理的な補正を行うことで音声認識精度の向上を図る会話文音声理解システムを構築することを目的とする。

音声理解システムについては、これまでに認識結果補正知識ベースによる一致度補正方式<sup>[1]</sup>という手法が提案されている。これは複数の音響モデルの認識結果を同時に取得し、知識ベースの情報をを用いて補正を行うため、音響モデルから外れる音声や利用環境による揺らぎが補正可能である。しかしながら、現在の音声理解システムは3文節以下にしか対応しておらず、4文節以上で構成された文章は補正が行えない。しかし、人間が会話をするときに、話す内容が3文節以下であることはほとんどなく、人間とコンピュータとの自然な会話を実現するためには、音声理解システムを4文節以上の文章にも対応させなければならない。

そこで、本稿では4文節以上の文章の補正が行える会話文音声理解システムの手法を提案し、音声認識装置と比較評価を行い、その有効性を検証する。

## 2. 関連技術

本章では、本研究において使用した技術について述べる。

### 2.1. 概念ベース

概念ベース<sup>[2]</sup>とは語（概念）と意味（属性）のセットを約9万語蓄積してあるものである。この概念と属性のセットには、その重要性をあらわす重みが付与される。ある「概念A」は、その概念と関連が深く、その概念の意味となると考えられる「属性 $a_i$ 」と、その属性の重要性を表す「 $w_i$ 」の組の集合で表される。概念Aは以下のように表される。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)\}$$

属性 $a_i$ を概念Aの一次属性と呼ぶ。一次属性 $a_i$ を一つの場合と見なせば、 $a_i$ からさらにその一次属性を導くことができ、 $a_i$ の属性 $a_{ij}$ を概念Aの二次属性という。

これを展開していくと、一つの場合Aはn次属性まで持つことができる。概念ベースの例を図1に示す。

語	属性 (1次属性)
雪	(雪, 0.61), (白い, 0.30), (下る, 0.27)
白い	(雪, 0.16), (白地, 0.14), (色, 0.14)

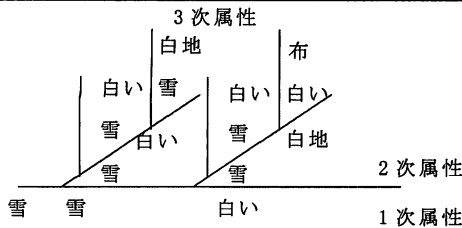


図1 概念ベースの例

### 2.2. 関連度計算方式

概念Aと概念Bの関係の深さを定量的に表すのが関連度計算<sup>[3]</sup>という方法である。関連度は、0以上1以下の連続的な数で表され、概念同士の関連が大きいほど関連度は高くなる。この関連度を求める計算は、それぞれの概念を二次属性まで展開し、その重みを利用した計算によって最適な一次属性の組み合わせを求め、それらが一致する属性の個数を評価することで算出する。関連度計算の例を表1に示す。

表1 関連度計算の例

基準概念	対象概念	関連度
自動車	車	0.920
	飛行機	0.116
	バス	0.040

## 3. 音声認識環境

本研究では、無線ハンドマイクを利用し、音声認識専用のパソコンで入力音声を受信する方式を取る。ま

た、入力音声を複数の音声認識専用マシンで同時に受信する。受け取った入力音声を音声認識装置で解析し、認識結果をネットワークを介して1台の補正マシンへ送信して補正を行う。本稿では、2種類の音声認識装置DragonSpeech<sup>[4]</sup>、Julius<sup>[5]</sup>が組み込まれた2台の受信端末で合計4つの認識結果を取得し、それらを補正マシンで補正する方式を取る。音声認識装置の精度評価は8章で行う。

## 4. 音声理解システム

本章では、認識結果補正知識ベース、一致度補正方式、及びこれらを用いた音声理解システムの処理について説明する。

### 4.1. 認識結果補正知識ベース

入力語Aに対する音声認識装置の出力した認識結果の読み $a_i$ と、その出現割合 $w_i$ の対の集合を求める。

ここで $a_i$ を「属性」、Aを「見出し語」とし、このよ

うな属性が定義された見出し語の集合を認識結果補正知識ベース（表2）と呼ぶ。認識結果補正知識ベースの作成には人と音声合成の認識結果を併用しており、多数の人からデータを取得する手間を軽減し、自動的に属性数を増加することを可能としている<sup>[6]</sup>。また、現在の登録語彙数は2272語である。

表2 認識結果補正知識ベース（一部）

見出し語	属性1, 重み	属性2, 重み	...
上	うえ, 186.7	うで, 7.7	...
腕	うで, 139.3	ので, 27.4	...

### 4.2. 一致度補正方式

これは、音声認識装置から出力された複数の認識結果を一組とし、認識結果補正知識ベースを参照して補正を行う方式である。具体的には、まず認識結果群の読みを、認識結果補正知識ベース中の任意の見出し語の属性集合と比較する。認識結果が属性集合の中に含まれる場合、各々の属性の重みの総和をその見出し語の一致度とする。そして全ての見出し語に対して一致度を計算し、一致度が最大となる見出し語を最終的な結果として出力する、という処理を行うものである。以下に例を挙げて説明する（図2）。

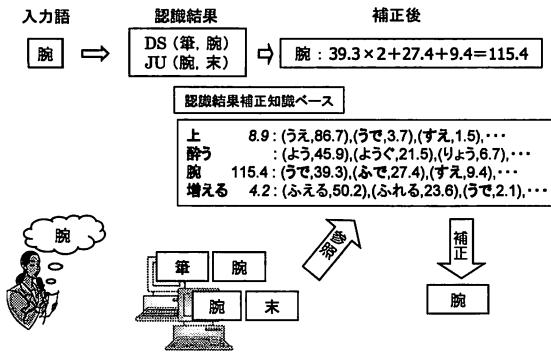


図 2 一致度補正方式

### 4.3. 音声理解システムの処理の流れ

まず、話者の入力に対してDS、JU合わせて4つの認識結果を取得する。次に、認識結果ごとに茶筌<sup>[7]</sup>を用いて形態素解析を行い自立語と助詞に分け、各々に対して補正を行う。自立語については3.2節で説明した一致度補正を行い、助詞については多数決処理による補正を行う。これは、音声認識装置から得た複数の認識結果から形態素解析によって助詞を取得し、最も多く出力された助詞を補正結果とする方法である。この流れを図3に示す。

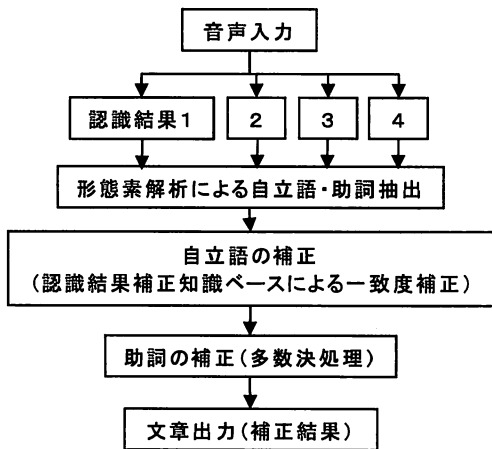


図 3 音声理解システムの流れ

## 5. 問題点とその解決案

### 5.1. 問題点

従来の音声理解システムの補正対象が3文節以下であったのに対し、本研究の補正対象は4文節以上と長くなったために、音声認識装置が認識できない部分が出てきた。つまり、4つの音声認識結果の自立語の数が揃わないという問題点が浮かび上がった。以下にその例を示す。

入力文：来週提出予定のレポートがあるけどまだできていない

認識結果①：来週提出予定のレポートがあるけどまだできていない ⇒自立語数 8

認識結果②：来週提出予定のレポートがあるけどまだできていない ⇒自立語数 8

認識結果③：来週必要とのレポートがある程度をできていない ⇒自立語数 7

認識結果④：買収低調でのレポートとはあるけど元できていない ⇒自立語数 7

このように自立語数が揃わないと、前から順番に自立語を対応させ一致度補正を行う際に、上手く対応付けが出来ない部分がある。上記の例の場合、1番目の「来週、来週、来週、買収」は対応させられるが、3番目になると「予定、予定、レポート、レポート」となり、本来は「レポート、レポート、レポート、レポート」を対応させなければならないところを、間違った対応付けをしてしまうことになる。

そこで、この問題点を解決するために考えた手法が、次節で説明する「読み一致」である。

### 5.2. 読み一致

この手法は、音声認識装置の認識ミスや認識できなかった場合でも、自立語を上手く対応させることができる手法で、言葉の読みに着目し、その関連性を調べるものである。

方法は、漢字やカタカナを平仮名に直し、平仮名表記でどの程度一致するかを比べるというものである。例えば「来週」と「買収」の場合、読みが「らいしゅう」、「ばいしゅう」となり、読みの的に80%一致しており関連性が高いと考えられる。同様に、「レポート」と「レポート」の場合も読みの的に75%一致しており、関連性が高いため対応付けが出来る。

この手法を用いて、一致度補正を行う際の自立語の対応付けを行う。詳しい使用方法は6章で説明する。

## 6. 会話文音声理解システム

### 6.1. 会話文音声理解システムの流れ

本節では、本研究で構築した会話文音声理解システムの全体の流れを述べる。

まず、話者の入力に対してDS、JU合わせて4つの認識結果を取得する。次に、認識結果ごとに茶筌を用いて形態素解析を行い自立語と助詞に分ける。

そして、自立語の数が4文とも一致する場合は従来の音声理解システムの処理で、自立語は一致補正方式を用いて補正し、助詞は多数決処理で補正を行い、補正結果を結合する。

また、自立語の数が4文とも一致しない場合は、補

正のベースとなる文を1文選択し（詳細は6.2節で説明する）、読み一致による自立語の対応付けを行い（詳細は6.2節で説明する）、自立語を一致度補正方式を用いて補正し、助詞の選択（詳細は6.2節で説明する）を行い、自立語の補正結果と助詞を結合する。

そして最後に、一致度補正を行った結果を更に概念ベースと関連度計算を用いて補正し、その補正結果を最終結果として出力する。以上述べた流れを図4に示す。

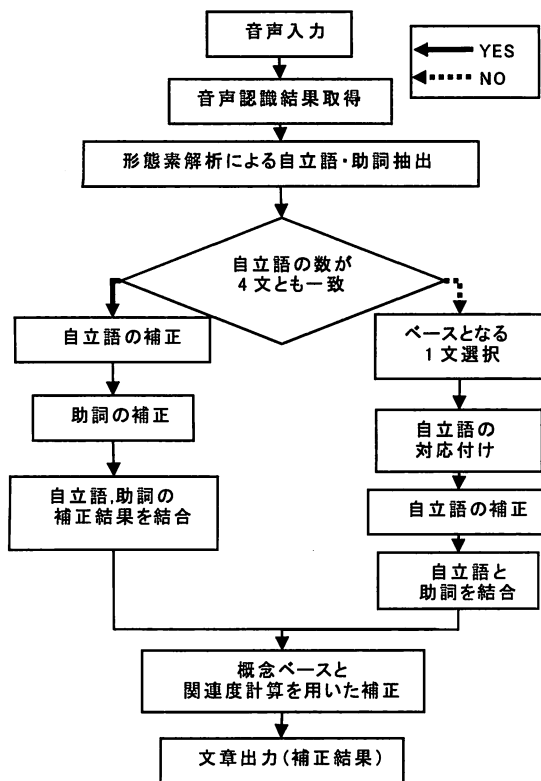


図4 会話音声理解システムの流れ

## 6.2. 自立語数が一致しない場合の処理

### 6.2.1. 補正のベースとなる1文の選択方法

音声認識結果4文の自立語数が一致しない場合、自立語を対応させるときに、読み一致を用いることを5.2節で説明したが、その4文の中には必要な自立語が抜けていたり、間違っ認識した結果全く意味の通らない文章になっていたりする文が存在する。その中で、必要な自立語だけを抜き出し、対応付けするのは難しいと考える。

そこで考えたのが、補正を行う際のベースとなる文を1文選ぶことである。ベースとなる文は、常識的に考え意味の通る文で、最も入力文に近い（最も精度が高い）文を選択する。これにより、読み一致を行うの

がベースとなる1文と他の3文の自立語の間だけになり、上手く自立語同士を対応させられると考える。そこで用いるのが、2.1節で述べた関連度計算である。関連度計算は語と語の関係の深さを定量的に表すものであり、我々が普段使う文章は語と語が密接に関係してできているものであると考えるので、関連度計算を用いれば、最も常識的で、より入力文に近い文が選択できると考える。

その選択方法は、音声認識結果4文の各文ごとに全ての自立語同士（副詞は除く）で関連度計算を行い、最も高い値が出た語を含む1文を補正のベースにするというものである。

### 6.2.2. 読み一致による自立語の対応付け

5.2節で自立語の対応付けに読み一致を用いることを説明したが、本節では、その対応付けの方法を詳しく説明する。

まず、6.2.1節で説明したように、補正のベースとなる1文を選択する。そして、ベースとなる文の全ての自立語と、残りの3文の全ての自立語とで読み一致を行い、自立語同士が50%以上一致していれば、対応させ一致度補正の組にする。なお、50という値は閾値で、実験的に定めたものであり、詳細は6.3節で説明する。

### 6.2.3. 助詞の処理

音声認識結果4文の自立語数が一致しない場合の助詞の処理は、補正のベースとなる文の助詞をそのまま採用することにする。

## 6.3. 読み一致に用いる閾値の検証

本節では、6.2.2節で説明した自立語の対応付けの際に用いるために定めた閾値について、その検証方法を説明する。

検証には、単語と単語が読みのどの程度一致するかという読み一致度（読み一致の値）を用いる。例えば、「来週」と「買収」の読み一致度は80となる。この読み一致度は「他の3文の自立語/ベースとなる文の自立語」で求める。検証に用いるテストデータは3.3節と同じものである。検証方法は、閾値を10~100まで10刻みで増やしていき、読み一致度が閾値を超えていけば一致度補正の組にするという方法で、その組の中に正しい候補が2つ以上あれば成解、2つ未満または間違っ候補が1つでもあれば不正解とし、その正解率を調べた。結果を図5に示す。

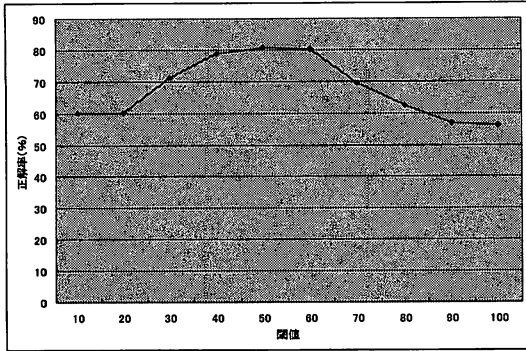


図 5 読み一致に用いる閾値検証の結果

図 5 より、閾値 50 のときに正解率が最大となっている。よってこの閾値 50 を自立語の対応付けの際の閾値に決定する。

## 7. 概念ベースと関連度計算を用いた補正

4 章で述べた従来の音声理解システムの一致度補正方式には語と語の関連性が考慮されていない。そのため、例えば「綺麗な景色を見て感動した」という入力文に対して、その補正結果は「紀元景色を見て感動した」となってしまふ。これは、表 3 のように「紀元」の一致度が「綺麗な」の一致度より大きくなるために「紀元」が出力されてしまふ。人間であれば、「景色」という語から関連性を推測し、表 4 の候補中に存在する正解語「綺麗な」を選ぶことができる。しかし、従来の補正では補正結果の候補中に正解語があっても正しい結果を得ることが出来ない。

表 3 「綺麗な」の一致度補正結果

	綺麗な
第一候補	紀元,114.09
第二候補	昼,109.1
第三候補	綺麗な,100.61

そこで、本章では従来の補正に加え、語と語の関連性を考慮した補正を行うことを提案する。このときに用いるのが、2 章で説明した概念ベースと関連度計算である。

### 7.1. 信頼語

本節では、概念ベースと関連度計算を用いた補正を行うときに用いる「信頼語」について説明する。

信頼語とは、以下の 2 つの場合の第一候補の語のことで、信頼性が高い語として特定する。表 4 の場合、それぞれ「景色」、「魚」が信頼語となる。

①一致度補正結果の候補が 1 つのみ（第一候補のみ）の場合

②（第一候補の一致度/第二候補の一致度）が閾値 7 以上の場合

表 4 「景色」、「魚」の一致度補正結果

	景色	魚
第一候補	景色,432	魚,301.4
第二候補		サッカー,20.9
第三候補		サラダ,18.2

ここでの閾値も実験的に定めたもので、7.2 節で詳細を説明する。

### 7.2. 信頼語取得の際の閾値の検証

検証に用いるテストデータは、3.3 節と同じ入力文から得られた音声認識結果をシステムに通し、得られた補正結果 100 文の自立語である。検証方法は、閾値を 1~10 まで 1 刻みで増やしていき、ある語の（第一候補の一致度/第二候補の一致度）が閾値を超えていれば（候補が第一候補のみの語は除く）、その語の第一候補を信頼語として取得し、その信頼語が補正結果として正解か否かを調べた。正解率は（正解した信頼語の数/全自立語数）で求めている。結果を図 6 に示す。

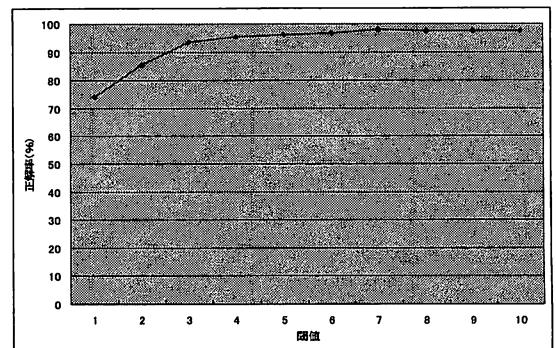


図 6 信頼語取得の際の閾値検証の結果

図 6 より、閾値 7 のときに正解率が最大となっている。よってこの閾値 7 を信頼語取得の際の閾値に決定する。

### 7.3. 補正の流れ

概念ベースと関連度計算を用いた補正の流れは以下の①~④のようになる。

- ①補正結果の中で信頼語ではない語の前後の自立語を抜き出す
- ②その語が信頼語で、かつ名詞 or 動詞 or 形容詞ならば、その語の一次属性を取得する（動詞・形容詞は原形にする）
- ③取得したすべての一次属性と信頼語ではない語の補正結果の第一候補~第三候補との間で関連度計算を行う
- ④最も高い関連度が閾値 0.1 を超えていれば、語の

一致度を2倍にする

#### 7.4. 信頼語取得の際の閾値の検証

この閾値の検証のテストデータは、補正結果100文の中から信頼語ではない語が含まれる文を用いた。そして、閾値を0.01~0.15まで0.01刻みで増やしていき、7.3節で説明した補正が正しく行えるか（正解率）を調べた。結果を図10に示す。

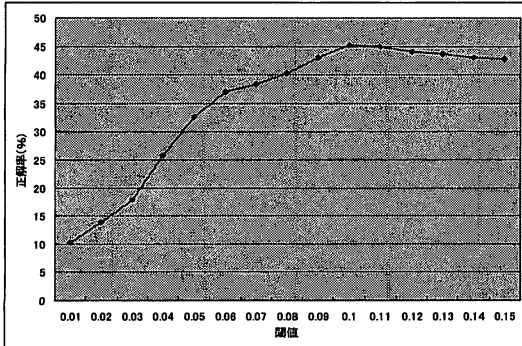


図7 信頼語取得の際の閾値検証の結果

図7より、閾値0.1のときに正解率が最大となっている。よってこの閾値0.1を一致度を増加する際の閾値に決定する。

#### 8. 会話文音声理解システムの精度評価

入力文は4文節以上で構成された文で、全部で100文ある。評価方法は音声認識結果（入力文1文につき4つ）をシステムに掛け、補正結果として出力された1文と入力文を文節区切りで対応させ、一致するなら正解、一致しないなら不正解とし、全文節に対する正解の文節の割合を精度とした。全文節数は444個で、結果は被験者5人（男性3人、女性2人）の平均を取った。音声認識装置の精度と共に図8に示す。なお、音声認識装置の精度評価は、本システムの評価方法と同じで、同じ入力文を用いて、入力文と音声認識結果を文節区切りで対応させ、一致するなら正解、一致しないなら不正解とし、被験者5人の平均を取っている。

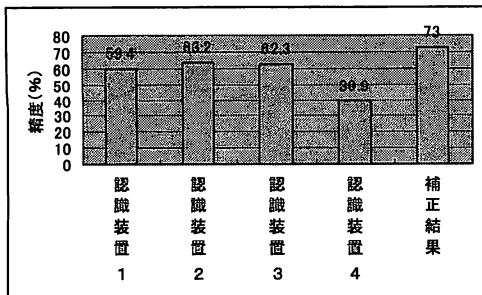


図8 信頼語取得の際の閾値検証の結果

図8より、4台の音声認識装置の最大精度が63.2%で、構築した会話文音声理解システムの精度が73%となり、本システムを用いることで音声認識装置の最大精度より9.8%精度を向上することができた。これにより、提案した手法は有効であると考えられる。

#### 9. おわりに

人間とコンピュータとが自然な会話を行うためには、高精度の音声認識が必要不可欠である。本稿では、それを実現するために、既存の音声認識装置による音声認識後の補正によって音声認識精度の向上を目的とし、既存の音声理解システムを拡張して、4文節以上で構成された文章の補正が行える会話文音声理解システムを提案した。また、それを構築するに当たり、既存の認識結果補正知識ベースを用いた一致度補正方式に加えて、概念ベースと関連度計算を用いた文章補正を提案した。これにより、音響モデルから外れる音声に対しても高い正解率を得ることができるだけでなく、文章中の単語間の常識的な関係を考慮することでさらに正解率を向上することができた。以上により、本稿で構築した会話文音声理解システムの有効性が示されたと結論付けることができる。

#### 謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクトにおける研究の一環として行ったものである。

#### 文献

- [1] 葛谷紳, 渡部広一, 河岡司, “誤認識データベースを用いた単語音声認識方式”, 信学技報, NLC2004-35, pp.71-76, 2004.
- [2] 広瀬幹規, 渡部広一, 河岡司, “概念間ルールと属性としての出現頻度を考慮した概念ベースの自動精練手法”, 信学技報, TL2001-49, pp.109-116, 2002.
- [3] 渡部広一, 河岡司, “常識的判断のための概念間の関連度評価モデル”, 自然言語処理, Vol.8, No.2, pp.39-54, 2001.
- [4] ScanSoft, “DragonSpeech7”, <http://www.scansoft.com/>
- [5] 京都大学, “情報処理振興事業協会(IPA), “大語彙連続音声認識デコーダJulius”, <http://julius.sourceforge.jp/>
- [6] 空野皇司, 渡部広一, 河岡司, “単語音声認識のための音声合成装置を用いた誤認識データベースの自動構築”, 言語処理学会第12回年次大会発表論文集, B1-3, pp.34-37, 2006.
- [7] 奈良先端科学技術大学院大学情報科学研究科, “形態素解析システム 茶筌”, <http://chasen.naist.jp/hiki/ChaSen/>