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

2008－ICS－151（2）
2008／3／5

常識判断を用いた音声理解システムの構築

合田 輝幸† 渡部 広一‡ 河岡 司‡

†‡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: †dtg0707@mail4.doshisha.ac.jp, ‡{hwatabe, tkawaoka}@mail.doshisha.ac.jp

あらまし 人とコンピュータの円滑な会話を実現するには音声認識技術が必要である．しかし現在の音声認識装置は話者や語彙などに制約があり，音声による柔軟な会話に用いることは難しい．本稿では，既存の音声認識結果に対して大規模知識ベースに基づく言語処理による補正を行うことで高い精度を得られる音声理解システムを構築する．まず認識結果補正知識ベースを用いた一致度補正方式によって，複数の音響モデルの認識結果に対して知識ベースの情報を用いて補正を行い，音響モデルから外れる音声や利用環境による揺らぎを補正する．さらに，文章を構成する単語はそれらの間で常識的な関係を持っていることが多いと考え，常識判断システムを用いて語と語の常識的な関連性を考慮した補正手法を提案する．

キーワード 音声理解システム，認識結果補正知識ベース，一致度補正方式，常識判断システム

Construction of Speech Understanding System using Commonsense Judgment

Teruyuki GOUDA† Hirokazu WATABE‡ and Tsukasa KAWAOKA‡

†‡Graduate School of Engineering, Doshisha University
1-3 Miyakodani Tatara Kyotanabe-shi, Kyoto, 610-0394 Japan
E-mail: †dtg0707@mail4.doshisha.ac.jp, ‡{hwatabe, tkawaoka}@mail.doshisha.ac.jp

Abstract The speech recognition technology is necessary to achieve a smooth conversation between a person and a computer. However, it is difficult to use flexible conversation by voice in a present speech recognition system, with restrictions on the speaker, the environment, and the vocabulary. Then, in this paper, the Speech Understanding System is constructed to obtain high accuracy by correction of language process based on large-scale knowledge base for existing speech recognition result. First, it corrects the speech that misses the sound model and the difference in the user environment by the Revision Method by the Degree of Coincidence using the Knowledge Base to Correct Recognition Result that corrected using information on the knowledge base for the recognition result with some sound models. Second, it thinks that some words in sentence often have a commonsense relation between them and it proposes the method for correcting using the Commonsense Judgment System to consider a commonsense relation between the words.

Keyword Speech Understanding System, Knowledge Base to Correct Recognition Result, Revision Method by Degree of Coincidence, Commonsense Judgment System

1. はじめに

近年，実用的な秘書ロボットや介護ロボットなど，人間との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．これらのロボットの実現に伴い，福祉介護や娯楽など多くの分野で活躍できると考えられる．このようなコンピュータとの円滑なコミュニケーションには音声による会話が望ましい．そのため，知能ロボットが話者の言葉を正確に聞き取り内容理解を行う音声認識技術が必要不可欠となる．しかし現在行われている対話形式の音声認識においては，話者や語彙数に制限をかけた上で認識を行っており，音声による人間との自然な会話が実現できているとは言い難い．

そこで，音声認識装置から得られた認識結果に対してさらに大規模知識ベースに基づく言語処理による補正を行うことで，高い精度を得ることのできる音声理解システムを構築する．その補正手法として，認識結果補正知識ベースを用いた一致度補正方式[1]という手法を提案している．これは複数の音響モデルの認識結果に対して知識ベースの情報を用いて補正を行うことで，音響モデルから外れる音声による揺らぎを補正する手法である．また，我々は話を聞く際音の情報からだけでなく文章として自然かどうかといった言語の情報も加えて内容を理解している．そこで本稿では，一般的に文章を構成する単語はそれらの間で常識的な関係を持っていることが多いと考え，常識判断システムを用いて文章レベルで補正を行う手法を提案する．

2. 音声認識環境

一般的に音声認識装置で使用されている音声入力方法は，有線ハンドマイク・ヘッドセットマイクによるものが多い．しかし知能ロボットが動作・移動を行いながらコミュニケーションを行う上で，有線による音声入力がその妨げとなる可能性がある．また，ヘッドセットの着用やピンマイクのセットの利用はユーザビリティに優れているとは言えない．そこで本稿では，無線ハンドマイクを利用し，2種類の音声認識装置DS，JUが組み込まれた2台の受信端末で同時に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（図 1）．ここで，エンロール，音響モデルとは，予め大量の学習データから入力音声の特徴を性別ごとに定義したモデルのことである．本稿では各音声認識装置において男性モデル，女性モデルを使用する．

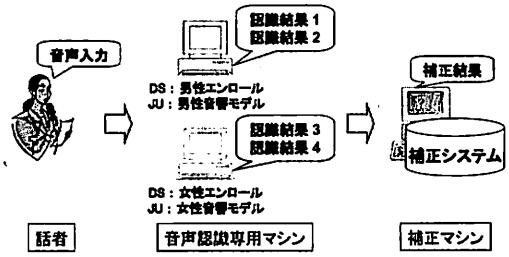

図 1．音声認識環境

3. 音声理解システム

本章では，これまでに構築された音声理解システムの処理の流れについて述べる．

3.1. 処理の流れ

まず2章で述べたように，話者の入力に対して4つの認識結果を取得する．そして認識結果ごとに茶筅[2]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という流れになっている（図 2）．

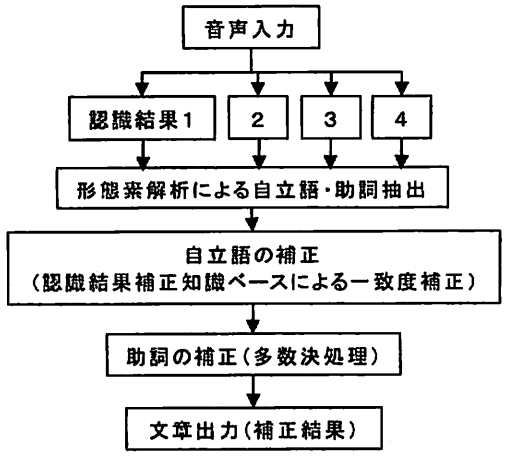

図 2．音声理解システムの流れ

3.2. 自立語の補正

本節では，自立語の補正方法について述べる．その際に用いる，認識結果補正知識ベースと一致度補正方式について説明する．

3.2.1. 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．

$$A=\{(a_1, w_1), (a_2, w_2), \cdots, (a_i, w_i)\}$$

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 1）と呼ぶ．

本稿では，特定語彙とする単語が認識結果補正知識ベースの見出し語として登録されており，現在の登録語彙数は 1140 語となっている．また認識結果補正知識ベースに登録する語は音声認識装置の単語辞書から選んでおり，現在は名詞 647 語，動詞 373 語，形容詞 106 語，副詞 14 語で構成されている．

表 1．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	属性3,重み
上	うえ,86.7	うで,3.7	すえ,1.5
酔う	よう,45.9	ようぐ,21.5	りょう,6.7
腕	うで,39.3	ふで,27.4	すえ,9.4
増える	ふえる,50.2	ふれる,23.6	うで,5.1

また，認識結果補正知識ベースは，以下の 3 種類の認識結果によって構成されている．

・正解語（見出し語の読み）
　⇒ 重みを最大として追加
・人の認識結果
　⇒ 男女各 1 名×DS・JU の認識結果
・音声合成の認識結果
　⇒ 4 キャラクタ×DS・JU の認識結果

音声合成と人の認識結果を併用することで，多人数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[3]．

3.2.2. 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，3.2.1 節で説明した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．図 3 に，「腕」と音声入力した場合の例を示す．

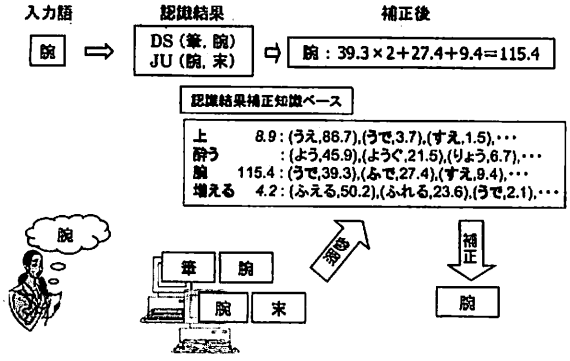

図 3. 一致度補正方式

3.3. 助詞の補正

助詞の補正には多数決処理による補正を用いている．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．

4. 音声理解システムの評価

本章では，3 章で説明した音声理解システムについて評価を行う．

4.1. 評価結果

3.2.1 節で説明した認識結果補正知識ベースを用いて，被験者 10 名が入力したテストデータ 100 文の補正を行った．テストデータは，認識結果補正知識ベースに登録されている単語から作成された文章である．結果は図 4 のようになった．

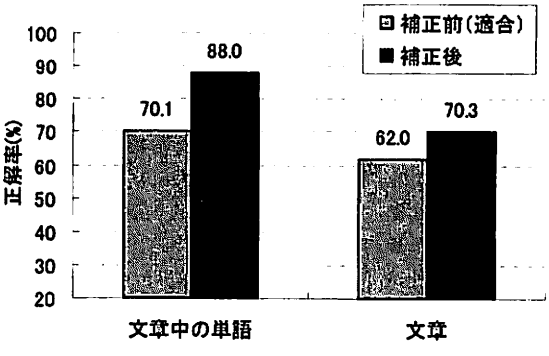

図 4. 音声理解システムの評価結果

図 4 では，文章中の単語と文章のそれぞれの正解率を示しており，以降，これらを「文章中の単語正解率」と「文章正解率」と表記する．また評価結果は「補正前（適合）」，「補正後」の比較を示している．「補正前（適合）」とは，被験者 10 名における，以下の 2 つの正解率の平均値である．

・音声認識装置 DS：
「男性エンロール」と「女性エンロール」から得られた認識結果における高い方の正解率
・音声認識装置 JU：
「男性音響モデル」と「女性音響モデル」から得られた認識結果における高い方の正解率

結果から，補正を行わない時に比べて文章中の単語正解率で 17.9%，文章で 8.3%正解率が向上しており，認識結果補正知識ベースを用いた一致度補正が有効であることがわかる．

4.2. 一致度補正方式の問題点

本節では一致度補正方式の問題点を，例を挙げて説明する．

① 入力音声「港で船に乗る」
② 認識結果を取得し，形態素解析
　認識結果 1 「港 で 船 に 取る」
　認識結果 2 「里 で 胸 に 乗る」
　認識結果 3 「港 で 根 降りる」
　認識結果 4 「港 へ 脛 に 乗る」
③ 一致度補正方式による自立語の補正
　「港」「里」「港」「港」 ⇒「港」
　「船」「胸」「根」「脛」 ⇒「胸」
　「取る」「乗る」「降りる」「乗る」 ⇒「乗る」
④ 多数決処理による助詞の補正
　「で」「で」「で」「へ」 ⇒「で」
　「に」「に」「に」 ⇒「に」
⑤ 補正結果「港で胸に乗る」

上の例から，「船」が「胸」と出力され，補正が失敗していることが分かる．この時，単語ごとに一致度の大きい順から補正候補を並べると表 2 のようになり，「胸」の一致度が「船」の一致度より大きくなっていることがわかる．

表 2. 一致度補正の補正候補

	港	船	乗る
第 1 候補	港,332.1	胸,120.5	乗る,225.1
第 2 候補	箸,25.1	船,115.5	取る,118.4
第 3 候補	寒い,13.2	絵,18.5	降りる,110.1

我々人間であれば，「港」や「乗る」という語から関連性を推測し，候補中に存在する正解語「船」を選ぶことができる．これは，人間が長年の中で蓄積した「常識」という知識を持っているからである．そこで，この補正候補中の語と語の関係を考慮した補正方法を新たに追加し，第 2 候補以降に正解がある例を補正することを提案する．

5. 常識判断システムを用いた文章補正

本章では，4.2 節で述べた一致度補正方式の問題点を解決するため，常識判断システムを用いた補正手法を提案する．これは，文章中の単語間の常識的な関係を考慮し補正を行うものである．例えば「八百屋で林檎を買う」という入力に対して一致度補正方式で「八百屋でリングを買う」というように補正されたとする．この結果を見た時，我々人間であれば「八百屋で売っているのは野菜・果物」という常識を持っているので「リング」は「林檎」の間違いではないかと考えることが出来る．このような処理を，常識判断システムを用いて実現する．この手法を用いることで文章中の単語間の関連性を見出すことができ，従来システムの問題点を解決できると考える．

5.1. 常識判断システム

常識判断システムとは，人間が普段，意識的または無意識のうちに用いている常識をコンピュータに理解させるシステムである．我々が日常的に用いる常識には様々なものがあり，例えば時期や季節などの「時間」，重さや速さとなどの「量」，学校や教室などの「場所」，暑いや美味しいなどの「感覚」，嬉しいや悲しいなどの「感情」に関する常識が挙げられる．本稿では，会話において特に重要だと思われる「場所」と「時間」に関する常識を新たな補正手法として組み込むこととする．以降より，本手法で使用する場所判断システム[4]，場所連想システム[5]，時間判断システム[6]について説明する．

5.1.1. 場所判断システム

場所判断システムとは，ユーザが入力した語が場所であるかどうかを判断し，場所を表す語（場所語）である場合は，その場所に存在する人や物とその場所で行われる事象を出力するシステムである．前者を主体語，後者を目的語と呼び，これらをユーザが入力した語の関連語とする．

例：入力「学校」
⇒ 主体語「学生，教師，生徒・・・」，
目的語「勉強，授業，教育・・・」

5.1.2. 場所連想システム

場所連想システムとは，ユーザが入力した語からそれに関係する場所語を出力するシステムであり，5.1.1 節の場所判断システムと対称の役割を持つ．この出力された場所語をユーザが入力した語の関連語とする．

例：入力「電車，乗る」
⇒ 「駅，線路，鉄道・・・」

5.1.3. 時間判断システム

時間判断システムとは，時間に関する表現を理解し，ユーザが入力した語（時語）が表現している時期や季節を出力するシステムである．この時期や季節を代表語と呼ぶ．本手法で用いる場合には，出力された代表語を持つ時語を時語知識ベースから取得し，これらをユーザが入力した語の関連語とする．

例：入力「クリスマス」
⇒ 代表語「冬」
⇒ 「正月，雪，12 月・・・」

5.2. 処理の流れ

処理の流れは以下の 1～4 のようになる．

1. 一致度の値から信頼語を取得
2. 信頼語が場所語か時語かを判定
3-1. 場所語であれば，場所判断システムから関連語として主体語，目的語を取得
3-2. 時語であれば，時間判断システムから関連語として時語を取得
3-3. どちらでもなければ，場所連想システムから関連語として場所語を取得
4. 補正候補中で，関連語と一致する語の一致度を増加し，最も一致度の大きい語を補正結果として出力

【処理 1】

一致度補正によって得た補正候補から信頼性の高い語（信頼語）を取得する．これは一致度補正の段階で，正解と見なせる語を決定するということである．この信頼語を以降の常識判断の処理に用いる．具体的な信頼語の求め方は以下の式の通りである．

$$M_1 / M_2 > N$$

M_1：第 1 候補の一致度
M_2：第 2 候補の一致度
N ：閾値

ここで閾値 N は実験的に定めた値（5.3 節で説明）を用いる．表 2 の例だと，信頼語は「港」となる．

【処理 2】

処理 1 で決定した信頼語を場所判断システム，時間判断システムに入力し，信頼語の中から場所語，時語があるかどうかを判断する．この結果によって，処理 3 で使用する常識判断システムを決定する．信頼語が場所語であった場合は処理 3-1 を，時語であった場合は処理 3-2 を，どちらでもない場合は処理 3-3 を行う．

【処理 3-1】

処理 2 で場所語を取得した場合，場所判断システムから，関連語として主体語，目的語を取得する．

例：信頼語「港」
⇒ 主体語「船，魚，海・・・」
目的語「出港，乗る，降りる・・・」

【処理 3-2】

処理 2 で時語を取得した場合，時間判断システムから代表語を取得する．そして取得した代表語と同じ代表語を持つ時語を時語知識ベースから出力する．

例：信頼語「正月」

⇒ 代表語「冬」

⇒ 関連語「雪，冬至，初春･･･」

【処理 3-3】

処理 2 で場所語，時語が両方とも取得できなかった場合は場所連想システムを用い，信頼語から関係する場所語を取得する．

例：信頼語「船」

⇒ 関連語「港，駅，空港」

【処理 4】

処理 3 で取得した関連語全てを補正候補に参照し，同じ語があればその語の一致度の値を増加する．増加値は実験的に求めた 2 倍（5.4 節で説明）とする．そして一致度順に候補を並べ替え，最終的に第 1 候補となる語を補正結果として出力する．

以下の例の場合，一致度補正候補は表 2 のようになり，関連語から一致度を増加して一致度順に候補を入れ替えると表 3 のようになる（表中の太字は一致度が変化した語）．

例：「港で船に乗る」

⇒ 信頼語「港」

⇒ 関連語「船，海，乗る，降りる･･･」

表 3. 一致度増加後の補正候補

	港	船	乗る
第 1 候補	港,332.1	**船,231.0**	**乗る,450.2**
第 2 候補	箸,25.1	胸,120.5	**降りる,220.2**
第 3 候補	寒い,13.2	絵,18.5	降りる,110.1

この結果から一致度が最大となる第 1 候補を取ることで，「港で船に乗る」という正しい補正結果を導くことが出来る．

以上のように，一致度補正によって取得した補正候補とその一致度から，場所語を含む文章については場所判断システム，場所連想システムを，時語を含む文章については時間判断システムを用いることで，文章中の単語間の常識的な関連性を見出して補正を行う．この様な手法を用いることで，入力音声を音の情報だけでなく言語の情報も加えて認識することができ，より人間に近い音声認識を実現することができる．

5.3. 信頼語決定のための閾値の検証

本節では，5.2 節の処理 1 で説明した，信頼語を求める際の閾値を定めるための検証を行う．5.2 節でも述べた様に，閾値は $M_1 / M_2 > N$ で表される．

検証方法として，閾値 N を 1～10 で変化させ，上記の式で求めた単語ごとの信頼語の正解率を比較することで適切な閾値を決定する．テストデータに文章 100 文の一致度補正結果を対象として検証を行ったところ，結果は図 5 のようになった．

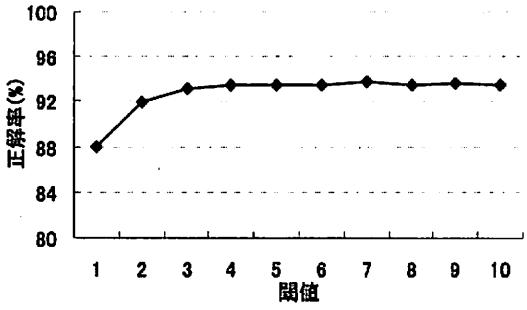

図 5. 信頼語の閾値の検証結果

結果から，正解率は閾値 4 以降で一定となっていることがわかる．よって閾値 4 以降で正解率が最大となる閾値 7 を，信頼語決定のための閾値として決定する．

5.4. 一致度の増分の検証

本節では，5.2 節の処理 4 で説明した，一致度を増加する際の増分を定めるための検証を行う．

検証方法として，増分を 1～10 倍で変化させ，改良補正を用いた場合の正解率を比較することで適切な増分を決定する．テストデータに文章 100 文を対象として検証を行ったところ，結果は図 6 のようになった．

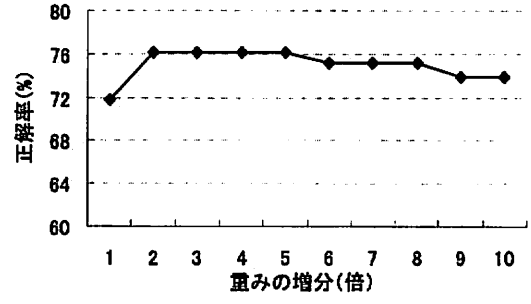

図 6. 一致度の増分の検証結果

結果から，増分が 2～5 倍の時に最大となり，それ以降は低下していることがわかる．よって，2 倍を一致度の増分として決定する．

6. 改良した音声理解システムの評価

本章では，5 章で提案した常識判断システムを用いた文章補正を追加し，改良された音声理解システムについて評価を行う．

3.2.1 節で説明した認識結果補正知識ベースを用いて，被験者 10 名が入力したテストデータの補正を行った．テストデータは認識結果補正知識ベースに登録されている単語から作成された文章であり，以下の 3 種類を作成した．評価方法として，補正前（適合），従来システムの正解率，改良システムの正解率を，文章中の単語，文章についてそれぞれ比較する．

【テストデータ】

① 場所語を含んだ文章 100 文

例：「公園で遊ぶ」「港で船に乗る」

② 時語を含んだ文章 100 文

例：「クリスマスに雪が降る」

③ 場所語，時語を含まない文章 100 文

例：「前に進む」，「頭を撫でる」

それぞれのテストデータの結果は以下の図 7，図 8，図 9 のようになった．

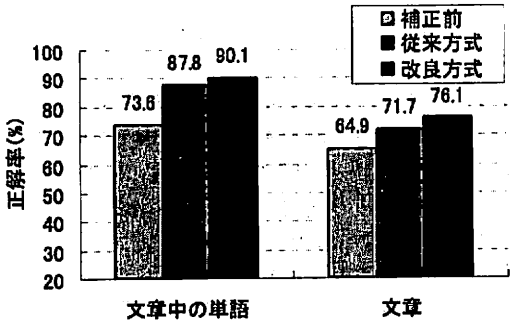

図 7．テストデータ①の評価結果

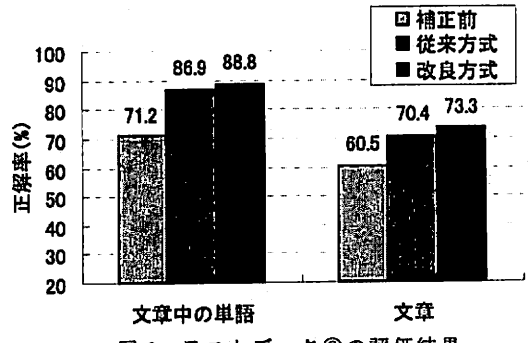

図 8．テストデータ②の評価結果

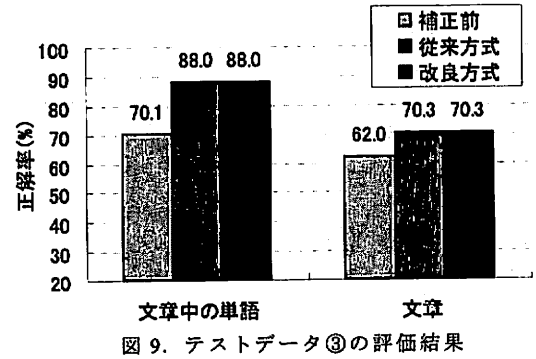

図 9．テストデータ③の評価結果

図 7，図 8 において文章の正解率を見ると，テストデータ①については従来方式より 4.4%正解率が向上し 76.1%に，テストデータ②については従来方式より 2.9%正解率が向上して 73.3%となった．また図 9 においてテストデータ③の評価を行った理由は，単語間に常識的な関連が無い文章において提案手法が適用された時誤った補正が行われるかどうかを検証するためである．しかし結果より正解率に変化は見られなかったため，単語間に常識的な関連が無い文章への影響はほとんど無いと考えられる．以上より，一致度補正方式に常識判断システムによる文章補正を加えた，音声理解システムの有効性を示した．

7. おわりに

ロボットと人間が自然なコミュニケーションを行う際に最も重要となるのは会話のための正確な音声認識である．本稿では，既存の音声認識装置による音声認識後の補正によって高精度の音声認識を得ることを目的とし，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常識判断システムを用いた文章補正を提案した．一致度補正方式においては，音響モデルから外れる音声に対して高い正解率を得ることができた．また提案した常識判断システムを用いて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．以上より，本稿で構築した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出来る．本稿では常識判断の利用例として場所，時間についての常識を用いたが，他にもいくつかの常識判断システムが構築されている．これらを上手く利用できれば，より人と人とのやり取りに近い音声認識を人とコンピュータの間で実現できると考えられる．

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．

文　献

[1] 三谷健，渡部広一，河岡司，“認識結果補正知識ベースを用いた音声理解方式”，信学技報，NLC2006-75, Vol.106, No.517, pp.13-18, Jan.2007.

[2] 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　茶筅”，http://chasen.naist.jp/hiki/ChaSen/

[3] 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，Mar.2006.

[4] 杉本二郎，渡部広一，河岡司，“概念ベースを用いた常識場所判断システムの構築”，情報処理学会自然言語処理研究会資料，2003-NL-153-11，Vol2003，No.4，pp81-88，Jan.2003.

[5] 手原信太郎，渡部広一，河岡司，“概念ベースを用いた場所連想システムの構築”, FIT2006, E-034, pp.227-229，Sep.2006.

[6] 土屋誠司，奥村紀之，渡部広一，河岡司，“連想メカニズムを用いた時間判断手法の提案”，自然言語処理，Vol.12，No.4，pp.111-129，Oct.2005.