音声言語情報処理 45−19
（２００３．２．８）

音声認識の信頼度と対話履歴を利用した最尤推定型言語理解

水谷 誠† 伊藤 敏彦‡ 甲斐 充彦†† 小西 達裕‡ 伊東 幸宏‡

† 静岡大学 情報学研究科 ‡ 静岡大学 情報学部 †† 静岡大学 工学部
〒 432-8011 静岡県浜松市城北 3-5-1

Email: †,‡ {cs7088,t-itoh,konishi,itoh}@cs.inf.shizuoka.ac.jp
†† kai@sys.eng.shizuoka.ac.jp

あらまし 音声対話インタフェースのひとつとして、カーナビゲーションシステムが注目されているが、自然発声であることや走行ノイズなどの影響による誤認識によって対話がスムーズに進まなくなり、ユーザに不快感を与えることが多い。そこで、本研究では音声認識結果の信頼度や対話履歴を利用して言語理解や応答生成を行うことで、スムーズな対話と高いユーザ満足度を得られる対話システムの構築を目指している。本稿では、その対話システムにおける音声言語理解手法について示す。単純に音声認識の信頼度を使うのではなく、発話の種類や対話履歴の情報も利用して生成されるスコアを使用する事で、対話全体において尤もらしい言語理解が可能である。評価実験結果からは、単純に音声認識結果 (n-best) の第一候補を再優先する言語理解手法よりも 10%程度言語理解性能が高いことが示された。

キーワード 音声言語理解, 音声認識信頼度, 音声対話システム

Maximum-Likelihood Spoken Language Understanding Using CSR Confidence Measure and Dialogue History

Makoto MIZUTANI†, Toshihiko ITOH‡, Atsuhiko KAI††,
Tatsuhiro KONISHI‡, and Yukihiro ITOH‡

†Graduate School of Informatics ‡Faculty of Information, Shizuoka University
††Faculty of Engineering, Shizuoka University
Johoku 3-5-1,Hamamatsu,Shizuoka,432-8011 Japan

Email: †,‡ {cs7088,t-itoh,konishi,itoh}@cs.inf.shizuoka.ac.jp
†† kai@sys.eng.shizuoka.ac.jp

Abstract Although the car-navigation system attracts attention as one of the spoken dialogue interfaces, a dialogue will not progress smoothly by miss recognition under the influence of a natural speech and a run noise, and a user will feel displeasure. Thus, this research aims at the construction of a dialogue system which can obtain a smooth dialogue and the high degree of user satisfaction by performing language understanding and response generation using the confidence measure(CM) based on continuous speech recognizer(CSR) and the dialogue history. This paper shows the spoken language understanding technique in the dialogue system. The CM is not used alone, but it is used for generating an integrated score which is generated by the CM ,the speech type and the dialogue history, and it will be possible to achieve a spoken language understanding which is more plausible for a dialogue. As the result of evaluation experiment , it was shown that 10% higher in the language understanding performance than the one in a simple language understanding technique which simply gives priority to the first hypothesis of a speech recognition result(n-best).

Keywords spoken language understanding,CSR confidence measure,spoken dialogue system

1 はじめに

近年、音声認識処理技術の高精度化により、さまざまな音声対話システムが実用化されている。カーナビゲーションシステムもその１つとして注目されている。運転という主となる行為と並行する操作としては、「手」や「目」を使わなくてはならないリモコン操作よりも、音声入力操作の方がより安全である。しかしながら、自然発声であることや走行ノイズ等の影響により、現在の音声認識処理技術では、誤認識を回避することは困難である。誤認識が起きると、システムはユーザ発話を正しく理解できないため、ユーザの意図とは異なる応答をすることになる。その結果、正しく理解された場合よりも対話がスムーズに進まなくなり、ユーザに不快感を与えることになる。そのため、音声認識の誤認識に対する様々な研究がなされている [1][2]。また、音声認識結果の信頼度を利用し

た対話制御に対する研究もなされてはいるが [3][4][6]、文脈情報と併用して言語理解を行っている研究はあまりない。

そこで、本研究では音声認識の信頼度と文脈情報を利用して言語理解や応答生成を行うことで、対話をスムーズに進行させ、高いユーザ満足度を得られる対話システムの構築を目指している。本稿では、音声認識の信頼度と対話履歴を利用した音声言語理解の手法とその評価実験結果について示す。単純に音声認識の信頼度を使うのではなく、発話の種類や対話履歴の情報も利用して生成されるスコアを使用する事で、より対話的に尤もらしい言語理解が可能であると。このような考えに基づいて言語理解手法を検討し、実際の対話を想定した評価実験結果を行った結果、提案する方法の効果を確かめることができた。

2 タスクと発話タイプ

本研究で構築する対話システムの扱うタスクは、カーナビゲーションシステムの目的地 (ランドマーク [以下、LM]) 設定である。LM は、インター名, 駅名, 市区町村名を指し、LM には県名, 自動車道名, 鉄道路線名を付加できる。これらは、図 1 のように 3 カテゴリ (PR,HR,LM) に分類され、基本的に木構造になっている。また、カテゴリの各要素はクラスと呼ばれ、クラスは全 6 種類である。

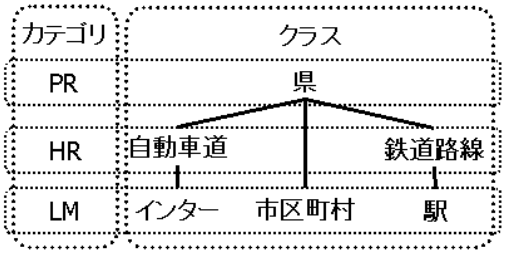

図 1: 発話内容の分類

ユーザ発話は、一度で全てのカテゴリを入力、複数回に分けて入力の両方を想定している。その発話は以下に示す通り、詳細化, 訂正, 回答, 再入力の 4 つに分類することができる。

詳細化 応答された内容に情報を追加する発話

訂正 応答された内容に対して訂正を行う発話

回答 質問を含む応答に対して回答する発話

再入力 再入力要求の応答に対する発話

また、肯定語や否定語を発話することも可能である。対話例を図 2 に示す。

U1: 静岡県の
S1: 静岡県
U2: 浜松西 IC (**詳細化発話**)
S2: 静岡県の何 IC ですか
U3: 浜松西 IC です (**回答発話**)
S3: 静岡県の浜松 IC ですか
U4: いいえ、浜松西です (**訂正発話**)
S4: もう一度発話してください
U5: 浜松西 IC です (**再入力発話**)
S5: 静岡県の浜松西 IC ですか
U6: はい
S6: 目的地に設定しました

図 2: 対話例

3 言語理解プロセス

本研究で構築しているシステムの構成は図 3 の通りである。音声認識器は SPOJUS[7] を用いている。2 節で示した発話タイプに基づいて 1600 発話 (400 発話 × 4 人) を収集し、SPOJUS で認識させた結果、単語 (キーワード) 認識率は 76%であった。

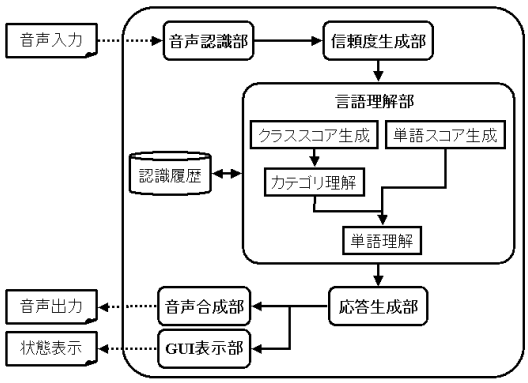

図 3: システムの構成

音声認識の結果は、n-best と呼ばれる音響的な尤度で順位付けられた複数の候補から成る。音声認識の信頼度は、単語とクラスの 2 種類について、認識結果から得られる音響的な尤度 $P(x|w)$ と n-best 中の出現頻度から事後確率に基づく尺度 $P(w|x)$ として計算される [3]。

システムの言語理解部では、音声認識結果の信頼度と対話履歴を用いてクラス・単語スコアを生成する。次に、クラススコアからカテゴリ理解を行い、最後に単語理解が行われることで言語理解内容が生成される。

スコアとは、あるクラスや単語がこれまでの文脈の中で、どれくらい発話されているかの可能性を示す値である。スコアは、言語理解の他に応答生成においても使用される。ユーザが新たな情報を追加した場合、システムはその情報だけでなく、以前に発話された情報も正しく理解しなくてはならない。これに対し、訂正発話が行われた場合には、以前のスコアを修正する

枠組みが必要である。生成されたクラス・単語スコアは、履歴に基づく統合認識結果として認識履歴に残される。

カテゴリ理解では、認識履歴のクラススコアと最新の認識結果のクラス信頼度の両方に対して、カテゴリスコアを計算する。カテゴリスコアは、同じカテゴリに属するすべてのクラスのスコア (信頼度) を足したものである。それぞれのカテゴリスコアは閾値で判定され、PR,HR,LM の 3 カテゴリに対して判定結果の論理和を計算する。そこで得られた結果が、現在までに発話されたカテゴリの組合せを示している。図 4 にカテゴリ理解の例を示す。

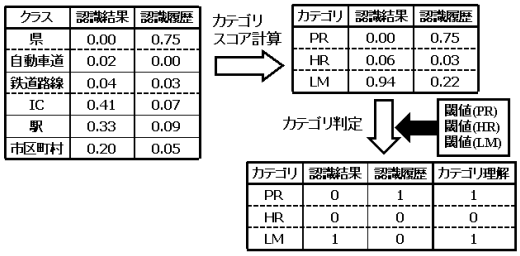

図 4: カテゴリ理解

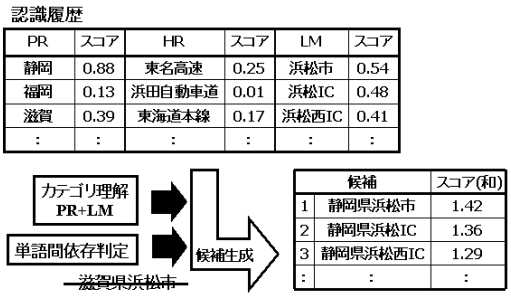

図 5: 単語理解

単語理解では、カテゴリ理解結果をもとに、認識履歴の中から最もスコアの和が高い単語の組み合わせを決定する。このとき、“滋賀県浜松市” のように実際には有り得ない組み合わせを生成しないように、単語間の依存関係を考慮している。図 5 に単語理解の例を示す。

3.1 クラススコア生成

クラススコアの生成では、認識履歴と最新の認識結果から発話タイプを判定し、発話タイプごとにスコア生成式を適用する。発話タイプは以前の情報に新しい情報を追加する働きが必要な詳細化と回答と、以前の情報を訂正する働きが必要な訂正と再入力の 2 種類に分類した。

現在、発話タイプの判定には、表 1 の 4 つの判定材料を用いている。これらの判定材料は発見的に求めたものであるため、これ以外の判定材料も存在するはずである。暫定的なものではあるが、この判定方法における判定率は、詳細化・訂正で 87.7%, 訂正・再入力で 91.4%であった。ここで用いたデータは、4 節の対話データ A である。

表 1: 発話タイプ判定

判定材料	判定結果
応答が “もう一度発話してください”	訂正・再入力
認識結果に否定語が存在する	訂正・再入力
別のカテゴリが発話された	詳細化・回答
それ以外	訂正・再入力

■詳細化・回答のクラススコア生成式

詳細化・回答は前述の通り、以前の情報に新しい情報を追加する必要がある。これらの発話タイプと判定された場合、クラス c のスコア生成式は以下の通りである。

$$Score(c) = Score(c) * weight_s + Conf(c)$$

但し、$Score$: 認識履歴のクラススコア
$Conf$: 最新認識結果のクラス信頼度
$weight_s$: 重み $(0.0 < weight_s < 1.0)$
c : スコアを生成するクラス

以前に発話された内容を考慮するために、認識履歴のクラススコアと最新認識結果のクラススコアを足す。重み $weight_s$ により一定の割合で認識履歴のクラススコアを下げているのは、“情報が古くなるごとに信頼性が低下する” という戦略を適用しているためである。この重み $weight_s$ は、4 節で説明する対話データ A を用いて、カテゴリ理解精度が最も高くなる値を実験的に求めた。生成されたクラススコアは統合認識結果として、認識履歴に残される。

■訂正・再入力のクラススコア生成式

訂正・再入力の生成式も、基本的には詳細化・回答と同じである。異なる点は、同カテゴリ異クラスの信頼度をマイナスしていることである。これにより、クラスを間違っていた場合にスコアが修正されやすくなる。

$$Score(c_a) = Score(c_a) * weight_t - Conf(c_b) + Conf(c_a)$$

但し、$Score$: 認識履歴のクラススコア
$Conf$: 最新認識結果のクラス信頼度
$weight_t$: 重み $(0.0 < weight_t < 1.0)$
c_a : スコアを生成するクラス
c_b : c_a と同じカテゴリで異なるクラス

3.2 単語スコア生成

単語は、1) 認識履歴中の単語と 2) 最新の認識結果中の単語の 2 つに分類し、それぞれ異なる

戦略を用いてスコアを生成する。後者の場合の単語は、最新認識結果の複数候補 (n-best) に含まれる全単語が対象となる。スコア生成は、言語理解部が最新の認識結果を獲得するたびに、1),2) の順で行う。

1) の認識履歴中の単語は、単語の新しさ, システムの応答内容とユーザ発話タイプ (詳細化, 訂正, 回答, 再入力) から、既存の単語スコアを上下させて、新しい単語スコアを生成する。現在は、次の 5 種類の戦略を使用している。

戦略 1 古い情報は信頼性が低くなるという仮定のもとに、新しい認識結果が入力されるたびに、認識履歴中のすべての単語スコアを下げる。

戦略 2 認識履歴中の単語 A と認識結果の単語 B が詳細化の関係にあった場合、単語 A のスコアを上げる。

戦略 3 認識履歴中の単語 A と認識結果の単語 B が訂正の関係にあった場合、単語 A のスコアを下げる。

戦略 4 認識結果に肯定語 (はい, うん 等) が含まれていた場合、応答に含まれていた単語のスコアを上げる。

戦略 5 認識結果に否定語 (いいえ, 違う 等) が含まれていた場合、応答に含まれていた単語のスコアを下げる。

2) の最新の認識結果中の単語は、応答内容と発話タイプ,n-best の順位, 発話長により、音声認識の信頼度を上下させて単語スコアを生成する。現在は、次の 4 種類の戦略を使用している。

戦略 6 認識結果の単語 A と応答に含まれる単語 B が詳細化の関係にある場合、単語 A のスコアを上げる。

戦略 7 システム応答が質問 (例. 何インターですか) で、認識結果の内容が回答である場合、認識結果の単語のスコアを上げる。

戦略 8 認識結果の上位に正解単語が多く含まれているため、上位に含まれる単語のスコアを上げる。

戦略 9 発話長が長い発話 (短い) は認識されやすい (認識されにくい) ため、1 カテゴリの結果はその単語のスコアを下げ、2 カテゴリ以上の単語はその単語スコアを上げる。

戦略 2 が適用される場合の単語 w_A の単語スコア生成式は以下の通りである。

$$Score(w_A) = Score(w_A) + weight_2 * Conf(w_B)$$

但し、$Score$: 認識履歴の単語スコア
$Conf$: 最新認識結果の単語信頼度
$weight_2$: 重み $(0.0 < weight_2 < 1.0)$
w_A : 認識履歴の単語 A
w_B : 最新認識結果の単語 B

単語スコア生成式では、この重み $weight_n$ を用いて単語スコアを計算する。すべての重みは、言語理解精度が最も高くなる値を採用している。この値は、4 節のデータ A を用いて実験的に求めた。

以上の戦略による操作を認識履歴や認識結果に対して行い、単語スコアを生成する。生成された単語スコアは、新しい統合認識結果として認識履歴に残される。この手法により、認識結果の n-best の順位に関係なく、単語スコアの高い、つまり対話的に最も可能性の高い単語がより理解内容となりやすくなる。ただし、ここに挙げた戦略は発見的に導いたものであるため、他の戦略も存在するはずである。それらの適用については今後の課題として挙げられる。

4 評価実験

本稿で示してきた言語理解手法と、音声認識結果 (n-best) の第一候補を最優先する言語理解手法の性能を比較するために評価実験を行った。実験を行うために、2 種類のシステムを作成した。一方は、最新の認識結果の n-best の第一候補を最優先する言語理解手法を採用したシステム (以下,SYS-α) で、他方は、本稿で示した言語理解手法を採用したシステム (以下,SYS-β) である。これらのシステムの言語理解部以外の性能はすべて同等である。

SYS-β には、スコア生成式に複数の重みが存在する。重みの最適な値の組合せを調べるために、全値の組合せに対して対話データを与え、言語理解精度が最も高くなるものを最適な組合せとして採用した。この対話データは、以下で説明する対話データ A を用いた。

予備実験として、情報系学部生・大学院生 5 名にあらかじめ用意しておいた浜松西インターに関する 14 発話を一人 3 回ずつ読み上げで発話してもらい、音声認識を行った。そこで、得られた発話データから、U1-S1-U2 対話データ **(対話データ A)** を生成した。そして、生成された 3909 対話データを SYS-α,SYS-β に与え、U2 発話に対するシステムの言語理解内容 (以下, 理解内容) とそれまでに発話された内容 (以下, 正解内容) を比較して、完全一致率, 単語正解精度を求めた。完全一致率とは理解内容と正解内容の各クラス単位での単語が一致している割合を示し、単語正解精度は理解内容のある単語と正解内容の単語の正誤関係を示したものである。完全一致率、単語正解精度を総称して言語理解精度と呼ぶ。

次に、SYS-β で被験者実験を行った。被験者は工・情報系学部生・大学院生 10 名である。実験は、1) 発話練習,2) システム練習,3) システム本番という順で行った。

発話練習では、あらかじめ用意しておいた浜松西インターに関する 14 種類の発話を 2 回ずつ、計 28 発話を読み上げで発話してもらい、音声認識を行った。ここで得た 10 人分の発話

データから、予備実験と同様に U1-S1-U2 対話データ (**対話データ B**) を生成した。生成された対話数は 7833 対話であった。

システム練習では、被験者は実際にシステムと対話をして、ランドマーク 2 箇所を設定した。

システム本番では、あらかじめ用意しておいたランドマーク 10 箇所をシステム練習と同様に設定した。この本実験により、10 人で 100 対話 (タスク)、361 発話のデータ (**対話データ C**) を得た。ここで得られた対話データを SYS-α に与えたときに生じる理解内容の違いについて分析・評価を行った。

4.1 発話タイプ別の言語理解精度の比較

ここでは、全く同一の対話データを 2 システムに与えることによって、それぞれのシステムの発話タイプ別の言語理解精度の違いを調べた。予備実験、本実験の発話練習により、それぞれのべ 210,280 発話を収集したが、認識ミスもしくは発話ミスにより、認識候補が 1 つもないデータは除いた。それぞれの単語認識率は、80.5,52.0%であった。予備実験は音声対話実験の経験がある被験者のみであり、本実験では経験のない被験者が含まれていたため、2 つの発話データには認識率の違いが現れた。それぞれの発話データから生成された U1-S1-U2 対話データ A,B を SYS-α,SYS-β に与え、U2 における理解内容を言語理解精度を用いて比較を行った。結果を表 2 に示す。結果では、SYS-α よりも SYS-β のほうが完全一致率, 単語正解精度ともに高いことから、本稿で示した言語理解手法は有効であることが言える。対話データ A,B で大きな違いがあったのは、それぞれの評価データの認識率に 30%の違いがあったことに原因があると考えられる。2 つの対話データの認識率に違いはあったが、どちらの対話データに対しても SYS-β は SYS-α よりも有効であった。また、置換, 脱落が以前の結果に比べて増加していることから、以下の影響を受けている可能性があると考えた。

置換の増加
置換が増加している要因としては、1) カテゴリ理解失敗,2) 単語理解失敗の 2 つが挙げられる。

脱落の増加
脱落が増加している要因として考えられるのは、主にカテゴリ理解失敗である。

そこで、カテゴリ理解失敗の影響がどの程度あるかを調べるために、SYS-β のカテゴリ理解精度を 100%のシステム (以下,SYS-γ) を作成し、対話データ B を与えて言語理解精度を調べた (表 2)。結果から、カテゴリ理解精度が上がることにより、完全一致率が上がることが分かる。よって、カテゴリ理解の枠組みについては今後も検討する必要があると言える。

単語正解精度では、SYS-γ の単語脱落率が SYS-β よりも増加したために精度が低下した。現在の単語理解の枠組みでは、カテゴリ理解結果に合う単語の組合せが生成できない場合は、単語理解失敗である、つまりシステムは “何も理解できない” という状態になる。SYS-γ ではカテゴリ理解精度を無理矢理 100%にしたため、単語理解失敗という状態が起こりやすくなる。そのため、結果的に SYS-β よりも SYS-γ の方が正解が減少し、脱落が増加したと考えられる。単語理解失敗の状態の対処法については、現在の枠組みでは解決できないため、今後検討すべき点である。

4.2 発話タイプの出現頻度を考慮した言語理解精度の比較

SYS-β を用いて、ランドマーク設定を行った対話データ C を用いた評価を行った。このデータを用いることにより、発話タイプの出現頻度を考慮した言語理解精度の比較をすることが可能である。この評価では、対話データを SYS-α,SYS-β に与え、理解内容と正解内容を発話毎に比較し、最終的に対話データ全体でどちらのシステムの理解内容の言語理解精度が高いのかに注目する。また、SYS-γ でも同様に言語理解精度を求めることで、カテゴリ理解失敗の影響がどの程度あるのかを調べる。

結果 (表 3) から、対話データ A,B での結果と同様に、本稿で示した言語理解手法を採用したシステムの方が、完全一致率, 単語理解精度ともに有効であることが分かる。また、SYS-γ の結果から、カテゴリ理解精度の向上により、言語理解精度の更なる向上を見込むことができる。SYS-β よりも SYS-γ の精度が向上している要因として、置換, 挿入数が減少したことが挙げられる。SYS-β が置換, 挿入を引き起こす原因は以下の通りである。

置換 カテゴリ理解失敗により正解と異なるカテゴリの組合せで単語理解を行うために生じる。

挿入 カテゴリ理解失敗により必要ないカテゴリを含んだ組合せで単語理解を行うために生じる。

これらは、そのカテゴリが発話されたか否かを調べるカテゴリ判定が失敗しているために発

表 2: 対話データ A,B における各システムの言語理解精度

全体 [**対話データ** A:3909 **対話** (9537 **単語**), **対話データ** B:7833 **対話** (18422 **単語**)]							
システム	データ	完全一致 (%)	正解 (%)	置換 (%)	挿入 (%)	脱落 (%)	単語正解精度 (%)
α	A	**2649(67.8)**	8623(90.4)	661(6.9)	632(6.6)	253(2.7)	**83.8**
β	A	**3255(83.3)**	8830(92.6)	484(5.1)	95(1.0)	223(2.3)	**91.6**
α	B	**2601(33.2)**	11191(60.8)	4563(24.8)	664(3.6)	2668(14.5)	**57.1**
β	B	**3403(43.4)**	12964(70.4)	2909(15.8)	382(2.1)	2549(13.8)	**68.3**
γ	B	**4150(53.0)**	12995(70.5)	2806(15.2)	0(0.0)	2621(14.2)	**70.5**
詳細化 [**対話データ** A:2896 **対話** (7188 **単語**), **対話データ** B:4462 **対話** (11287 **単語**)]							
システム	データ	完全一致 (%)	正解 (%)	置換 (%)	挿入 (%)	脱落 (%)	単語正解精度 (%)
α	A	**1911(66.0)**	6510(90.6)	479(6.7)	533(7.4)	199(2.8)	**83.2**
β	A	**2384(82.3)**	6620(92.1)	376(5.2)	78(1.1)	192(2.7)	**91.0**
α	B	**1507(33.8)**	6770(60.0)	2603(23.1)	326(2.9)	1914(17.0)	**57.1**
β	B	**2060(46.2)**	8096(71.7)	1404(12.4)	201(1.8)	1787(15.8)	**70.0**
γ	B	**2345(52.6)**	7971(70.6)	1528(13.5)	0(0.0)	1788(15.8)	**70.6**
訂正 [**対話データ** A:208 **対話** (444 **単語**), **対話データ** B:1595 **対話** (2996 **単語**)]							
システム	データ	完全一致 (%)	正解 (%)	置換 (%)	挿入 (%)	脱落 (%)	単語正解精度 (%)
α	A	**110(52.9)**	367(82.7)	72(16.2)	26(5.9)	5(1.1)	**76.8**
β	A	**140(67.3)**	379(85.4)	45(10.1)	16(3.6)	20(4.5)	**81.8**
α	B	**506(31.7)**	1975(65.9)	890(29.7)	260(8.7)	131(4.4)	**57.2**
β	B	**553(34.7)**	2001(66.8)	726(24.2)	171(5.7)	269(9.0)	**61.1**
γ	B	**895(56.1)**	2246(75.0)	638(21.3)	0(0.0)	112(3.7)	**75.0**
回答 [**対話データ** A:700 **対話** (1785 **単語**), **対話データ** B:1336 **対話** (3614 **単語**)]							
システム	データ	完全一致 (%)	正解 (%)	置換 (%)	挿入 (%)	脱落 (%)	単語正解精度 (%)
α	A	**603(86.1)**	1636(91.7)	100(5.6)	0(0.0)	49(2.8)	**91.7**
β	A	**642(91.7)**	1723(96.5)	55(3.1)	0(0.0)	7(0.4)	**96.5**
α	B	**403(30.2)**	2141(59.2)	862(23.9)	28(0.8)	611(16.9)	**58.5**
β	B	**559(41.8)**	2536(70.2)	654(18.1)	2(0.1)	**424(11.7)**	**70.1**
γ	B	**568(42.5)**	2364(65.4)	563(15.6)	0(0.0)	**687(19.0)**	**65.4**
再入力 [**対話データ** A:105 **対話** (120 **単語**), **対話データ** B:440 **対話** (525 **単語**)]							
システム	データ	完全一致 (%)	正解 (%)	置換 (%)	挿入 (%)	脱落 (%)	単語正解精度 (%)
α	A	**25(23.8)**	110(91.7)	10(8.3)	73(60.8)	0(0.0)	**30.8**
β	A	**93(88.6)**	108(90.0)	8(6.7)	1(0.8)	4(3.3)	**89.2**
α	B	**185(42.1)**	305(58.1)	208(39.6)	50(9.5)	12(2.3)	**48.6**
β	B	**231(52.5)**	331(63.1)	125(23.8)	8(1.5)	69(13.1)	**61.5**
γ	B	**342(77.7)**	414(78.9)	77(14.7)	0(0.0)	34(6.5)	**78.9**

表 3: 対話データ C における各システムの言語理解精度

対話データ C:100 **対話** (361 **発話**,901 **単語**)							
システム	データ	完全一致 (%)	正解 (%)	置換 (%)	挿入 (%)	脱落 (%)	単語正解精度 (%)
α	C	**212(58.7)**	656(72.8)	101(11.2)	26(2.9)	144(16.0)	**69.9**
β	C	**257(71.2)**	776(86.1)	**54(6.0)**	**17(1.9)**	71(7.9)	**84.2**
γ	C	**301(83.4)**	789(87.6)	**34(3.8)**	**0(0.0)**	78(8.7)	**87.6**

生する。言語理解精度の向上を図るためには、このカテゴリ理解についての検討が必要である。

5 まとめ

本稿では、音声認識結果の信頼度と対話履歴を利用した言語理解手法を示した。その手法では、n-best と呼ばれる複数候補を持つ音声認識結果とその信頼度、認識結果の履歴、応答内容との関係などから、認識結果中の全単語に対してスコアを生成し、そのスコアを用いて、最尤な理解内容を生成する。評価実験結果から、n-best の第一候補を再優先する言語理解手法よりも 10%程度精度が高いことを示した。しかし、本稿の手法でもカテゴリ理解の失敗により全体の精度を低下させていることが分かった。

今後の課題としては、カテゴリ理解の枠組みについて再検討し、全体の言語理解性能の向上を図ることが挙げられる。

参考文献

[1] 甲斐充彦, 石丸明子, 伊藤敏彦, 小西達裕, 伊東幸宏, “目的地設定タスクにおける訂正発話の特徴分析と検出への応用", 日本音響学会全国大会論文集,2-1-8,pp.63-64,2001.

[2] 平沢純一, 宮崎昇, 相川清明, “質問-応答連鎖からの音声対話システムの誤解の検出", 情報処理学会研究報告,SLP-34-41,pp.239-244,2000.

[3] 駒谷和範, 河原達也, “音声対話システムにおける音声認識結果の信頼度の利用法", 日本音響学会講演論文集,3-5-2,pp.73-74,2000.

[4] 新美康永, 小林豊, “音声認識の信頼性に基づいた対話制御方式" , 信学技法,SP96-30,pp.75-80,1995.

[5] X.Pouteau,E.Krahmer and J.Landsbergen “Rubust Spoken Dialogue Management for Driver Information Systems", Proc. Eurospeech ’97,pp.2207-2210,1997.

[6] D.J.Litman,M.A.Walker and M.J.Kearns, “Automatic detection of poor speech recognition at the dialogue level", Proc. 37th Annual Meeting of the Association of Computational Linguistics(ACL99),pp.309-316,1999.

[7] 中川聖一, 甲斐充彦, “文脈自由文法制御による One Pass 型 HMM 連続音声認識法", 電子通信学会論文誌,Vol.J76-D-II,No.7,pp.1337-1345,1993.

[8] 甲斐充彦, 伊藤克亘, “対話システムにおける音声認識”, 情報処理学会研究報告,SLP-33-2,pp.7-12,2000.