Vol.2010-SLP-82 No.21
2010/7/24

バージイン許容音声対話システムにおけるユーザ発話の分析と指示対象同定への応用

松山匡子[†1] 駒谷和範[†2] 武田龍[†1]
尾形哲也[†1] 奥乃博[†1]

本稿は，バージイン許容列挙型音声対話におけるユーザ発話の分析と，分析結果を応用した指示対象同定手法の拡張について報告する．バージイン許容音声対話では，個々のユーザやシステムの発話内容によってユーザの発話タイミングや発話表現が異なる．そこでこれらを事前確率として反映させ，発話意図解釈の性能向上を図る．我々はまず，ニュース読み上げとクイズの 2 つの列挙型対話システムで収集したユーザ発話 1584 発話を分析し，ユーザの参照表現発話率が個々のユーザやシステムの列挙項目長に依存することを明らかにした．さらに，これらの特性を指示対象同定の枠組みに組み込み，タイミングと音声認識結果の解釈の事前確率として反映させる．この事前確率の推定には，ロジスティック回帰を用いる．事前確率として一定値を用いた場合に比べて，指示対象同定精度が最大 6.2 ポイント向上することを実験により確認した．

Analysis of User Utterances and Application to Identify User's Referent in Barge-in-able Spoken Dialogue System

KYOKO MATSUYAMA,[†1] KAZUNORI KOMATANI,[†2]
RYU TAKEDA,[†1] TETSUYA OGATA[†1] and HIROSHI G. OKUNO[†1]

This paper reports the extension of identification method based on analyses of user utterance in barge-in-able spoken dialogue system which reads out items. Generally, user's behaviors such as barge-in timing and utterance expressions vary in accordance with the user's preference and the content of system utterances. To interpret users' intention robustly, first, we analyze 1584 utterances collected by our systems with quiz and news-listing tasks and reveal that the ratio of using referential expressions depends on individual users and average lengths of listed items. Second, we incorporate this tendency as a prior probability into our probabilistic framework for identifying user's intended item. This prior probability is calculated by logistic regression. Experimental results show that our method improves the identification accuracy by as many as 6.2 points in the best case over the non-informative prior.

1. はじめに

我々は，ユーザのバージインタイミング情報と音声認識結果を確率的に統合する手法を開発し，列挙型音声対話における指示対象同定問題へ応用してきた[1)]．バージインとはユーザのシステム発話への割り込み行為を指し，ユーザのバージインには，システム発話の内容の指示参照，ユーザ発話の強調などの意図が含まれる．そのため，音声認識結果だけでなく，バージインタイミング情報も活用することで，音声認識誤りに頑健な対話遂行が可能となる．従来バージインに関する研究は，そのほとんどがバージインの高速かつ正確な検出[2),3)]や，検出後のシステムのしかるべき挙動の検証[4),5)]が目的であり，バージインをユーザの意図解釈に用いた研究はなかった．バージインタイミング情報をユーザの意図解釈に有効に活用できる対話に，システムが項目を列挙し，ユーザがその中の一つを指示する列挙型対話（図 1）がある．列挙型対話は，(1) 情報検索タスクの結果出力部では必須の対話であり，頑健な対話遂行が要求される，(2) ユーザのバージインタイミングを生かした直感的なインタラクションが実現できる，(3) バージインタイミングは比較的簡単に検出できるので，音声認識が困難な環境での音声対話にも応用できる，という点から重要である．バージインタイミングの利用に加えて本稿では，個々のユーザやシステムの列挙項目の内容に応じてユーザの発話行動が異なる点に注目し，これらを新たに事前確率として解釈に反映させることで発話意図解釈性能のさらなる向上を図る．

本稿ではまず，2 つの異なるタスクの列挙型対話における，ユーザの振る舞いを分析した結果を報告する．タスクの内容はニュース読み上げタスク，クイズタスクの 2 種類である．前者は，ユーザはシステムが読み上げる列挙項目を吟味してから指定できる．対して，後者は，クイズの答えが分かった時点で即座に指定できる．そのため，これらのタスクでは，ユーザのバージインタイミングや発話表現の傾向は異なる可能性がある．ここで，タイミングまたは項目内容の発話だけで意図を伝えるような，ユーザの特性を示す共通の特徴や，列挙項目の内容など，タスク毎に異なる特徴を調査する．

次に，分析により得られた特徴を利用し，バージインタイミング情報と音声認識結果に対

†1 京都大学大学院 情報学研究科 知能情報学専攻
Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
†2 名古屋大学大学院 工学研究科 電子情報システム専攻
Dept. of Electrical Engineering and Computer Science, Graduate School of Engineering, Nagoya University

Vol.2010-SLP-82 No.21
2010/7/24

System: 詳しく聞きたいニュースがあれば，指定してください．
『東京ペットショップ，毒ヘビ販売事件』,『東京江戸川区で不審火が 5 件相次ぐ ･･･』
User: それ！
System: 2 つ目のニュースですね？東京江戸川区で，･･･

図 1 対話例（ニュース読み上げタスク）
Fig. 1 *Dialogue example in news-listing task*

する事前確率という形で，指示対象同定の枠組みに反映させる．ユーザやタスク毎の違いを反映した特徴を用いて事前確率を計算することにより，それらに応じて，音声認識結果を重視したり，バージインタイミングを重視した指示対象同定が実現できる．

本稿は 2 章で指示対象同定のフレームワークについて説明し，3 章でユーザ発話の分析について述べる．4 章でロジスティック回帰による事前確率推定の最適化と評価実験について説明する．最後に 5 章で本稿のまとめを示す．

2. バージインタイミング情報と音声認識結果を用いた最尤推定による指示対象同定

本章では，列挙型対話における，バージインタイミング情報と音声認識結果を用いた指示対象同定の枠組みについて説明する．図 2 は，指示対象同定のフローを表している．我々は指示対象同定問題を，$P(T_i|U)$ を最大化する T_i を求める問題として定式化する．ここで，T_i はシステムが列挙する i 番目の項目，U はユーザ発話を表す．事前確率 $P(T_i)$ は等確率とし，$P(U)$ は i に依存しないとすると，以下式が成り立つ．

$$T = \underset{T_i}{\operatorname{argmax}}\, P(T_i|U) = \underset{T_i}{\operatorname{argmax}}\, \frac{P(U|T_i)P(T_i)}{P(U)} = \underset{T_i}{\operatorname{argmax}}\, P(U|T_i). \quad (1)$$

式（1）中の $P(U|T_i)$ は，バージインタイミング情報により解釈される場合 C_1 と音声認識結果により解釈される場合 C_2 の 2 つの解釈から算出される．

$$P(U|T_i) = \sum_{k=1}^{2} P(U|T_i, C_k)P(C_k|T_i) \quad (2)$$

$$= (1-\alpha)P(U|T_i, C_1) + \alpha P(U|T_i, C_2). \quad (3)$$

T_i に対して，C_k の解釈の場合に U が発生する確率を $P(U|T_i, C_k)$ で表す．$P(U|T_i, C_1)$ は，ユーザの発話タイミングに対して仮定したガンマ分布を用いて計算する．ここで，発話タイミングは，各項目毎にシステムの発話開始時間とユーザの発話開始時間の差である（図 3）．$P(U|T_i, C_2)$ は，ユーザ発話の音声認識結果と，各項目 T_i をベクトル空間モデルで表現した際のコサイン距離を用いて計算する．$P(U|T_i, C_k)$ の詳細は文献[1] を参照されたい．どちらの解釈を重視するかの事前確率 $P(C_k|T_i)$ を α で表すと，式（3）が成り立つ．

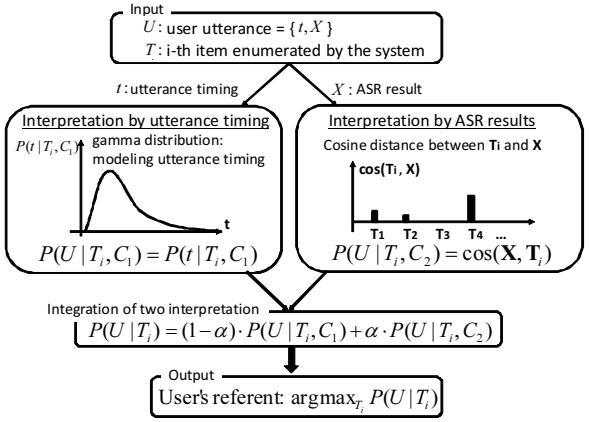

図 2 指示対象同定フロー
Fig. 2 *Flow of identifying user's referent*

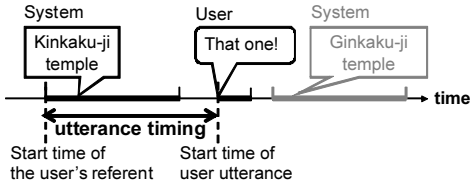

図 3 発話タイミングの定義
Fig. 3 *Definition of utterance timing*

3. 列挙型対話における発話表現の分析

タイミングの解釈 $P(U|T_i, C_1)$ と音声認識結果の解釈 $P(U|T_i, C_2)$ の事前確率 $P(C_k|T_i)$ をユーザの発話毎に計算することで，指示対象同定精度の向上を図る．これは，以下のような状況に応じて事前確率を変化させることに相当する．例えば,「それ！」と発話してタイミ

System: サッカー選手は誰でしょう？
ジャッキー・チェン，デビッド・ベッカム，ペ・ヨンジュン ···
User: ベッカム！
System: ベッカムですね？正解です.

図 4 対話例（クイズタスク）
Fig. 4 *Dialogue example in quiz task*

ングで指定するのを好むユーザがいれば，列挙項目をいくつか聞いた後，列挙項目中の内容語（助詞以外の語）を用いて指定することを好むユーザもいる. また，システムの列挙する項目が長く，内容語を用いて指定しづらい場合は，ユーザは簡潔に「それ！」などとタイミングを用いて指定することもある. このように，タイミングによる指定を多用するユーザや，タイミングで指定されやすいシステムの列挙内容に対しては，音声認識結果よりタイミングを重視することで，頑健な指示対象同定が可能である．本章では，実際に対話システムを実装し，運用したデータを用いて，ユーザの振る舞いとシステム発話との関係を分析する．

3.1 2 つのタスク間のユーザの発話表現の傾向

まず，2 つの異なるタスクをもつ列挙型対話の仕様と，分析対象データを示す．1 つ目のタスクでは，システムは 10 種の RSS フィードから自動的にニュースを取得し，そのニュースタイトルを列挙する（図 1）．ユーザは，自分が興味を持った項目を指定することで，そのニュースの内容を詳しく知ることができる．2 つ目のタスクでは，システムは 1 つのクイズに対して 8 つの選択肢を列挙し，ユーザが正解だと思ったものを指定する．クイズは全部で 40 種類用意した．対話例を図 4 に示す．各タスクのシステム発話の特徴を表 1 に示す. ニュース読み上げタスクでは，各列挙項目はニュースの内容を示す 1 文であり，ユーザにとっての未知の単語や予測不能な単語が含まれる．一方，クイズタスクでは，クイズの答えとなる選択肢はほとんど 1 単語である．それぞれの列挙項目長の平均は，ニュース読み上げタスクでは 5.65 秒，クイズタスクでは 1.59 秒である．ユーザが指定する際の違いとして，ニュース読み上げタスクは，ユーザが自分の興味のある内容を吟味して発話できるのに対し，クイズタスクは，正解候補が列挙された時点で即座に発話できる．前者では，20 名の被験者から 400 発話を収集し，後者では 31 名の被験者から 1184 発話を収集した.

次に，これらの 2 つのタスクにおける，ユーザが列挙項目を指示する発話表現の傾向について分析する．2 つのタスクにおける参照表現発話と内容表現発話の内訳を表 2 に示す．ここで，参照表現発話を「それ」などの発話とし，発話の書き起こしからは指示対象を同定できない発話と定義する．一方，内容表現発話は，システム発話に含まれる名詞などのキーワードを含む発話で，書き起こしから指示対象を同定できる発話と定義する．前者の場合，ユーザはタイミングを用いて指定しており，後者の場合は発話内容を用いて指定している．表 2 から，2 つのシステムにおける参照表現発話率が異なることがわかる．ニュースタイトルの読み上げタスクにおいては，ユーザは参照表現発話を用いることが多く，クイズタスクにおいては，ユーザは内容表現発話を用いることが多い．これは，ニュースタイトル読み上げタスクは，列挙項目が 1 文であり，未知語が含まれることがあるので，ユーザがタイミングを用いて指定しやすいからである．また，クイズタスクでは，クイズの答えとなる選択肢は既知の 1 単語であることが多く，ユーザにとって内容語を用いて指定しやすい．この結果から我々は，タスクによる違いを各列挙項目長の違いとして捉え，事前確率 $P(C_k|T_i)$ の推定の際の特徴として用いる.

表 1 2 つの列挙タスクのシステム発話の特徴
Table 1 *Number of user utterances in two tasks*

タスク	平均列挙項目長	列挙項目の特徴
ニュース読み上げ	5.65 秒	未知語が多い
クイズ	1.59 秒	既知語が多い

表 2 2 つの列挙タスクにおける発話表現数の内訳
Table 2 *Number of user utterances in two tasks*

タスク	参照表現	内容表現	合計
ニュース読み上げ (被験者数 20 名)	263 (65.7%)	137 (34.3%)	400
クイズ (被験者数 31 名)	434 (36.7%)	750 (63.3%)	1184

3.2 参照表現発話率の傾向と仮説

2 つのタスクにおける，ユーザの参照表現発話の使用率を図 5，7 に示す．横軸は被験者数を表し，縦軸は参照表現発話率（参照表現を用いた回数/ユーザの全発話数）を示す．図 5 では 8 名の被験者が，図 7 では 2 名の被験者が 80 %の頻度で参照表現発話を用いる．このように，同一タスク内でも，ユーザによって参照表現発話の使用率は異なる.

次に，各ニュースや各クイズ毎の参照表現発話率を分析する．参照表現発話率とシステムの平均列挙項目長の相関を図 6，8 に示す．横軸はシステムが 1 ニュースおよび 1 クイズにおいて列挙する項目の平均長，縦軸は参照表現発話率を示す．図中の 1 点はそれぞれ 1 ニュースや 1 クイズを示す．図から，列挙項目長が長ければ長いほど，ユーザは参照表現発話を多用することがわかる．相関係数は，図 6 において 0.51，図 8 において 0.81 である．ニュー

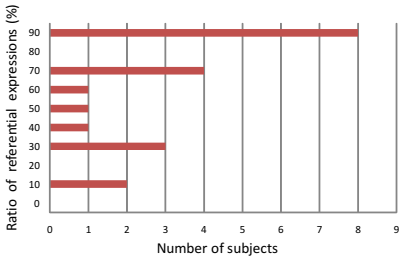

図 5 ニュース読み上げタスクにおける，被験者の参照表現発話率
Fig. 5 *Ratio of referential expression for subjects in news-listing task*

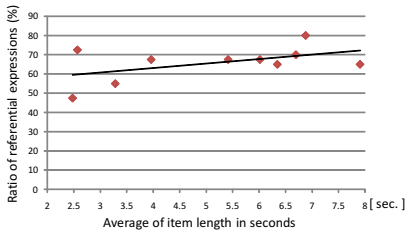

図 6 ニュース読み上げタスクにおける，列挙項目長と参照表現発話率
Fig. 6 *Ratio of referential expression for item lengths in news-listing task*

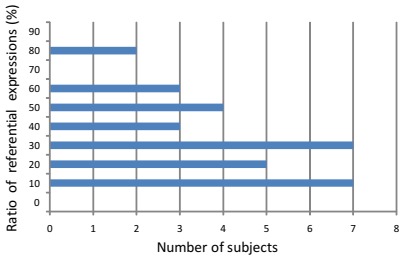

図 7 クイズタスクにおける，被験者の参照表現発話率
Fig. 7 *Ratio of referential expression for subjects in quiz task*

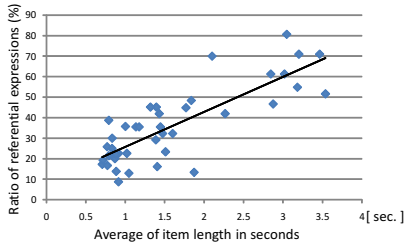

図 8 クイズタスクにおける，列挙項目長と参照表現発話率
Fig. 8 *Ratio of referential expression for item lengths of items in quiz task*

ス読み上げタスクで相関が小さいのは，システムの列挙項目長が比較的短くても，ユーザにとっての未知語が多く含まれることがあり，参照表現を多用するユーザが多かったからである．

ユーザ発話の分析により明らかになったことは，(1) 同一タスク内でもユーザによって参照表現発話率は異なること，(2) 参照表現発話率はシステムの列挙項目長に比例することの 2 点である．つまり，ユーザの参照表現発話率やシステムの列挙項目長などを特徴として，タイミングと音声認識結果の解釈の事前確率を推定することで，ユーザやシステムの特性に応じた指示対象同定が可能であることが示唆される．

表 3 ユーザ特性やシステムに関する特徴
Table 3 *Features of user's and sysetm's characteristics*

F_1:	ユーザの音声認識精度
F_2:	ユーザの参照表現発話の割合
F_3:	システム発話の平均列挙長 [秒]

4. ユーザの発話傾向の事前確率値推定と評価

本章ではまず，個々のユーザやシステムの列挙項目の特性をタイミングと音声認識結果の解釈の事前確率に反映させるための，ロジスティック回帰を用いた α の推定手法について述べる．次に，実験条件と評価実験について説明し，ユーザやシステムの特性が回帰学習時に反映されているかを考察する．

4.1 ロジスティック回帰による α の推定

我々は，ロジスティック回帰[6] を用いて α を推定する．ロジスティック回帰では，ユーザの特性やシステムの平均列挙項目長などから得られる特徴を用いて，式 (4) のように計算する．

$$\alpha = \frac{1}{1+\exp(-(a_1F_1 + a_2F_2 + a_3F_3 + b))}. \quad (4)$$

ロジスティック回帰により，それぞれの特徴の関係を自動的に推定し，α を計算できる．式 (4) 中の係数 $a_1, \cdots, a_3$ は学習データから推定する．学習の際の教師信号は，0 か 1 であり，音声認識結果だけでユーザの指示対象が同定可能ならば 1，それ以外は 0 とした．

回帰に用いた特徴 $F_1, \cdots, F_3$ を表 3 に示す．学習データを利用し，これらの特徴は事前に平均 0，分散 1 となるように正規化している．F_1 は当該ユーザの音声認識結果の単語正解精度を表し，一発話ごとに逐次的に取得されるものではなく，各ユーザ毎に収集した全発話データから算出する．F_2 は参照表現発話率であり，これはユーザの発話毎に逐次的に更新される．参照表現発話とは，音声認識結果に列挙項目中の内容語が含まれていない発話である．特徴 F_3 は，各列挙対話においてシステムが列挙する項目の平均長とした．F_1，F_2 はユーザに関する特性であり，F_3 はシステムに関する特性である．

4.2 実験条件

評価データはニュース読み上げタスクにおいて収集した 400 発話とクイズタスクにおいて収集した 1184 発話である．被験者は，ヘッドセットなどの接話型マイクロフォンではなく，ロボットに組み込まれたマイクロフォンに話しかけるように発話していたため，ICA に

表 4 ニュース読み上げタスクにおける指示対象同定精度 [%]
Table 4 *Identification accuracy [%] for user utterances in news-listing task*

α	参照表現 (#:263)	内容表現 (#:137)	合計 (#:400)
0.0 （タイミングのみ）	86.7 (#:228)	32.1 (#:44)	68.0 (#:272)
0.3 （ベースライン）	80.6 (#:212)	47.4 (#:65)	69.3 (#:277)
1.0 （音声認識結果のみ）	0.38 (#:1)	46.0 (#:63)	16.0 (#:64)
Oracle	84.8 (#:223)	71.5 (#:98)	80.3 (#:321)
本手法	83.7 (#:220)	59.9 (#:82)	75.5 (#:302)

表 5 クイズタスクにおける指示対象同定精度 [%]
Table 5 *Identification accuracy [%] for user utterances in quiz task*

α	参照表現 (#:434)	内容表現 (#:750)	合計 (#:1184)
0.0 （タイミングのみ）	98.4 (#:427)	84.3 (#:632)	89.4 (#:1059)
0.1 （ベースライン）	97.9 (#:425)	87.1 (#:653)	91.0 (#:1078)
1.0 （音声認識結果のみ）	9.45 (#:41)	59.2 (#:444)	41.0 (#:485)
Oracle	98.8 (#:429)	94.1 (#:706)	95.9 (#:1135)
本手法	97.9(#:425)	88.2 (#:662)	91.8 (#:1087)

基づく音源分離手法[7] を用いてユーザ発話を分離した．音声認識器は Julius[8] を用いている．各タスクにおいて用いた言語モデルと音声認識精度を以下に示す．

ニュース読み上げタスク 言語モデルは，10 種の RSS フィード毎に統計的言語モデルを作成した．それぞれ列挙するニュースタイトルとコマンド発話 115 発話を用い，10 種の言語モデルの平均語彙サイズは 230.5，音声認識精度は 58.6 %である．

クイズタスク 言語モデルは，40 種のクイズ毎に統計的言語モデルを作成した．それぞれ列挙項目とコマンド発話 118 発話を用い，40 種の言語モデルの平均語彙サイズは 123.1，音声認識精度は 39.6 %である．

コマンド発話として,「それ教えて」,「今のニュース聞きたい」など，想定されるユーザ発話を事前に人手で記述した．音声認識精度が低い原因として,（1）ユーザ発話を分離する際の雑音の消し残りや残響,（2）被験者は音声対話システムに不慣れなため，早口であったり音量が小さい，の 2 点がある．本実験は音声認識精度が低いが，実際の運用に適した状況で実験した結果である．ロジスティック回帰における係数は，10 分割交差法で学習した．

4.3 評 価 実 験

表 4，5 に，事前確率として一定値を与えた場合の指示対象同定のベースライン，α を最適化した場合の指示対象同定精度のオラクル，さらにロジスティック回帰により α を推定した際の指示対象同定精度を示す．ここで，同定精度のオラクルは以下の手順で計算される．

(1) すべてのユーザ発話に対して，α を 0.0 から 1.0 まで 0.1 刻みで変化させた場合の $P(U|T_i)$ を計算する．

(2) いずれかの α で指示対象を正しく同定できてる場合，その発話は同定可能発話とみなす．

表 6 本手法により指示可能となった発話例（クイズタスク）
Table 6 *Example of identifiable utterance by our method*

ユーザ発話の書き起こし:「4 番目に言ったやつ」,
ユーザ発話の音声認識結果:「4 番目って」，正解番号: 4
割り込んだタイミング: 5 番目の項目の直後
各特徴値: $F_1 = 0.57$，$F_2 = 0.07$，$F_3 = 3.08$

α	$P(U\|T_3)$	$P(U\|T_4)$	$P(U\|T_5)$	同定番号
0.0 （タイミングのみ）	1.03E-4	6.63E-3	**2.68E-1**	5
0.1	9.28E-5	7.63E-2	**2.41E-1**	5
1.0 （音声認識結果のみ）	0.00	**7.02E-1**	0.00	4
本手法 ($\alpha = 0.47$)	5.46E-5	**3.35E-1**	1.42E-1	4

表 4 は，ニュース読み上げタスクにおける指示対象同定精度を示している．$\alpha = 0.0$ の場合はタイミングによる解釈のみの場合に相当し，$\alpha = 1.0$ の場合は音声認識結果による解釈のみの場合に相当する．表中の $\alpha = 0.3$ は，上記の手順（1）で，最も精度が高い場合であり，事前確率として一定値を与えた場合のベースラインである．オラクルでは，ベースラインと比べて，同定精度が 11.0 ポイント向上している．ロジスティック回帰により事前確率を推定した本手法では，ベースラインに比べて新たに 25 発話同定可能となった．

表 5 は，クイズタスクにおける指示対象同定精度を示している．$\alpha = 0.1$ は，上記の手順（1）で最も精度が高く，事前確率として一定値を与えた場合のベースラインである．オラクルでは，ベースラインと比べて，同定精度が 4.9 ポイント向上している．さらに，本手法では新たに 9 発話が同定可能となった．クイズタスクにおいて，同定可能となった発話例を表 6 に示す．この発話に付与される特徴は，$F_1 = 0.57$，$F_2 = 0.07$，$F_3 = 3.08$ である．F_2 からわかるように，この発話をしたユーザは，参照表現をあまり用いず，発話内容で項目を指示する傾向にある．また，F_3 から，システムの列挙項目長は比較的長い対話状況であったことがわかる．この発話は，ベースラインである $\alpha = 0.1$ の場合には，タイミングの解釈の事前確率値が大きく，5 番目の項目が指示対象とみなされる．しかし本手法では，ユー

表 7 ロジスティック回帰時の係数の平均と標準偏差（ニュース読み上げタスク）
Table 7 *Average and standard deviation of each coefficient in news-listing task*

係数	平均	標準偏差
a_1	0.05	0.08
a_2	-1.40	0.12
a_3	-0.01	0.07
b	-1.67	0.10

表 8 ロジスティック回帰時の係数の平均と標準偏差（クイズタスク）
Table 8 *Average and standard deviation of each coefficients in quiz task*

係数	平均	標準偏差
a_1	0.12	0.06
a_2	-0.76	0.04
a_3	0.19	0.03
b	-1.40	0.04

ザやシステムの特性を反映するように音声認識結果の解釈の事前確率値が大きく推定され，正しく 4 番目を同定できた．

4.4 考　　察

本節では，3 章におけるユーザ発話傾向の分析結果が，ロジスティック回帰における係数値に反映されているかを検証する．表 7，8 はロジスティック回帰において，10 分割交差法で学習した係数の値の平均と標準偏差である．特徴 F_1 の係数 a_1 は正の値であり，これは F_1 が大きければ音声認識結果を信頼する度合いが高く，音声認識結果の解釈を多く見積もるべく α を大きくする傾向に一致する．特徴 F_2 の係数 a_2 は負の値である．これは，参照表現発話の可能性が高ければ，α を小さくすべき傾向と一致する．さらに，a_2 は他の係数に比べて値が大きく，回帰の際に比較的重視される特徴である．F_3 の係数 a_3 は，ニュースタスクでは負の値であり，クイズタスクでは正の値であった．ユーザ発話の分析から，F_3 の値が大きい場合は参照表現発話の割合が高いため，α を小さく見積るべきであり，a_3 は負の値を取ると予想された．クイズタスクにおいては，ユーザが音声認識結果で指定しても，ユーザの発話タイミングから指示対象を同定できることが多かった．例えば，図 4 の対話例では，システムが「ベッカム」と発話した直後にユーザが「ベッカム」と内容語で指定した場合，音声認識結果が誤っても，タイミングで指示対象を同定できる．そのため，列挙項目長が長ければタイミングによる解釈，短ければ音声認識結果による解釈の事前確率を大きく見積もるべきだという事前に考えられた予想とはずれた係数になった. これは，参照表現発話ではタイミングでしか認識できないが，内容表現発話ではタイミングと音声認識結果を用いて認識可能だからである．つまり，ユーザがタイミングで指定しやすいクイズのようなタスクでは，内容表現発話の場合でもタイミングに対する事前確率を高く設定することで，音声認識誤りに頑健な同定が可能である．

5. おわりに

本稿では，指示対象同定において，ユーザやシステムの特性に応じた事前確率を設定するために，2 つの列挙タスクからユーザ発話 1584 発話を収集し，ユーザやシステムの特性と参照表現発話率の傾向を調査した. 調査の結果から，実際にユーザやタスク毎に参照表現発話率が異なることを確認し，これらに応じて事前確率を変化させる本手法の着眼点の妥当性が示された．さらに，調査結果をもとに，ユーザやシステムの特徴とロジスティック回帰を用いて，指示対象同定におけるタイミングと音声認識結果による解釈の事前確率を設定した．評価実験から，ユーザやシステムの特性を導入した本手法は，事前確率として一定値を与える場合に比べて，同定精度が向上することを確認した. また，ユーザ毎の参照表現発話率が回帰時の学習において重要な特徴であることを確認した．

謝辞　本研究の一部は，科研費若手研究 (B), 基盤研究 (S), 特定領域研究の支援を受けた．

参 考 文 献

1) Matsuyama, K., Komatani, K., Ogata, T. and Okuno, H.G.: Enabling a User to Specify an Item at Any Time During System Enumeration – Item Identification for Barge-In-Able Conversational Dialogue Systems –, *Interspeech*, pp.252–255 (2009).
2) Rose, R.C. and Kim, H.K.: A hybrid barge-in procedure for more reliable turn-taking in human-machine dialogue systems, *Proc. ASRU*, pp.198–203 (2003).
3) Ljolje, A. and Goffin, V.: Discriminative training of multi-state barge-in models, *Proc. ASRU*, pp.353–358 (2007).
4) McTear, M.F.: pSoken Dialogue Technology: Enabling the Conversational User Interface., *ACM Computing Surveys*, pp.90–169 (2002).
5) Ström, N. and Seneff, S.: Intelligent Barge-in in Conversational Systems, *Proc. ICSLP*, Vol.2, pp.652–655 (2000).
6) 丹後俊郎，山岡和枝，高木晴良：ロジスティック回帰分析，朝倉書店 (1996).
7) Takeda, R., Nakadai, K., Komatani, K., Ogata, T. and Okuno, H.G.: Barge-in-able Robot Audition Based on ICA and Missing Feature Theory under Semi-Blind Situation, *Proc. IEEE/RSJ IROS*, pp.1718–1723 (2008).
8) 河原達也，　李晃伸：連続音声認識ソフトウェア Julius，人工知能学会誌，Vol.20, No.1, pp.41–49 (2005).