音声対話システムにおける ユーザのふるまいの経時的変化の分析

駒谷 和範　　河原 達也　　奥乃 博

京都大学大学院 情報学研究科 知能情報学専攻

komatani@kuis.kyoto-u.ac.jp

1　はじめに

音声対話システムの性能を向上させるうえで，ユーザのふるまいは考慮されるべき重要な要素である [1]．我々は，京都市バス運行情報案内システム（075-326-3116）によりデータを収集しており，ここでの 34ヶ月間の対話データにおけるユーザのふるまいを分析した [2]．これにより，タスクを達成する際のターン数に個人差があることや，バージインにより入力された発話が音声認識誤りかどうかを予測するのに，ユーザがバージインを行う度合が有効である可能性を確認した [2]．このように，現実の使用条件下でのユーザのふるまいは多様である．これを適切にモデル化したうえで，音声認識や対話管理を適応させることが，システムの性能向上のためには必須である．

ユーザのふるまいの多様性は，個人間の差にとどまらず，同じ個人内でも，慣れによる変化が無視できない．つまり，ユーザはシステムに慣れるにしたがって，どの程度バージインを行うかなどといったふるまいを変えることが予想される．本稿では，ユーザがシステムを使用するにつれて，これらのふるまいがどのように変化したかを調査する．

2　分析対象データ

京都市バス運行情報案内システムにより収集した，2002 年 5 月から 2005 年 2 月まで（34ヶ月間）のデータに対して分析を行う．システムは 3 つのスロット（乗車場所，降車場所，系統番号）を持ち，このうち乗車場所を含む 2 つの内容が得られると，バスの接近情報を出力する [3]．システムの語彙サイズは，バス停名が 652，名所や施設の名前が 756 である．音声認識は FSA ベースで行う．また，ユーザはシステムからのプロンプトの途中に，それを遮って発話することができる（バージイン： barge-in）．もしユーザがシステム発話の内容を既に知っており，それを最後まで聞かなくても次の発話を行える場合には，ユーザはバージインを行うことで，タスクを早く終了させることができる．

システムのログには，コールが行われた時刻や音声認識結果の他に，発信者番号，システムプロンプトが最後まで再生されたか，システムプロンプトの時間などが記録されている．システムプロンプトが最後まで再生されなかった場合，前述のバージインが起きていたとわかる．発信者番号は，ユーザが番号非通知で電話をかけた場合には記録されていないが，全体 7,988 コールのうち 5,927 コールで発信者番号が記録されていた．本稿ではこれをもとに，個々のユーザ（発信者番号）ごとのふるまいを分析する．

得られた各コール／各発話に対して，人手でラベルを付与した．ラベルの付与は 2 名の学生が分担して行った．ラベルの内容は以下である．

1. 発話内容の書き起こし
2. 音声認識結果が誤りかどうか
3. タスクごとの成功／失敗
   タスク成功，タスク失敗，中断，システム調整中
4. その他コメント

2. は，ユーザが発話した内容語が正しく音声認識結果に含まれていた場合，正解とした．3. では，ユーザの音声を人間が聞いたうえで，システムが出力したバスの接近情報がユーザの意図したものであった場合には「タスク成功」とした．

3　分析結果

人手で付与したラベルとシステムログに基づき，以下を調査した．

1. ユーザ間でふるまいが異なるかどうか
2. そのうちコール数が多いユーザについて，経時的にふるまいが変化するかどうか

本稿では特にユーザのバージインに着目して分析を進める．これはバージインが音声対話システム特有の

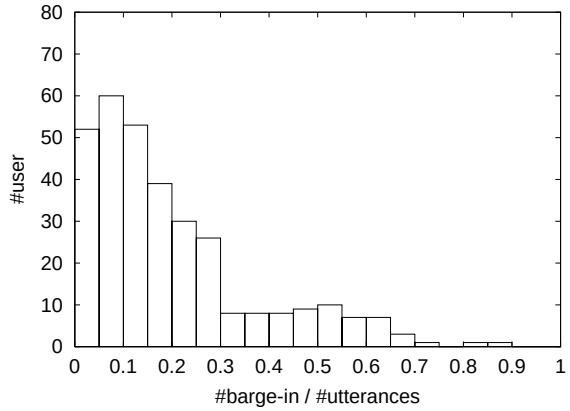

図 1: バージイン率ごとのユーザ数

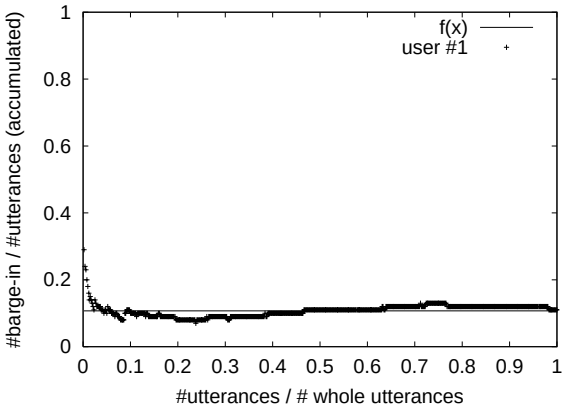

図 2: ユーザ#1 のバージイン率の経時的変化

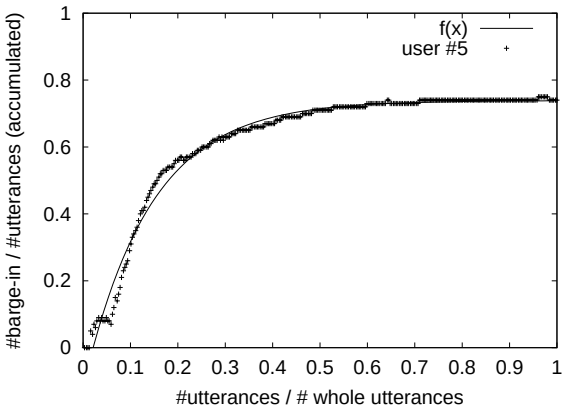

図 3: ユーザ#5 のバージイン率の経時的変化

現象であり，定量的な分析があまり行われていないことや，実ユーザに対するデータの分析結果では音声認識率が 50%程度と通常発話よりも低く [2]，システムの性能向上の鍵となると考えるためである．

3.1 ユーザごとのバージイン率の違い

まずユーザのふるまい（ここではバージイン率）がユーザにより異なるかどうかを，全データにおいて調査した．バージイン率は，当該ユーザの発話数のうち，ユーザがバージインにより入力を行った発話数，として定義している．

ユーザごとに，全データに対するバージイン率の平均を計算した．その分布を図 1 に示す．ただしここでは，2 回以上コールを行った 323 ユーザを対象として計数した．図 1 からわかるように，バージインを行う度合はユーザにより大きく異なる．これをユーザのプロファイルの一つとして用いることで，バージイン時の誤り検出の精度向上が期待できる．

3.2 ユーザのふるまいの経時的変化

次に，ユーザのふるまいの経時的変化を，以下の 3 つの尺度から分析する．

- バージイン率
- 音声認識率
- タスク達成率

これらの分析は，全データで 50 回以上コールを行っていた 12 名（電話番号）を対象とした．時間軸として，当該ユーザのある時点までのコール回数を，全コール回数で割った値を x 軸とした．したがって $0 < x \leq 1$ である．y 軸には，そのコールまでのバージイン率，音声認識率，タスク達成率を，それぞれプロットした．

またこれらの値の変化を以下の関数で近似した．

$$f(x) = c - a \cdot \exp(-bx)$$

c は，ユーザがシステムに十分に慣れた際に，対象とする尺度が収束する値，b は収束の速度，a はこの区間 $(0 < x \leq 1)$ での変化量，におおよそ対応する．これら a, b, c を最小二乗法により算出した．ただし $a \geq 0$ とした．

3.2.1 バージイン率の経時的変化

図 2，図 3 にそれぞれ，ユーザ#1，#5 に対するバージイン率の経時的変化を示す．ユーザ#1 の場合は，バージイン率は一定で，ほぼ変化していない．一方ユーザ#5 の場合，システムを使うにつれてバージイン率が上昇している．このように，ユーザによって

表 1: 多頻度ユーザ（50 コール以上）のふるまいの経時的変化 ($\Delta = f(1) - f(0)$)

user ID	バージイン率			音声認識率			タスク達成率		
	$f(1)$	Δ	x_I	$f(1)$	Δ	x_I	$f(1)$	Δ	x_I
#1	.11	0	-	.88	.20	.25	.95	.28	.21
#2	.19	0	-	.89	.24	.47	.94	.19	.25
#3	.60	.60	> 1	.89	.03	< 0	.96	.06	< 0
#4	.17	0	-	.78	.60	.46	.89	.89	.52
#5	.74	.74	.58	.94	0	-	.98	0	-
#6	.10	.06	< 0	.89	0	-	.92	.40	.11
#7	.04	.04	.06	.94	0	-	.93	.09	.08
#8	.71	0	-	.89	0	-	.87	.77	.37
#9	.49	.47	.62	.81	.27	.10	.93	0	-
#10	.10	.10	.29	.90	0	-	1	0	-
#11	.15	.04	.13	.72	.20	.17	.79	.30	.19
#12	.23	0	-	.79	.37	.21	.80	0	-

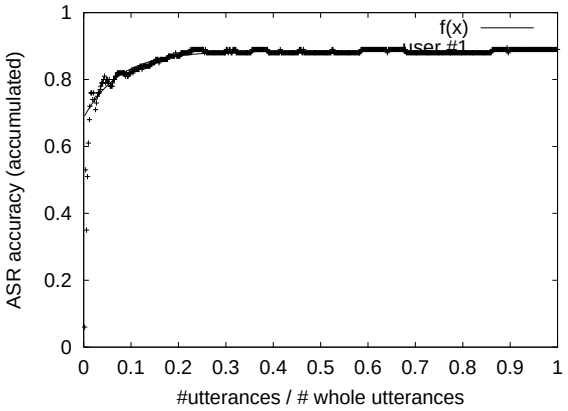

図 4: ユーザ#1 の音声認識率の経時的変化

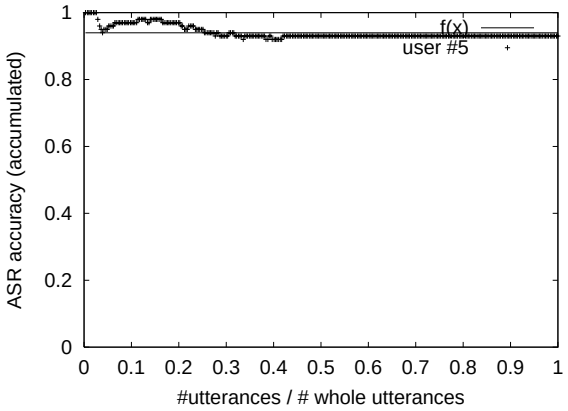

図 5: ユーザ#5 の音声認識率の経時的変化

バージイン率に変化がみられるユーザと，一定のまま変わらないユーザが見られる．

ここでグラフの概形を表すために，$f(x)$ の変化量がある一定値を下回る際の x を求めた[1]．本稿では，$\frac{df(x)}{dx} = 0.1$ となる x を x_Iとして求めた．例えば図 3 に表される，ユーザ#5 のバージイン率の変化では $x_I = 0.58$ であった．つまり，$x = 0.58$ の時点でバージイン率の変化がおおよそ収束したことを表している．

表 1 に，50 回以上のコールがあった 12 名のユーザに対する経時的変化をまとめた．バージイン率で説明すると，表中の $f(1)$ は各ユーザの最終的なバージイン率，つまり全区間での平均バージイン率であり，図 1 のデータに対応する．また $\Delta = f(1) - f(0)$ とし[2]，当該ユーザの，初期のバージイン率と最終的なバージイン率との変化量を求めた．x_Iは先述したようにバージイン率がほぼ収束する時の x の値である．$\Delta = 0$ の場合は $f(x)$ の値に変化がないため，x_Iは存在しない．

バージイン率については，表 1 の左側より，まず $f(1)$ の値にばらつきが多いことがわかる．これは図 1 での傾向と一致し，タスクを遂行する際のユーザのふるまいの多様性を示している．また，#3, #5, #9 など一部のユーザで，バージイン率の大幅な上昇が見られる．一方，残りのユーザでは大きなバージイン率の変化は見られない．このように，ふるまいの変化の度合もユーザによって異なることがわかる．

3.2.2 音声認識率・タスク達成率の経時的変化

音声認識率，タスク達成率についても，バージイン率と同様に経時的変化を調べた．ユーザ#1，#5 についてそれぞれ，図 4，図 5 に音声認識率の変化を，図

[1] $\frac{df(x)}{dx} = ab \cdot e^{-bx}$であるため，$ab > 0$ の場合 $f'(x)$ は単調減少する．

[2] 関数近似の結果 $f(0) < 0$ となった場合は，$f(0) = 0$ として算出した．

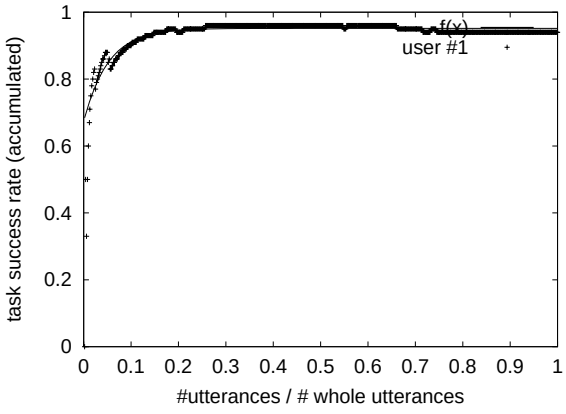

図 6: ユーザ#1 のタスク達成率の経時的変化

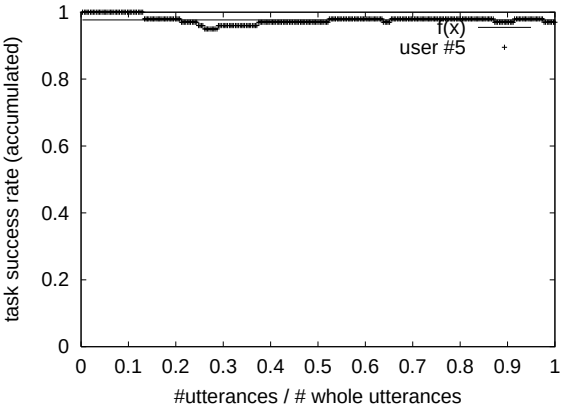

図 7: ユーザ#5 のタスク達成率の経時的変化

6，図 7 にタスク達成率の変化を示す．音声認識率，タスク達成率の両方において，ユーザ#1 では徐々に値が増加し，$x = 0.2$ を越えたあたりでほぼ一定値に収束している．一方，ユーザ#5 は両方の場合で初めから値が高く，x が増加しても値に変化はほぼない．

表 1 の中央および右側に，12 名のユーザに対する音声認識率，タスク達成率の変化を表す数値を示す．50 回以上システムを利用したユーザについて集計しているため，全区間での音声認識率やタスク達成率の平均 ($f(1)$) は総じて高く，ばらつきも少ない．しかし一部のユーザでは Δの値が大きく，使用するにつれて音声認識率やタスク達成率が向上していたことがわかる．つまり，およそ初めからシステムの使い方を知っていたユーザと，システム使用するにつれて徐々にシステムに習熟していったユーザがいることが，定量的に示されている．

3.2.3 バージイン率と音声認識率・タスク達成率との経時的変化の関係

まず表 1 からわかるように，音声認識率とタスク達成率の変化の相関は大きい．次に，#3, #5, #9 といったバージイン率の変化が大きい（Δが大きい）ユーザについては，音声認識率やタスク達成率におけるΔが小さい．これは，バージイン機能を使いこなせるようになるのは，最初からある程度音声認識率の高かったユーザに限られるという可能性を示唆している．

一方，音声認識率に関するΔが比較的大きいユーザ（#1, #2, #4, #9, #11, #12）については，バージイン率が比較的低い．つまり，システムを使い始めた頃に音声認識誤りが多かったユーザは，システムプロンプトを逐一聞いたうえでタスクを達成するようになる可能性を示唆している．以前の分析では，一定のタスク達成方法にユーザが固執する傾向が示唆されており [2]，これらの傾向とターン数との相関も調査の必要がある．

4 おわりに

本稿では，京都市バス運行情報案内システムにより収集したデータにおいて，ユーザのふるまいの経時的変化を分析した．音声対話システムでの対話管理において，ユーザのふるまいのモデル化は非常に重要な要素である．その一部として，バージイン率，音声認識率，タスク達成率の 3 つの尺度について，ユーザ毎の経時的な変化を調査した．今後これらの傾向を特徴として，音声認識誤りの判別や対話管理に活用する方法について検討する．さらに，より多くのデータに対する分析を行い，本稿で述べた傾向の一般性についても検討する予定である．

参考文献

[1] Komatani, K., Ueno, S., Kawahara, T. and Okuno, H. G.: User Modeling in Spoken Dialogue Systems to Generate Flexible Guidance, *User Modeling and User-Adapted Interaction*, Vol. 15, No. 1, pp. 169–183 (2005).

[2] 駒谷和範, 河原達也, 奥乃博: 京都市バス運行情報案内システムにおける実ユーザのふるまいの分析, 言語処理学会第 12 回年次大会発表論文集, pp. 42–45 (2006).

[3] 安達史博, 河原達也, 奥乃博, 岡本隆志, 中嶋宏: VoiceXML の動的生成に基づく自然言語音声対話システム, 情報処理学会研究報告, SLP-40-23, HI-97-23 (2002).