

# 1E-6 音声自由対話システムにおける対話データ収集

新地 秀昭\* 瀬戸 重宣\*\* 竹林 洋一\*\*

\*東芝ソフトウェアエンジニアリング(株) \*\* (株)東芝 研究開発センター

## 1.はじめに

音声メディアの自然でユーザフレンドリーな特徴を活かした音声対話の研究が活発になってきている[1-4]。また、音声対話の研究のための模擬対話による対話データの収集や、実システムを用いた対話音声データ収集について検討されている[5,6]。筆者らは、音声対話システムを人工メディアとしてとらえ、ユーザ中心のヒューマンインタフェースの実現を目指して研究を行っており、先に音声応答キャンセル機能を加えた音声自由対話システム TOSBURG II(Task-Oriented dialogue System Based on speech Understanding and Response Generation II)を試作した[7]。本稿では、本システムを用いた対話データの収集とその評価結果について述べる。

## 2.音声対話システムの評価

### 2.1.音声自由対話システム TOSBURG II

音声自由対話システム TOSBURG IIは、雑音や不要語に強い雑音免疫法によるキーワード検出、キーワードに基づく音声理解、応用分野の知識を利用した対話制御、音声メディアと視覚メディアを利用したマルチモーダル応答により、通常の話し言葉による計算機との自然な対話を実現した。また、音声応答キャンセル機能を加えシステムの応答中にもユーザの割込み入力を可能とし、より制約の少ない対話を実現した。TOSBURG IIの構成を図1に示す。

### 2.2.音声自由対話システムの評価

人間は対話の相手やその状況に応じて言葉遣いや話し方などを変え対話を円滑に進めているが、対話システムを相手にした場合のユーザの反応に関する検討は充分になされていない。実際に音声対話システムを用いて対話を行いデータ収集を行う試みがなされている[6]。筆者らは大規模な実験を行うためには、実時間システムの開発を行い実対話の音声データ収集を行う

とともに音声理解と対話処理方式の開発を考慮したデータベース管理が必要であると考えている。

特に自由発話を扱う音声対話システムの評価において、システムの実時間動作や応答の解り易さなどはユーザの自然な発声を得るためにシステムに必須の条件である。これらは認識性能とともに、ユーザから見たシステムの印象に関係し、ユーザの口調や言葉遣いなどに影響すると考えられる。また、システムの応答として音声だけでなく他のメディアを併用し様々な視点からの情報を呈示することによりユーザの発話のタイミングや対話の進行方法なども変化すると考えられる。こうした人間と計算機との間の人工メディアを利用した対話における問題はシミュレーションではなく実際にシステムを使い対話を行うことにより初めて明らかとなる。

以上の観点から筆者らはTOSBURG IIをベースに、実時間でシステムを動作させながら、対話中のユーザおよびシステムの発話データと、その処理過程で得られる情報を記録する対話データ収集ツールを試作した。

対話データとして記録する情報は、単語検出部でスポッティングした単語ラティス、構文・意味解析部で得られた文候補のリスト、対話制御部で選択した発話意味表現とそれに対して生成した応答意味表現、応答生成部で生成された応答文の内容、および、各部の処理時間などである。また、ビデオカメラで対話の様子を記録し、音声には現れなかったユーザの反応を解析に利用できるようにした。

## 3.評価実験

### 3.1.音声認識・理解性能の評価

連続音声からのキーワード検出、文認識、文理解などの性能について、被験者が自由に選んだ品目の注文をタスクとして実際に対話を行い、その対話データについて評価を行った。被験者は男性4名で各自3回ずつ対話を行った。単

「ハンバーガーとコーヒーを3つください」

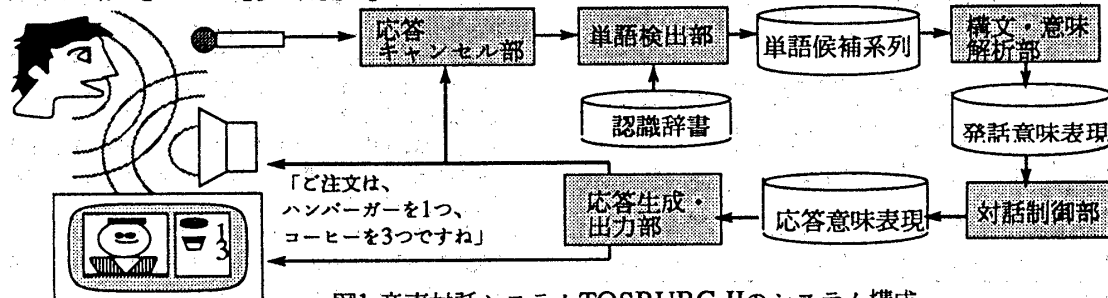


図1.音声対話システム TOSBURG IIのシステム構成

Dialogue Data Collection using a Spontaneous Speech Dialogue System  
 Hideaki SHINCHI\*, Shigenobu SETO\*\*, Yoichi TAKEBAYASHI\*\*  
 (\*Toshiba Software Engineering, \*\*Toshiba R&D Center)

語検出用の辞書は男性話者34名の連続発声から取り出したキーワードを学習セットに用いて作成したものを使用した。なお、入力レベル過小の2発話は評価から除いている。

ユーザの発話中のキーワードの数に対して、単語検出部で正しく検出されたキーワードの割合を単語検出率、対話制御部で理解結果として選択した文候補に含まれる正しいキーワード数の割合を単語認識率、理解結果として選択した文候補に含まれるキーワードが全て正しい発話数の全発話数に対する割合を文認識率と呼ぶことにする。4名の被験者の全68発話中に含まれるキーワード数185に対し、単語検出部で検出されたキーワードの総数は780、そのうち正しく検出したキーワード数は175で、単語検出率は96.2%、対話制御部で選択した文候補に含まれる正しいキーワード数は170で、単語認識率は91.9%、正解となった発話数は52で、文認識率は76%であった。

また、キーワードが完全に一致しなくても理解結果として得られた意味表現がユーザの発話の意味内容と一致する発話数の割合を文理解率とすると、正解となった発話数は56で、文理解率は82%であった。

### 3.2.マルチモーダル応答の効果

マルチモーダル応答の効果を確認するために以下の実験を行った。TOSBURG IIの応答メディアのうち、表1に示す4つの場合について、あらかじめ定めた5つの品目の注文をタスクとする対話を行い、1回の対話に要する時間を調べた。被験者は男性4名である。なお被験者には各対話を始める前に欠落するメディアが何であるかを知らせている。実験結果を図2に示す。

トレイが欠けている条件2や条件3では、すべての応答を出力する条件1にくらべて対話時間が増加する傾向がある。また、条件1や条件2ではシステムの音声応答が出力し終わらないうちにユーザが割込んで次の入力を始めようとするケースがよく観察されたが、条件3の対話ではユーザはシステムの応答出力が終わるのを待つ傾向があった。これらの結果は、音声だけでなく視覚メディアを用いて補助的な情報を示すことが対話の円滑な進行に有効に機能していることを意味している。

音声応答のない条件4の対話ではシステムが入力待ちの状態になっているにもかかわらず

表1.マルチモーダル応答の評価実験の条件

|     | 音声応答  | 応答文テキスト | トレイ   |
|-----|-------|---------|-------|
| 条件1 | 出力する  | 出力する    | 出力する  |
| 条件2 | 出力する  | 出力する    | 出力しない |
| 条件3 | 出力する  | 出力しない   | 出力しない |
| 条件4 | 出力しない | 出力する    | 出力する  |

注:トレイとは、注文品目とその個数を現すアイコンを意味する

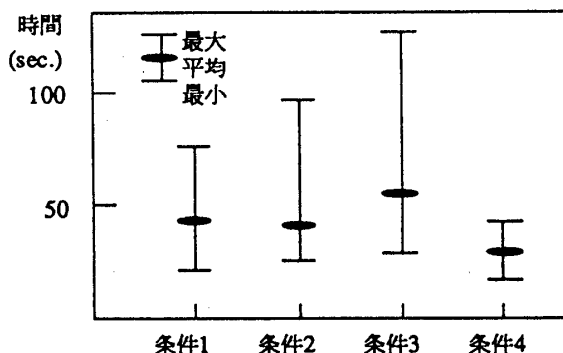


図2.マルチモーダル応答の評価

ユーザがすぐに次の入力をしない場合が見られた。視覚的なシステム応答に対して音声入力をするために入力のタイミングがつかみづらいつと感想を述べた被験者もいた。

### 4.おわりに

音声自由対話システムTOSBURG IIを用いた対話データの収集とその評価について述べた。

音声応答キャンセル機能により常時ユーザの入力を可能とし、実時間処理による対話を行い、対話に要する時間を比較した。それにより、音声だけでなく視覚メディアを用いて補助的な情報を示すことが対話の円滑な進行に有効に機能していることが明らかになった。

今後は、構築した対話データ収集ツールを改良するとともに、多数のユーザによる音声対話データの収集と評価を行い、ユーザにとって使いやすい人工メディアとしての音声対話システムの実現につき検討していきたい。

### 参考文献

- [1] Allen, J. : "Recognition Intention from Natural Language Utterances", Computational Model of Discourse, MIT Press, pp.107-166 (1983)
- [2] Grosz, B. : "Attention, Intentions, and the Structure of Discourse", Computational Linguistics, Vol.12, No.3, (1986)
- [3] 飯田, 有田 : "4階層プラン認識モデルを使った対話の理解", 情報処理学会論文誌, Vol.31, No.6, pp.810-821 (1990)
- [4] 山本, 太田, 山下, 溝口 : "音声理解システムSPURT-Iのための対話管理機構:MASCOTS", 人工知能学会全国大会(第4回)論文集, 13-2, pp.385-388 (1991.7)
- [5] 板橋, 速水, 小林, 竹沢 : "音響学会会誌連続音声データベース", 日本音響学会春季講演論文集, 2-2-17, pp.285-286 (1992.3)
- [6] 速水, 伊藤, 田中 : "音声対話システムの構築とそれを用いた会話音声収集", 電子情報通信学会音声研究会資料, SP91-101 (1991)
- [7] 瀬戸, 永田, 新地, 橋本, 竹林 : "音声自由対話システムTOSBURG IIの試作", 電子情報通信学会音声研究会資料, SP92-109, pp.47-54 (1992.12)