

音声自由対話システム TOSBURG II の評価

7R-4

館森 三慶 金沢博史 竹林洋一

(株) 東芝 研究開発センター

1. はじめに

音声により人間と計算機との自然な対話実現するために、自由発話研究と共に、大規模なデータベースの構築、実システムによる対話データの収集、評価が進められている [1, 2, 3]. 筆者らは、先にユーザに何らの制約も設けないというコンセプトのもとに音声自由対話システム TOSBURG II を構築した [4]. また、大量のデータを効率良く収集、評価するために評価ツールを作成した [5]. 本稿では評価ツールを用いた対話システムの評価について述べる.

2. TOSBURG II による対話データ収集・評価

2.1 実対話システムによる対話収集

人間対人間の対話では、話し手は自分の意思を自然言語で伝えるだけでなく、相手の表情や応答を考慮に入れながら、声の調子、発話速度、言葉遣いを変化させている. 計算機との対話においても、システムの対話制御や応答文や画面表示によってユーザの発話が変化し、システムに対する印象もユーザの発話に影響する [6].

このため実システムを用いて対話データを収集し、評価を行なうことの重要性が認識されてきている. しかし、実時間対話システムの構築例は少なく、十分な検討はなされていない. また、大量の対話音声データ収集においては、データベース化のためのトランスクリプションに多大な労力を必要とする. そのため、効率の良い対話データ収集方法の確立が重要課題となっている.

2.2 対話データ収集・評価ツール

筆者らは、以上の考えに基づいて TOSBURG II に、図 1 に示す対話データ収集・評価ツールを追加した [7].

TOSBURG II では、自然な話し言葉による対話を可能にするため、キーワードに基づいて発話を理解する. 一般に、連続音声対話システムでは、話し言葉の現象を文法として記述する必要があるため発話の詳細な記述を必要とする. これに対してキーワードベースのシステムでは、発話をキーワードのみで記述し、トランスクリプションの手間を大幅に軽減できる.

対話データの収集の際には、ユーザ発話やシステム応答の他に、キーワードラティス、キーワード文候補、発話

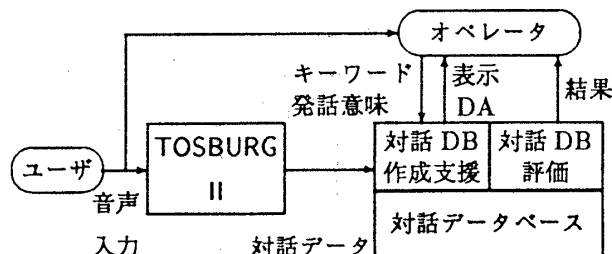


図 1: TOSBURG II によるデータ収集・評価

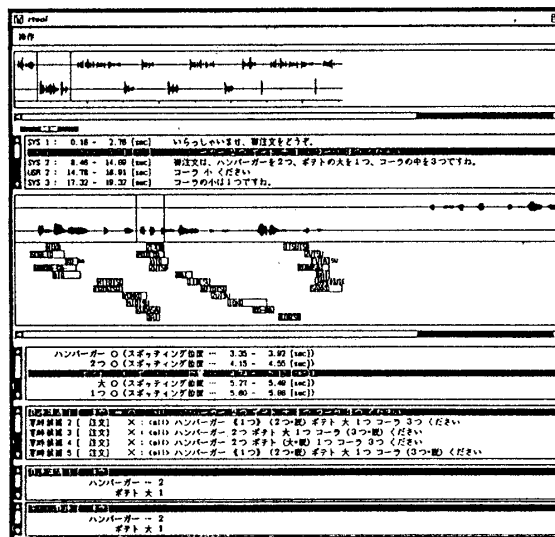


図 2: 評価ツール

意味表現候補、応答意味表現、対話履歴、システム状態などのデータを記録する.

図 2 は評価ツール表示画面である. このツールを用いることにより、キーワード検出結果、文候補結果、文理解結果などシステムの発話処理の過程を表示し、大量のデータを効率良く解析し、評価することを可能とする.

3. 実験と考察

3.1 実験条件

男性 25 名 (システムに慣れたユーザ 7 名, 初心者 30 名) を被験者とし、注文品目を指定し対話実験を行なった. さらに男性初心者 12 名を被験者に加えた計 37 名により、注文品目を指定しない場合の対話実験を行なった. また、数、サイズは発声せずに、注文の際には品目指定するだけという条件のもとで、先の 2 つの実験とは独立に、男性 20 人による対話実験を行なった.

Evaluation of Spontaneous Speech Dialogue System
TOSBURG II

Mitsuyoshi TACHIMORI, Hiroshi KANAZAWA and
Yoichi TAKEBAYASHI

Research and Development Center, Toshiba Corp.

tacci@isl.rdc.toshiba.co.jp

表 1: 品目指定なしの場合の評価結果

| 被験者 | 単語検出率 (%) | 文認識率 (%) | | 文理解率 (%) | |
|-----|-----------|----------|------|----------|------|
| | | 1位 | 2位まで | 1位 | 2位まで |
| 熟練者 | 95.9 | 62.5 | 72.9 | 70.8 | 89.5 |
| 初心者 | 94.5 | 57.1 | 70.9 | 62.8 | 79.1 |

表 2: 品目指定ありの場合の評価結果

| 被験者 | 単語検出率 (%) | 文認識率 (%) | | 文理解率 (%) | |
|-----|-----------|----------|------|----------|------|
| | | 1位 | 2位まで | 1位 | 2位まで |
| 熟練者 | 95.6 | 72.3 | 80.8 | 76.5 | 87.2 |
| 初心者 | 94.4 | 56.7 | 71.4 | 61.9 | 79.2 |

3.2 中間処理過程の評価

表 1, 表 2, 表 3 に評価結果を示す。どの表においても単語検出率ももっとも高かった。文認識率は発話中の全てのキーワード系列を正しく認識している割合である。文理解結果は発話の意味が正しく理解している割合である。すべてのキーワードを認識しなくとも発話の意味が理解できる場合があるので、文理解率が文認識率を上回った。

この評価ツールでは、正解キーワードを与えることにより、単語検出率、文認識率、文理解率などのシステムの中間処理結果が自動的に求められる。対話データ収集の際には、対話の状態や発話意味表現なども記録しているので、上述した結果の他に、注文、追加などの発話内容による分類、システムの内部状態による対話処理過程も自動的に調べることができる。

3.3 発話指定の影響

TOSBURG II はファーストフードの注文をタスクとしているので、使用法の指定やユーザの馴れにより対話認識性能が変動する。

まず、発話品目の指定の有無による認識率への影響について述べる。表 1, 表 2 を比較すると、文認識率、文理解率については、熟練者の場合には品目を指定したときに認識率が向上した。初心者の場合にはシステムに馴れておらず、品目指定の有無によらず対話が円滑に進まなかったため、認識率に有意差が現れなかったと考えられる。

数とサイズを指定しない場合の評価結果を表 3 に示す。単語検出率は 98.9%, 文認識率は 88.7%, 文理解率は 90.7% であった。認識の難しい数詞、サイズを指定しないことにより認識率が大幅に向上した。

以上の結果から、システムの性能評価を行なうためには、話者の熟練度や発話の自由度を含めた評価基準を確立することが重要であることが確認できた。

3.4 認識誤りの要因

品目の指定をせずに対話を行なった初心者 12 人の

表 3: 数詞・サイズなし、品目自由の評価結果

| 単語検出率 (%) | 文認識率 (%) | | 文理解率 (%) | |
|-----------|----------|------|----------|------|
| | 1位 | 2位まで | 1位 | 2位まで |
| 98.9 | 88.7 | 90.7 | 90.7 | 96.9 |

表 4: 誤認識の原因と内訳

| | |
|-----------------|-----------|
| ・ いい淀み | 28 発話 |
| ・ キーワード不足 | 8 発話 |
| ・ 上記 3 種の混合した発話 | 4 発話 |
| 誤認識総数 / 総発話数 | 67/159 発話 |

159 発話について認識誤りの原因を調べた。文認識誤りを起こした 67 発話のうち、40 発話が表 4 に示す発話であった。ユーザがいい淀んだ場合、システムは発話を途中で打ち切り、処理を行なうため対話が混乱する。間投詞は「えー」「えーっと」が多く、キーワードの挿入が起こり認識率の低下の原因となる。また、このシステムでは 49 単語のキーワードを設定したが、「(注文は) 以上です。」や「(コーヒーは) なし」などのキーワードとすべき単語が出現した。これは模擬対話や人間同士の対話では現れなかった単語であり、実対話データ収集の必要性を示している。

4. むすび

音声自由対話システム TOSBURG II をベースに構築した対話データ収集ツールにより多様な実対話データを効率良く収集できることを確認した。また、評価ツールを用いてシステムの評価を行ない、認識誤りの要因について調べた。

今後は、使い勝手の良さに関してユーザの意見を採り入れ、主観評価も加えて TOSBURG II の評価を行っていく予定である。

参考文献

- [1] V. Zue et al: "The Voyager Speech Understanding System: Preliminary and Its Evaluation," ICASSP-90(1990), S2.9.
- [2] 速水, 他: "音声対話システムの構築とそれを用いた会話音声収集," 信学技報 (1993), SP91-101.
- [3] 黒岩, 他: "内線電話の接続をタスクとする電話対話収集システム," 信学技報 (1993), SP93-8.
- [4] 瀬戸, 他: "音声自由対話システム TOSBURGII の試作," 信学技報 (1992), SP92-109.
- [5] 金沢, 他: "音声対話システム TOSBURG II におけるデータ収集と評価環境," 信学技報 (1993), SP93-114.
- [6] 新地, 他: "音声自由対話システムにおける対話データ収集," 情報学会第 46 回全国大会集, 分冊 2(1993).
- [7] 瀬戸, 他: "音声自由対話システム TOSBURG II による実対話データ収集", 音講論 (1993), 3-7-10.