音声言語情報処理 8－2
（1995．10．20）

# 音声対話の収集について

小林 豊　　新美康永
京都工芸繊維大学

筆者らは先に、音声模擬対話収集支援システムを開発して、対話を収集し、システム発話の内容や形態がユーザ発話に影響を与えることを観察し、システムの対話制御戦略の重要性を指摘するとともに、階層的な対話制御方式を提案した。また、情報機械との音声対話における情報伝達能率と快適性についても検討してきた。対話システム構築の機運が高まっている昨今であるが、これまで音声対話システム構築の指針は十分に検討されてきたとは言えない。本稿では、WOZ 法を利用した音声対話システム構築および対話データの収集において考慮すべき問題点について、EAGLES SLWG 資料を参考にして考察した。ここに挙げたほかにも考慮すべき項目は多岐にわたるので、対話システム設計および評価の指針とともにさらに整備していく必用があると考える。列挙した項目の多くは、マルチモーダル対話システムの場合にもほぼそのまま当てはまると考えるが、各入出力装置の特性と相互関係をよく検討する必要があるだろう。

# Collecting Spoken Dialogues

Yutaka Kobayashi and Yasuhisa Niimi
Kyoto Institute of Technology

In order to realize a practical human-machine spoken-dialog system, each aspect of human dialog must be well studied and modeled in the system. The authors have pointed out the fact that dialog control strategies affect user's attitude and utterances to the system, and proposed a hierarchical dialog control strategy. In spite of enthusiasm to build spoken- and/or multimodal-dialog systems here and there, the guide line to their designing has not been well investigated. In this paper, we review the problems which arise when designing a spoken-dialog system and using it in order to collect dialog data.

## 1　はじめに

近年、音声対話の研究が活性化する中で、人間同士の音声対話の特性を研究する試みとともに、人間と機械の音声対話に特有の問題を詳しく検討する研究が行なわれている [1]。そのためには、人間同士の対話音声、音声対話システムあるいは人間が真似て行なう模擬対話システムを用いて多量の対話音声データを収集して、対話の特性を解析し、言語現象をモデル化することが行なわれてきた [2–15]。

さらに、音声対話に限定せず、他の入出力メディアを加えたマルチモーダルな対話に関する研究も活性化している [16–20]。このように対話システムの様々な側面を総合的に検討して、人間と機械の快適な音声対話を実現しようとする試みが増えている。

筆者らは、人間と機械の音声対話の特性について研究し、能率的で快適な対話制御の実現を目指して研究を進めている。先に、システム発話の内容や形態がユーザ発話に影響を与えることを観察してシステムの対話制御戦略の重要性を指摘し、階層的な対話制御方式を提案した [21]。また、音声認識誤りがある状況での対話制御方式の数学的モデル化の 1 方式を提案した [22]。

さて、単語音声認識システム、連続音声認識システ

ムに関しては、比較のための評価方法がある程度確立してきた感があるが、対話システムの設計方針や評価方法については、これまで十分な検討がなされて来たとは言い難い。対話システムへの関心が高まる中、第7回音声言語情報処理研究会のパネル討論を契機に新田氏(東芝)らを中心に具体的な検討WGが発足して、まず、「対話システムに関する研究報告の標準的枠組み」を提案し、続いて「システム設計方針」を検討する段取りと聞いている。このとき、EAGLES SLWGの資料[23]をたたき台として参照しているが、同資料では対話システムの設計・評価の指針に加えて、模擬対話収集に関する留意事項が整理・検討されている。

本稿では、人間と機械の音声対話システムのrapid prototypingの観点から、どのように模擬対話を収集し、システムを設計していくべきかについて、EAGLESの資料の教訓を紹介しながら考察する。したがって、ここでは人間同士の自然な対話の収集は話題としない。

## 2 対話システムの設計方針の概要

(1) 対話システムの仕様を明確にする。
- システムの目的、対話の目標。
- 利用者。
- 対話状態(状況)の記述(dialog grammar)。
- システム構築の基盤(計算環境; hard and soft)。
- 通信プロトコル、DB access, GUI, AUIなど。

(2) 直感的に設計することのできるのは以下の時:
- ドメイン内のすべてのタスクが固定の順序に構造化される。
- システムが常に対話の主導権をもっている。
- システムのプロンプトの設計が良いか、あるいはドメインの特性のために、ユーザの発する文の種類・構造が制限される。

(3) 観察に基づく設計:
- 音声技術者はせっせと音声データを集めていたというのに、計算言語学者は相変わらずnative speakerの直感を頼りにして研究を進める傾向にあった。corpus-based NLPを研究している機関は非常に少なく、IBM T.J.Watson研究所くらいのものであった。ようやく最近になって、生の言語データの重要性が叫ばれるようになった。しかし、データ収集には費用も時間もかかる。頻度の高い単語は直感で思いつくが、100番目の単語となると無理である。単語選択より上位の問題についても同様のことが言える。
- 人間同士の対話が対話システムの設計にどの程度役に立つか考え直す必要がある。
- 多くの場合、NLPは標準語についてだけ開発されるが、実対話では、方言、相手との関係など様々な要因が表現に影響を与える。
- 自然言語による人間と機械との対話の経験はほとんどないので、人間同士の対話データはシステムの語彙、対話状態オートマトンなどの初期値を定めるのに使う程度にするのがよい。
- 機械との対話データが得られたら、システム情報を更新していく必要がある。

## 3 WOZ technique

WOZ (the Wizard of Oz) 法では、一般に人間がコンピュータを演じる。それにはいくつかの条件がある。

(1) 本当らしく振る舞えること。

(2) 近未来のコンピュータを想定して、システムがどの程度の性能を示すのか仕様を明確にすることができること。認識誤りはどれくらいかなど。WOZの目的の1つは、そのような仕様を明確にするため。

(3) ユーザは、wizardがコンピュータであると思い続けられるようなタスクであること。変声機で声を加工するなどの工夫が必要。また、コンピュータが知り得ると思われる知識だけを用いて、システム出力を作ること。

次に、WOZ実験の仕様決定で考慮すべき項目を詳しく検討する。

### 3.1 ユーザの認識性能に関する項目

(1) システム音声の了解性。合成音声の音質はいろいろであり、伝送系にも影響される。また、ユーザの学習効果により、時間とともに了解性が上がってくることが知られている。

(2) 語彙の問題。wizardが使う単語の意味が分かるかどうか。

### 3.2 ユーザの発話に関する項目

以下の変形などに対して、システムはどのような振る舞いをすることにするのか。

(1) 強いアクセント。

(2) 音質。

(3) 方言。発音、単語、言い回し。

(4) 発話の丁寧さ。

### 3.3 ユーザの知識に関する項目

(1) ドメイン知識の豊富さ。novice or expert.

(2) システムに対する慣れの程度。

(3) Wizard に関する情報。丁寧さが変わるだろう。未来技術に対する信頼度にも影響をもつだろう。

(4) 対話の相手をコンピュータと思ったかどうか。

### 3.4 Wizard の性能に関する項目

(1) 音響、語彙、構文、語用論レベルでの誤認識率。正確に誤認識性能を実現するのは困難。wizard タイピストを間に置いて back-end NLP システムで誤認識をシミュレートするか、音声認識部の出力を wizard が解釈して NLP システムに渡すのがよいだろう。しかし、対話の場面によっては認識誤りを模擬するのが困難なケースもあるので、正確に誤認識 5% というようなことは不可能。我々の模擬対話システムでは、wizard 用端末に認識誤りを起こすべきタイミング、確認発話のタイプを適宜表示する。誤認識は、意味レベルの単語置換を実現している。

(2) すばやい話者適応化。

(3) 音質、イントネーション、構文、文脈知識。

(4) 応答までの時間。人間の専門家より少し遅い程度が望ましい。Newell の WOZ 実験では特殊なキーボードを使って毎分 180 単語以上入力したとか。

### 3.5 談話モデルに関する項目

(1) 最初に談話モデルを確定するのは危険。シミュレーション、分析、再設計を繰り返して改善していくことが望まれる。

### 3.6 開発ステップに関する項目

(1) 訓練。wizard は、application domain, システムの能力、WOZ 実験中に利用する道具をマスターする必要がある。

(2) 道具。データベース、紙など。必要なら wizard 助手を採用すること。

(3) wizard は人間か。一部をマシンが受け持つ場合、bionic wizard と呼ぶ。漸化的にシステム開発を行う場合、システムコンポーネントを順次コンピュータに置換していく。

### 3.7 通信チャネルに関する項目

(1) ユーザ音声の音質。あまり劣化した音声だと wizard の負荷が大きくなる。

(2) システム音声の音質。機械的な感じを出したいので、vocoder 等で wizard の音声を劣化させるか、音声合成器を用いる。将来の対話システムの仕様をどのレベルに置くかが問題になる。

(3) 相互作用。発話権の交代、発話の重なり、割り込みなどが問題になる。TOSBURG II [14] では、ユーザ発話を検出するとその時点でシステム発話をキャンセルする手法を用いている。

(4) 一度に伝える情報の数。EAGLES では指摘していないが、音声だけを用いた対話の場合、伝えたいことが多数あっても、それらを一度にまくしたてるのは得策ではない。誤認識があることを考えると、ユーザ、システム双方が 1 発話の長さとそれに含める情報の数を適度に押さえることによって、全体として効率のよい意志伝達を図るのが望ましいと考える。情報を数える単位、伝達効率の尺度は [21] で基礎的検討を行ったが、タスク依存の側面もあり、今後とも検討を続ける必要がある。音声以外のチャネルをもつ対話システムでは、チャネルの切り替えや併用も含めて対話戦略を考えなければならない。

### 3.8 繰り返し WOZ 法

一般的に次の 3 つの段階に分けられる。

(1) pre-experimental phrase。ドメイン知識、対話シナリオなどの整備と wizard の訓練。

(2) first phase。 制約を少なくして対話実験を実施。

(3) second phase。 wizard はシステムとして言ってはいけないことが仕様で決まっている。問題点がなくなるまで、システムの知識などを修正・改良して再び second phase を実施。

# 4 音声対話収集実験の概要

つぎに、我々が実施している音声対話収集実験の概要を説明する。

## 4.1 模擬対話収集支援システム

京都観光案内システムを含む模擬対話収集実験の環境では、ユーザ（被験者）用とシステム用にそれぞれ別の部屋を用意して、システム担当者の存在を気付かれないように注意した。対話は音声のみによって行ない、ユーザはヘッドホンと接話型マイクロホンを利用する。ユーザに臨場感を与えたり、知識を同程度に統一するための地図、対話によって達成すべき目標を明確化するためのプランシートを渡した。一方、システム側は、誤認識率といった音声認識性能や確認発話を用いた対話制御戦略を模擬的に変化させながら対話を行ない二十歳前後の学生 43 名から 67 対話収集した。

## 4.2 対話タスク

音声対話システムの構築を考えるとき、比較的小さく限定されたタスク世界における目的指向対話の実現を当面の目標と考えてよい。ユーザ、システム双方が互いに相手から情報を引き出すようなタスクとして、本研究では「京都観光案内システム」を想定した。ユーザはシステムとの対話を通じて京都の 1 日の観光プランを作成する。

## 4.3 システム発話制御

音声対話においては意志伝達には、発話テキストの内容に加えて、イントネーション、アクセント、発話速度に代表される韻律情報の果たす役割が大きいことが知られているが、互いに相手の話し方に影響を受けるので、人間と機械の音声対話を考える場合も、ユーザが話し易く快適で、機械にとってはユーザ音声を認識し易い発話制御が望まれる。様々な制御因子を考えることができるが、ユーザ音声の認識誤りを完全になくせないことを容認して、その上でどのようなシステム発話を生成するのが望ましいのか検討する必要がある。

認識誤りをもつシステムではユーザの入力に対して何らかの確認作業が必要となる。そこで 2 つの方法を考えた。例えばユーザの発話が例 1 の u のような質問であったとする。確認の手段としてはまず直接的な確認発話をする方法が考えられる (s1)。また、単に答だけを返す (s2) 代わりに、認識結果を応答に含ませる間接的な確認発話が考えられる (s3)。これらをそれぞれ直接確認、間接確認と呼ぶことにする。

例 1

| |
|---|
| u ： 清水寺の拝観料はいくらですか。 |
| s1 ： 清水寺の拝観料ですか。[直接確認] |
| s2 ： 300 円です。 |
| s3 ： 清水寺の拝観料は 300 円です。[間接確認] |

本研究では、ユーザ発話の誤認識率とシステムの直接・間接確認発話の生起確率を種々設定して、模擬対話を収録し、対話の効率と快適性について検討した。

## 4.4 確認発話と快適性

模擬対話収録後、被験者へのアンケートの結果、確認方法に関して以下のような知見を得た。

- 間接確認は確認作業を含みながらユーザ発話に応えているため、自然な対話の流れが保てるが、認識誤りしたときはユーザに不必要な（誤った）情報を長々聞かせることになる。
- 直接確認は文脈が大きく変わる時は有効であるが、そうでない時に多く出現すると、すぐに煩わしくなる。
- システムが正しく認識していても直接確認が発話されると、その度にユーザは何らかの応答を返さねばならない。「文脈の大きく変わるところで直接確認、ひとつの文脈内では間接確認を主に用いる」というようにそれぞれの異なる特徴を考慮した対話制御戦略の検討が必要であることがわかった。

## 5 まとめ

本研究では、WOZ 法を利用した音声対話システム構築および対話データの収集において考慮すべき問題点について、EAGLES SLWG 資料 [23] に基づいて考察した。その結果、構築に先立ってシステムの仕様を詳細化するとともに、よく制御された環境を作ってデータ収集を行わなければならないことがわかった。ここに挙げたほかにも考慮すべき項目は多岐にわたるので、対話システム設計および評価の指針とともにさらに整備していく必用があると考える。列挙した項目の多くは、マルチモーダル対話システムの場合にもほぼそのまま当てはまると考えるが、各入出力装置の特性と相互関係をよく検討する必要があるだろう。

一方、当研究室で開発した音声模擬対話収集支援システムでは、音声認識性能や対話制御戦略を模擬的に変化させながら音声対話を収録し、対話の快適性や情報伝達の能率を分析、検討している。今後、

(1) 対話音声資料を増やすとともに、

(2) 音声対話における情報伝達能率のモデル化、応答に含む情報の数 (選択肢) など新たな変量、

(3) 対話の快適性とイントネーション、間投詞の挿入など (自然さ)、

ついても検討を進めてゆく。

## 参考文献


[1] 堂下, "音声・言語・概念の総合的処理による対話の理解と生成に関する研究", 文部省科研費重点領域研究成果報告書, 京都大学 (1994.3 および 1995.3)

[2] R.Moore and A.Morris, "Experiences collecting genuine spoken enquiries using WOZ techniques", *Proc. Speech & Natural Lang. Workshop*, pp.61–63, (1992.2)

[3] J.Mariani, "Spoken language processing in the framework of human-machine communication at LIMSI", *Proc. Speech & Natural Lang. Workshop*, pp.55–60 (1992.2)

[4] 竹沢, 田代, 森元, "音声言語データベースを用いた自然発話の言語現象の調査", AI 学会研資, SIG-SLUD-9403-3, pp.13–20 (1995.2)

[5] 田窪, "音声対話の言語学的モデル — 談話管理標識としての感動詞の分析 —", 情処研報, 94-SLP-1-3 (1994.5)

[6] N.M.Fraser and G.N.Gilbert, "Simulating speech systems", *Computer, Speech and Language*, 5(1), pp.81–99 (1991)

[7] 内藤, 他, "大規模内線電話受付システムの試作", 信学技報, SP94-90, pp.37–42 (1995.1)

[8] L.Hirschman, et al., "Multi-site data collection for a spoken language corpus — MADCOW", *Proc. Speech & Natural Lang. Workshop*, pp.7–14, (1995)

[9] S.Springer, et al., "The MONEY TALKS interactive speech technology assessment: a report from the field", *Proc. EUROSPEECH-95*, pp.1939–1942, (1995.9)

[10] L.Lamel, et al., "Development of spoken language corpora for travel information", *Proc. EUROSPEECH-95*, pp.1961–1964 (1995.9)

[11] J.Bertenstam, et al., "The WAXHOLM application database", *Proc. EUROSPEECH-95*, pp.833–836 (1995.9).

[12] T.Morimoto, et al., "A speech and language database for speech translation research", *Proc. ICSLP-94*, pp.1791–1794 (1994)

[13] J.F.Allen, et al., "Spoken dialogue and interactive planning", *Proc. Spoken Lang. Systems Technology Workshop*, pp.202–211 (1995.1)

[14] 金沢, 他, "音声自由対話システム TOSBURG II におけるデータ収集と評価環境", 信学技報, SP93-114, pp.1–6 (1993.12)

[15] 青野, 他, "地図課題コーパス (中間報告)", AI 学会研資, SIG-SLUD-9402-5, pp.25–30 (1994.10)

[16] 伊藤, 他, "音声・視覚・画像をもつインタラクションシステム", 情処研報, 95(16), 95-SLP-5-5, pp.31–38 (1995.2)

[17] 伊藤, 他, "音声対話システム構築のためのデータ収集実験", 音響講論集, 1-Q-15, pp.139–140 (1995.3)

[18] K.H.Loken-Kim, et al., "翻訳通信環境におけるマルチモーダル入力の分析と統合", 情処研報, 95(73), 95-SLP-7-12 (1995.7)

[19] 綿貫, 外川, "対話における感情の変化の解析", 情処研報, 95(73), 95-SLP-7-13 (1995.7)

[20] 安藤, 畑岡, "マルチモーダルなエージェント型ユーザインタフェースの評価と対話制御の検討", 情処研報, 95(73), 95-SLP-7-15 (1995.7)

[21] 小林, 新美, "対話制御の処理レベルとモデル化", 情処研報, 93-SLP-3-13, pp.59–61 (1994.2)

[22] 新美, 小林, "音声認識の誤りを考慮した対話制御方式のモデル化", 情処研報, 95(16), SLP-5-7, pp.47–54 (1995.2)

[23] R.Moore, "Interactive dialogue", *The EAGLES Spoken Language Working Group: Progress Report*, EAG-SLWG-IR.2, Chap.5, pp.116–165 (1994.10)