音 声 言 語 情 報 処 理 2－6
（1994．7．7）

音声対話システム構築のための実対話データ収録実験

伊藤克亘　秋葉友良†　長谷川修　速水　悟　田中和世

電子技術総合研究所　†東京工業大学

音声対話システムを構築する知見をえるための実対話データを収録する実験について述べる。従来より、対話データを収録する手法として、よく知られる Wizard of Oz (WOZ) 方式と呼ばれる手法がある。WOZ 方式は、システムになりすました人 (この人を Wizard とよぶ) と被験者が対話をする手法である。従来おこなわれてきた WOZ 方式でのデータ収録は、文法や語彙の設計など、ある特定の音声対話システムを構築するための知見をえるためのものが多かった。本稿では、人間と機械がやりとりをおこなううえで、そのやりとりを可能にする要素を明らかにするためのデータを収録することを目的とした。WOZ 方式による実験で、40 人の被験者による対話データを収録した。実験では、被験者ひとりあたり、20 分から 70 分程度のデータを収録した。

Experiments for Collecting Human-Machine Interaction Data for development of Speech Dialog System

ITOU Katunobu, AKIBA Tomoyosi†, HASEGAWA Osamu,
HAYAMIZU Satoru, and TANAKA Kazuyo

Electrotechnical Laboratory, †Tokyo Institute of Technology

This paper describes collecting and analyzing nonverbal elements for maintenance of dialogue using a Wizard of Oz (WOZ) simulation. The Wizard of Oz technique is that a human (who is called the wizard) plays the role of the system in a simulated interaction between human and system. Many WOZ experiments have been reported to date, but those were mainly designed to develop specific spoken dialogue applications. Our WOZ experiment is designed to collect data in order to analyze the elements which enable the interaction between human and machine. We have done experiments for forty subjects. Each subject talked with the WOZ system from twenty to seventy minutes.

1 はじめに

知的な主体としての人工物が実世界の一員となったとき、やはり、便利で手軽なコミュニケーションの手段としては、言語による対話が一番にあげられるのではないだろうか。この目標を達成するための基礎として、実際の対話のデータが必要になる。

「実際の対話のデータ」として、どのような対話をどのように収録すればよいかについては、様々な試みがなされてきた。一番、手軽なのは、人間対人間の対話を収録することである。しかし、人間は、相手のふるまいによって態度を変える。したがって、知識など様々な点で制限のある人工物を相手にする場合には、人工物に比べれば制限の少ない人間相手の場合に比べると、対話が変化する可能性がある。

そこで、われわれは、人工物である音声対話システムと人間との対話データを収録し、分析した [1,2]。しかし、現状程度の実システムでは、余りにも制限が大きすぎるため、対話の流れや発話などのバリエーションがほとんど得られない。その結果、収録されたデータの大半は、間投詞などについては、それなりの知見がえられるものの、より知的なレベルである言い回しなど現象については、ほとんどシステムの範囲を越えるものがなかった [2]。これは、ある意味で当然のことで、システムで扱えない発話というのは、システムの設計段階では対処できないから扱わないのであって、そういった発話はなるべく起こらないように設計するため、対話が成立するような場合には、結果としてシステムで扱えない発話がおこらないのである。

そこで、今回は、表層の発話を意味構造に変換したもの程度を入力とする対話プログラムを作成し、そのプログラムの性能を越えるような発話の解釈と全ての発話の聞き取りは人間がおこないながら、システムが対話をしているようにみせかけるシステム (なりすまし方式: Wizard of Oz simulating[3]) を設計し、構築した。本稿では、その設計方針と具体的な対話実験の概要について述べる。

2 システムの設計方針

筆者らの目標は、知的な主体との一対一の対話の実現である。したがって、人間どうしの対面時の対話での音声によるコミュニケーションをなりたたせる要因を明らかにするようなデータの収集を、今回の目的とする。また、筆者らは、非言語的な情報のやりとりも、対話に不可欠な要素であると考えている。そのうちの視覚的な情報についても検討するため、今回は、画像データも収録する。

このような目的でのデータ収録をおこなうという前提にたつので、このシステムでは、道案内を題材に対話をおこなうようになっているが、それは、システム側の知識や推論能力の基準としてのみとらえる。つまり、道案内システムの開発を目的としているのではない。したがって、道案内の戦略などの最適性などは全く問題にしていない。そこで、以下のような方針でシステムを設計した。

1. システムの応答は、合成音声以外では出力しない。(地図なども表示しない。)
2. 前もっての、システムの利用方法などの指導は極力おこなわない

システムの内部状態がわからないというのは、アプリケーションとしてそのシステムを利用する場合には、不自然な状態であるが、実世界に存在する主体の内部状態は直接観測できないという立場にたつため、むしろ、実際の状況に近いと考える。

3 実験システムの構成

実験システムの構成図を図 1 に示す。Wizard とかかれたオペレータ側と、Subject とかかれた被験者側は、それぞれ別の部屋であり、オペレータがいる部屋は、被験者から全く見えないようになっている。

被験者側には、ハンドマイク (マイクスタンド付き) (g) とワークステーション (f) を用意する。ワークステーションには、課題の文章を表示するためのディスプレイとシステムの応答を合成するための音声合成器が取り付けられており、音声合成器の出力はスピーカー (e) から出される。被験者の正面からには、被験者の様子を録画するためのビデオカメラ (d) が 3 台設置されている。カメラは、レンズ以外の部分をカーテンでおおって、なるべく、カメラ自体や、カメラのオペレータが被験者の視界に入らないようにした。

記録するデータは、音声データと画像データである。音声データは、被験者の発話と音声合成器の出力を DAT の別チャネルに同時に収録する。画像データは、3 台のカメラを用いて音声信号と同時に VTR に収録する。カ

メラは被験者の正面と左右に配置し、正面のカメラでは被験者の視線動きや表情の変化を、また左右のカメラでは上半身の挙動を収録する。

対話プログラムのオペレータ側には、被験者の様子を観察するために、DAT で収録している音声のモニタ用スピーカ (a) と、左前方から被験者の上半身を撮影しているビデオカメラからの映像をモニタするディスプレイ (c) を用意した。対話プログラムの操作は、ワークステーション (b) のキーボードを利用しておこなう。

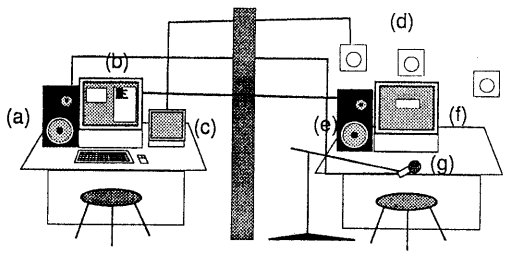


図 1 実験システムの構成

4 データ収録実験の概要

実験は、1994 年 2 月から 1994 年 3 月末までにわたっておこなわれた。

この本実験に先だって、本実験の被験者とは別の数名の被験者で予備実験をおこない、課題の適切さ・指導の度合などを決定した。基本的には、そのまま一定の条件で全実験をおこなったが、本実験を開始した後で、対話が成立しなくなるような不都合が対話プログラムなどに発見された場合などは、その不都合がなくなるように修正した。

被験者は、図 2 に示したパンフレットを利用して募集した。個別に連絡があった場合も、システムに関する質問には、このパンフレットにのっている以上のことは答えなかった。

コンピュータとの音声対話実験
参加ボランティア募集

電総研 音声研究室・画像研究室

当研究室では、人間と機械の究極のコミュニケーションを目指した研究をおこなっております。その目標に向け、現在、簡単な音声対話システムを構築していますが、人間のふるまいは非常に複雑で、変化に富んでいるため、現在の研究の水準では、まだまだ、人間同士の対話にはほど遠い簡単な対話しかできません。また、人間が対話の相手に対してどのようなふるまいをするのかもよく知られていません。

この水準を向上させるためには、たくさんの人にシステムを試してもらって、その様子を記録して、システムが対応できなかった部分をひとつづつ解消していく作業が必要だと考えています。

そこで、このシステムを試しに使った様子をデータとして収録する実験に協力していただけるボランティアの方を募集いたします。収録するデータは、

- システムに質問したり、こたえたりする時の音声 (DAT で録音します)
- その時の表情の変化、頭・視線の動き、上半身の動きなどの画像データ (ビデオに録画します)

です。これらのデータは、工学的、科学的な解析を加えられます。また、これらのデータは上記で示した目的で収録するため、**原則として、学術目的に限っては公開することを前提とさせていただきます。**公開の形態としては、

- 学術刊行物などに、実観測データとして公表する
- 研究用データベースとして公開する

などが考えられます。

実験は、**2 月の中旬から 3 月上旬に、電総研の音声研究室でおこなう予定です (電総研は東大通り沿い、並木大橋の西側あたりです)。所要時間は、説明、収録、アンケートへの回答の全体で 1 時間 30 分から 2 時間くらいの予定です。**

なお、協力者の方には、交通費等の謝金を検討しております。ご協力していただける方は、下記までご連絡下さい。詳細についてご案内させていただきます。みなさま、よろしくお願いいたします。

連絡先 〒305 茨城県つくば市梅園 1-1-4
通産省 工業技術院 電子技術総合研究所
知能情報部 音声研究室
伊藤 克亘 (いとう かつのぶ)
Office-Phone: 0298-58-5940
Fax: 0298-58-5939
E-mail: kito@etl.go.jp

図 2 被験者募集パンフレット

実験の最初に、実験をおこなう部屋とは別室で、次の文章を見せて、被験者にシステムの説明をおこなう。

今回、使っていただくシステムは、渋谷の道案内をするシステムです。ディスプレイに案内できる場所 (飲食店、デパートなど) の一覧が表示されています。このシステムには、これらの場所について、タウンマップから抽出した情報が入力されています。

このシステムを使って、5 つの課題を順番におこなっていただきます。課題に必要な情報は、全て、このシステムから聞き出せるようになっています。

各々の課題は、10 分程度です。また、課題と課題の間には、数分間程度休憩があります。また、実験の最初に、カメラ・マイクの調整と学術データとしての写真撮影 (証明写真程度) をおこないます。終了後には、簡単なアンケートをおこないます。なお、最後の課題では、それまでの課題で説明した場所について質問します。

図 3 被験者指導用紙

被験者への指導は、最初の段落の文章でおこなっている。実験システムの設計方針としては、システムの説明

を行わないとなっているが、全く何もないと、実験が成立しない可能性があるので、イメージを喚起させるため最後の1文を付け加えている。システムが知っている内容や、システムが扱えるいい回しなどは全く教えない。

[開始]
まず、課題にさきがけて、今日の日付けとあなたの名前をマイクに向かって、しゃべって下さい。

OK の場合は、課題を開始します。

[課題 1]
あなたからシステムに話しかけて下さい。
そして、東急ハンズの場所をたずねて下さい。

[課題 2]
システムから、はなしかけます。

それから、喫茶店、ファーストフードなどの店の行き方を、いくつかお尋ね下さい。

[課題 3]
あなたからシステムにはなしかけて下さい。

そして、QUINCAMPOIX(カンカンポア)、ダル・ボロニェーゼ、ORRB(オーブ)の場所を尋ねて下さい。これらの店の駅からの距離、取り扱っている物などから、自分の好みの店を探してみて下さい。

[課題 4]
自由に、何か所でも構いませんから、お尋ね下さい。(案内できる場所の一覧を下に表示します。しばらくしても一覧が表示されない場合は、「出ません」とマイクにむかってしゃべって下さい。)

[課題 5]
これまでに、システムがあなたに案内した場所のうち、よく覚えている 2 か所について、システムに行き方を説明して下さい。

図 4 課題の一覧 (例)

このあと、実験室に移動する。まず、基礎画像データとして、被験者の顔のアップと上半身の写真撮影をおこない、また、システム (ワークステーションのディスプレイ) に対面着座して、ディスプレイの四隅を注視した視線位置をビデオに収録した。

撮影が終了したら、被験者の座る位置をビデオカメラの位置にあわせ、マイクの位置あわせをおこなって課題を開始する。はじめは、図 4 の [開始] の部分にかかれている文章がディスプレイ上のウィンドウに表示されている。その状態で、説明者が、「日付けと名前が認識されればウィンドウが消えて、課題を示したウィンドウが表示されるので、その指示にしたがって、対話をおこなってください。課題が終了したら、ウィンドウが消えて、次の課題を示したウィンドウが表示されるまで、数分間の休憩になります。」とだけ指示して、課題をおこなってもらう。

課題は、図 4 からもわかるように、この対話システムが説明できることを被験者に徐々に知らせるように並べられている。被験者によって、この課題中の「喫茶店」などのキーワードや尋ねるべき場所の固有名詞を変更して課題をおこなった。[課題 5] は、1 から 4 までの課題と違って、システムから被験者に場所をたずねるようになっいる。この課題 5 に関しては、対話プログラムを使わず、オペレータが全て対話をおこなっている。

この例では、[課題 1,3] では、ユーザから話しかけるように指示してあり、[課題 2] では、システムから話しかけるとなっている。被験者ごとに、課題によって、ユーザから話しかける場合と、システムから話しかける場合を適当に設定している。[課題 4] では、そのような指示をせず、ユーザから話しかければそのまま対話を開始し、ユーザがなかなか話しかけない場合はシステムから対話を開始する。

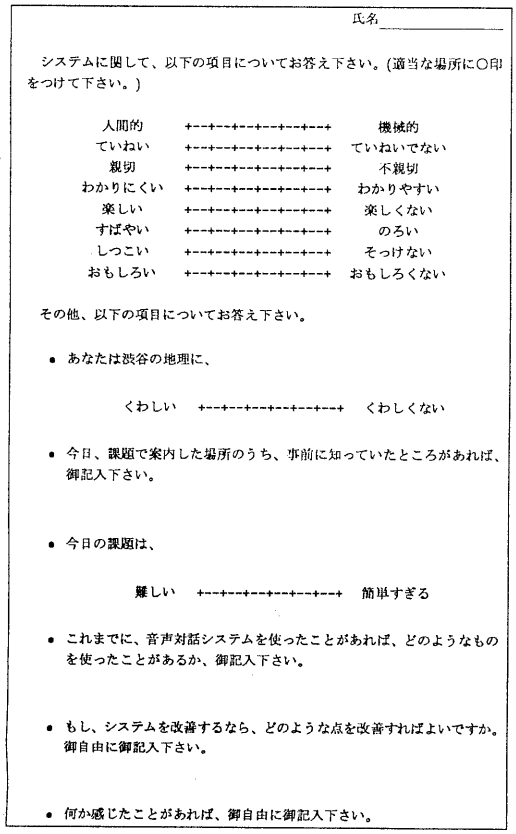
課題中には、被験者には、メモなどをとることは許していない。これは、メモをとることを許すと、被験者がメモの方を見がちになって、顔の表情を収録することが困難になるからである。

課題中に被験者が指示以外の対話をはじめても、対話プログラムがあつかえるものについては、そのまま対話を続けるようにした。したがって、同じような課題でも被験者によって、対話の長さはかなり異なる。

被験者の発話からシステムの応答までの時間は、最短で数秒、最大で 20 から 30 秒程度要する。

終了後、アンケートをおこなう。アンケート用紙を図 5 に示す。アンケートでは、領域知識の有無や、音声対話システムの利用の経験の有無、システムや課題の印象について尋ねた。

実験は、説明の開始からアンケートの終了まで、ほぼ 1 時間 30 分を要した。



氏名＿＿＿＿＿＿＿＿

システムに関して、以下の項目についてお答え下さい。(適当な場所に○印をつけて下さい。)

人間的	+--+--+--+--+--+--+--+	機械的
ていねい	+--+--+--+--+--+--+--+	ていねいでない
親切	+--+--+--+--+--+--+--+	不親切
わかりにくい	+--+--+--+--+--+--+--+	わかりやすい
楽しい	+--+--+--+--+--+--+--+	楽しくない
すばやい	+--+--+--+--+--+--+--+	のろい
しつこい	+--+--+--+--+--+--+--+	そっけない
おもしろい	+--+--+--+--+--+--+--+	おもしろくない

その他、以下の項目についてお答え下さい。

- あなたは渋谷の地理に、

  くわしい +--+--+--+--+--+--+--+ くわしくない

- 今日、課題で案内した場所のうち、事前に知っていたところがあれば、御記入下さい。

- 今日の課題は、

  難しい +--+--+--+--+--+--+--+ 簡単すぎる

- これまでに、音声対話システムを使ったことがあれば、どのようなものを使ったことがあるか、御記入下さい。

- もし、システムを改善するなら、どのような点を改善すればよいですか。御自由に御記入下さい。

- 何か感じたことがあれば、御自由に御記入下さい。

図 5 アンケート用紙

5 対話プログラムの概要

5.1 処理の流れ

対話プログラムは、以下の形式でそれぞれの地点のデータを保持する。例えば、「東急ハンズ」の場合、次のようになる。

```
[ id: bu0007, class: 建物, famous: +,
  指示表現 (1): "東急ハンズ",
  指示表現 (2): "ハンズ",
  建物のタイプ: デパート,
  街路との関係: 3,
  街路 (1): [ id: st0004,
    区間: [ from: 10, to: 11, 位置: 右 ],
    近接 (1): [ id: bu0080, 関係: 対面],
    近接 (2): [ id: bu0081, 関係: 対面]
  ],
    ...
  説明 (1): [ 起点: cr0005,
    応答: ("~s の手前右側の角に~s があります",
          [cr0005, bu0007]),
    基準: [ street: st0004, direction: +],
          street: st0004, direction: +,
          distance: 10
  ],
    ...
]
```

このシステムは、約 150 個所分のデータを持っている。場所として、交差点・建物・通り・店舗などが用意されている。

ひとつの場所は、複数の呼び方に対応できるようになっている。それぞれの場所は、隣接する街路とさらにその街路を通じて近接する別の場所との関係を保持している。この関係を利用して、ある場所からある場所への経路を評価する。近接する場所との関係に応じて、道案内のための説明文を用意してある。説明は、用意された説明文を連結していくことでおこなう。ただし、場所の名称が入る ~s となっている部分には、文脈によって、固有名詞と、「その建物」といった指示表現を使いわけられるようになっている。

プログラムへの入力は、「東急ハンズの場所を教えて」という発話の場合、以下のような形式であたえられる。

```
want('$you',inform_ref('$i','$you',
                       '$place'(bu0007)))
```

$you はユーザ、$i はシステムをあらわし、$place (bu0007) は「東急ハンズの場所」をあらわす。オペレータの入力にかかる時間を短縮するために、この入力と同様の内容をあらわす rrp(東急ハンズ). という形式のコマンドを作成し、全部で約 80 種類ほど用意した。

対話プログラムは、この入力から、ユーザモデルを構築する。

```
not(know_ref('$you', '$place'(bu0007)))
know_ref('$you', '$name'(bu0007))
```

このふたつの式は、「ユーザはパルコの場所は知らないが、パルコの名前は知っている。」ことを表わしている。このふたつの式と、このときのユーザの発話そのものが、ユーザモデルとして登録される。また、入力からシステムのゴールを決める。

```
inform_ref('$i', '$you', '$place'(bu0007))
```

このゴールにしたがって、プランニングをおこなってシステムの行為を決定する。システムの行為には、「入力受け付け」「応答」「開始/終了」がある。

対話をおこなっている間、システムはユーザモデルを蓄積し、蓄積されたユーザモデルからプランニングをおこなうことで、文脈的な知識をある程度利用できるようになっている。このとき、ある発話までに蓄積されているユーザモデルと、新たにその発話で生成されたユーザモデルが矛盾をおこすときには、最新のユーザモデルを優先して更新していく。

説明は何種類か可能になっており、1) ユーザが知っている場所かどうか、2) 目標地点までの距離が近いかどうか、3) 説明をおこなうのに中間地点が少ないかどうか、の三つの基準でスコアづけし、最もスコアの高い説明を採用するようにしている。説明に使用する場所をユーザが知っているかどうかわからない場合は、まず、その場所を知っているかどうか確認する対話をおこなう。説明の例を以下に示す。(U: は被験者、S: はシステムをあらわす。)

U: 東急ハンズの場所を教えて下さい。
S: では、パルコパート 1 を知ってますか。
U: いいえ。
S: では、西武百貨店 A 館を知っていますか。
U: 西武百貨店 A 館ですか。
S: はい。
U: 知っています。
S: それでは、西武百貨店 A 館からの道を案内します。

また、このシステムでは、行き方以外にも、その店舗が、どういった種類の店か (イタリア料理店など) とか、距離、移動に必要な時間などを答えられるようになっている。

システムの発話は、漢字仮名まじり文として生成される。そのあと、漢字のよみや、単語や文節のアクセントを付与するプログラムで処理して、市販の音声合成装置から出力される。この処理を数文の説明文ごとにおこなう。システムが合成音声を出力している間は、オペレータがそれを中断したり、割り込んだりすることはない。

5.2 オペレータの役割

オペレータの役割のうち、最も重要なものは、ユーザの発話から対話プログラムの入力を作成するところである。対話プログラムへの入力は、被験者の発話が省略などを含む場合や不完全な場合でも、それらをオペレータが解消することはなく、対話プログラムがそれらの解消をおこなう。したがって、自然言語処理の面から見ると、課題 1 から課題 4 は、これまでのなりすまし方式で自動化の度合が高いことを特徴としているシステム (Bionic Wizard[4]) と比較しても、非常に自動化の度合が高いといえる。

被験者の音声やディスプレイモニタを介して観察できる様子から、非言語的な情報がえられれば、それに相当する入力をオペレータの判断で、適当に入力することを許している。たとえば、システムが確認を求めるような発話をした場合に、肯定するような表情もしくは雰囲気でうなづいた様子であれば、被験者が明示的に、言語的な発話をおこなわなくても、オペレータの判断で「はい」と同等の入力を行うことができる。

発話が対話プログラムの能力をこえる場合には、「理解できません」といった内容の発話を生成するようにしている。ただし、それ以上詳しく、具体的に、何が理解できないとか、何なら理解できる、などといった発話はおこなわない。

課題 5 は、推論もオペレータがおこなっているが、そのときの質問の方針としては、まず、「どこでもいいですから説明してください」と発話し、被験者が説明をおこなったら、「もう少し詳しく説明してください」「途中に何か目印はありますか」「その建物にはどんな店がありますか」「それは、何の店ですか」「その店のそばには、どんな店がありますか」などの質問をおこない、説明を続けさせた。また、被験者が、説明してくれない場合は、「だいたいでいいですから、説明してください」と質問して、説明をうながせた。また、質問に対して答えてくれないような場合には、「(東急ハンス) の行き方を説明してください」「渋谷には、いったことがありますか」「渋谷について、何でもいいですから教えてください」といった質問をして、被験者になんとかしゃべらせるようにした。

5.3 システムからの積極的なやりとり

対話を円滑にすすめる要素について調べるために、システムからの間投詞や確認の発話、視覚の存在を示すような発話を、積極的に被験者になげかけるようにした。

間投詞については、「えーと」と「えー」というふたつの間投詞が、説明文の読点の後と発話の最初に任意に挿入できるようにしておき、被験者ごとに、課題によって、間投詞の生成確率を、0 ないし 0.5 に設定して実験をおこなった。

確認の発話としては、ユーザが「東急ハンズの場所を教えて下さい。」などの固有名詞を含む発話をおこなった場合に、次のシステムの発話の最初に「東急ハンズですね。」という発話を挿入できるようにした。確認の発話についても、被験者ごとに課題によって生成確率を、0, 0.5, 1.0 のいずれかの値に設定して実験をおこなった。

間投詞と確認の発話を発話の最初におこなう場合は、通常の応答を生成する場合に比べて、推論などを介さずになるべく早く応答するようにした。したがって、間投詞や確認の発話が発話の最初におこなわれる場合は、システムの反応時間が短めになっている。

6 収録データ

6.1 被験者

被験者 40 名の、年齢、性別の構成を **表 1** に示す。10 代、20 代の被験者は全て学生なので、ほとんどの被験者が学生であった。

表 1 被験者の構成

	男性 (人)	女性 (人)
10 代	2	2
20 代	20	11
30 代	0	1
40 代	1	3

アンケートを集計した結果、音声対話システムを利用したことがないと答えた被験者は 38 名で、あると答えた被験者も、一人は、科学万博での自律ロボットデモンストレーション (コスモ星丸) をあげているが、もう一人は、合成音声を発する自動販売機をあげているため、一人をのぞいては、音声対話システムを利用するのは、初めてだったといえる。

また、人間がシステムになりすましていることに気付いた被験者は皆無であった。

6.2 対話例

実際にシステムと対話している時間は、17 分から 68 分で、平均約 33 分であった。以下に収録した対話の例を示す。

U: ロマーナのお店の行き方を教えて下さい。
S: では、パルコパート 1 を知ってますか。
U: あのー、知らないです。
S: では、丸井本館を知っていますか。
U: いえ。
S: では、東急ハンズを知っていますか。
U: わかります。
S: こっちを向いて下さい。
U: どっちだ… 知ってます。
S: では、東急ハンズからの行き方を説明します。
U: お願いします。
(略)
S: 交差点を右に曲がり、20m ほど進むと右側に建物があります。
U: 建物は何という建物ですか。
S: ええと… 表現が曖昧です。
U: 建物の名前はなんですか。
S: 表現が曖昧です。
U: 東急ハンズからの道をもう一度お願いします。
S: もう一度、説明します。
(略)
S: ロマーナは、その建物の一階です。
U: 建物の一階ですね。
S: えー…
U: 建物の一階ですか。
S: ロマーナはその建物の一階です。
U: わかりました。

6.3 データ解析方針

収録したデータの解析は、対話システムとしての実装を第一の目標とするため、客観的な物理量を中心におこなう予定である。現在のところ、以下の要素について解析する予定である。

1. 音声データ

   物理量: 単位時間当りの発話回数、発話までに要する時間、発話速度

   言語現象: 間投詞、言い誤り、言いよどみ、ひとりごと

2. 画像データ

   表情の変化: 眉、目などの特徴点の移動量

   頭の挙動: 首を中心とした座標系の x 軸、y 軸、z 軸回りの回転角

   手の挙動: 左右の手の挙動

   肩の挙動: 腰を中心とした座標系の x 軸、y 軸、z 軸回りの回転角

7 まとめ

人間と機械の円滑なやりとりの実現を目標とした、機械と対話する場合の人間のふるまいを明らかにするためのデータ収録実験の方法について述べた。この試みでは、人間がシステムになりすます方法 (Wizard of OZ 方式) を用いているが、従来の方法と比較すると、対話を円滑にすすめる表面的な行為は人間が代行するが、言語処理や推論などの行為は完全にシステムがおこなう点が大きくことなっている。また、対話を円滑にすすめる視覚的な情報について解析するために、音声データと同時に被験者の表情や上半身の動きといった画像データを収録した。1994 年 2 月から 3 月までの二か月にわたる実験で、被験者 40 人分、合計約 1300 分間のデータを収録した。今後は、音声、画像それぞれのデータの解析をおこなうとともに、聴覚的な行為と視覚的な行為の対応などを考えた解析をおこなう予定である。

謝辞

本研究の一部は RWC 計画の一貫としておこなわれたものである。関係各位に感謝いたします。貴重な御意見をいただいた、電総研新情報計画室と音声研究室、画像研究室の皆様に感謝いたします。対話プログラムのデータは東工大の上條俊一氏が作成してくださったものです。感謝いたします。

参考文献

[1] K. Itou, S. Hayamizu, K. Tanaka, and H. Tanaka. System design, data collection and evaluation of a speech dialogue system. *IEICE Trans. INF. & SYST.*, Vol. E76-D, No. 1, pp. 121–127, 1993.

[2] 伊藤克亘, 速水悟, 田中穂積. 音声対話システムに対する利用者の発話の解析. 日本音響学会講演論文集, pp. 21–22, 1992.

[3] N. M. Fraser and G. N. Gilbert. Simulating speech systems. *Computer Speech and Language*, Vol. 5, No. 1, pp. 81–99, 1991.

[4] C. MacDermid. Features of naive callers' dialogues with a simulated speech understanding and dialogue system. In *Proc. EuroSpeech 93.*, pp. 955–958, 1993.