

# 音声対話システムの応答部における 間投詞の利用とユーザのふるまいの分析

伊藤敏彦 中川聖一

豊橋技術科学大学 情報工学系

〒441 豊橋市天伯町字雲雀ヶ丘 1-1

著者らは、対話に存在する間投詞に着目し「発話中の間投詞は聞き手に対してどのような働きを持つか」「協調的な応答文生成において間投詞は有効・必要なのか」という観点から、聴取実験による検討を行ない、その結果、間投詞に関するいくつかの知見を得ることができた。

本稿では、これらの知見に基づき、対話システムにおいて「より自然なシステム応答」及び「情報検索・応答文生成によって生じる無音が引き起こす不自然さの軽減」を目的として、システム応答音声中に間投詞を挿入することを考案し、WOZ(Wizard of OZ)による音声対話システムを用いて、間投詞が付与されたシステム応答に対する評価実験を行なった。実験結果より、間投詞挿入による効果が確認され、間投詞が音声対話システムにおける応答文生成時間の確保や、発話権の維持、及びシステムが動作中であることを示すサインとして有用であることが分かった。

また、聞き手(認識・理解)が人間か機械かというユーザの意識によって、ユーザのふるまいがどのように変化するかという被験者実験も行ない、意識の違いだけによってもユーザのふるまいが変化することが分かった。

## Application of the generation of filled pauses to a dialogue system and analysis of a user's behavior

Toshihiko ITOH and Seiichi NAKAGAWA

Department of Information and Computer Sciences

Toyohashi University of Technology, Tenpaku, Toyohashi, 441, Japan

We investigated filled pauses found in lecture speech and dialogue speech from the following viewpoints; 1) the role of the filled pauses in listener's understanding, 2) the necessity or effectiveness of generating filled pauses to make its responses more cooperative. And a series of listening tests were carried out. As a result, we obtained several findings on the above issues.

Based on the findings, in this paper, we propose that a speech dialogue system should insert filled pauses in response sentences to increase the naturalness of the responses and to exclude unnatural silent segments (corresponding to no actions) inevitably caused by retrieving a specific database or producing response sentences.

To test this proposal, we conducted evaluation experiments using a WOZ-based dialogue system. Results showed that the filled pauses in the system's responses work adequately, that, by presenting any kind of filled pauses, response sentences can be produced without causing any unnatural events and that a filled pause can be a sign indicating that the system is preserving the turn or working normally.

Finally, to test if the user's utterances are influenced by a hearer, i.e., human or machine, we also did evaluation experiments using a WOZ-based dialogue system. As a result, the user's behavior was changed by these situations.

### 1 はじめに

現在の音声対話システムの研究対象は自然な発話に関するものが多い。その場合「書き言葉」には見られない、「話し言葉」特有の言語現象(言い直し、間投詞など)を扱う必要がある[6]。最近では、対話音声中の各種の現象を分析した研究が行なわれている[7]が、分析結果を実験者の主觀によって分類する作業が一般的であり、実際に音声を聴取している被験者の動向/判断に基づく研究例は少ない。

我々は、協調的な問題解決の対話音声中に存在する間投詞に着目し「発話中の間投詞は聞き手に対してどのような働きを持つか。」、「協調的な応答文生成において間

投詞は有効又必要なのか。」という観点から、まず、知覚実験による検討を行なった[8]。

次に、上に述べた知覚実験より得られた知見を検証するため、WOZ方式の音声対話システムを用いて、被験者実験を行なった。実験結果より、間投詞挿入による効果が確認され、間投詞が音声対話システムにおける応答文生成時間の確保や、発話権の維持、及びシステムが動作中であることを示すサインとして有用であることが分かった。

しかしながら、被験者実験におけるユーザの発話を分析してみると、自然な会話に比べて、間投詞や言い直し、割り込みの出現が少ないなどの現象があった。この

現象は、「機械と対話しているという意識」のために起こる現象であるのか調べる必要があった。そこで、同じ応答性能を持った対話システムにおいて、聞き手(認識部と言語処理部)が人間であるか、又は、機械であるかという意識によってユーザのふるまいは変化するのか、先の実験で用いた音声対話システムを用い実験を行なった。

関連研究としては、以下のものがある。R.Amalbertiら[9]は、システムがコンピュータであると被験者に信じ込ませたコンピュータグループとシステムは人間が操作していると被験者に教えたオペレーターグループの2つを用意し、WOZ法を用いて被験者のふるまいや主観を比較している。その結果、コンピュータグループの被験者は使用する言語がより簡素化し、1対話当たりの単語数が多くなり、システムを単にツールとして使用する傾向があるのに対し、オペレーターグループの被験者はより対話者に協力的であると述べている。同様に、新美ら[10]は、京都観光案内をタスクとしたWOZシステムによって、機械らしく対応するモード1と人間らしく対応するモード2による応答の違いによる被験者のふるまいの違いを検討している。結果としては、モード1では、「対話全体の長さがあまり変化しないが、被験者の発話の数は少なくなる」、「不要語が少なくなる」、「単語もしくは名詞で終る句を用いた発話が多くなる」ことを示している。

新美らの方法は、応答のふるまいの違いによるユーザの発話の変化を対象としているが、我々の方法は認識・理解・応答は全く同じであり、ユーザの意識の違いによるユーザのふるまいの変化を対象としている。この考え方はR.Amalbertiと同じであるが、我々は同じタスクに対して人間同士の対話も比較対象としている。

## 2 聽取実験

### 2.1 予備的考察

従来の音声認識等では間投詞は冗長語として忌み嫌われ、特別な検出機構を設けることによって[1]最終的には不要語として扱われてきた。間投詞は、「通常の発話行為を継続できない/したくない」心理状況を相手に提示する術として無音を埋めるべく使用していると考えられる。もちろん長い無音を呈することでも同様な意図伝達は可能であるが間投詞の方がより効果的に働くと思われる(filled pauseといわれる所以)。

一方、間投詞の存在により、聞き手の理解が助けられる(発話内容を理解するための時間が与えられる)ことも考えられる。しかし、聞き手に時間を与えられることが間投詞の主な働きであるとするならば、単なる無音と置き換えるても何ら支障を来さないことになるが、果たしてそうであろうか。

以上のような考察から、以前行なった講演調音声に対

する知覚実験[2]と同様に以下のような項目が実験の検討項目としてあげられる。

間投詞の必要性(応答生成の立場から) 間投詞を抜き出して切り詰めた音声試料は聞き取りやすいのか/難いのか。

間投詞の種類の必要性(応答生成の立場から) 間投詞が発話を自然にするなどの理由から文生成において必要であるなら、いろいろな種類の間投詞を使い分ける必要はあるのか。

間投詞と無音の等価性 間投詞は無音と置換しても何ら支障はないのか。その等価性は対話速度などに依存しないのか。等価性があるならば違和感を感じない無音の長さはどの程度なのか。

間投詞の有無などによる無音の許容時間の変化 違和感を感じない無音時間は、無音後の第一発声が間投詞か否かによって変化するのか。またその無音時間は対話速度よりも変化するのか。

### 2.2 音声試料

協調的対話の音声試料として文部省重点領域研究「音声・言語・概念の統合的処理による対話の理解と生成に関する研究」の対話音声コーパスから「地図課題(主観的に対話理解難易度は高く、対話速度は速い)」と「地理・旅行案内(理解難易度は低く、対話速度は遅い)」の対話を用いた[3]。

これらの音声データにおける間投詞の位置とその前後の無音区間の位置を視察/聴取によって決定し記録した。これらの位置データを利用し以下のような処理を施した。

- (1) 間投詞を切り出して切り詰める。
- (2) 間投詞を同一時間長の無音(バックグランドノイズ、以下同様)と置換する。
- (3) 間投詞を異発話の間投詞と時間長が等しい無音と置換する。
- (4) 別の場所の同一種類の間投詞と置換する。
- (5) 異なる種類の間投詞と置換する。
- (6) 間投詞をランダムな無音区間と置換する。
- (7) 発話権の移行に伴う無音区間と語頭の間投詞をランダムな無音長と置換する。
- (8) 発話権の移行に伴う無音区間をランダムな無音長と置換する。

無音に置換する場合は、背景雑音を考慮し実際の対話中の無音部分の波形を用いて置換した。他の間投詞と置換する場合も、ピッチの連続性を考慮して置換対象の間投詞を決定した。異なる間投詞と置換する場合にはその間投詞の意味的素性等は考慮していない。

「間投詞の必要性や無音との等価性」を調査するためには、実験に用いる各対話を5等分し各々について(1)・

(2)・(4)・(5)・(7)の処理を施したもの用意する。この中から処理が重ならないようにランダムに選択し全対話を構築する。このようにして作られた試料(音声試料1)を4種類作成した(タスク毎に2種類)。「間投詞の自然性や聞きやすさ、理解しやすさ」の調査には、上述8種の処理から2つをランダムに選択し、約10ターン程の長さでランダムに抜き出した2つの同じ音声区間に對して適用したものを使用した(音声試料2,6 サブ対話×2種類×2タスク)。最後に「許容される無音時間の変化」を調査するために、対話音声の間投詞が含まれている1~2ターンを抜きだし(間投詞の位置「文頭・文中」と発話権の移行“あり・なし”の条件で選択)、a)間投詞の前の無音区間、b)処理(7)による無音区間、c)間投詞の後の無音区間の長さを250m秒刻みでいろいろ変化させた音声試料を作成した。

### 2.3 実験方法

まず、絶対的な評価として前節で示した音声試料1の内2対話(1対話×2タスク)を本大学の学生(6名・本実験+4名・予備実験)にスピーカを通して提示した。被験者には「提示音声には人工的な操作が加えられているので、不自然を感じる場所があったら知らせて欲しい」という指示を与えた。被験者が知らせてきた不自然を感じた場所は、実験者が書き起こしデータ上にマークし、音声試料提示終了後にマークされた位置において違和感を感じた理由をインタビューした。その後、簡単なアンケートに記入してもらった。

次に相対的な評価として、音声試料2を1サブ対話(2データ)毎聞かせては、その2種類の音声データを聞いて、自然性や聞きやすさなどに関してアンケートを記入させた。この場合も先ほどと同様に違和感を感じる場所があればその位置を知らせるように指示した。

最後に、対話の2~3ターンを抜きだし、間投詞とその前後の無音の長さをいろいろ変化させ、被験者が違和感を感じる無音時間の長さを調べた[5]。

### 2.4 実験の結果

知覚実験の結果より以下のような知見が得られた。

1. 間投詞に対する感じ方は、タスクや話速によって左右されない。
2. 間投詞の種類や有無、長さによって、文に対して違和感を感じることはない。
3. 対話において、無音で許される時間は基本的に2.0秒未満である。
4. 間投詞の有無やその発声時間によって、間投詞前後や間投詞部分に許される無音時間の長さは変化しない。
5. 無音で許される時間は、文頭、文中といった間投詞の位置によって変化しない。
6. 無音で許される時間は、文脈的な内容によってある程度変化する。

## 3 音声対話システムによる被験者実験

### 3.1 無音と間投詞の比較実験

前節で述べたような知見を検証するためと、対話システムにおいて検索や生成に時間が必要な時、間投詞は有効なのかということ調べるために被験者を用いた実験を行なった。タスクは電話番号案内とし、氏名・住所・年齢・職業から電話番号を検索してもらった。音声対話システムはWOZ方式のもの(図2)を作成した。実際の実験ではマイクからの被験者の発話をWizardが聞き、それに対する応答を予め録音されたPCMデータからスピーカで出力する方法で行なった。応答音声は実音声と合成音声の2種類を用意し、それぞれに間投詞の含まれているもの(全ての応答に間投詞が含まれている訳ではない)と含まれていないもの(間投詞部分が無音になっている)を用意した。間投詞は実音声の場合は、別の発話に含まれていた間投詞を数種類切り出し、応答文に種類はランダムに挿入した。合成音声の場合、規則合成では間投詞らしく音声を出力できないため、ピッチ変換などの音声処理を行なって間投詞に似せている。間投詞の発生規則は以下に示すような2種類の発生規則を用意した。これらのパターンに従ってを2秒・3秒・4秒の無音時間があいている応答文(無音部分は文頭・文中の2種類がある)に間投詞を挿入した。

#### 【間投詞発生規則1】

1. 1.5秒経過しても文(文節)生成が完了していない場合、とりあえず間投詞を発生
2. さらに1.5秒経過しても文(文節)生成が完了していない場合、次の間投詞を発生
3. 以降、2の繰り返し

#### 【間投詞発生規則2】

1. 0.5秒経過しても文(文節)生成が完了していない場合、とりあえず間投詞を発生
2. 2秒経過しても文(文節)生成が完了していない場合、次の間投詞を発生
3. さらに1.5秒経過しても文(文節)生成が完了していない場合、次の間投詞を発生
4. 以降、3の繰り返し

被験者は学部大学生9名で、対話システムに関する知識は全くない。被験者には最初に、実験の内容、検索データ、アンケートの書かれた紙を渡し、熟読してもらった。そして、実音声/合成音声、間投詞/無音の組合せである合計4種類の音声対話システムを用いて電話番号を検索してもらった。使ってもらうシステムの順番は被験者毎にランダムである。その後、被験者にはアンケートに答えてもらった。

### 3.2 実験の結果

まず、被験者の9人中6人が、間投詞無しの音声対話システムで、応答の文中の無音部分で自分の発話を始めてしまうことがあった。これは間投詞有りの音声対話システムではほとんど見られなかったことから、間投詞が発話継続の印として非常に有効であることを示している。しかし、合成音声の応答の場合、ほとんどの被験者が間投詞有りのシステムに非常に否定的であった。そのほとんどが非常に不自然な間投詞部分に注目していた。つまり、間投詞があまりに間投詞に聞こえないために、無音区間の違和感より間投詞の違和感の方が強いようであった。そのため、自然さ・聞きやすさ共に間投詞無しのシステムが良かった。一方、実音声では、間投詞有りのシステムは自然さ・聞きやすさ共に間投詞無しと変わらない(若干、人によっては良い)が、間投詞自体の違和感が無くなることによって、無音で代用する場合は、応答出力までの長い無音区間や応答途中の無音区間にによる違和感が強くなり、アンケートにもシステムがきちんと動作しているか不安になるという意見が多かった。また間投詞に対して、人間らしさ(被験者によってなれなれしいと感じる)という意見が多かった。実音声と合成音声ではほとんどの被験者が実音声が良いと答え、間投詞に関しては、システムがきちんと動作している感じがして安心できたという肯定的な意見や、なれなれしく聞こえるといった否定的な意見もあった。

アンケートの結果をまとめると以下のようになる。

- 間投詞は相手に発話継続のサインとして、非常に効果がある。つまり、発話生成にかかる時間を確保するのに有効である。
- システムとしての利点は、相手にシステムがきちんと動作していることを示す効果もある。これは、電話のような音声のみの対話に対しては非常に有効であると考えられる。
- 間投詞は、人間らしさや親しみと言ったものを相手に与えることができ、対話の自然性を向上させる効果がある。
- しかし、間投詞の効果を示すにはある程度の音声の品質が必要である。

### 3.3 ピープ音と間投詞の比較実験

間投詞は対話相手への発話継続やシステム動作の印として有用であるが、合成音声では音声の品質のために違和感が強い。それでは間投詞の代わりに電子音(ピープ音)を発話継続やシステム動作の印として使用できないか検証するために追加実験を行なった。応答音声は、合成音声(無音/ピープ音)と実音声(間投詞/ピープ音)の4種類の応答音声を用いて先の実験と同様な実験を行なった。被験者は学部大学生5名で音声対話システムの知識は全くない。

まず合成音声でのピープ音と無音の比較では、ピープ音使用的場合、発話の終りや動作への不安感がなく(5人中4人)、無音では不安があった(5人中3人)。これから、発話継続や動作中の印としてピープ音は効果があることが分かる。しかし、若干自然らしさの値が低かった。さらに、アンケートでは3人が無音の、1人がどちらでも、1人がピープ音のシステムを使いたいという意見であった。やはりピープ音自身の違和感がシステムの動作に対する不安感より強いためと考えられる。

実音声でのピープ音と間投詞の比較では、ピープ音が間投詞に比べ、自然らしさがほんの少し悪く、間投詞のシステムの方が人間と対話している感じがするという意見は一致している。しかし、ピープ音と間投詞への印象は、好みによって意見が分かれた(対話システムに人間らしさを求めるか求めないかで意見が別れる)。また、間投詞がいつも同じであることに違和感を感じることがあった。多数の間投詞の種類は用意する必要はないことは、聴取実験より得られているが、ピッチや長さの変化は付ける必要があると思われる。

### 3.4 ユーザのふるまいの分析

間投詞に関する被験者実験において発話された9人のユーザ発話、全572発話を分析した。まず、間投詞の出現回数は84回であり、これは全発話の約15%であった。しかし、その出現回数のうち被験者2人によって発話されたものが70(25+45)回にも及び(出現頻度は138発話中70回で51%)、それを除いた場合、434発話中14回で約3%となる。これは、自然な対話の分析結果に比べるとかなり少ない[7]。

また、その内訳を見てみても、「驚き」や「疑問」の間投詞が多く、場つなぎ的な間投詞は非常に少ない。今回の実験では対話の間を制御していること、タスクが簡単であることも原因と考えられるが、機械と意識しているためにこのような現象が現れているとも思われる。なぜなら、どのような話し方をしてもシステムはきちんと応答できることは被験者には伝えているにも関わらず、被験者はシステムにどう話し掛けで良いか悩み、被験者はシステムの対応できると思われるような発話の仕方で対話を行なおうとしていたためである。特にシステムとの最初の対話には顕著にそれが現れ、被験者は単語やそれに近い単位で話し掛けようとすることが多かった。

次に、言い淀みや言い直し、言い替えに関しては、4回と非常にその出現頻度は少なかった。

最後に、割り込みや発話の衝突に関しては、システムの発話に割り込んでの相槌が3回(相槌までの平均時間は176ms、最低107ms、最高213ms)、発話の終りを間違えたための発話の衝突が14回(衝突までの平均時間182ms、最低71ms、最高5670ms)、自分の発話の間違いに気づいたための独り言による衝突が1回出現した。特に発話の終りを間違えたための発話の衝突は、先に述

べたように、ほとんどが間投詞による無音の補間を行なわなかつた発話において出現し、その出現場所も全てが同様な場所で発生している。例えば「お名前は山本英二さんですね。」といった確認の発話の「お名前は」のあとに無音がきた場合である。このことから、その無音までの部分で一つの発話として成り立ってしまう場合には、間投詞をすぐ発生するなどして、無音をできるだけなくす必要があることが分かる。つまり、このような場合(発話の途中)には間投詞発生規則2の方法が有効であり、文頭には間投詞発生規則1が有効であると思われる。また、発話の衝突の回避の別のアプローチとしては、ある時点までに生成の完了した部分が一つの発話として成り立ってしまう場合は、次の部分が生成完了するまで出力しないといった考え方もある。

## 4 ユーザのふるまいに関する被験者実験

### 4.1 実験

前節で被験者の発話のふるまいが、実際の自然な対話とかなり異なっていると述べた。ではその原因は何にあるのか。被験者が機械(人間とは異なるもの?人間に劣っているもの?)と対話しているという意識によってこのような現象が現れるのか。もし仮に人間と全く同じふるまいをするシステムがあっても、ユーザは人間とは同じふるまいをしないものなのかな。これらに関して考察するために、3節の実験で用いたWOZの音声対話システムを用いて被験者実験を行なった。

被験者は大学生12名、タスクは電話番号案内であり、応答音声は、間投詞や無音を挿入していない実音声だけを使用した。被験者が対話する相手は以下の3種類である。A)人間のオペレーター[図1]、B)聞き手はオペレーターで、応答は最適な録音音声をオペレーターが選択[図2]、C)聞き手も応答も機械[図3](実際のシステム構成は“B”)と全く同じ、つまり、被験者の意識だけが異なる)。また、B)とC)のシステムの違いを被験者に感じさせるために、C)のシステムには被験者の発話の開始と同時に「認識開始」、発話の終了と同時に「認識終了」の文字をディスプレイに表示している。応答はリアルタイム性を重視し、できるだけ早く応答するようにした。被験者には、聞き手が人間や機械であることを図などで説明し、その上で自由に話してもらった。3人の被験者には3つの対話相手全てと対話してもらい(順番は、“C”→“A”→“B”)、残りは3人ずつ、それぞれの対話相手とだけ対話してもらった(参考データとして他の対話相手とも対話はしてもらっている)。

### 4.2 実験結果

まず発話数は、A)で120発話(参考データを含め224発話)、B)で106発話(227発話)、C)で123発話(265発話)であった。C)のシステムは、B)と認識・理解・

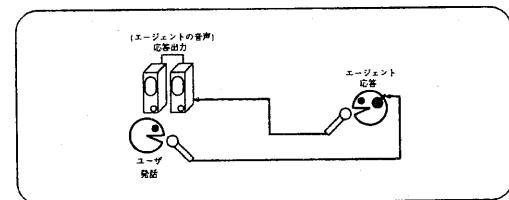


図1: 人間のオペレーター“A”

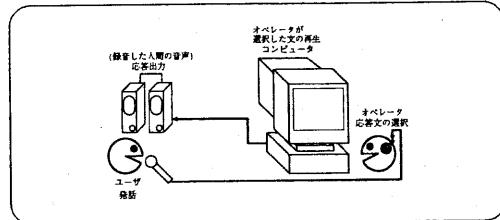


図2: 聞き手がオペレーター“B”

応答能力は全く同じであり、聞き手が人間か機械かという意識の違いだけで、総発話数が増えている。これは被験者が、人間では一度に伝えられることを細切れに伝えたり、最初に何を言って良いか分からぬために発生する不要な発話のためであった。

間投詞の出現回数に関しては、A)で24回(56回)、B)で20回(47回)、C)では20回(26回)と明確な差はでなかった。これはC)の被験者の一人が13回も間投詞を発声しているためである。そのため、参考データも加えた結果でみると、やはりB)とC)のシステムの間で明確に間投詞の出現回数に差が出た。

言い直しや言い替え、言い淀みは、A)で5回(9回)、B)で2回(2回)、C)で2回(2回)となった。これに関しては、人間との対話においてのみ出現回数が多かった。割り込みや発話の衝突などは、A)で7回(8回)、B)で0回(1回)、C)で2回(2回)であった。特に人間との対話では、相槌による発話の割り込みが多かった。これは、応答を録音音声で行なうため、文単位で応答しているB)やC)では見られなかった。割り込みなどに関しても今回の実験では一切制限していないが、やはり録音音声の応答方法自体が割り込みしにくい雰囲気を作りだし、割り込みを許してしまった可能性がある。

1対話の平均時間に関しては、A)で34秒(33秒)、B)で48秒(49秒)、C)で58秒(57秒)であった。1ター

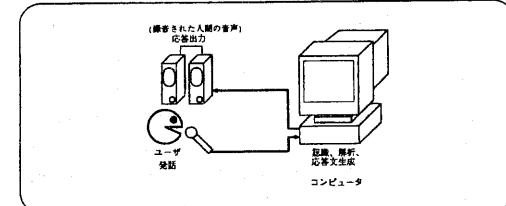


図3: 音声対話システム“C”

ンの平均時間は A)5.3 秒 (5.4 秒)、B)8.2 秒 (7.8 秒)、C)9.0 秒 (8.0 秒) であった。録音音声の場合はゆっくりと丁寧に発話していることや、どうしても応答の選択に必要な時間があるため、人間同士に比べて対話時間は全体に伸びている。B) と C) で 1 ターン平均時間はあまり違いがないことから、発話数が B) と比べ C) が多いために対話時間がかかっていることがわかる。

他の特徴としては、ユーザが機械と意識してしまうと、どのように言って良いのか悩む傾向が強いことが分かる。そのため、システムとの最初の発話で「片岡幹雄」や「花井俊一、32 歳」といったような、あるシステムに特化した単語の発話をすることがある。これは他の対話相手の場合は、ほとんど出現しない。また、C) のシステムとの対話においてだけ、そばに待機している実験監視者に話し方に関する質問することが多かったことからも分かる。

アンケート結果でも、12 人中 7 人が機械、つまり C) に対して、意識してゆっくりと丁寧に話したと記入した。さらに、B) と C) では応答までの時間は実際にはほとんど一緒であるにも関わらず、B) のシステムは応答が遅く感じる記入した。また機械の利点としては、人間に比べ質問する時の圧迫感がない(ゆっくり考えられる)という意見が幾人かあった。逆に人間の利点としては、間違えても何とかなる安心感があるという意見であった。対話システムに対して、人間らしさを望むか、機械らしさを望むかというアンケートには、6 人が人間らしさ、5 人が機械らしさを望み、1 人が中途半端な人間らしさならいらないと答えた。

今回の実験では、システムの聞き手の部分(認識・理解)に対するユーザの意識を人間であるかと機械であるかの 2 種類に変化させ、応答部分は録音音声という制限された応答能力とした。実際の人間同士の対話(A)と聞き手が人間である対話(B)を比較すると、応答能力の制限によってもかなりユーザの発話が変化することがわかる。このことから、人間同士が行なうような雑談的な自然な対話を行なうためには、認識・理解が完全で、ユーザの発話に対してただ最適な応答するだけでは駄目で、もっと、抑揚のような人間らしさ、対話のリズムや、相槌といったシステムからの割り込みなどの要素も自然な対話には重要であることが分かる。また、性能としては人間と全く違いはなくとも、ユーザは機械と対話していると意識するだけで、聞き手が人間の場合と同じように対話することができず、どこか不自然な、基本的には機械に協力的な発話になることもわかった。これは、コンピュータや機械に対する知識があるほどどの傾向は強いと思われる。このような意識がある限り、機械と話すこと 자체が目的のような雑談的な対話は難しいだろう。しかし、協調作業的な対話システムを構築するような場合、機械と意識した場合に起こり得る発話パターン、つまり、機械にできるだけ協力的に発話しよ

うとするユーザの意識を利用することはシステム開発において有用であると考えられる。また、実際のシステム開発においてもユーザの機械と対話しているという意識を考慮する必要があるだろう。

## 5 まとめ

対話音声中の間投詞に関する聴取実験で得られた知見を検証するために、音声対話システムの応答文を制御した被験者実験によって、間投詞が発話継続やシステムの正常動作の印として非常に有効であること、人間らしさを与えることなどが分かった。またビープ音でも同様の効果が得られるが、自然なしさを若干損なうことが分かった。しかし、合成音声の品質では、間投詞自体の不自然さのため、間投詞の発話における利用は非常に困難である。

さらに、ユーザの意識によってユーザのふるまいが変化するのかどうかに関して被験者実験を行ない、ユーザは機械と意識するとどうしても機械に協力的な発話になってしまうことが分かった。また、ユーザによっては、機械には機械らしさを求めていることも分かった。今回は言語的な現象に関してのみ分析したが、音響的にも変化するのか今後見ていきたい。

## 参考文献

- [1] 甲斐、中川：「冗長語・言い直し等を含む発話のための未知語処理を用いた音声認識システムの比較評価」、電子情報通信学会論文誌, Vol.80-DII, No.10, pp.2615-2625 (1997)
- [2] 峯松、片岡、中川：「講演調の話し言葉に対する分析」、情報処理学会・音声言語情報処理研究会資料、95-SLP-8-7, pp.39-46 (1995)
- [3] 板橋、山本、河原：「重点領域研究「音声対話における音声コーパス」、人工知能学会研究会資料, SIG-SLUD-9701-5 (1997)
- [4] 田嶋：「音声言語の言語学的モデルをめざして—音声対話管理標識を中心に—」、情報処理学会会誌, Vol.36, No.11, pp.1020-1026 (1995)
- [5] 伊藤、峯松、中川：「対話音声中の間投詞の分析と音声対話システムにおけるその応用」、人工知能学会、第 12 回全国大会, 29-03 pp.499-502 (1998)
- [6] 森元、村上：「音声対話における言語現象」、日本音響学会誌, vol.50, No.7, pp.558-562 (1994)
- [7] 中川、小林：「自然な音声対話における間投詞・ポーズ・言い直し出現パターンと音響的性質」、日本音響学会誌, vol.51, No.3, pp.202-210 (1995)
- [8] 伊藤、峯松、中川：「対話音声聴取時における間投詞の働き」、情報処理学会、第 55 回全国大会論文集, 3J-2, pp.27-28 (1997)
- [9] Rene Amalberti, Noelle Carbonell, Pierre Falzon: User representations of computer systems in human-computer speech interaction, Int.J. Man-Machine Studies, pp.547-566 (1993)
- [10] 新美、小林：「音声を用意にする対話制御方式」、文部省重点領域研究“音声対話”公開シンポジウム予稿集, pp.49-58 (1993)