

# 雑談可能な目的達成型音声対話システム

大村祐司<sup>†</sup> 川端豪<sup>†</sup>

目的達成および雑談的な対話制御機能を併せ持つ音声対話システムを構築した。目的達成のための対話制御はスロットフィリングに基づいて行われる。一方、雑談的な対話制御は刺激-応答型の事例ベースで行われる。刺激-応答に基づく対話制御部は雑談のきっかけとなる単語を監視し、発見すればその単語に応じた雑談を開始する。数回のやり取りの後、制御は目的達成型対話制御部に戻る。11名の被験者にこのシステムと対話をして貰い、アンケート結果を集計したところ、雑談機能を持つ音声対話システムのほうが「面白い」「興味深い」「人間的」などの項目でよい評価を得た。

## Implementation of the chat function into the task-oriented spoken dialog system

YUJI OMURA<sup>†</sup> TAKESHI KAWABATA<sup>†</sup>

This paper describes the spoken dialog system which has the task-oriented dialog-flow controller and the chat capability. The task-oriented part of this system is based on the slot-filling semantics and controls the dialog to accomplish the given task. On the other hand, the S-R (Stimulus-Response) example-based dialog controller which drives chat behaviors is implemented into this system. The S-R dialog controller watches for trigger words, and takes a dialog initiative for beginning the chat. After some words, the control is returned to the task-oriented dialog controller again. The subjective evaluation shows that the dialog system which has chat functions is more “fun”, “interesting” and “human-like”.

### 1. はじめに

近年、音声認識技術の発達とともに、それをを用いたインターフェースの需要が高まり、音声対話システムの開発が行われ、社会で用いられ始めている[1]。また、それらのシステムの多くは扱うタスクが定まっており、そのタスクを達成する上で必要のない会話を交わすことはない[2]。つまりシステムとユーザの間で、目的遂行のために必要な情報の伝達のみが行われる。

しかし、実生活において人間対人間で会話をを行い、何らかのタスクを達成する際には、タスクを達成する以外の会話を交わすことがよくある[3]。特に親しい者同士で会話による情報伝達を行う場合には、雑談を交わしている中で、必要な情報を織り交ぜることも考えられ、対話の中におけるタスク遂行に必要な会話の割合が小さくなることさえ考えられる。例えば、夕食の献立の確認を行う際に、ある品物が安売りされていた、テレビでレシピを紹介していた、などの理由やその料理にまつわる思い出など様々な話題があり、そういった雑談を行う中で、「今晚の夕食はこれでもいい？」という確認が行われる。

このように人間の日常会話においては目的遂行以外の情報伝達が行われるが、これは雑談を織り交ぜることによって対話を円滑に進めたり、あるいは対話自体を楽しむためである。

もし、対話システムが人間と同じように、目的遂行に必要な内容以外の発言を織り交ぜて会話を進めた場合、ユーザはそれに対しどのような印象を持つのだろうか。より人間らしさを感じ、会話がしやすくなるのか、逆に違和感によって会話が妨げられるのか、結果を容易に想像することができない。

本報告では、タスク遂行型の音声対話システムに、タスクと関係の深い話題に関する雑談機能を与え、あるタイミングで会話内容を雑談へと振り向けようとする行動をさせてみた。そして、このようなシステムの振る舞いに対し、ユーザがどのような印象を持つかを事後アンケートによる主観評価によって観察することを試みた。

<sup>†</sup> 関西学院大学  
Kwansei Gakuin University

## 2. 雑談可能な目的達成型システム

### 2.1 対話システムの構成

タスク遂行（目的達成）と雑談を同時に行うための音声対話システムの構成を図1に示す。

図1の音声認識部ではJulius[4]の、ディクテーションモードによる音声認識を行う(2.2節)。

対話制御部は全体の対話制御を行い、タスク遂行時、雑談制御時にそれぞれに応じた動作を行う。対話制御部はタスク進行のための意味理解部を持っており、スロットフィリングによるタスクの進行管理を行う。基本的にはまだ埋まっていないスロット値を埋めるための発声を要求する。一方、対話制御部は雑談を行うために、刺激-応答型の事例ベース行動制御機構を併せ持っており、雑談時にはユーザが入力した音声に対して決められた応答を返す。

応答生成部では、エージェントがアニメーションによって反応を返す画像的な応答と、事前に用意した音声ファイルを再生する対話の応答が行われる。

### 2.2 音声認識

音声認識部にはJuliusのディクテーションモードによる音声認識を用いている[4]。

音声対話において、目的の達成、および本報告における雑談を行うために必要であると考えられる語彙や文法を含む辞書ファイルを手作業で作成した。辞書は単語数336、文法数85で構成されている。入力を受け付ける文例を図2に示す。

今回音声認識はシステムが発声している際にも動いており、システムが発声している間の入力音声は次のユーザの応答として扱われているが、互いの音声が行き交うことによる混乱を防ぐため、被験者にはできるだけシステムの発声が終わるのを待ってから、入力を行ってもらった。

### 2.3 対話制御

#### 2.3.1 目的達成対話制御

システムが与えられたタスクを達成するためにユーザと対話をし、必要な情報を取得していく上でどのような会話をすべきかを想定した。今回はタスクを映画のチケット販売を想定してシステムを構築した。映画のチケットを販売するシステムであった場合、予約を完了させるために必要な情報は「映画名」「希望する上映時間」「座席番号」の3つで、選択することができる映画名、上映時間、その時間の空席情報等は予め設定しておき、対話開始前に、ユーザには映画名とその映画が上映されている時間が開示されている。

実際の対話の流れとしては、まずシステムはユーザに対し「希望する映画」を質問し、映画名の情報を取得しようとする。しかし、ユーザによっては映画名と同時に上映時間までを入力することが考えられるため、映画名、上映時間が同時に入力された場合でも正しく反応するように設計した。そしてユーザが希望する映画の情報を得られたとき、同様に希望する上映時間と座席、あるいは希望する座席のみをユーザに対して尋ねる。

また、対話において三回以上同じ項目で入力ミスがあった場合は、ユーザが入力すべき情報を誤解している可能性を考慮し、どういった内容をどのように発声すればよいか指定するようにした。

- U1: こんにちは  
 U2: (映画名) でお祈いします  
 U3: (映画名) の(時間) で  
 U4: どんない映画をやってるの?  
 U5: 評判がいいって聞いたから  
 U6: 友達に勧められたので

図2 認識を受け付ける文書例

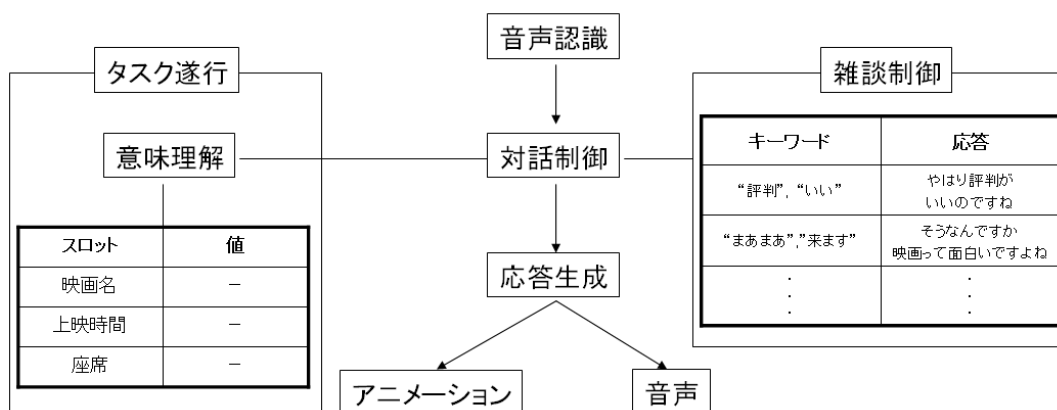


図1 システム構成図

- S1:** それは面白いと評判の映画ですね。  
どうしてその映画を見ようと思ったのですか？
- U1:** ネットで面白いと見かけたので
- S2:** やはり評判がいいのですね  
よく映画を見られるのですか？
- U2:** あまり見に来ませんね
- S3:** そうなんですか？では今日は何か特別な日ですか？
- U3:** 今日はデートなんです
- S4:** それはすてきですね。すばらしい一日にしてください  
それでは次はご希望の上映時間を選択してください

図3 雑談例1

- S1:** それは有名な海賊映画ですね  
どうしてその映画を見ようと思ったのですか？
- U1:** 友達に勧められたんです
- S2:** なるほど、そのご友人とはよく映画を見に来られるのですか？
- U2:** あまり見に来ません
- S3:** そうですか、ぜひ一緒にいらしてください。  
お一人ではよく見に来られるのですか？
- U3:** まあまあ見に来ます
- S4:** そうなんですか、映画って面白いですよ。  
それではお席の予約に移らせていただきます

図4 雑談例2

### 2.3.2 雑談対話制御

雑談対話は刺激-応答型の行動制御を行っており、事前に設定されているキーワードが入力されたとき、事例ベースに基づいてシステムが決められた応答を返すようになっており、雑談が開始される。そしてある程度の回数会話を重ねると、雑談を終了させ再び目的達成に必要な情報をユーザに尋ねてくるよう設計されている。

雑談例を図3、図4に示す。



図5 システムエージェント

### 2.4 応答生成

システムが起動してから目的を達成するまでの間、モニタには女性をモデルとしたキャラクターを表示し、入力や状態に応じて、いくつかの仕草をとらせた(図5)。

キャラクターの表示にはあらかじめ用意しておいたビットマップ形式の画像ファイルを、順番に表示させることでアニメーションとして表示させている。用意した仕草は、「ユーザからの入力待ち状態時、一定間隔で瞬き」、「ユーザからの入力システムが期待していたものであった時、頷く」、「すべてのタスクが遂行されたとき、別れの挨拶をする」の3つである。

またシステムは対話の発声としてあらかじめ男性が録音した音声ファイルの速度を早くすることで音程を変化させたものを再生している。

### 3. 実験条件

本報告では、目的を達成するシステムが雑談を織り交ぜる事により、ユーザに対しどのように影響を与えるかを調べるために、今回は映画のチケット販売を想定し、システムを設計した。これは映画のチケット販売であれば交わされる雑談が明確であり、雑談の内容が設定しやすいと判断したためである。

実験には雑談を織り交ぜるシステムと、雑談を織り交ぜないシステムとの会話を用意し、被験者に2つのシステムを使用してもらい、対比較実験によって確かめることとした。なお今回の実験では、必ずシステムと雑談を行ってもらうため、被験者にキーワードとなる言葉の1つを伝えておき、発言してもらった。

実験の対話では被験者にヘッドセットを利用してもらい、ヘッドフォンから出力されるシステムの音声を聞き、対話をしてもらった。

行われる対話はチケット販売に必要な情報である映画名、上映時間、座席番号を取得するための会話をベースとし、雑談可能なシステムではシステム側が話題を提供する数発話程度の雑談を加えたものとなっている。

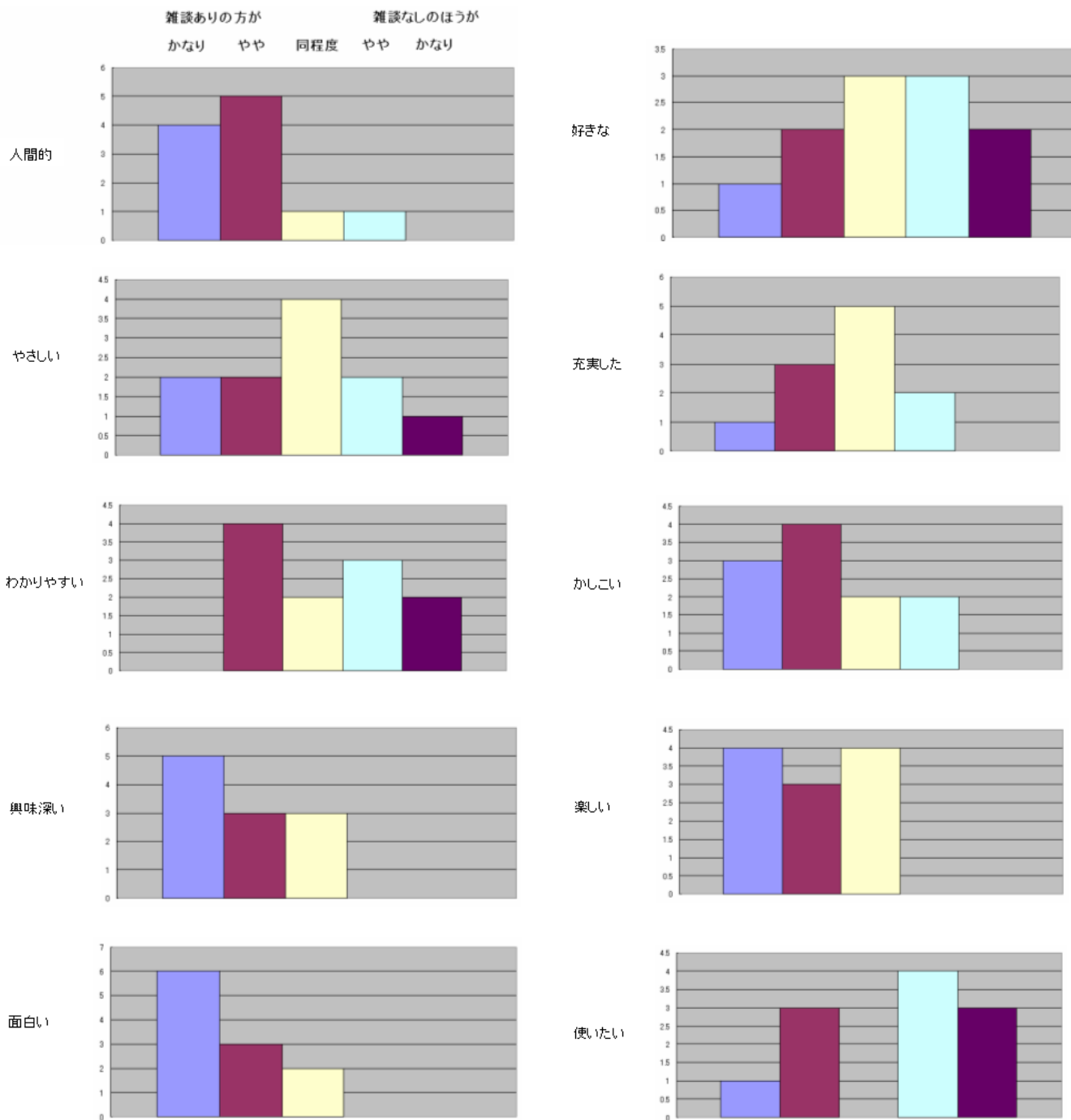


図6 アンケート結果

表1 質問項目

人間的	好きな
やさしい	充実した
わかりやすい	かしこい
興味深い	楽しい
面白い	使いたい

実験には男性8名、女性3名の計11名に協力してもらい、2つのシステムそれぞれと対話を行なってもらった後、表1の10個の項目について、どちらがより当てはまるかを5段階で評価してもらった。なお雑談あり、雑談なしの対話をしてもらう順番は被験者ごとにランダムとし、偏りがないように実験した。

#### 4. 結果

アンケート結果をまとめたものを図6に示す。

この結果から、まず分かることは今回雑談可能なシステムとして用意した対話システムは、「面白い」、「興味深い」、「楽しい」といった会話としての面白さをあらわす項目で高い評価を得ていることである。これはただタスクを遂行するよりも雑談を混ぜたものの方がより楽しいと被験者が感じたことを示していると考えられる。

次に「人間的」という項目についても雑談を交えたもののほうがより高い評価を得ていることから、タスクを遂行する対話であっても、その中で雑談を挟む対話のほうがより人間的であると感じられるということがわかった。この結果から人間が日常の対話行動において目的のある対話であっても、対話とは直接関係のない会話を交えながら会話をしているのではないかと考えられる。

さらに、「使いたい」に着目してみると、この項目では雑談のないシステムのほうが高い評価を得ている。つまり、被験者は雑談を交えないシステムをより使いたいと感じていることが分かる。この結果と面白い、人間的の項目を見て考えられることは、ユーザは目的達成型システムを使う際に、面白さや人間らしさを求めておらず、面白いから、人間的であるからといってそのシステムを使いたいと感じるわけではないと言うことが分かった。

#### 5. まとめ

本報告では、目的達成型の音声対話システムが、その過程においてユーザと雑談を交わした場合、ユーザはシステムにどのような印象を受けるのかを対比較実験によって確かめた。

実験に用いたのは映画のチケット販売を行う音声対話システムで、ユーザの希望する「映画名」「上映時間」「座席番号」を取得することを目的とする。

音声認識には Julius を使い、ディクテーションモードに

よって音声認識を行った。対話の間ディスプレイには女性キャラクターのエージェントが表示され、ユーザから正しい入力があればわずく等のアニメーションを行った。雑談はシステムの主導で行われ、システムの提示する話題、質問に応答を返していれば会話が成立するよう設計されている。

実験は被験者11名に協力してもらい、アンケートでシステムから受けた印象を評価してもらった。

結果、雑談を交えるシステムは、より人間的で面白い、という評価を受けたが、雑談を交えないシステムの方がより使いたいと多くの被験者が感じたことから、目的遂行型のシステムに、人間らしさや面白さを求めている人の方が少なく、面白いからと言って使いたいと思うわけではないということが分かった。

#### 参考文献

- 1) 河原達也: 音声情報処理技術の最先端 : 4. 話し言葉による音声対話システム, 情報処理学会誌 vol. 45, No. 10, pp. 1027-1031 (2004)
- 2) 栗野健太郎, 伊藤仁, 伊藤彰則, 牧野正三: マルチタスク型システムの対話制御のための発話識別, 情報処理学会研究報告, vol. 2009-SLP-79 No. 7 (2009).
- 3) 尾崎健太郎, 小暮悟, 甲斐充彦, 小西達裕, 伊藤幸宏: 複数の車内機器操作と雑談を扱えるマルチタスク音声対話システムのユーザビリティの向上, 情報処理学会研究報告, vol. 2010-SLP-80 No. 6 (2010).
- 4) 京都大学河原研究室, 名古屋工業大学 Julius 開発チーム: 汎用大語彙連続音声認識エンジン Julius, <http://julius.sourceforge.jp/>, (参照 2012-11)