

# 韻律・表層的言語情報を発話タイミング制御に用いた 音声対話システム

A spoken dialog system activating the natural response timing using  
prosodic and linguistic information

豊橋技術科学大学 情報工学系

Department of Information and Computer Sciences, Toyohashi University of Technology

中川聖一

Seiichi NAKAGAWA

< 研究協力者 >

豊橋技術科学大学 情報工学系

Department of Information and Computer Sciences, Toyohashi University of Technology

北岡教英

Norihide KITAOKA

竹内真士

Masashi TAKEUCHI

If a dialog system can respond to a user as natural as a human, the interaction will be smoother. It is important to generate natural timing of responses to user's utterances to make a natural conversation like human to human. In this paper, we developed a dialog system which could generate timing of *aizuchi* and turn-taking based on a decision tree in real time. We considered the pitch and the power contours at the last 100 ms of an utterance, part-of-speech of the last word and the last content word in the last utterance as the features of the decision tree.

We constructed a decision tree based on the features related to the information described above, and developed a timing generator using rules derived from the decision tree. We evaluated the timing generator by a subjective evaluation method in which subjects actually talked to the system.

Key words: Prosodic information, Spoken dialog system, Resoponse timing

## 1 研究の目的

人間と機械が対話を行うことを考えるとき、機械が人間同士の会話と同様にあいづちや割り込みなどの応答を返すことができれば、より円滑な対話を行うことが期待できる。

人間同士の会話においては、話者は互いにあいづちやうなずきでその発話を理解していることを明示する。メイナードによればあいづちは「続けて」というシグナル、内容理解を示す表現、相手の意見や考え方に賛成の意思表示をする表現などを表すものであるとし [1]、水谷は、あいづちは「そこまでわかったから、つぎどうぞ」という合図で、いわば「進め」の青信号のようなものであるとしている [2]。田窪はあいづちの種類毎にその機能を分類している [3]。また、発話者が交替するタイミングについては、相手の発話の終了から適切な間で発話を開始したり、時

にはオーバーラップしたり割り込んだりするなどして会話はスムーズに進行する。

特に雑談のような対話に着目すると、その中ではたわいもない内容でありながら、あいづちや話者交替をタイミングよく繰り返すことによって継続されていく。すなわち、親しみやすい対話インターフェースを構築するには、このようなタイミングの考慮が不可欠である。

これまでもリアルタイム応答として、ユーザの発話にあいづちを打つという研究は行われてきている。Wardらは低ピッチ区間が一定時間続く箇所においてあいづちを生成するシステムを構築している [4]。平沢らは連続音声認識アルゴリズムの中間結果を用いて言語情報からあいづちを打つこと検討している [5]。岡登らは、コーパス上で実際にあいづちが打たれた箇所の直前の韻律特徴のテンプレートを作成することで、あいづちの生成を行っている [6]。佐藤らは、

実際の人間同士の会話でのあいづちの出現頻度からあいづちの生成フローチャートを作成し、あいづちの生成を行っている [8]。さらに、野口らは、韻律情報、品詞情報があいづちに与える影響を調べており、局所的な韻律情報のあいづちシグナル性は品詞情報のそれに比べて低いとしている [9]。

本研究においても、韻律情報を用いたあいづち生成システムの試作を行い [10]、このシステムで生成されたあいづちのうち、その 70% がほぼ自然なものであるという評価が得られている。

一方、発話交替については、Sacks らは話者の交替は発話者が質問や確認の発話を行って聞き手に対して返答を求めた箇所で発生するものが望ましいとしている [11]。

これまでユーザの使い勝手などを考えてシステムの発話に対して割り込む（パージン）ことができるシステムは存在するが [5][12]、ユーザに対するシステム発話のタイミングを考慮した例はない。

そこで本研究では、まず自然な雑談対話をする上で最も重要であるタイミング生成、すなわち、あいづち、割り込みのタイミングの判定を、人間同士の対話の特徴から学習した決定木を用いて行うことを考える。この決定木は、韻律情報と言語情報を素性として用いるが韻律情報は発話句音声末のおよそ 100 ms の変動、言語情報は発話終端単語の品詞や発話の最後に現れた自立語の品詞情報などを考える。この決定木により生成されたタイミングを人間との一致度合い、および被験者による聴取実験により評価する。

得られた決定木で生成されたルールで実際に検出された話者交替、あいづちの検出を行った例を示し、また 3 つのタスクの対話を決定木によって分類したときのそれらの再現率と適合率を算出し、ルールの評価を行う。そして決定木から得られた話者交替とあいづちのタイミングの自然性を被験者が聴取することにより評価する。

さらに、このタイミング生成器を用いて実際に音声対話システムを構築する。このシステムは、リアルタイム音声認識と、タイミング生成器の出力から、発話内容のみでなくタイミングも考慮して応答する。本システムのユーザによる使用実験を行い、あいづちや返答タイミング、返答内容の自然性の評価を行う。

## 2 人間の対話の分析

### 2.1 対話コーパス

分析に用いた対話コーパスは、人工知能学会コーパス利用研究グループの談話タグ付き対話コーパス（全 29 対話） [13] を用いた。コーパスは、対話の音声（L, R チャンネル）とそのタグ付き対話コーパスからなる。

本研究では、この対話コーパスのうち 3 タスク 11 対話（1842 発話）をトレーニングとテストに用い、タスクは雑談タスク、旅行案内タスク、テレフォンオペレータとの対話の 3 つとした。

### 2.2 対話コーパスの分析

#### 2.2.1 韻律情報

小磯ら [14] は発話句音声末 1 モーラ分のピッチ、パワーが表 1, 2 のように変動パターンによって、発話継続、終了、あいづちが出現する傾向が異なると分析している。この知見より、句音声末の 100ms（およそ 1 モーラ分）のピッチおよびパワーの変動を回帰係数により調べることで判別したパターンがあいづちや話者交替に関係するものと考えられる。

また、Geluykens ら [15] や Hirschberg [16] もあるフレーズにおけるピッチの変動とその長さが、発話終了と発話継続に関係があるとしている。本研究においてもこれらの現象が確認できた。

#### 2.2.2 言語情報

話者交替がおきるとき、その発話の終端単語の品詞は、助詞、助動詞・動詞、感動詞であるケースが多く見られた。雑談タスクの場合は助詞「ね」、それ以外のタスクの場合は「か」の後に話者交替が見られる傾向にあり、終端の助詞の種類は話者交替に関係するものと考えられる。

人間同士の会話では、あいづち、話者交替が会話の焦点となるフレーズ、相手が発すると予想される語が現れた直後、もしくはそのようなフレーズを含む発話の終了時に出現することが多い。特にテレフォンオペレータとの対話タスクであれば、注文商品名などがユーザから発せられた直後にあいづちが打たれている。

また岡登ら [6] は、あいづちが打たれる場合は発話長が長くなる傾向にあると分析しており、ユーザの発話長に大きく影響されるものであると考えられる。

表 1: ピッチパターンの変化による発話の継続・終了、相槌の頻度（小磯ら [14] による）

ピッチパターンの変動	継続	終了	相槌回数
平坦型	36	15	3
上昇型	0	1	1
下降型	7	6	1
平坦長音型	4	0	0
下降長音型	3	0	0
平坦下降型	11	3	7
上昇下降型	13	3	10

表 2: パワーパターンの変化による発話の継続・終了、相槌の頻度（小磯ら [14] による）

パワーパターンの変動	継続	終了	相槌回数
平坦型	39	7	17
平坦下降型	25	5	4
下降型	4	16	0
平坦長音型	6	0	0

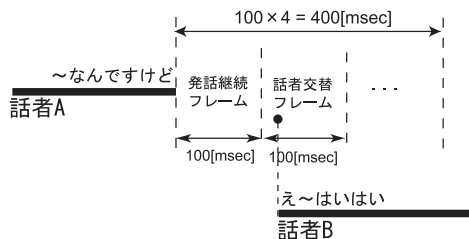


図 1: 発話の無音部の分析方法

### 3 応答タイミングを生成する決定木の構築

#### 3.1 決定木の構築

前節の分析結果に基づいた特徴を素性とした決定木を用いて話者交替、あいづちのタイミングを検出することを考える。

決定木の生成は帰納学習システム C4.5[7] を用いて行う。C4.5 は与えられた学習データで初期決定木を構築、その後枝刈りを行う。

ユーザの発話からポーズはリアルタイムに検出できるものとし、ポーズ区間に入ってから、100ms から 400ms までの範囲を 100ms 毎に、それらをあいづち / 話者交替 / 発話待ち 1 / 発話待ち 2 のいずれ

かに分類して出力することを考える。ここで、発話待ち 1 とは次発話に同じ話者が続くときに先行するポーズのクラス、発話待ち 2 とは、次発話にあいづちもしくは話者交替が発生するときに先行するポーズのクラスである。

そこで図 1 のようにコーパス中のユーザ発話の無音部を 100ms 毎に区切って 400ms まで各フレーム毎に素性の値を求め、正解を付与して学習データとした。素性としては、

1. ユーザの発話長
2. 発話終端単語の品詞
3. 特に終端単語が助詞であった場合に「ね」、「か」であるかどうか
4. 発話終端からの時間長（～400ms）
5. 発話中の最後の自立語の品詞
6. その自立語の時間長
7. その自立語から発話終端までの時間長
8. 発話句音声末 100ms のピッチの変動
9. 発話句音声末 100ms のパワーの変動

の 9 つを考える。このうち、1,4,6,7,8,9 が連続値である。ここで、ピッチ、パワーの変動は発話句音声末 100ms を 3 分割し、それぞれの区間の回帰係数を計算することで求めた。

タスクは前述の雑談タスク、旅行案内タスク、テレフォンオペレータとの対話タスクであり、それぞれ 3 対話 (92 発話)、2 対話 (127 発話)、3 対話 (318 発話) をトレーニングに用いた。さらに決定木の学習は、これらすべてを同時に用いた場合と、タスク別に分けて用いた場合を行った。

#### 3.2 決定木の分析

3.1 節で述べた手法で構築した決定木の一部を図 2 に示す。ここで構築された決定木に現れた主要な素性は、発話長、ピッチやパワーの変動といった韻律情報がほとんどであった。

言語情報は構築された決定木にはほとんど現れなかった。発話中の最後の自立語は主要な素性であったものの、言語情報は自立語であるかどうかという情報のみであり、韻律情報があいづちと話者交替を検出する重要な要因となっていると言える。

また、タスクの違いの影響を調べるため、タスク毎に作成した決定木においても同様の傾向を示し、言語情報よりも韻律情報があいづちと話者交替を分

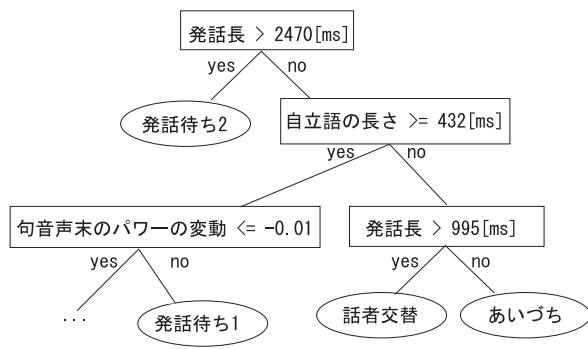


図 2: 学習された決定木の一部

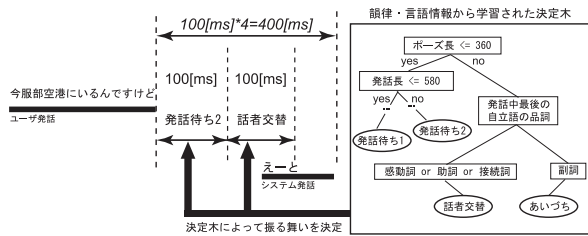


図 3: 決定木によって検出されたあいづち，話者交替タイミング例

類する上で重要な素性となっていた．このことからあいづちや話者交替の検出には，すべてのタスクで学習した場合の決定木の結果と同様，韻律情報に加え自立語であるという判定などの表層的な言語情報が大きく関係しているといえる．また，韻律情報が主要な素性であることから，あいづちは先行する発話内容にはとらわれずに打たれているものと考えられる．

## 4 決定木による応答タイミング生成

3.1 節で構築した決定木を用いたタイミング生成を行う．図 3 に示すように，ポーズ区間を 100ms 毎に区切っていき，各区間において決定木により動作を決定する．

### 4.1 決定木の評価

テストデータはそれぞれのタスクごとに無作為に取り出した 1 対話 (それぞれ 25 発話，195 発話，102 発話) をオープンテストデータに，トレーニングデータから 1 対話ずつを取り出したものをクローズドデータとした．表 3 にクローズドとオープンの

場合の人間の対話との一致度合いを示す．

人間同士の対話で，あいづちが打たれていない箇所や話者交替が起こっていない箇所は，必ずしもあいづちを打つべきでないわけではなく，話者交替するのは不自然な場所とは限らないので，人間の対話との一致は一つの目安に過ぎないことに注意を要する．

オープンデータでは，あいづちのクラス分類は再現率 57.4%(クローズドデータでは 53.8%)，適合率 24.8%(82.1%)，話者交替のタイミングのクラス分類においては再現率 44.8%(76.0%)，適合率 71.6%(74.1%) といずれもクローズドテストデータの結果を大きく下回るものとなった．この原因として，学習データの不足，あいづち，話者交替の話者依存性が影響ではないかと考えられる．

しかし，あいづちには絶対的な正解箇所は存在せず，完全なあいづち箇所の判定は困難であるのであいづちの被験者による違いを調べる必要がある．そこで，コーパスからあいづちのみを除いた対話音声被験者 4 人に聴取してもらい，あいづちをどこで打つべきかを指摘させる実験を行った．その結果を表 4 に示す．被験者のあいづち指摘箇所とコーパスのあいづち箇所では大きく違いがあることが容易に見て取れる．被験者毎に適切と考えるあいづち挿入箇所に違いがあり，いずれも誤りではない．従ってあいづちを被験者が実際に聞くことにより主観的な評価を行うことが重要であると考えられる．

### 4.2 被験者による主観的な評価

決定木によって生成されたあいづちと話者交替タイミングの自然性を調べるため，被験者による主観的な評価を行った．

評価のためのサンプル音声は，テストコーパス中の発話に対して決定木が生成したあいづちタイミングに，同じ話者が別の箇所で発声したあいづち音声を挿入して作成した．話者交替に関しては同様に決定木が話者交替と判定した箇所に対して，同じ話者の“えーっと”という音声をつけた音声を作成し実験サンプルとした．

サンプル数はあいづち 16，話者交替 17 である．比較対象として，同じコーパス中で実際の対話の中で発せられたあいづちや話者交替の発話を，同様の別の音声で置換することで人間のタイミングのみを再現したサンプルも用意した．あいづちと話者交替はそれぞれ 16 サンプル，18 サンプルである．このサンプル音声を被験者 5 人が聴取し，1:早すぎる，2:早い，3:よい，4:遅い，5:遅すぎる，6:論外という 6

表 3: ポーズ内のフレーム毎の分類結果

(a) 分類結果 [総数]

task	in \ out	class (closed / open set)			
		あいづち	話者交替	発話待ち 1	発話待ち 2
雑談	あいづち	10 / 3	2 / 0	5 / 0	3 / 0
	話者交替	0 / 2	7 / 8	2 / 9	0 / 3
	発話待ち 1	0 / 8	0 / 0	21 / 16	1 / 13
	発話待ち 2	2 / 5	0 / 0	6 / 11	17 / 14
旅行案内	あいづち	56 / 2	13 / 0	5 / 3	17 / 0
	話者交替	2 / 0	63 / 43	2 / 14	9 / 4
	発話待ち 1	1 / 4	3 / 4	206 / 82	17 / 32
	発話待ち 2	22 / 0	11 / 41	1 / 58	111 / 40
電話通販	あいづち	15 / 10	9 / 2	5 / 10	1 / 9
	話者交替	0 / 6	70 / 35	10 / 48	24 / 38
	発話待ち 1	0 / 20	1 / 14	61 / 32	0 / 18
	発話待ち 2	1 / 5	14 / 2	12 / 71	127 / 82

(b) 分類精度 (すべてのタスクの平均)

尺度	あいづち	話者交替	発話待ち 1	発話待ち 2
REC [%]	53.8 / 57.4	76.0 / 44.8	94.9 / 49.5	75.7 / 42.2
			92.8 / 81.7	
PRE [%]	82.1 / 24.8	74.1 / 71.6	75.8 / 38.8	78.9 / 51.7
			85.4 / 79.3	
F	65.0 / 34.6	75.0 / 55.1	84.3 / 43.5	77.2 / 46.5
			88.9 / 80.4	

REC:Recall PRE:Precision F:F-measure

つの評価基準で評価を行った。その結果を、表 5, 6 の 1, 2 行目 (Rep) に示す。

これらの結果でシステムと人間との間に大きな違いは見られなかったことから、あいづちと話者交替タイミング生成に韻律情報と表層的な言語情報を用いることは有用であると考えられる。またこの中で、人間同士の対話におけるあいづちと話者交替タイミングでありながら不自然と判定されるケースも見られた。

次に、音声を置換しない実際のコーパスで出現したあいづちと話者交替音声それぞれ 19, 16 サンプルを被験者が聴取し、その自然性を評価する実験も行った。評価結果を表 5 と表 6 の 3 行目 (Real) に示す。

あいづちの場合は、実際の人間のあいづち音声と同等以上の評価が得られており、あいづちのタイミングの自動生成は良好に動作していると言える。一方、話者交替の場合は、別の音声に置き換えられた場合とは同等であったが、実際の対話には大きく及ばない。これは、タイミングの良し悪しの評価であるものの、前発話との内容的つながりの影響も無視できないためであると考えられる。

## 5 音声対話システム

現状の音声対話システムは、発話してもシステム側からの応答がなく、発話途中でシステムに入力さ

表 4: 被験者とコーパスのあいづちタイミングの一致度合 [%]

被験者	タスク	recall	precision
被験者 1	雑談	50.0	47.6
	旅行案内	36.8	35.0
	電話通販	37.5	8.6
被験者 2	雑談	20.0	57.1
	旅行案内	5.3	14.3
	電話通販	0.0	0.0
被験者 3	雑談	5.0	4.8
	旅行案内	61.4	33.3
	電話通販	25.0	3.8
被験者 4	雑談	20.0	50.0
	旅行案内	29.8	38.6
	電話通販	12.5	6.7
平均	雑談	23.8%	39.9%
	旅行案内	33.3%	30.3%
	電話通販	18.8%	4.8%

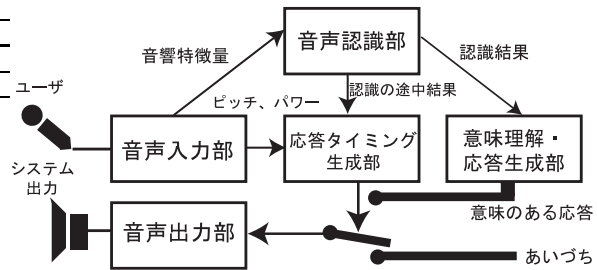


図 4: システムの構成

れているかどうか分からないといった問題がある。前述のリアルタイム応答生成手法を対話システムの応答タイミング生成部に適用し、雑談のような対話の中でリアルタイムにあいづちと話者交替タイミングを検出して応答できるシステムの構築を行った。これにより、ユーザの発話に適度にあいづちを打ち、システムの発話タイミングを自動で検出してときにはユーザの発話に対して割り込むことが可能なシステムの構築を目指す。

### 5.1 想定タスク

本システムは、システムと人間とがより円滑にしかも雑談のように会話できることを目的とするため、システムの想定タスクを天気案内タスクとした。天気案内は、対話の中でユーザが天気を尋ねるだけでなく、その地域は暑いかどうかなどの雑談に発展することが容易で、しかも特定の語彙に限定しやすいタスクであると考えられる。

表 5: 被験者の主観によるあいづちタイミングの評価。

	1 早い	2	3 よい	4	5 遅い	6 論外	自然率 [%]
コーパス (Rep)	0	12	66	9	0	3	73.3
システム (Rep)	0	15	64	0	0	1	80.0
コーパス (Real)	1	13	69	6	0	6	72.6

”Rep”はあいづち音声は置換されている, ”Real”は実際の対話音声を使用したことを示す。

表 6: 被験者の主観による話者交替タイミングの評価。

	1 早い	2	3 よい	4	5 遅い	6 論外	自然率 [%]
コーパス (Rep)	0	15	54	7	0	4	67.5
システム (Rep)	3	21	55	1	0	5	64.7
コーパス (Real)	0	6	72	2	0	0	90.0

”Rep”は話者交替音声は置換されている, ”Real”は実際の対話音声を使用したことを示す。

## 5.2 システムの構成

構築したシステムを図 4 に示す。ユーザから入力された音声は音声入力部によって特徴パラメータ抽出, ピッチとパワーの抽出が行われ, 特徴量は音声認識部, ピッチとパワーは応答タイミング生成部へと送られる。音声認識部は本研究室で開発された SPOJUS[17] で 300 語程度の単語辞書を用いて, リアルタイムに認識の途中結果を出力することが可能である。

得られた認識の途中結果とピッチとパワーを用いて応答タイミング生成部で 100ms 毎にあいづち, 話者交替, 発話待ちの判定を行う。応答タイミング生成部によりあいづちと判定された場合はあいづちを, 話者交替と判定された場合は, 応答生成部がユーザの発話から天気案内タスクに関連する単語を抽出し, それらの単語を元に ELIZA 式 [18] に応答文を作成し応答を行う。

また, システムは, ユーザからの発話がない状態が 6 秒続いた場合にはユーザに発話を促すプロンプトを発声する, システムが連続して 4 回発話継続であると判定した場合にはあいづちを挿入するというルールを設けることにより, ユーザとの会話が途切れないようにしている。

## 5.3 システムとの対話例

実際のシステムと人間との対話例を図 5 に示す。「愛知県の気温はどうか?」という発話後のポーズをシステムは話者交替であると判定し、「どこの天気が知りたいんですか?」という返答を行っている。また, ユーザの「愛知県」という発話に対しては, その後のポーズを応答タイミング生成部があいづちと判定し, 実際に生成している。引き続き, ユーザは発話を継続している。

## 5.4 被験者との対話による評価実験

あいづちと話者交替タイミングやその内容を評価するため, 本システムと被験者が実際に対話を行い, 被験者が評価する実験を行った。

実験は各被験者に対し, システムが実際に応答タイミングで応答音声を送出するもの, 応答内容は本システムが生成したものであるがオペレータの示すタイミングで応答音声を送出するもの(すなわち, 図 4 の応答タイミング生成のみを人間が行う)の 2 通りでタイミング生成し, さらにそれぞれあいづちがある場合とない場合を合わせ, 総計 4 通りの組み合わせで行う。

各対話の後, 被験者にはアンケートに答えてもらい評価する。被験者には返答内容よりも, あいづちや返答のタイミングの評価に重点をおくよう説明を行ったほか, あいづちがない場合とある場合の違いやあいづちが対話に与える影響を特に注意して評価してもらった。

また, タイミングの評価については, ”悪い”(期待したタイミングに返答がない, もしくは会話をさえぎる箇所が多い), ”普通”(多少無理矢理割り込まれる部分はあるものの大体はよいタイミングである), ”よい”(タイミングはすべてよい)の 3 段階で評価を行うよう指示した。

ここで, 被験者は男性 4 人でうち 3 名は音声対話システムとの対話が未経験である。被験者は天気の話に対して自由に発話をする事ができ, 特に制限を設けていない。被験者はシステムから天気情報を聞き出すまで行うか, もし天気情報を引き出せなかった場合は 10 発話程度の発話後に対話を打ち切っている。被験者に流す音声は女性の合成音声 [19] を用いており, あいづちは「はい」のみである。また, それぞれの対話開始前に, 被験者にはシステムのあいづちのあり/なしを伝えてある。

被験者によるシステムの評価結果を表 7(a) (あい



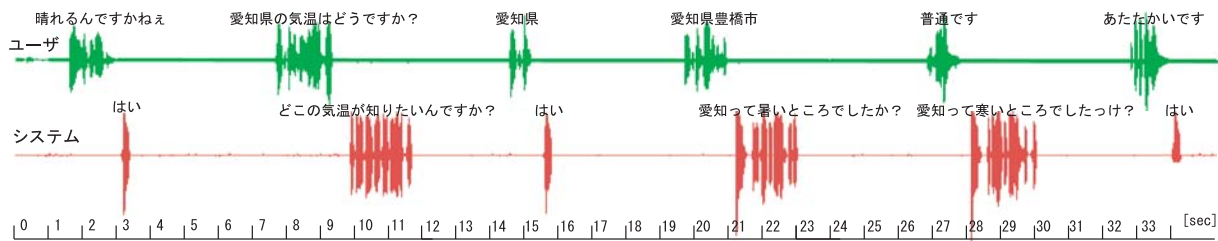


図 5: システムとの対話例

表 7: 被験者の評価

(a) タイミング生成：システム/人間（あいづちなし）

評価項目	評価結果（タイミング）				
返答タイミング	悪い 3 / 0	普通 1 / 2	よい 0 / 2		
評価結果（対話内容）					
親しみやすさ	悪い 0 / 0	少し悪い 0 / 0	どちらとも 0 / 0	ややよい 3 / 3	よい 1 / 1
返答回数	少ない 1 / 0	少し少ない 1 / 0	適量 2 / 4	少し多い 0 / 0	多すぎる 0 / 0
返答内容	悪い 1 / 0	少し悪い 3 / 4	どちらとも 0 / 0	ややよい 0 / 0	よい 0 / 0

(b) タイミング生成：システム/人間（あいづちあり）

評価項目	評価結果（タイミング）				
返答タイミング	悪い 3 / 1	普通 1 / 2	よい 0 / 1		
評価結果（対話内容）					
親しみやすさ	悪い 0 / 0	少し悪い 1 / 0	どちらとも 1 / 0	ややよい 1 / 3	よい 1 / 1
返答回数	少ない 0 / 0	少し少ない 0 / 0	適量 4 / 4	少し多い 0 / 0	多すぎる 0 / 0
返答内容	悪い 0 / 1	少し悪い 4 / 1	どちらとも 0 / 1	ややよい 0 / 0	よい 0 / 1
評価結果（あいづち）					
どちらががよいか	あいづちあり 2 / 4	あいづちなし 1 / 0	わからない 1 / 0		
あいづちのタイミング	悪い 1 / 1	普通 1 / 3	よい 2 / 0		
あいづちの頻度	少ない 0 / 0	少し少ない 0 / 0	適量 2 / 4	少し多い 0 / 0	多すぎる 2 / 0

づちがない場合）および (b)（あいづちがある場合）に示す。

表 7(a) の中で、返答タイミングと返答内容は悪いとしながらも、システムの親しみやすさとしてはシステムと人間の両方でよい評価を得ることができた。表 7(a) と (b) に共通してシステムの返答タイミングが悪いと判定された理由としては、発話終端で話者交替とされるべきポーズが発話継続と判定され、発話終了から 6 秒後に適用されるルールによりシステムが応答を行うことがしばしば見られたためと考えられる。

表 7(b) であいづちを導入した後では、被験者の半数が親しみやすいと回答した一方で、一部親みにくいという評価があるが、これはあいづちタイミングの誤検出が原因として考えられる。被験者の発話をさえぎるような悪いあいづちは、対話の親しみやすさに大きく悪影響を及ぼすことを示した結果であるといえる。

あいづちに関して、半数以上の被験者がタイミングはよいと評価した。また、被験者の半数からあいづちがあることによって発話が入力されていることが確認できる、対話全体が自然と感じるという回答が得られた。しかし、あいづちが「はい」だけでは物足りない、質問しているのに「はい」というあいづちで対話が終了してしまうことがあるとの意見も

得られており、これは本システムの改善項目の一つである。

本システムのあいづち以外の応答に関しては、言ったことに対して答えてくれない、同じことを繰り返されるという回答が得られた。この点に関しては対話のバリエーションを増やすなどの対応を行う必要がある。

## 6 まとめ

本研究では、リアルタイムにあいづち、話者交替タイミングを検出できる音声対話システムの構築を行った。

システムで最も重要な役割を担うタイミングの対話制御に韻律情報、言語情報を素性に用いた決定木を用いた。この決定木は韻律情報は発話句音声末のおよそ 100ms の変動、言語情報は、発話終端単語の品詞や発話の最後に現れた自立語の品詞情報などを素性に使用した。

この決定木を人間同士の対話でのあいづちや話者交替があった箇所と比較し、再現率と適合率を用いて評価したところ、オープンテストで、あいづちの分類において再現率 57.4%、適合率 24.8%、話者交替の分類において再現率 44.8%、適合率 71.6% の結

果を得た。また、この分類木を対話システムに導入し、被験者が主観的に評価した結果、システムの返答内容に関しては改良が必要であるものの、あいづちタイミング自体はよいという評価が得られた。

## 参考文献

- [1] メイナード K. 泉子, 会話分析, くろしお出版 (1993) .
- [2] 水谷 信子, “あいづちとポーズの心理学” (2001) .
- [3] 田窪 行則, “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”, 情報処理学会研究会報告, SLP-1-3, pp15-21 (1994).
- [4] N. Ward, “Prosodic features which cue back-channel responses in English and Japanese”, *Journal of Pragmatics* 32, pp.1177-1207 (2000).
- [5] 平沢 純一, 川端 豪, “音声対話システム Noddy—ユーザ発話途中でのうなずき・相槌生成—”, 情報処理学会研究会報告, SLP-20-4, pp.51-52 (1998) .
- [6] 岡登 洋平, 加藤 佳司, 山本 幹雄, 板橋 秀一, “韻律情報を用いた相槌の挿入”, 情報処理学会論文誌, Vol.40, No.3, pp.469-477 (1999).
- [7] R. J. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann (1992).
- [8] 佐藤 康将, 井上貴雄, 目良和也, 相沢輝昭, “自然言語対話システムのための多様なあいづち生成手法の改良”, 言語処理学会 第 8 年次大会発表論文集, pp.248-251 (2002) .
- [9] 野口 広彰, 片桐 恭弘, 伝 康晴, “心理実験を用いたあいづち応答の手がかり特徴の検証”, 人工知能学会研究会資料, SIG-SLUD-A002-13 (2000).
- [10] 竹内真士, 北岡教英, 中川聖一, “韻律情報を用いた相槌生成システムとその評価”, 情報処理学会第 64 回全国大会, Vol.2, pp.101-102 (2002).
- [11] H. Sacks, E. A. Schegloff, G. Jefferson, “A simplest systematics for the organization of turn-taking for conversation”, *Language*, 50, pp.696-735 (1974).
- [12] C. Kamm., S. Narayanan., D. Dutton., R. Ritenour, “Evaluating spoken dialogue systems for telecommunication services”, *Eurospeech-97*, Rhodes, Greece, pp.2203-2206 (1997).
- [13] 人工知能学会 談話・対話研究におけるコーパス利用研究グループ, “様々な応用研究に向けた談話タグ付き音声対話コーパス”, 人工知能学会研究会資料, SIG-SLUD-9903-4 (1999).
- [14] 小磯 花絵, 堀内 靖雄, 土屋 俊, 市川 薫, “先行発話断片の終端部分に存在する次発話者に関する言語的・韻律的要素について”, 電子情報通信学会技術報告, NLC95-72, pp.25-30 (1996).
- [15] R. Geluykens, M. Swerts, “Prosodic cues to discourse boundaries in experimental dialogues”, *Speech Communication*, 15, pp.69-77 (1994).
- [16] J. Hirschberg, “Communication and prosody: functional aspects of prosody”, *Speech Communication*, 36, pp.31-43 (2002).
- [17] 甲斐 充彦, 中川 聖一, “日本語連続音声認識システム SPOJUS-SYNO の改良と評価”, 電子情報通信学会技術報告, SP93-20 (1993).
- [18] J. Weizenbaum, “ELIZA - A computer program for the study of natural language communication between man and machine”, *communications of the ACM*, vol.9, no.1, pp.36-45 (1965).
- [19] 東 芝, 日本語音声認識 / 合成ソフト LaLaVoice2001, <http://www2.toshiba.co.jp/pc/lalavoice/>.