応答タイミングを考慮した雑談音声対話システム

A spoken dialog system for chat-like conversations considering response timing

西村良太　北岡教英　中川聖一

Ryota NISHIMURA　Norihide KITAOKA　Seiichi NAKAGAWA

豊橋技術科学大学情報工学系

Department of Information and Computer Sciences, Toyohashi University of Technology

{nishimura, kitaoka, nakagawa }@slp.ics.tut.ac.jp

Abstract

If a dialog system can respond to a user as natural as a human, the interaction will be smoother. In this research, we aim to develop a dialog system which can make various behaviors appearing in chat-like dialogs. In this paper, we developed a dialog system which could generate chat-like responses and their timing using a decision tree. The system could make "synchronous utterances", "*aizuchi*" and so on. The decision tree utiliezed the pitch and the power contours in the last 100 ms of user's utterance, recognition hyposethes, and the prepared response contents at that time as features. This system also reacts user's overlapping utterances and disfluencies robustly.

1　はじめに

近年，音声認識技術を用いたインターフェースが発展してきている．具体的な用途としては，情報検索 [1] や観光案内 [2] など，様々な対話システムが検討されている．しかし，これらのシステムにおいては，ユーザがシステムに話しかけた際に，途中にシステムからの反応が全く無く，システムがきちんとユーザ発話を聞いているのか分からないといった問題があり，これが音声認識を利用した音声対話システムに壁を感じる一因となっている．これからは，音声対話システムがより身近なものになり，生活の中に入り込んでくることが予想される．その際には，より自然な対話を実現する必要がある．これには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し，より円滑な対話を行うことが重要になってくる．

人間同士の会話においては，話者は互いにうなづきやあいづちによって相手の発話を理解していることを明示しており，それにより会話がスムーズに進行する．これを音声対話システムにも応用し，ユーザの発話に対して，システムがあいづちなどの反応を返す事が出来ればユーザは人間と対話している場合と同じように自然に対話が行えるのではないかと考えられる．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ている [3]．また，話者交替も，対話の自然性を考える上では重要である．相手の発話が終わったのか，まだ続くのかをしっかりと把握できなければ，円滑に対話を進めることは困難になってくる．

あいづちや話者交替を行う場合には，そのタイミングも重要であり，相手の発話に応じて適切なタイミングで応答を返し，時にはそれらをオーバーラップさせることによって，スムーズに会話が進行していく．

このことから，ユーザの発話をリアルタイムで分析し，後に続く現象を予測することが出来れば，適切な応答を適切なタイミングで返すことが出来るようになると考えられる．

これらのことをふまえ，本研究では，音声対話システムにおいてあいづちや，システムからユーザへの割り込み発話など，種々の現象を考慮しそれらを適切なタイ

ミングで行う天気予報を話題とする雑談システムを構築した．目的指向の対話システムではなく，雑談を通して，ユーザの知りたい情報を提供するシステムにすることで，ユーザが対話自体を楽しみながらシステムを使用する事ができる．雑談には，様々な現象があるが，今回構築したシステムは，様々な雑談現象を扱うことができるように設計されている．本稿では「同調発話」と「あいづち」を実装した．ここで言う “同調発話” とは，ユーザとシステムが同じ内容の発話を同じタイミングで発話したり，ユーザの発話の途中からシステムが補完をして発話することを指している．システム応答出力のタイミングの決定は，対話相手（ユーザ）の発話の種々の特徴やシステムの理解状態，対話履歴などを考慮いたルールによって決定する．以降ではシステムの設計や実装した天気情報案内における対話例を示す．

2　雑談に関する関連研究

あいづちについては，これまでにも研究がなされている．小磯ら [13] は，発話句音声末 1 モーラ分のピッチ，パワーの変動のパターンが話者交替やあいづちに強く関わっていると分析している．分析では，ピッチの変動が平坦型，平坦下降型，上昇下降型であり，またパワーの変動が平坦型，平坦下降型である場合にあいづちが多く打たれている．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]．話者交替については，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 [10]．またこれらにおいては，言語情報の方が効果的であるという結果を報告している研究が多いが，大須賀らは，韻律情報が話者交替の予測に有効にはたらき，発話末まで待たなくても予測できるとしている [11]．

これまでにも，リアルタイム応答として，ユーザ発話にあいづちを打つという研究は行われてきている．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を検討している [6]．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から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 [7]．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]．Ward らは低ピッチ区間が一定時間続く箇所にあいづちを生成するシステムを構築している [5]．

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（バージイン）ことができるシステムは存在するが [6][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少ない．また，これまでのシステムのほとんどは，音声認識結果とポーズの検出によってシステムが応答を返すようなものになっている．この方式では，ポーズの検出の為に少なからず遅延が生じる．またユーザの発話にシステムが割り込むことが出来ないため，発話が重複するような対話の実現ができなかった．しかし，実際の対話においては，あいづちや話者交替が重複して発生することは少なくない．あいづちについては，全体の 3 割ほどが重複して発生しており，さらに雑談対話においては，半数以上が重複して発生している [4]．このことから，円滑に対話を行う上で，重複発話を無視することは出来ない．

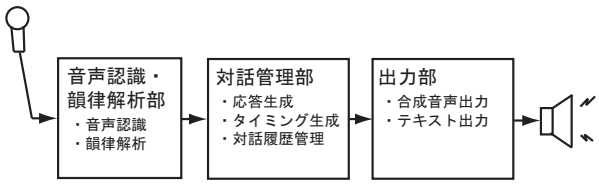

図 1: システムの構成図

今回構築したシステムでは，逐次的に音声認識と韻律解析を行うことにより，ポーズの検出による遅延の問題を解消している．また，ポーズを検出する必要ないので，ユーザ発話への割り込みが可能になっている．これにより，より実際に近い対話を実現することが可能である．

3　対話システムの実現

1 節，2 節のことをふまえ，実際に対話システムを構築した．まず，今回構築した音声対話システムの概略を説明する．システムの構成は，図 1 に示すように，大きく分けて 3 つの部分からなっている．音声認識・韻律解析部は，入力された音声を解析し，音声認識とピッチ・パワーの計算をする．管理部は，認識結果と韻律情報から応答文と応答タイミングを生成する．出力部は，応答文を音声にて出力する．この流れに沿って，リアルタイムに処理が行われている．各部の通信には，TCP/IP 通信を使用しており，それぞれを別のマシン上で動作させることも可能である．以下に，それぞれについての詳細を示す．

3.1　音声認識・韻律解析部

ここでは，ユーザがマイクにより入力した音声を音声認識器にて認識している．それと同時に韻律解析器

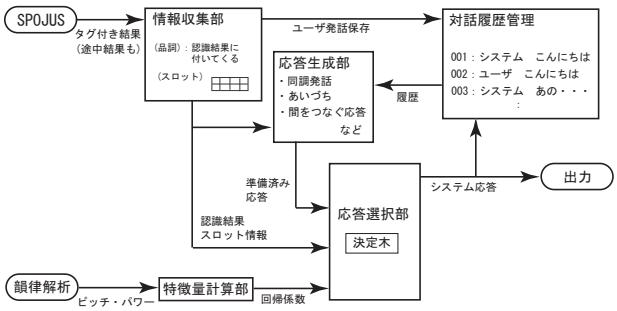

図 2: 対話管理部の構成図

[14][15] にて解析し，ピッチ (F0)・パワーを解析する．音声認識には本研究室で開発された SPOJUS[16][17] を用い，300 語程度の単語辞書を用いて認識を行っている，認識の途中結果をリアルタイムに出力することも可能であり，今回のシステムにおいても，認識途中結果を用いている．また，単語辞書に予め品詞をつけておくことで，認識途中や最終結果に形態素解析をかけなくとも品詞情報が得られる．ピッチ・パワーは，フレームシフトを 5ms，フレーム幅を 64ms として計算をしており，音声認識・韻律解析のどちらも，用いている音声のサンプリング周波数は，16kHz である．

3.2 対話管理部

対話管理部は，さらに図 2 のように 5 つの部分に分かれている．音声認識・韻律解析部より送られてくる認識結果と韻律解析データを，管理部内の各部分で順に処理していき，応答文を生成する．応答文を適切なタイミングで出力するために，決定木によってどの応答をどのタイミングで出力するかを決定する．韻律情報であるピッチ・パワーについては，回帰係数によって傾きを求め，その値を決定木の素性として用いる．

音声認識器の出力である認識途中結果や認識結果は，情報収集部に送られ，そこで必要な情報を収集し，それを情報スロットに格納しておく．また，収集された情報は応答生成部にも送られ，そこで応答文を生成する．応答文は複数用意され，生成された応答文の中から適切なものを決定木により選択し出力する．応答文の生成には，認識結果と情報スロットを元に ELIZA[18] 方式にて生成する方法を採用している．応答文のテンプレートは，各雑談現象ごとに用意されており，認識結果と情報スロットを見て，それぞれに対して応答文が生成される．ユーザ発話とシステム応答は，対話履歴管理部にて保存される．認識結果，情報スロットの値，応答文の準備状況を応答選択部の決定木に入力することで，応答タイミングと応答に用いる応答文を決定する．決定木により選択された応答文は出力部へ送られる．

タスクとしては，天気情報案内を想定しているため，天気情報を対話に取り入れる必要がある．その為の天気情報は，Web よりダウンロードする．

以下に，それぞれの部分を詳しく記載する．

3.2.1 応答生成

応答生成には，ELIZA 方式を用いている．この方法は，ユーザ発話の中のキーワードから，システムがそれに応じた応答をテンプレートを用いて生成するものである．

応答生成は，各雑談現象ごとに行われ，それぞれに対して応答文が用意される．現時点では雑談現象としてあいづちと同調発話を扱っている為，あいづちで応答する場合と，同調発話で応答する場合とで，別の応答を用意する．応答文は各雑談現象について，1 つずつ保持され，どの文を用いて応答するかは，応答選択部にて決定される．応答文は，ユーザ発話の認識結果を受けて，それに応じて生成している．しかし，認識途中や認識後に，どの雑談現象においても応答が準備出来ていないという場合がある．この場合は，雑談現象には関係なく返す為の応答を用意しておき，それを用いることとする．例えば,「心配」というキーワードが発話されていれば「心配事があるんですか？」と応答する．キーワードが入っていなければ「よく聞こえませんでした.」などと応答するようになっている．これによって，発話者が交代するタイミングに全く何も応答がないということが避けられる．あいづちについては，認識結果によらずどのような場合にも対応ができる “はい” や “うん” を準備している．

今回のシステムの場合には，雑談でのやり取りのなかに，天気の情報を織り交ぜて，対話を進めていくという戦略をとっている．ユーザとの対話で，今の話題が天気についてであると考えられる場合には，地名などが入力されると，それを記憶しておき，実際の天気予報のデータを対話に織り交ぜて生成し出力する．

応答生成に用いる具体的なデータ形式は，図 3 のようになっている．これは “@行” にキーワードもしくは正規表現を含んだ文を用意し，直下の “＝行” には，“@行” に対応する応答文のセットを用意する．このまとまりをルールセットとする．ルールにない発話が現れた場合には,「何が知りたいのですか？」「意味がわかりません。」といった応答を返すこととした．図 3 で，“＝行” は，スロット名とスロット値とのマッチング用正規表現，および応答文が記述されている．2 行目では，tenki スロットの値が空ではなく，今日の天気スロット

@ 今日 (.*) 天気 (.*)(だ|です)
= tenki:.+,tenki_1day:.+,$0 は、$1 ですよ。
= tenki:.+,city:(), 場所はどこですか？
@ 明日 (.*) どう
= tenki:.+,city:.+,tenki_2day:.+,$1 の$0 は、$2 のようですよ。
= tenki:.+,city:(), 場所はどこですか？:どこの$0 が知りたいんですか？

図 3: 応答テンプレート

(tenki_1day) も空ではない時に，応答文を返す．応答文の中にある “$0” は，左側に記述されたスロットの値を参照するためのもので，左側から 0 番，1 番 ···，と名前がついている．下から 4 行目の “=行” では，tenki スロット，都市名スロット (city)，明日の天気スロット (tenki_2day) が用いられている．認識結果が “@行” とマッチし，さらにスロットの値が “=行” で示されたようにマッチした場合に，その行の応答文を返すようになっている．

ここで，ユーザが「明日はどう？」と発話した場合に応答を返すためには，天気スロットが埋まっている必要がある．天気スロットは，ユーザ発話から “天気”，“空模様” というキーワードを格納するようになっている．それに加えて，ユーザがこれまでに地名も発話していた場合には，その土地の天気が応答文として出力される．例えば，天気スロットに “空模様”，都市名スロットに “豊橋” が入っており，また豊橋の明日の天気が “晴れ” であった場合には，システム応答は “豊橋の空模様は，晴れのようですよ。” となる．

最後の “=行” では，2 種類の応答文がコロンで区切られて記述されている．複数の応答を用意する場合にはこのように記述しておくことで，この中からランダムに応答を用意することができる．

3.2.2 応答選択部・決定木

応答選択部では，決定木により，応答文候補の中から応答に用いる応答文と，応答のタイミングを決定する．決定木の素性には，言語情報と韻律情報を用いる [4]．言語情報としては音声認識結果を用いる．韻律情報としては発話開始からの時間，発話終了からの時間，直前 100ms 区間のピッチ・パワーの傾きといった情報を用いる．また，どの雑談現象を用いた応答が準備出来ているかの情報も用いる．100ms 毎に各素性を決定木へ入力すると，決定木が何らかの応答を返すタイミングであると判断すれば，どの雑談現象を用いて応答するかも決定し，出力されるので，応答生成部はそれに応じた応答を出力する．応答を返すべきでないタイミングにおいては，決定木からは，待ち状態であるとする結果が出力され，応答は出力されない．また，応答は 1 ユーザ発話につき 1 応答としている．ただしあいづちについてはその制限はなく，何度でも打つことが出来る．

なお，現状のシステムでは，学習用のデータが準備できていないため，先行研究などによる分析結果を元に人手で作成した決定木を用いている．

3.2.3 情報スロット

情報スロットは，ユーザ発話に含まれる重要なキーワードを格納するために用意されている．今回は，天気情報タスクであるので，天気に関するキーワードを格納するようになっている．具体的には，天気の話題であることを確認するために，ユーザが天気に関することを発話した場合には，そのキーワードを格納しておくためのスロットがあり，ユーザ発話中に “天気” や “空模様” といった言葉が含まれている場合には，その言葉をスロットに格納する．また，“今日，明日，明後日” のように，日時が示された場合にそれらを格納するためのスロットが用意されており，“県名，地域名，都市名” を格納しておくためのスロットや，都市名のスロットが埋まった際に，その都市の天気を格納しておくためのスロットなどが用意されている．応答文を生成する際には，3.2.1 節で示したように，これらのスロットの中身を対話に織り交ぜることが可能である．

3.2.4 天気情報

天気情報は，国際気象海洋株式会社天気案内 WebPage[19] より自動でダウンロードしている．このページには，全国主要都市の今日・明日・明後日の 3 日分の天気情報と，今日・明日の降水確率，気温が記載されている．このページから，全国の天気情報を取得しプログラム内に読み込んでおく．ユーザが地名を示した場合には，その土地の 3 日分の天気情報を天気スロットに格納しておくことで，それ以降のシステム応答に天気情報を盛り込むことが可能になる．

3.3 出力部

出力部は，システムが生成した応答を，音声とテキストにて出力する．音声合成には，擬人化音声対話エージェントのツールキット Galatea Toolkit[20] に含まれる音声合成器の GalateaTalk を用いている．この音声合成器は，発話者タイプ（男女など）の変更や抑揚，話速を自由に変更が出来るため，今後，音声対話において，抑揚や話速を変化させるような応用も可能である．

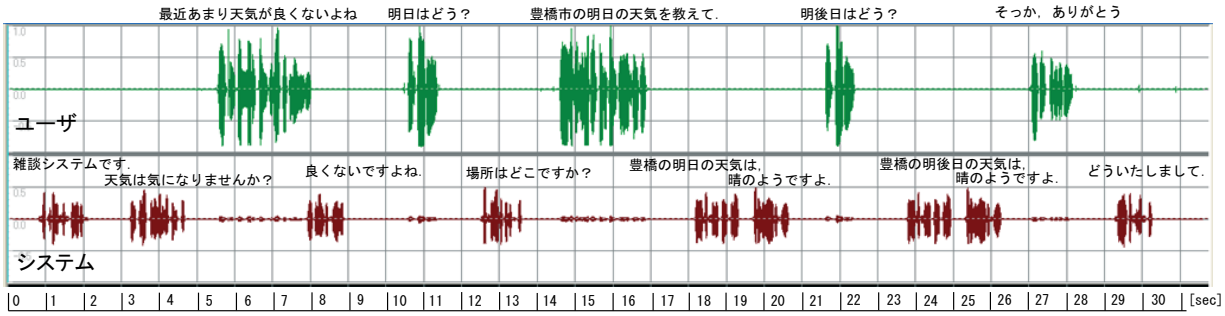

図 4: 対話例 1

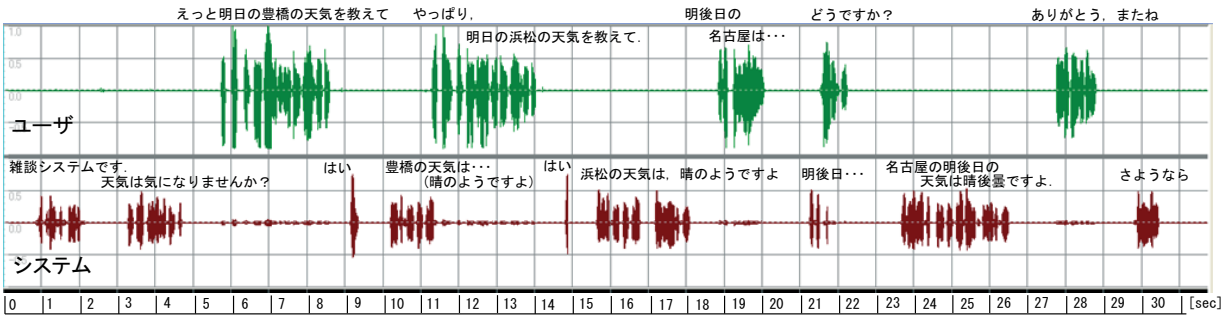

図 5: 対話例 2

4　システムの動作例

ここでは，実際のシステムとユーザの対話例を図 4, 5 に示す．各図においては，上がユーザ発話であり，下がシステム応答である．システム応答の波形に，ユーザ発話区間と対応した振幅の小さい波形が見られるが，これは収録環境によるノイズであり，合成音ではない．

現段階のシステムでは，1 ユーザ発話に対しては，1 システム応答としている．ただし，あいづちに関しては，その制限は外している．対話の進行を，システムの動作と共に解説していく．

どちらも最初は，システムの「天気は気になりませんか？」という呼びかけの発話から始まっている．まず図 4 を見る．システムの呼びかけに対してのユーザ発話「最近あんまり天気が良くないよね」に対しては，“最近” と “あんまり” というキーワードと今現在天気についての話題であるということから，“良くない” という**同調発話**を，**オーバーラップ**して応答している．次のユーザ発話「明日はどう」に対しては，“明日” と “どう” というキーワードと現在天気についての対話であるということから，システムは天気を応答しようとするが，場所が確定できていないため，天気予報スロットが空の状態になっている．そこで,「場所はどこですか？」と応答することにより，ユーザから場所を聞き出そうとしている．次のユーザ発話にて，“豊橋” ということが分かったので，豊橋の明日の天気を応答として返している．次のユーザ発話「明後日はどう？」に対しては，場所も確定しており天気情報がスロットに入っている状態であるので，明後日の天気を応答している．最後に，ユーザ発話「そうかありがとう」に対しては,「どういたしまして」というシステム応答を返している．この対話で，システム応答の「良くないですよね」が同調発話であり，その他の応答は，特に焦点を絞った現象ではなく，“その他の一般応答” として扱っている．これには，天気情報応答用のテンプレートや，その他の雑談対話用テンプレートが含まれている．

次に，図 5 を見る．最初のユーザ発話「えっと明日の豊橋の天気を教えて」では，最初の部分に「えっと」という**フィラー**が入っている．そして，この発話に対して，システムは「はい」という**あいづち**を返している．そのあと，システムは，ユーザ発話に対して応答をしようとしているが，ユーザが途中でシステム発話に対して割り込んだ為，システムは応答を停止し，ユーザに発話権をうつした．ユーザの**割り込み**発話である「やっぱり明日の浜松の天気を教えて」に対して，システムはあいづちを打ち，そして，天気の情報も返すことが出来ている．次のユーザ発話は,「明後日の名古屋はどうですか？」という発話であるが，ユーザが**言い淀み**をした為に，システムは 2 つの文として認識をした．そして，

ユーザ発話の前半部分である「明後日の名古屋は...」に対して，システムは「明後日は，何かあるんですか？」という応答を返そうとするが，ユーザの**発話が継続**したため，システムは途中で**応答を停止**している．ユーザ発話後半の「どうですか？」に対しては，これまでの対話により,「天気」「名古屋」「明後日」という情報がスロットに入っているために，“明後日の名古屋の天気はどうですか？” という意味であると理解し，正しい応答を返すことが出来ている．このように，ユーザの割り込みや，言い淀みがあっても，頑健に応答することが出来ている．

5 まとめと今後

本研究では，リアルタイムにあいづち，話者交替などの応答タイミングを検出し種々の雑談現象を扱い応答することが出来る雑談に向けた対話システムの構築を行った．タイミングの検出と応答の種類の決定には決定木を用いており，その決定木の素性としては，言語情報と韻律情報を用いている．ポーズを検出せずに逐次的に処理をして応答を返すことから，オーバーラップした応答なども返すことが可能であり，実際の雑談に現れる様々な現象を実現できる．

現在は，決定木を人手で作成しているが，今後は人間による対話実例を用いてそこから学習をさせた決定木を作成し，それを用いて評価実験を行っていく予定である．

謝辞

本研究にて韻律解析器として使用しているプログラムには，産業技術総合研究所の後藤真孝先生が提案されたアルゴリズムを用いて，早稲田大学の藤江真也先生が実装されたものを利用させて頂いております．ここでお礼を申し上げます．

参考文献

[1] 西崎博光，中川聖一.: “音声文書を対象とした音声入力型情報検索システムに関する研究”, 豊橋技術科学大学大学院 博士論文，(2003.2).

[2] 小暮悟，中川聖一.: “音声対話システムにおける頑健な意味理解と対話システムの移植性に関する研究”, 豊橋技術科学大学大学院 博士論文，(2002.2).

[3] 泉子・K・メイナード .: 会話分析, くろしお出版，(1993).

[4] Kitaoka, N., Takeuchi, M., Nishimura R., Nakagawa S.: “Response Timing Detection Using Prosodic and Linguistic Information for Human-friendly Spoken Dialog Systems”, 人工知能学会論文誌, Vol. 20, No. 3, pp.220-228 (2005) .

[5] Nigel Ward.: “Prosodic features which cue back-channe responses in English and Japanese”, *Journal of Pragmatics 32, pp.1177-1207*

[6] 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).

[7] 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).

[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] Sacks,H.，Schegloff,E.A.，& Jefferson,G.: “A simplest systematics for the organization of turn-taking for conversation”，Language，50，pp.696-735(1974).

[11] 大須賀智子, 堀内靖雄, 西田昌史, 市川熹.: “音声対話での話者交替/継続の予測における韻律情報の有効性”. 人工知能学会誌 Vol. 21 No. 1, pp.1-8, (2006).

[12] C.Kamm.，S.Narayanan.，D.Dutton.，and R.Ritenour.: “Evaluating spoken dialogue systems for telecommunication services”，Eurospeech-97，Rhodes，Greece，pp.2203-2206(1997)

[13] 小磯花絵，堀内靖雄，土屋俊，市川熹.: “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).

[14] 後藤真孝, 伊藤克亘, 速水悟.: “自然発話中の有声休止箇所のリアルタイム検出システム”, 電子情報通信学会論文誌 D-II, Vol.J83-D-II, No.11, pp.2330–2340, (2000).

[15] 藤江真也, 福島健太, 柴田大輔, 小林哲則.: “FST と韻律情報を用いた相槌・復唱機能を有する対話ロボット”, 人工知能学会研究会資料, SIG-SLUD-A401-03, Jun., (2004).

[16] 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 (SP-93-20)，(1993).

[17] 豊橋技術科学大学情報工学系中川研究室.: “日本語連続音声認識システム SPOJUS-SYNO”, http://www.slp.ics.tut.ac.jp/SPOJUS/.

[18] Weizenbaum.,J.: “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] 国際気象海洋株式会社.: “IMOC Weather Page”, *http://www.imocwx.com/*.

[20] 嵯峨山茂樹，川本真一，下平博，新田恒雄，西本卓也，中村哲，伊藤克亘，森島繁生，四倉達夫，甲斐充彦，李晃伸，山下洋一，小林隆夫，徳田恵一，広瀬啓吉，峯松信明，山田篤，伝康晴，宇津呂武仁.: “擬人化音声対話エージェントツールキット Galatea“, 情報処理学会研究報告 (2002-SLP-45-10), pp.57-64, Feb. (2003).