Vol.2009-SLP-77 No.22
2009/7/18

応答タイミングを考慮した音声対話システムとその評価

西 村 良 太[†1]　中 川 聖 一[†1]

人間と機械が対話を行う際に，機械が人間同士の会話と同じように，自然な応答を返すことが出来れば，より円滑な対話を行うことが期待できる．本研究では，人間同士の雑談対話中にて生じる対話現象を模倣する音声対話システムを構築した．本システムでは，応答として，あいづち，復唱，共同補完などを扱っており，決定木を用いて応答種類と応答タイミングを決定している．また，本システムはユーザからのオーバーラップ応答（バージイン）や，非流暢な発話に対しても頑健に応答することが可能になっている．被験者実験の結果から，オーバーラップを含む通常応答やあいづちに対して高い自然性が示され，被験者の多くがあいづちに対して親しみを感じており，また，バージインは便利であると評価された．

Spoken Dialog System Considering The Response Timing and The Evaluation

Ryota Nishimura[†1] and Seiichi Nakagawa[†1]

If a dialog system can respond to a user as naturally as a human, the interaction will appear smoother. In this research, we aim to develop a dialog system that emulates human behavior in a chat-like dialog. The proposed system makes use of a decision tree to generate chat-like responses at the appropriate times. These responses include "*aizuchi*" (back-channel), "repetition", "collaborative completion", etc. The system also reacts robustly to the user's overlapping utterances (barge-in) and disfluencies. The subjective evaluation shows that there is a high degree of naturalness in the timing of ordinary responses, overlap, and *aizuchi*, and that the dialog system exhibits user-friendly behavior. Many subjects felt familiarity with *aizuchi*, and the barge-in was also useful.

†1 豊橋技術科学大学　情報工学系
Department of Information and Computer Sciences, Toyohashi University of Technology

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の需要が高まっており，音声対話システムも開発されている．しかし，現在の一般的なシステムでは，ユーザの入力に対して，途中でシステムからの反応が全くなく，システムがユーザ発話を聞いているのかが分からないという問題があり，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．これまでに幾つかの音声対話システムが開発されているものの，対話は堅苦しくユーザに対して不自然な印象を与えている．この問題を解決し，より自然な対話を実現することが重要である．

人間同士の会話においては，あいづちや話者交替が自然になされており，これらの生起にはピッチ（F0）やパワーなどの韻律情報が大きく関係している．これまでに，あいづちや話者交替をリアルタイムに生成するシステムがいくつか開発されており[1),2)]，韻律情報（F0やポーズ長など）を用いて応答の予測・制御を行っている．しかし，これらの先行研究では，個々の応答現象しか扱っていない．

本研究の目的は，あいづち，復唱，共同補完などの多様な応答現象を扱い，応答タイミングを考慮して自然な応答を生成し，ユーザと円滑に対話を行う音声対話システムを開発することである．応答種類選択と応答タイミングの決定には，人間同士の対話コーパスをもちいて学習した決定木を用いており，素性には韻律情報と表層的な言語情報を用いている．このタイミング生成手法を用いることで，ヒューマンフレンドリな音声対話システムを実現する．提案システムは，ユーザ発話に対してのオーバーラップ応答も可能であり，また，ユーザからのオーバーラップ発話（バージイン）にも対応している．

本稿では，最初に対話システムでの応答タイミング生成の実装について述べ，その後で，構築した対話システムの被験者評価の結果を示す．

2. 応答タイミングのモデル化

実際の人間同士の対話コーパスを用いて，応答種類選択と応答タイミングの決定をモデル化し，それをシステムに実装する．本システムでは，人間による対話を模倣するために，あいづち・復唱・共同補完と，その他の一般的な応答を扱う．あいづちは，対話相手に対して対話を聞いていることを示したり,「続けて」という意思を示したりする場合に用いられる．復唱は，対話中に内容を確認するために用いられ，共同補完は，対話中にお互いに情報を補完しあって一体感をもって対話を進めていくものである．

Vol.2009-SLP-77 No.22
2009/7/18

2.1 応答タイミング生成の為の素性

小磯ら[3] や大須賀ら[4] によると，ピッチ・パワーの概形パターンは，応答生成に関連している．例えば，発話の最終モーラのピッチ・パワーの概形が特定の型である場合には，対話相手によるあいづちや話者交替が誘発される．このように，ピッチ・パワーの概形により，対話相手の応答選択に影響がある．このことから，今回提案するシステムでの応答タイミング生成には，発話の最後 100ms 部分のピッチ・パワーの一次線形回帰係数を用いる．対象となる 100ms の区間を，3 つの区間に分け（窓幅は 55ms，フレームシフトは 25ms），その区間での傾きを計算し，素性として用いる．また，発話の最後の 500ms の区間に関しても，同様にピッチ・パワーの傾きを計算し，素性として用いており，この場合には，窓幅 100ms，フレームシフト 100ms の 5 つの区間に分割している．これらの素性については，計算コストが低く，リアルタイムで計算することが可能である．

復唱と共同補完については，ユーザから対話中のトピックに関するキーワードが検出された場合に生成される為，認識結果（ユーザ発話中に得られる認識途中結果も含む）の単語の属性情報も素性として用いている．

これらのことから，応答タイミングの決定に関して，以下の素性を用いている[5]．

- ユーザ発話開始時点からの経過時間 (F1)
- ユーザ発話終了時点からの経過時間 (F2)
- システム発話終了時点からの経過時間 (F3)
- ユーザ発話の最後の 100ms 区間のピッチ・パワーの傾き (F4)
- ユーザ発話の最後の 500ms 区間のピッチ・パワーの傾き (F5)
- 認識結果（途中結果も含む）の最終単語の属性 (F6)

図 1 に，各素性の関係を図示する．

2.2 決定木による応答タイミング生成

以前に我々は決定木を用いた応答タイミング生成器を提案したが[6]，このタイミング生成器は，ユーザ発話終了後に応答を返すことを想定しており，ポーズを検出した後で応答を返すシステムであった．このことから，あいづちなどを含む全ての応答に対して，ポーズ検出後にしか応答することが出来ないという制約があった．今回構築したシステムでは，ユーザの発話中・ポーズ中に関わらず，全てのセグメント (100ms 毎) に対して，応答するかどうかの判定を行うことで，ユーザ発話にオーバーラップする応答を返すことが出来る．今回用いた決定木の一部を図 2 に示す．図に示すように，決定木の最初の判定ルールは，ユーザの発話長になっており，ユーザの発話長が，応答種類決定に大きく関係していることが分かる．また，各素性によるルール判定を行っていき，最終的な判定は，一番下の楕円で囲まれたものになる．

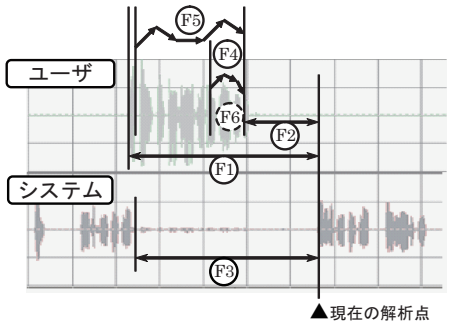

図 1 決定木で用いる素性

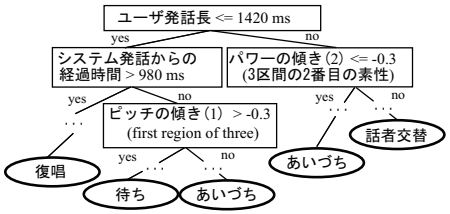

図 2 決定木の一部

応答タイミング生成器は，決定木にて 2.1 節に示した素性を用いて応答タイミングを生成する．また同時に，応答生成器にて生成された応答の中から適切な応答を選択する．決定木では，応答生成器にて応答が準備できているかどうかも素性として用いる．各応答種類毎に一つの応答が準備される．各素性は，100ms 毎に決定木に入力され，応答すべきかどうかの判定と，応答する場合には適切な応答種類の判定を行う．選択される応答の種類には,「あいづち・復唱・共同補完・一般的な応答・待ち」がある.「待ち」の場合には，応答を出力しない．応答の回数は，1 回のユーザ発話に対して 1 回のシステム応答に制限されているが，あいづちと復唱に関してはこの制限はない．つまり，1 回のユーザ発話に対して，共同補完と一般的な応答は 1 回応答することができ，あいづち・復唱は何度も応答することができる．

決定木の学習には，RWC コーパス[7] を用いており，このコーパスには，対話ドメインとして「自動車販売」「海外旅行計画」の 2 種類の対話が合計 48 対話含収録されている．データ量としては，コーパス全体で 6.5 時間分あり，各対話の時間は 10 分程度である．また，発話数は，16,399 発話である．一方の話者は，実際の自動車販売員，または旅行代理店員であり，もう一方の話者は，12 人の一般人である．このコーパスを用いて決定木の学習を行った．決定木の学習器には，C4.5[8] を用いた．

「あいづち，一般的な応答，待ち」に対する学習は，RWC コーパスを用いて行ったが，「復唱・共同補完」については，RWC コーパス中に十分なデータ量がなく学習することが

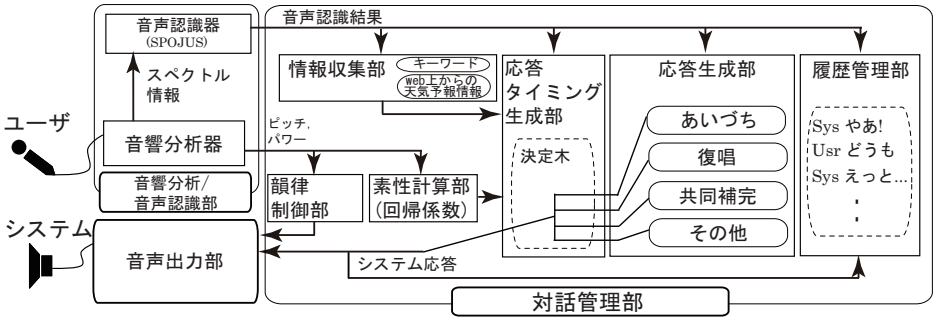

図 3 対話システムの概略図

困難であった為，人手によりルールを作成した．復唱に関しては，対話ドメインに関連するキーワードが検出された時点で応答を返すようにし，共同補完については，応答が可能になる入力がユーザからあった時点で応答を返すようにした．さらに，ユーザ入力がしばらくなかった場合への対処として，6 秒以上のポーズが検出された場合には，システムからユーザへ入力を促す発話を行うルールを追加した．

3. 対話システム

今回提案した，上述した現象を扱うことができるシステムの概略図を図 3 に示す．対話ドメインには天気情報に関する対話を選択した．この理由は，被験者がドメインに対して気楽に喋ることができ，WEB(http://www.imocwx.com/) から情報を収集することができ，また，ドメインや発話を自然に制限することができる為である．

3.1 音響分析・音声認識部

本システムで用いる音声認識器には，本研究室で開発された SPOJUS[9),10)] を用いる．SPOJUS には，2 つのバージョンがあり，一つは n-gram を用いた大語彙連続音声認識用のもの，もう一つは CFG (Context Free Grammar) を用いたものがあり，今回は，CFG 版の SPOJUS を用いている．

SPOJUS は，音響特徴量として 12 次元の MFCC (Mel-Frequency Cepstrum Coefficients) とその 1 次・2 次微分である ΔMFCC と ΔΔMFCC，Δ パワー，ΔΔ パワーを用いており，音声のサンプリング周波数は 16kHz，分析窓はハミング窓であり，フレーム長は 25ms，フレームシフトは 10ms である．音響モデルとしては HMM を用いており，4 状態，5 ループ，各状態には 4 混合のガウス混合モデルを用いており，全共分散行列を用いている．また，文脈依存型の音節 HMM モデルを用いており，モデル数は 928 モデルである．認識の途中結果もリアルタイムに出力させることが可能であり，システムはその結果を用いて復唱応答の準備などを行うことが可能となっている．

音声認識器に用いている語彙サイズは，約 300 単語であり，都市名，日付，天気の種類，フィラーなどが含まれており，全ての単語に対して，属性情報（単語クラス，品詞のようなもの）が付与されている．このようにすることで，認識結果に単語の属性情報が追加され，認識結果を形態素解析することなく，各単語の属性を知ることが可能になる．この情報を元にキーワード検出を行っている．例えば，音声認識結果には「浜松 [都市名] の [ ] 明日 [日付] の [ ] 天気 [天気関連] は [ ] どうですか [質問]」という風に情報が付与されている．本研究では，天気情報に関する内容を対話ドメインとしている為，キーワードとなる属性は，都市名・日付・天気情報などとなる．

音声認識と同時に，本システムでは，音響分析として韻律情報の抽出も行っており，ピッチ・パワー情報を抽出して応答タイミング生成部へ送信している．これは，2.1 節で述べた決定木の素性としての用途の他に，出力音声の韻律制御にも用いている．

3.2 対話管理部

対話管理部の詳細を，図 3 に示す．対話管理部は，6 つのサブコンポーネントで構成されており，音声認識結果と韻律情報を用いて応答を生成する．1 つのコンポーネントである応答タイミング生成部では，決定木を用いており，韻律情報から得られる素性に基づいて応答タイミングの決定を行っている．ユーザ発話中のピッチ・パワーの概形を素性として用いており，これらは，F0 と対数パワーの回帰係数として求められる．また，ピッチ・パワーはシステム応答の音声出力の際の韻律制御にも用いられている．

音声認識器 SPOJUS による音声認識結果は，情報収集部へと送られ，情報スロットに必要な情報を格納する．情報スロットに格納された情報は，応答生成部へと送られ，応答生成に用いられる．応答生成部では，音声認識結果と情報スロットを用い，ELIZA 型のテンプレートによる手法によって応答が生成される．そして，システムで扱う応答現象（あいづち，復唱，共同補完，一般的な応答）の為の応答を，同時に平行して用意し，用意された応答の中から適切な応答を，決定木によってリアルタイムに適切なタイミングで選択する．選択された応答は，音声出力部へと送られ，応答としてユーザに提示される．

Vol.2009-SLP-77 No.22
2009/7/18

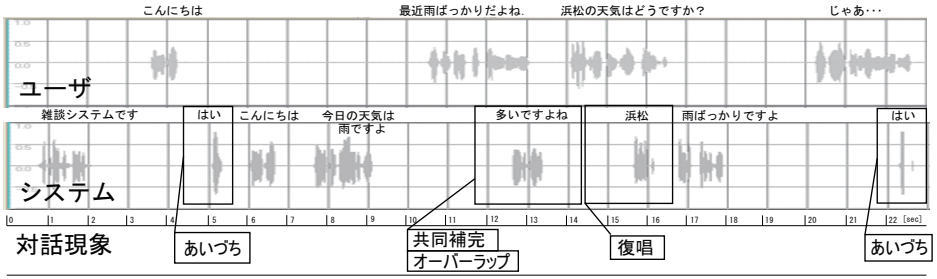

図 4 音声対話システムとの対話例

3.3 韻律制御部

韻律制御部では，システム応答出力に対して，韻律情報の制御を行う．制御対象としては，ピッチ・パワー・話速である．ここで，韻律制御は，システム応答の 1 発話の平均に対して行われるものであり，1 発話中のパターンを局所的に変更するものではない．韻律情報は，音声出力部に送られ，出力音声に反映される．本システムでの韻律制御に関する詳細は，文献[11] を参照されたい．

3.4 音声出力部

音声出力には，あらかじめ録音しておいた人間による録音音声を用いた．女性音声を用い，発話数は 3410 発話である．発話には，あいさつ，あいづち，天気情報などが含まれており，親しみのある楽しげなイントネーションにて発話を行っている．なお，合成音声を用いたシステムも開発しており，両者の比較も行っている．今回は，声質の違いによる比較は行わず，応答種類や応答タイミングに関する評価を，目標とする対話システム環境に近いもので実験を行うために，録音音声を用いた．

4. 評価実験と結果

構築した音声対話システムとユーザとの対話例を図 4 に示す．あいづち，復唱，共同補完，一般的な応答が含まれており，それらがオーバーラップもしながら応答されている．対話内容を見ていくと，最初はシステムからのプロンプト発話である「雑談システムです」から始まっている．ユーザの「こんにちは」発話に対しては，システムが「はい」とあいづちを返し，その後，一般的な応答として「こんにちは」と応答している．次に，システムから「今日の天気は雨ですよ.」という発話が生成されているが，これは，システム内に予めユーザ情報が入力されており，ユーザの居る地域の天気を案内している発話である．これらの情報は，情報スロットに格納されている．次に，ユーザからの発話「最近雨ばっかりだよね.」に対して，システムから共同補完として「多いですよね.」と応答を返している．そして，この応答は，ユーザ発話にオーバーラップをして出力されている．今回の共同補完に対しては，ユーザ入力中の「最近」「雨」がキーワードとなり，最近の天気情報から雨が多いことを確認し,「多いですよね」という応答を行っている．システムは共同補完に対するテンプレートも複数持っており，ユーザ入力がテンプレートにマッチした場合には，情報スロットの内容を元に，ユーザと同調する応答を生成する．次に，ユーザから「浜松の天気はどうですか？」と質問をされており，このユーザ入力からキーワードとして地名（浜松）を検出し，即座に地名の復唱を行っている．そして，実際の天気情報から，システムは「雨ばっかりですよ」という応答を生成している．このように，本システムが扱う対話には，あいづち・復唱・共同補完・一般的な応答が含まれており，これによって，人間同士が行うような対話を実現することが可能になっている．

この音声対話システムを用いた被験者実験による評価結果について，次節で述べる．また，韻律制御に関しては，評価していない．これは，韻律制御モデルによる制御が，韻律情報をゆっくりと変化させるものであり，ターン数の少ない短い対話に対しては変化が少ない為である．

4.1 対話システムの評価

4.1.1 準　　備

音声対話システムの評価は，被験者が実際に幾つかの音声対話システムと対話を行い，その後アンケート調査を行うことで行われた．用いた音声対話システムの設定を以下に示す．

system I: 基本システム (一般的な応答のみ，オーバーラップ無し)

system II: オーバーラップ応答を行うシステム（一般的な応答のみ）

system III: 全機能を扱うシステム（全種類応答，オーバーラップ有り）

被験者は男性 10 名であり，対話ドメインは,「天気情報に関する話題」である．対話に際しては，幾つかの都市の天気情報を聞きだすことを指示した．各被験者は，約 20 ターン（ユーザ発話が約 20 回）になるまで対話を行い，対話終了後にアンケートに答えた．質問は 5 段階評価によるものと，自由筆記のものを用意した．実験に際し，被験者には，システムの応答タイミングを重視して評価を行うように注意をした．

表 1 質問 “応答タイミングは自然だったか (5 段階評価)” に対するアンケート結果．“ポジティブ数” は，評価として 4 または 5 を付けた被験者数を表し，“ネガティブ数” は評価として 1 または 2 を付けた被験者数を表す．

応答現象	ポジティブ数	ネガティブ数	どちらでもない
一般的な応答 (オーバーラップ無し)	**8**	1	1
オーバーラップ応答	**4**	3	3
あいづち	**6**	4	0
復唱	2	6	2

表 2 質問 “その応答現象があると話しやすいと感じたか (5 段階評価)” に対する回答．

応答現象	ポジティブ数	ネガティブ数	どちらでもない
オーバーラップ応答	3	3	4
あいづち	**7**	2	1
復唱	3	4	3

4.1.2 評価結果

アンケート調査の結果として，“応答タイミングは自然だったか (5 段階評価)” に対する結果を表 1 に示す．表中の “ポジティブ数” は，評価として 4 または 5 を付けた被験者数を表し，“ネガティブ数” は評価として 1 または 2 を付けた被験者数を表す．結果から，一般的な応答に対して評価が高かった被験者数は 10 名中 8 名と多くなっており，一般的な応答は自然に応答できている．オーバーラップ応答とあいづちに関しても，評価が高くなっており，これらについても，自然に応答ができている．しかし，復唱に関しては，評価が低い被験者の方が多くなっている．

アンケート結果として，“その応答現象があると話しやすいと感じたか (5 段階評価)” に対する結果を表 2 に示す．結果から，被験者はあいづちがある対話では，話しやすいと感じている．

オーバーラップ応答に関しては，アンケートの結果から，10 名中 3 名の被験者が，オーバーラップ応答があると話しやすいと答えている．この場合の被験者の意見としては,「オーバーラップする応答があると，より自然な対話になる」などが挙がっていた．逆に，10 名中 3 名の被験者が，オーバーラップ応答があると話しにくいと答えている．この場合の被験者の意見としては,「ユーザの発話中にシステムからの発話があると不愉快である」などが挙がっていた．ネガティブ意見であった被験者 3 名のうち 2 名は，システムからのオーバーラップ頻度が非常に高く，0.769 と 0.421 であった（他の 8 名のオーバーラップ頻度の平均は，0.15 程度である）．この 2 名の被験者については，オーバーラップ頻度が高すぎたため，評価が悪くなったと考えられる．システムの改良点として，オーバーラップ頻度についても対話中に考慮して応答する機構が必要である，

あいづちに関しては，アンケートの結果から，10 名中 7 名の被験者が，あいづちがあると話しやすいと答えている．この場合の被験者の意見としては,「あいづち応答があると，親しみを感じる」,「ユーザの発話中に，システムが聞いているかどうかを確認することができる」などが挙がっていた．逆に，10 名中 2 名の被験者が，あいづちがあると話しにくいと答えている．この場合の被験者の意見としては,「あいづちのタイミングが悪く，話しにくい」などが挙がっていた．

復唱に関しては，アンケートの結果から，10 名中 3 名の被験者が，復唱があると話しやすいと答えている．この場合の被験者の意見としては,「ユーザの入力が終了する前に，システムがちゃんと認識できているか確認ができて良い」,「認識誤りの際に，早期に発見できて良い」などが挙がっていた．逆に，10 名中 4 名の被験者は，復唱があると話しにくいと答えている．この場合の被験者の意見としては,「ユーザの発話中に復唱されると話しにくくなる」,「認識誤りがあった際に，誤った単語を復唱されると不愉快になる」などが挙がっていた．復唱に対してポジティブな被験者とネガティブな被験者との間で，全く逆の意見になっている．被験者によって，復唱に対する好みがあり，意見が 2 つに分かれたことから，システムにて復唱を扱う場合には，ユーザの好みに合わせて応答制御を行う必要がある．

バージイン機能 (ユーザからシステムへのオーバーラップ発話) については，実際に使用した被験者は 10 名中 7 名であり，その全員がポジティブな評価であった．被験者からの意見としては,「認識誤りが起こった際に，正しい入力をすぐに発話でき，誤りを訂正できる」などが挙がっていた．

各被験者の音声認識率の差による違いも現れており，音声認識率が高い被験者は，オーバーラップ応答を自然であると感じており，逆に認識率が低い被験者は，オーバーラップ応答を自然だとは感じていない．また,「話しやすさ」についても違いが現れており，あいづちに関する質問に対しては，全ての質問事項に対する結果が異なる傾向を示していた．音声認識率が高い被験者は，あいづちが入ることで，システムと話しやすくなると感じており，認識率が低い被験者は，そのようには感じていない．復唱に関しては，認識率の違いによる，アンケート結果の差は現れなかった．

音声認識率が高い被験者は，システムからユーザが望む応答を返すことができる確率も高くなっており，応答タイミングや話しやすさに対する評価が高くなっていた．つまり，システムの音声認識率がよく，理解度もよく，応答内容も自然な対話ができるような，期待通り

表 3 応答現象と被験者の印象との間に高い相関が示されたもの

対話中の各種統計量	印象評価	相関
認識率	オーバーラップ応答の評価（ポジティブ）	0.765
オーバーラップ頻度	話しやすさ	-0.564
あいづち頻度	話しやすさ	0.643

のシステム性能になれば，今回提案したシステムを用いることで，被験者からオーバーラップ，あいづちに対して良い評価を得ることができると言える．

4.1.3 評価結果の分析

対話中の各種統計量（音声認識率，正しい応答率，オーバーラップ応答頻度など）とアンケート結果との間の相関関係を調査した．その中で高い相関関係が示されたものを表 3 に示す．

音声認識率 (accuracy) とアンケート結果の間の関係を調査すると，オーバーラップ応答の評価との間に有意な相関関係が示された（相関係数 0.765，$p<0.01$[⋆1]）．つまり，音声認識率が高い場合には，オーバーラップ応答が好まれていることが示された．

オーバーラップ頻度とアンケート結果の関係を調査すると，オーバーラップ応答がある場合の話しやすさとの間に，高い負の相関関係が示された（相関係数-0.564，$p=0.090$）．相関係数が負の値であることから，システムとの対話においては，オーバーラップ頻度が低いほうが好まれていることが示された，

あいづち頻度とアンケート結果の間の関係を調査すると，あいづち頻度と，あいづちがある場合の話しやすさとの間に，有意な相関関係が示された（相関係数 0.643，$p<0.05$）．システムとの対話においては，あいづちの頻度が高いほうが好まれていることが示された．

5. ま と め

本稿では，雑談対話によってフレンドリな対話を行い，リアルタイムに応答生成と応答タイミング生成を行う音声対話システムを提案し，実装，被験者評価を行った．本システムには，実際の人間同士の対話で扱われる様々な応答現象（あいづち，復唱，共同補完や，それらのオーバーラップ応答，またはユーザからのバージイン発話）が実装されている．

被験者実験では，被験者の多くが，あいづちに対して親しみを感じており，また，バージイン機能が便利であると感じていた．応答タイミングの自然性の評価は，一般的な応答・オーバーラップでの応答・あいづちに対して高い評価を得ていた．復唱に関しては，被験者の意見が 2 つに分かれており，システムが復唱を用いるべきかどうかは，ユーザの好みに合わせて制御すべきである．オーバーラップ応答に関しては，オーバーラップ応答の頻度が高くなると，被験者評価の結果が悪くなっていたものの，音声認識率が高い対話においては，オーバーラップ応答が好まれることが示された．

今後は，決定木の学習に用いるコーパスを，人間同士の対話コーパスではなく，システムを用いて収集する．また，複数の対話エージェントと人間による対話システムを開発し，多数話者対話の効果を調査・分析する．

⋆1 有意に相関が無いとする帰無仮説は，危険率 1%で棄却される．

参 考 文 献

1) Ward, N. and Tsukahara, W.: Prosodic features which cue back-channel responses in English and Japanese, *Journal of Pragmatics, 32*, pp.1177–1207 (2000).
2) 佐藤康将：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).
3) 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).
4) 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).
5) Nishimura, R., Kitaoka, N. and Nakagawa, S.: A Spoken Dialog System for Chat-Like Conversations Considering Response Timing, *TSD 2007*, pp.599–606 (2007).
6) Kitaoka, N., Takeuchi, M., R., N. and S., N.: 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).
7) 田中和世，速水 悟，山下洋一，鹿野清宏，板橋秀一，岡 隆一：RWC 計画における音声対話データベースの構築，情報処理学会音声言語情報処理１１－７ (1996).
8) J.Quinlan, R.: C4.5:Programs for machine learning, *Morgan Kaufmann* (1992).
9) 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).
10) Zhang, J., Wang, L. and Nakagawa, S.: LVCSR based on context dependent syllable acoustic models, *Asian Workshop on Speech Science and Technology, SP2007-200*, pp.81–86 (2007).
11) Nishimura, R., Kitaoka, N. and Nakagawa, S.: Analysis of Relationship Between Impression of Human-to-human Conversations and Prosodic Change and Its Modeling, *Proceeding of the Interspeech 2008*, pp.534–537 (2008).