社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－61 (8)
2006／5／11

対話に関するリズムや同調作用を考慮した音声対話システム

東海林 圭輔，髙橋 美佳，井原 誠也，伊藤 敏彦，荒木 健治

北海道大学 情報科学研究科
〒060-0814 札幌市北区北14条西9丁目

人間同士の対話における最適な対話のリズムは，その対話の中で生まれる．対話の中でそのリズムを意識することにより，より円滑で満足度の高い対話を行うことが期待できる．本稿では，人間と人間のコミュニケーションをできるだけ模倣する方向性の一つとして，対話のリズムを重視した音声対話システムを開発し，ユーザに自然な発話を促すことでユーザ満足度の向上を目指す．今回，我々が着目した対話のリズムの要素は，システムからの発話や相槌のタイミング，発話速度である．これらの自然なリズムを実現するために，ポーズ単位で言語理解を行いながら，発話途中でも相手のタスク意図を予測する言語理解部，ユーザモデルを用いて協調的でリズムを考慮した応答を生成する応答生成部，相槌判定を含む話者交代判定とリズムの同調をリアルタイムに行う対話リズム生成部の3つのモジュールを新たに作成し，タスク指向型の音声対話システムを構築した．

Spoken Dialog System considered Rhythm and Synchronized Tendency of Conversation

Keisuke SHOJI, Mika TAKAHASHI, Seiya IBARA,Toshihiko ITOH and Kenji ARAKI

Department of Information Science and Technology
Hokkaidou University, Hokkaidou, 060-0814, Japan

The best rhythm of the conversation between humans is developed during their conversation. It can be expected that users conscious of rhythm will perform smoother conversation. In this paper, as one of the methods is to copy human communication abilities as much as possible, we develop spoken dialog system which puts the importance to the rhythm of dialog aiming at improvement of user satisfaction by encouraging an user to utter naturally. Elements of rhythm of dialog that we paid our attention is speaking rate and timing of an utterance and backchanneling from a system. To realize such natural rhythm, we newly designed three modules - Understanding Component to predict user's task intention in the middle of his utterance while performing language understanding by a pause unit; Response Generator which generates the response considering rhythm and uses a user model; Rhythm Generator to perform a speaker change judgment including backchanneling judgment and rythm synchronization in real time. These components are to construct a task oriented spoken dialog system.

1 はじめに

「音声対話」は，我々人間のコミュニケーションにおいて最もプリミティブで効率的な手段であると言われている．実際，人間は対面対話においては，音声や表情，ジェスチャーなどの様々なモダリティを活用しながら，円滑にコミュニケーションを行っている．さらに，電話などの非対面対話であっても，音声のみで伝えることができる言語的・非言語的情報を活用し円滑なコミュニケーションを実現している．

しかしながら，円滑で効率的な音声対話を実現するために必要な要素や情報が何であるか，それらの要素や情報がどのように人間同士の音声対話に作用・影響しているかなどは，まだ完全に明らかにされていない．そのため，近年，人間同士の音声対話において円滑なコミュニケーションに活用されていると考えられる様々な要素に関する分析の研究が増えてきている [1][2].

人間と機械との対話である音声対話システムに関しても，数多くの研究が行われてきたが，これまでの研究の中心は相手の発話を正しく理解することや，効率的にタスクを遂行することなどの効率重視の研究が多く，円滑で自然なコミュニケーションに着目した研究はあまり行われていない．ここ最近の研究では，上述の人間同士の対話分析の知見をシステムに実装し，相槌 [3] やジェスチャー [4] などにより，ユーザに対してシステム側の意思のようなものを伝えたり，ユーザとシステムが互いに割り込むことが可能なシステムも構築され，それらの評価実験でユーザ満足度の上昇が見られたと報告されている．

さらに，我々が行った様々な状況下での人間対人間，人間対機械とのタスク指向対話の言語的・音響的特徴を分析した結果から，

- 機械的な発話を行うような対話相手に対して，人間は自然な発話 (対話) ができない
- 対話相手が人間か機械かという事実自体はそれほど重要ではない

- システムの対話リズム，音声の韻律（感情を含む）などの応答能力を人間に近づけることが自然な発話（対話）を促すために重要

などが分かってきている．

そこで現在，音声インターフェースに関して，我々は二つの方向性でシステム開発を行っている．一つは，人間同士のコミュニケーションを出来るだけ模倣したシステムを構築し，人間が機械とのコミュニケーションであることをできるだけ意識しないで使用できるインターフェイスを開発すること．もう一つは，機械に対する発話やコミュニケーションが，人間同士のものとは少し異なることを上手く活用し，システムとしての精度を向上させたり，新たなインターフェイスを構築することを考えている．

本研究では，人間と人間のコミュニケーションをできるだけ模倣する方向性の一つとして，対話のリズムを重視した音声対話システムを開発し，ユーザに自然な発話を促すことでユーザ満足度の向上を目指す．今回，本研究で着目した対話のリズムの要素は，システムからの発話や相槌のタイミング，発話速度である．

これらの自然な対話リズム実現のために，ポーズ単位で言語理解を行い発話途中でも相手のタスク意図を予測する言語理解部，ユーザモデルを用いて協調的でリズムを考慮した応答を生成する応答生成部，相槌判定を含む話者交代判定とリズム同調をリアルタイムに行う対話リズム生成部の3つのモジュールを新たに構築した．

2 対話タスク

本研究では，対話のリズムに着目した音声対話システムを実現するため，音声認識，言語理解，応答生成，対話リズム生成をリアルタイムで処理することが求められる．そのため，比較的小さなタスクである必要がある．そこで，今回は本大学の情報科学研究科の建物案内タスクとした．

建物案内タスクは，表1に示されるような情報に関して音声対話を用いて検索できる．

また，システムを利用する対象者も外部からの来客や在学生，教官など，建物やシステムに対する知識などがそれぞれ異なる利用者が対象となり，それらを考慮した協調的な応答が必要となるタスクとなっている．

3 音声対話システムの概要

本研究における音声対話システムは，各々のモジュールで生成された情報をどのモジュールからも参照できる枠組みにする為，黒板モデルを採用した．黒板モデルでは，通信はモジュール間で直接行なわれずに黒板を介して行なわれるため，各モジュールが独立した形となっている．本システムではこの黒板を中心に，韻律解析部，音声認識部，信頼度生成部と意図理解部を含む言語理解部，応答生成部，話者交代判定部と同調処理部を含む対話リズム生成部，音声合成部の6つのモジュールで構成される．

タスク	ユーザの発話例
建物案内タスク	・A-21教室はどこですか？ ・自習できる場所はどこですか？
人物情報タスク	・○○先生に会いたいのですが ・○○先生の連絡先を知りたいのですが
授業案内タスク	・現在出ている休講情報を知りたいのですが ・自然言語処理は何時からですか？
会議案内タスク	・○○委員会議は今日でしたか？ ・○○さんの講演はいつ開催予定ですか？

表 1: ユーザ発話例

このシステムの構成を図1に示す．

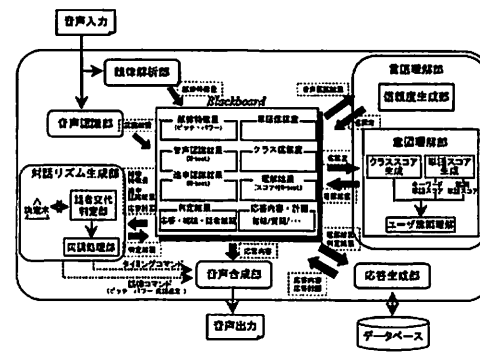

図 1: 音声対話システムの構成

ユーザから入力された音声は音声認識部と韻律解析部へ送られ，音声認識部は認識途中結果を，韻律解析部は韻律特徴量を逐次出力する．また，音声認識部は一定以上のポーズが検出されるたびに最終結果としてポーズ単位のn-best認識結果を出力する．

言語理解部は，信頼度生成部にてn-best認識結果を取得し，その認識結果より自立語や付属語の各単語の信頼度，クラスの信頼度を計算する．その後，得られた単語やクラスの信頼度と対話履歴を用いて，本タスクにおいて発話される可能性がある単語やタスク意図に関してユーザが実際に発話した可能性の度合いを単語スコアや意図スコアとして計算する．そして，それらのスコアにより順位付けられた単語・意図スコアつきのn-best言語理解結果を出力する．

応答生成部は，その言語理解結果を基にユーザの状態を推定し，ユーザモデルを構築する．その後，言語理解結果とユーザモデルを活用し，ユーザ満足度を考慮した協調的な応答を生成する．

対話リズム生成部は，ピッチやパワー等の韻律特徴量と認識途中結果を逐次取得し，話者交代や相槌の判定を行う．また，現在のユーザの発話速度，発話開始時間を計算し，それらを基に，ユーザに同調したシステムの発話速度，発話開始時間を決定し，応答内容をそのリズム

で出力する．

システムの対話例を図 2 に示す．

U1：伊藤先生の･･･
S1：はい．
U2：えー部屋の場所を知りたいのですが．
S2：言語メディア学研究室の伊藤助教授ですか．
U3：そうです．
S3：伊藤助教授の部屋は7-13ですが、現在授業中の為不在です．
U4：･･･えーっと･･･
S4：･･･
U5：えーいつごろ戻られますか．
S5：16時には戻られていると思います．
U6：分かりました．
S6：はい．またのご利用をお待ちしております．

図 2: 対話例

4 音声認識部・韻律解析部

ユーザが入力した音声は音声認識部で認識され，同時に韻律解析部でその音声のピッチ (F0) とパワーが解析される．音声認識部には，中川研究室で開発された音声認識器 SPOJUS[5][6] を用いる．認識に用いた単語語彙数は 519 語，文法数は 710 個である．音声分析条件は特徴ベクトルが MFCC（12 次元× 4 フレームを KL 展開で 24 次元に圧縮＋Δ MFCC ＋Δ Δ MFCC ＋Δ POW ＋Δ Δ POW の計 50 次元）で，16kHz のサンプリング周波数，10ms のフレーム周期，25ms フレーム長のハミング窓となっている．また，音響モデルは男性依存であり，116 音節の 5 状態 4 ループ（4 混合ガウス分布，全分散）連続出力分布型 HMM である．また，SPOJUS はリアルタイムに認識途中結果を出力することが可能である．

韻律解析部では，16kHz の周波数，6ms のフレームシフト，64ms のフレーム幅でピッチとパワーを計算する．

5 言語理解部

言語理解部では，リズム重視の対話を実現するために，発話内の小休止（ポーズ）区切りで出力される音声認識結果を用いて，発話全体が入力される前にある程度のタスク意図を推定しながら言語理解を行う手法を提案する．日本語ではタスク意図理解に重要な助詞が発話の最後に出現しやすいため，本手法では助詞の認識信頼度を利用する事で発話途中でのユーザのタスク意図を予測する．また，音声認識部の n-best 認識結果と対話履歴を用いることにより，音声認識誤りへの頑健性も考慮する．

5.1 処理の流れ

言語理解部は，最新の音声認識結果の単語やクラス，助詞の認識信頼度と対話履歴（認識履歴とシステム発話履歴），さらにタスク知識，対話理論などを活用し，本タスクにおいて発話される可能性のある全てのタスク意図に関して意図スコアを計算し，スコア順にランク付けされた n-best の言語理解結果を生成する．基本的な処理の流れは [7] と同様であるが，助詞を用いた意図推定を行う部分が新規に存在するなど若干処理過程が異なる．

言語理解部において用いられる名詞や疑問詞などのキーワード，助詞，クラスの認識信頼度は，各候補 w の尤度 $P(x|w)$ と認識結果（n-best）中の出現頻度から事後確率に基づく尺度 $P(w|x)$ として計算される [8]．

言語理解部ではまず，これまでの発話（一つのゴールを達成するまでの発話）の認識結果中に出現した全てのキーワードの単語について，どの程度発話された可能性があるか（以降，単語スコア）を求める．このスコアは，対話履歴中のキーワードの単語スコアと最新の認識結果中のキーワードの認識信頼度を用いて求められる．そして，単語レベルよりも上位概念のクラスについても，単語と同様に，対話においてどの程度発話された可能性があるか（以降，クラススコア）を求める．

次に，クラススコアと助詞スコアに基づき，本タスクにおいて発話される可能性がある各タスク意図ごとに意図スコアを求める．ここでのユーザのタスク意図とは，例えば，ある施設の開館時間を返答してほしい「開館時間返答」や，ある教官の部屋の所在を教えてほしい「教官の部屋返答」といった “大まかな” レベルでのタスク意図であり，キーワードを含んだ詳細なタスク意図のことではない．

最後に，各タスク意図ごとの意図スコアと各キーワードの単語スコアを組み合わせ，キーワードを含んだユーザのタスク意図を理解する．この理解結果は，例えば，「図書館 の閉館時間を知りたい」，「田中先生 の部屋の場所を知りたい」などである．

ポーズ単位の認識結果が言語理解部に入力される度にこれらの処理を行い，意図スコアでランク付けされた n-best の意図理解結果を生成する．

5.2 単語スコア・クラススコア

5.2.1 単語スコア・クラススコア

単語スコアやクラススコアは，単語やクラスの認識信頼度を基本スコアとして，様々な戦略によりそのスコアを上げ下げすることによって計算される．以下に単語スコアの算出における戦略を示す．

1. 古い情報は信頼性が低下する為，認識結果が得られる度に認識履歴中の全単語のスコアを下げる．
2. 認識履歴の単語 W_h と最新認識結果の単語 W_r が内容的に関係がある場合，単語 W_h のスコアを上げる．
3. 最新の認識結果に肯定語 (はい，うん等) が含まれていた場合，システム発話に含まれていた単語 W_h のスコアを上げる．
4. 最新の認識結果に否定語 (いいえ，違う等) が含まれていた場合，システム発話に含まれていた単語 W_h のスコアを下げる．
5. 認識結果の内容がシステムの質問 (ex. 専攻はどちらですか？) に対する回答の関係である場合，認識結果の単語 W_r を上げる．

6. 認識結果の上位に含まれている単語 W_r のスコアを認識結果の順位に応じて上げる．

クラススコアも上記と同様な戦略を用いて計算される．

5.2.2 助詞スコア

単語スコアは基本的に，対話でその単語が発話された可能性を表したものであるが，助詞の場合は，どのようなキーワードと助詞の組み合わせで発話されたかが重要となる．そのため，助詞スコアは認識履歴を反映させず，最新認識結果から求めた認識信頼度をそのまま用いる．

5.3 意図スコア

意図スコアは，今までの発話により，あるタスク意図がどの程度発話された可能性があるかを表す．この意図スコアは，"大まかな"タスク意図を推定するため，単語よりも上位の概念であるクラススコアと助詞スコアを用いて計算する．

このとき，タスク意図ごとに要求される情報のクラスやそのクラスの重要度は異なる．つまり，あるタスク意図に対しては付加的な情報であるが，別のタスク意図に対しては必ず必要な情報のクラスであったりする．

そのため，クラスに対する重要度を以下の3種類に分類し，重要度によりクラススコアを意図スコアへ加算する場合の重みを変化させた．

- あるタスク意図に対して必ず含まれる情報のクラス (必須クラス)
  例) 休講返答意図の場合の「学科・学年」クラス
- あるタスク意図に対して付加的な情報のクラス (オプションクラス)
  例) 休講返答意図の場合の「日時」クラス
- あるタスク意図を表現する単語のクラス (意図推定語クラス)
  例) 休講返答意図の場合の「休講」という単語

また，発話途中でも可能な限り早くタスク意図を推定するために，クラススコアを用いて求められた意図スコアを，さらに助詞スコアを用いて増減させる．最新の認識結果に「キーワード+助詞」の組が存在した場合，その組み合わせで表現される可能性があるタスク意図の意図スコアに関してはキーワードの単語スコアと助詞スコアに応じて増加させる．逆に可能性のないものについては，減少させる．ただし，ユーザの多様な言い回しや誤認識を考慮して，減少させる重みは現時点では小さな値にしている．

5.4 ユーザ意図理解

意図スコアによりランク付けされたタスク意図の各クラスには，複数の単語が代入候補として存在するので，タスク意図スコアと各キーワードの単語スコアから，最終的なタスク意図理解候補が生成される．例えば，「休講返答」のタスク意図とその必須クラスの単語「2年，4年」，さらにオプションクラスの単語「明日」の単語スコアから，

- 明日 の 2年生 の休講情報が知りたい
- 明日 の 4年生 の休講情報が知りたい

などのようにキーワードを含んだ複数の理解候補が生成される．これらの候補は意図スコア・単語スコアにより順序付けされ，n-best の言語理解結果として出力される．

6 応答生成部

応答生成部では，言語理解部が生成した n-best の言語理解結果からユーザの状態を推定し，ユーザモデルを構築する．また，言語理解結果の第1候補の意図スコアの値や第1候補と第2候補の意図スコアの差，各意図に含まれる単語スコアの値，競合する単語との単語スコアの差などのパターンからタスク遂行に関する状況を判断し，次発話の応答計画（情報提供，確認，質問など）を決定する．応答計画とユーザモデルにより応答テンプレートを選択し，データベースより検索された応答に必要な情報をテンプレートのスロットに代入することで応答文を生成する．なお，同一条件における応答テンプレートを複数用意し，ランダムに使うことにより自然性を考慮した．

6.1 応答に用いられるユーザモデルや知識

音声対話システムにおいて協調的な対話を実現するため，様々なユーザモデルが構築されてきた [9]．本研究では，ユーザモデルとしてシステムと建物に対する知識レベルと対話制御に関するユーザ嗜好を導入し，さらにクラス周知度を考慮することで協調的な対話を行う．

- 建物やシステムに関する知識レベル
  知識レベルはユーザに提供する情報の取捨選択に用いる．
- 対話制御に関するユーザ嗜好
  対話の主導方法はユーザごとに嗜好が異なるので，ユーザに合わせて変化させる．
- クラス周知度
  タスクに関連した様々な知識や情報は，ユーザに対する周知度が一定ではない．そこで，周知度の高い情報を用いて対話を行なうことで，ユーザ満足度を高める．

なお，現在のユーザモデルの判別は，学習データの不足から我々が手で作成したルールで行われる．

6.2 対話戦略と応答計画

本システムの対話戦略は，ユーザ満足度をできるだけ高くするために，誤認識・誤理解をユーザに悟らせず，タスクを遂行することを基本方針としている．

応答生成部が，現在想定している応答計画は，相槌，

時間確保（間投詞など），確認，質問，情報提供の５種類である．

相槌は対話リズム生成部で相槌と判定されたり，言語理解結果からタスク意図が全く判別できない（第１候補の意図スコアが閾値以下など）場合に次発話を促すために使用される．

時間確保は対話リズム生成部で話者交代と判定されたが，応答生成部の処理が終了していない場合，暫定的に間投詞や時間確保の発話が用意される．応答生成部の処理が終了次第，そちらの応答処理に移行する．

確認は意図スコア・単語スコアは確信できるほど高くはないが，他の関連情報の入力をこれ以上期待できない場合に使用される．確認方法は暗黙的確認，明示的確認の両方を想定し，状況によって使い分ける．

質問は必須な情報が入力されていない場合や，言語理解結果のタスク意図や単語において競合するタスク意図や単語が一定スコア内で存在し，他の関連情報を獲得することによりその競合が解決できそうな場合に使用される．質問項目に関しては競合解消度やクラス認知度，予測音声認識率を考慮して決定される．

情報提供は，意図スコア・単語スコアも一定以上であり，競合する意図や単語のスコア差も大きく，情報提供に必要な項目の情報が獲得できた場合に使用される．

建物やシステムに関する知識レベルは，情報提供時に提供する情報の内容や付加的な情報の有無に影響する．また，システムに関する知識レベルは対話制御における主導方法にも影響を与えるが，ユーザの主導方法に関する嗜好が優先される．

7 対話リズム生成部

対話リズム生成部は，リアルタイムに話者交代・相槌判定を行い，ユーザの発話開始時間と発話速度に対し同調することでシステムの対話のリズムをユーザに同調させる．

7.1 相槌判定を含む話者交代判定

相槌と話者交代の判定に関して様々な研究が行われており，その判別には発話終端１モーラの韻律情報が有効である [10] が，それを除いたそれ以前の韻律情報によっても判定できる [11] との報告もある．我々はこれらの知見を参考に，有効であると判断された素性を用いて，オーバーラップ的な応答を可能とするために，発話終了前に判定ができる枠組みで話者交代・相槌判定を行ないたいと考えている．

しかし，現段階では知見で有効と言われている一部の素性をリアルタイムに抽出する手段の実現が難しく，また，学習のための対話データも不足しているため，北岡らの研究 [3] で得られた決定木を参考に話者交代・相槌判定を行っている．

以下にこの決定木に用いる素性を示す．

- ユーザの発話長
- 発話終端単語の品詞
- 発話終端助詞の種類
- 発話終端からの時間長
- 発話中の最後の自立語の品詞
- その自立語の時間長
- その自立語から発話終端までの時間長
- 発話句音声末 100ms のピッチの変動
- 発話句音声末 100ms のパワーの変動

ピッチと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ている．

ここで，以上のように構築された決定木を用いた話者交代判定について述べる．まず，ユーザから音声が入力され始めた時点で，韻律特徴量と認識途中結果をリアルタイムに取得する．この認識途中結果は形態素解析により品詞情報やモーラ数などを含めた形に変換される．その後，これらの素性を基に決定木で話者交代・相槌判定を行う．但し，発話終端からの時間長など一部の素性は発話終了後にしか得られないので，その場合は発話終了前を意味する値に置き換えて判定を行う．この判定は対話の中で常に行い，判定結果を黒板に出力する．その判定結果が応答・相槌の場合は同調処理を行い，応答タイミングを生成する．

7.2 発話開始時間と発話速度の同調

対話のリズムの同調に関しても様々な研究が行なわれており，対話中に相手とコミュニケーション行動パターンが影響し合う現象は，同調傾向として知られている．

長岡ら [12][1] によれば，対話の場面において，相槌のパターンや発話の長さ，発話開始時間，発話速度などの韻律的特徴が対話者間で類似・連動するとし，これらの同調傾向は対話の早期から生じ，時間経過と共に程度を増し，特に対話が協調的な場合は類似化の傾向が強くなるとしている．この知見より，同調は会話の始まりから行い，直前の発話だけではなく過去３発話分の発話も考慮して処理を行う．

一方，同調は様々な形で表現されており，韻律的特徴が類似以外の変化をすることで，否定の意思や対話の局面の変化などを対話相手に感じさせ，対話がスムーズになる場合もある．しかし，これらの全てを実現するには分析結果が不足しているため，本研究では，まず発話速度と発話開始時間に着目し，システムのそれらがユーザに徐々に類似化することを同調と定義する．

また，人間が自分の発話開始時間や発話速度を基準として同調による変化を起こす様に，システム側も発話開始時間と発話速度の適切な基準が必要である．これについて，本研究室が収集した対話データ [2] を分析した結果によれば，発話開始時間と発話速度の大多数は一定範囲内に存在しており，また，２つの間で相関が見られなかった．よって，それぞれ平均値を基準値とし，大多数が分布

していた範囲を同調範囲として，システムの発話開始時間の基準値を 0.4[sec]，同調範囲を 0.0〜1.0[sec] に，発話速度の基準値を 8[mora/sec]，同調範囲を 4〜12[mora/sec] に設定した．

相槌のタイミングに関してはその同調に関する知見や分析が不十分なので，今回は [3] の決定木で判定されたタイミングをそのまま用いた．

以上の知見より本論文が扱う同調に使用するデータ (発話開始時間 RT, 発話速度 SR) を以下に示す．st は現在までのシステムの発話回数，ut は現在までのユーザの発話回数であり，n は 2 まで考慮する．

1. システムの標準設定 (SRT, SSR)
2. 過去 1 発話の同調処理結果 ($SyRT_{st-1}$, $SySR_{st-1}$)
3. ユーザの過去 3 発話 (URT_{ut-n}, USR_{ut-n})

2 は同調処理が行なわれるたびに，3 について，URT_{ut-n} はシステムからユーザに話者交代した時点で，USR_{ut-n} についてはユーザの発話が終了した時点で更新する．参照する過去のデータが存在しない場合は，システムの標準設定を代わりに用いて処理を行う．

話者交代・相槌判定で応答と判定された場合は，その時点でユーザに同調した発話開始時間 $SyRT_{st}$ と発話速度 $SySR_{st}$ を以下の式で計算する．

$$SyRT_{st} = 0.3 \times SRT + 0.3 \times SyRT_{st-1} + 0.2 \times URT_{ut} + 0.1 \times URT_{ut-1} + 0.1 \times URT_{ut-2} \quad (1)$$

$$SySR_{st} = 0.3 \times SSR + 0.3 \times SySR_{st-1} + 0.2 \times USR_{ut} + 0.1 \times USR_{ut-1} + 0.1 \times USR_{ut-2} \quad (2)$$

但し，この式は実際の同調傾向に基づくものではなく，今回定義した同調を表現する便宜上のものである．

基本的に，現時点のユーザの発話終了後のポーズが，この式により計算した発話開始時間を過ぎていなければ話者交代・相槌判定を継続し，到達した時点で応答する．

話者交代・相槌判定で発話終了後の時間長が大きく影響した場合は，判定時点で計算した発話開始時間を大きく超えている場合も有り得るので，その場合は即座に応答する．

一方，相槌と判定された場合はその時点で相槌を打ち，発話継続と判定された場合は話者交代・相槌判定を継続する．

8 音声合成部

音声合成部には株式会社アニモの商品質音声合成ライブラリ FineSpeech Ver.2[13] を用いる．このツールは XML で制御されるが，SSML(Speech Synthesis Markup Language) にも対応しており，細かな韻律制御が可能である．この XML の例を図 3 に示す．韻律的特徴は図 3 の prosody 内で設定可能であり，発話速度は speed と speed_rate の 2 つのパラメータを用いてそれぞれ 9 段階で制御される．speed は 1 段階変化させるごとに発話速度そのものを，speed_rate は speed の 1 段階ごとの変化量を変更させる．システムが応答する相槌と応答文は全てこのツールを用いて出力される．

```
<?xml version="1.0" encoding="Shift_JIS"?>
<speak>
  <anlmo:prosody speed="7" speed_rate="4">
    はい,伊藤先生は7階の助教授室にいらっしゃいます。
  </anlmo:prosody>
</speak>
```

図 3: 音声合成の使われる XML の例

9 まとめ

本研究では，ユーザに協調的でより自然なリズムの対話を実現するため，新たに言語理解部，応答生成部，対話リズム生成部の 3 つのモジュールを構築し，動作を確認した．言語理解部では，発話途中でも相手のタスク意図を予測する手法を，応答生成部では，ユーザモデルを導入し，協調的でリズムを考慮した応答を生成する手法を，対話リズム生成部では，リアルタイムに話者判定・相槌判定を行い，ユーザに対し対話リズムを同調する手法をそれぞれ提案し実装した．

今後は評価実験を行い，その結果を参考に各モジュールの改良を行う予定である．

参考文献

[1] 長岡千賀, 小森政嗣, 中村敏枝: 対話における交替潜時の対話者間影響, 日本人間工学会 2002, 38(6), pp.316-323, 2002.

[2] 山田真也, 伊藤敏彦, 荒木 健治: 対話相手の音声の品質を考慮した対話状況での目語的・音響的特徴の分析および様々な観点からの考察, 情報処理学会研究会報告, 2005-SLP-56, pp.101-106, 2005

[3] Norihide Kitaoka, Masashi Takeuchi, Seiichi Nakagawa: Response Timing Detection Using Prosodic and Linguistic Information for Human-friendly Spoken Dialog Systems, Journal of The Japanese Society for Artificial Intelligence, Vol.20, No.3 SP-E, pp.220-228, 2005.

[4] 藤江真也, 福島健太, 三宅梨帆, 小林哲則: 相槌生成/認識機能を持つ音声対話システム, 人工知能学会研究会資料, SIG-SLUD-A502-09, pp.41-46, 2005.

[5] 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, pp.49-56, 1993.

[6] 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/.

[7] 藤原敬記, 伊藤敏彦, 荒木 健治: 認識信頼度と対話履歴を用いた音声言語理解手法, 電子情報通信学会論文誌 D-II, 採録決定.

[8] 駒谷, 河原: 音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理, 情報処理学会論文誌, vol.43, no.10, pp.3078-3086, 2002.

[9] 駒谷和範, 上野晋一, 河原達也, 奥乃博: 音声対話システムにおける適応的な応答生成を行なうためのユーザモデル, 電子情報通信学会論文誌, Vol.J87-D-II, No.10, pp.1921-1928, 2004.

[10] 小磯花絵, 伝 康晴: 円滑な話者交替はいかにして成立するか—会話コーパスの分析にもとづく考察—, 認知科学, Vol.7, NO.1, pp.93-106, 2000.

[11] 大須賀智子, 堤内靖雄, 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, 人工知能学会論文誌, Vol.21, No.1, pp.1-8, 2006.

[12] 長岡千賀, 小森政嗣, Draguna Raluca Maria, 河瀬諭, 結城牧子, 片岡智嗣, 中村敏枝: 協調的対話における音声行動の 2 者間の一致ー意見固持型対話と聞き入れ型対話の比較ー, ヒューマンインタフェースシンポジウム 2003 論文集,pp.167-170, 2003.

[13] 株式会社アニモ: 音声合成ミドルウェア AnimoFinespeechV2.1, http:/www.animo.co.jp/products/tts/.