

機関番号：10101

研究種目：若手研究 (B)

研究期間：2008 年度 ～ 2010 年度

課題番号：20700150

研究課題名 (和文) 人間らしさを感じる音声対話システム

研究課題名 (英文) Human-like Spoken Dialog System

研究代表者

伊藤 敏彦 (ITO TOSHIHIKO)

北海道大学・大学院情報科学研究科・准教授

研究者番号：20313926

研究成果の概要 (和文)：

本研究は音声インターフェイスにおいて、対話のリズムが、ユーザの快適性や安全性にどれほどの影響を与えるか明らかにし、これらの要素を音声インターフェイスに導入するための新たな枠組みを提案することである。

そこで、これらに関する対話リズムを生成するためのモデル化のさらなる改良と、音声対話システムへの実装、システム処理速度向上などを行った。

結果、これまでの音声対話システムに比べ、人間らしさ、安心感などの評価を上げることができたが、制作システムの処理速度、タイミング認識精度、音声認識・言語理解精度などの不完全さにより、人間と同等の評価まで上げることはできなかった。

研究成果の概要 (英文)：

In this research we study how dialog rhythm influences user's comfort and reliability and propose a new framework for building spoken interfaces based on this framework. Although we confirmed user's increased satisfaction and smoothness of conversation, we have not reached the level of naturalness of human to human dialog.

To achieve this we have improved our model for generating rhythmical dialogs, re-implemented it into the system and increased processing speed.

In result, we achieved better human-likeness and reliability comparing to the previous system, but we could not reach evaluation similar to human's as our system's timing recognition performance level, voice recognition and language understanding performance levels were imperfect.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	1,100,000	330,000	1,430,000
2009 年度	1,000,000	300,000	1,300,000
2010 年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声対話システム、音声対話、対話リズム、発話タイミング、ユーザ満足度、発話意図、対話制御、韻律

1. 研究開始当初の背景

近年、バリアフリー情報機器といった高度で人に優しいインターフェイスの重要性が言われているなか、日常生活で使用する音声対話をマンマシンインターフェイスとして利用する音声対話インターフェイスは、特別な訓練を必要としない、他のモダリティ（手や目）を拘束しないなど次世代のインターフェイスとしての期待も大きい。しかしながら、現在、カーナビゲーションシステムの一部で利用されている程度であり、実世界のマンマシンインターフェイスとして活用されているとは言い難い。

大きな原因は音声認識精度の問題と思われるが、アプリケーションのタスクを限定し、ユーザの発話範囲を狭めれば、重要なキーワードの音声認識精度は一定以上に保つことができる。背景知識や状況など様々な言語的知識を活用すれば、ユーザ発話の理解精度はさらに向上する。また、システムが理解を間違えていれば、ユーザが直ちにそれを訂正できるといった音声対話の利点を用いれば、現段階でもある程度の実アプリケーションのインターフェイスとして十分な精度でタスクを達成することができる。しかし、評価実験においてもタスクを効率よく確実に達成できるだけでは、ユーザ満足度の高い音声対話インターフェイスにはならず、足りない要素が音声対話インターフェイスの普及を妨げている可能性が高い。

以前の研究で、非対面の人間-人間、人間-人間（応答能力を機械と同等に制限）、人間-機械（文単位理解、文単位応答、発話の割り込み禁止）のタスク指向対話を収集し、音響的・言語的特徴を分析した結果、人間は機械に対して通常の人間に話しかけるように発話することができず、音響的・言語的にも少し不自然な発話になった。また、応答能力を制限された人間相手に対しても、通常のような発話をするのができず、人間と機械の中間的な特徴を示した。これは人間と同等な音声認識能力と言語理解能力を持たせたシステムを用いても同様の結果である。つまり、人間は対話相手の応答内容のみならず応答能力（対話のテンポや応答に含まれる韻律的特徴）から対話相手の対話能力を推測し、それに合わせた対話を行っているとは推測される。

そのため、「ユーザ満足度の高い音声対話インターフェイス」を実現するためには、システムとユーザの対話リズム（対話テンポや発話・訂正タイミングの自由さ）を人間に近づけることにより、人間同士の対話に近い感覚（安心感）をユーザに与え、「いつ話せば?」、「何を話せば?」といったストレスや不安感を消し去ることである。

この目的のために対話リズムを考慮した

音声対話システムの基本システムを構築した。これは人間同士の対話データから発話タイミングを機械学習し、ユーザの音響的特徴と言語的特徴から音声対話システムの発話タイミングを決定する方法で実現した。しかし、評価実験からユーザ満足度や発話のしやすさなどの向上は確認できたが、人間同士の対話に近い感覚を与えるまでには至らなかった。

この原因を調査するために予備的な分析ではあるが人間同士の対話データを少量収集し、発話タイミングや韻律的特徴を発話意図（発話内容）の違いにより分類・比較した結果、対話における話し手の発話タイミングは対話相手の発話特徴のみで決定できるわけではなく、話し手の発話意図（発話内容）や発話の重要度、感情などに大きく影響を受けることが示唆された。つまり、音声対話システムがリズムカルに発話するだけでは人間は機械に対して人間らしさ（安心感）を感じるわけではなく、発話意図（発話内容）や発話の重要度、感情なども考慮した適切なタイミングで発話することが人間らしさ（安心感）を感じさせるために重要である。また、聞き手も話し手の発話タイミングの変化やずれなどから発話内容に隠された意図や感情情報を理解していると考えられ、音声対話システムにこれらの理解機構を組み込むこともユーザ満足度の高い音声対話インターフェイスの実現には不可欠である。

2. 研究の目的

本研究では、マンマシンインターフェイスとしての音声対話インターフェイスにおいて、人間と同等な対話リズム（即応性、ユーザの発話タイミングの自由度、システム発話の応答タイミングや話速、基本ピッチの適切さ）の実現が、ユーザの快適性や満足度にどれほど影響を与えるかを明らかにする。そのため、発話意図（発話内容）や発話の重要度、感情などを考慮した対話リズムを音声対話インターフェイスに実現するための新たな枠組みを提案する。その後、その枠組みを用いた対話リズムを重視した音声対話システムを構築し、被験者による評価実験によってその有効性を実証する。

3. 研究の方法

本研究の焦点は、発話意図（発話内容）や発話の重要度、感情を考慮した人間的な対話リズムがタスク指向型音声対話システムにおいてユーザの満足度や快適性にどれほど寄与するか調査し、それらの要素を音声対話システムに組み込むための新たな枠組みを提案するものである。そのため、対話リズムを人間にできるだけ近づける必要があり、音声認識部や言語処理部、対話制御部、応答生

成部のリアルタイム処理が要求される。現在の計算機の処理能力では複雑なタスクをリアルタイムで認識・理解することは難しいため、ユーザに自由な発話をさせてもある程度言い回しが単純で、システム想定外の発話が発生しにくいタスクにする必要がある。そこで、本研究では音声対話システムの対話タスクをカーナビゲーションの目的地設定・検索タスクとした。

研究の方法は以下の手順で行った。

(1) 大量の人間同士の自然なタスク指向対話の収録：

対話リズムの分析のために対面・非対面、協調的・非協調的、親密・疎遠（初対面）、喜怒哀楽といった対話状況やタスクの種類のような分析対象となる条件は変化させながらも、それ以外の対話条件はできるだけ揃えた効率的に分析が行える大量の自然なタスク指向対話を収録した。また、対話リズムには個人差（性格差）も存在すると考えられるため、発話者一人当たりの対話データもできるだけ大量に収録することを目指した。

(2) 収録した対話データの書き起こしと分析用タグの付与：

収録した音声対話データの書き起こしを行い、対話リズムの分析に必要なタグを付与した。必要となるタグは、発話タイミングタグ、発話意図タグ、感情タグ、重要度タグである。

(3) 人間の対話リズムの分析：

本研究では、対話リズムは発話タイミング、話速、基本ピッチによって構成されると定義し、これらに着目して、発話意図（発話内容）やタスク、感情などの対話状況や対話条件の違いによる対話リズムの変化を分析。対話リズムに関係する対話現象としては、話者交代、話者継続、通常発話、共同発話、相槌、復唱、割り込み特に重点をおきながら分析を行った。

(4) 対話リズムモデルの構築：

(3)の分析に基づいて、音声対話システムに人間的な発話を生成させるための対話リズムをモデル化した。

(5) 音声対話システムの構築と評価実験：

これまで構築した対話リズムを重視した音声対話システム (4)で構築した対話リズムモデルを導入し、より人間らしい対話リズムを生成できる音声対話システムを構築した。その後、通常システムと改良前、改良後のシステムを実際に被験者に使用してもらい、タスク達成率、対話時間などの効率、ユーザ満

足度などの客観的、主観的の両観点から比較した。

4. 研究成果

人間同士の対話データを分析した結果、対話における話し手の発話タイミングは対話相手の発話特徴のみで決定できるわけではなく、話し手の発話意図（発話内容）や発話の重要度、感情などに大きく影響を受けることが確認できた。つまり、音声対話システムがリズムカルに発話するだけでは人間は機械に対して人間らしさ（安心感）を感じるわけではなく、発話意図（発話内容）や発話の重要度、感情なども考慮した適切なタイミングで発話することが人間らしさ（安心感）を感じさせるために重要である。また、聞き手も話し手の発話タイミングの変化やずれなどから発話内容に隠された意図や感情情報を理解していることも被験者実験から示唆された。そこで、様々な発話タイミングに変化させた多数の音声対話を被験者に聞かせ、自然性や理解しやすさなどを評価してもらったところ、発話内容に即した発話タイミングの音声対話に対してもっとも自然性や理解しやすさなどの評価が高く、これまでの分析や仮説を裏付ける結果を得ることができた。そこで、これらに関する対話リズムを生成するためのモデル化のさらなる改良と、音声対話システムへの実装、システム処理速度向上などを行った。結果、これまでの音声対話システムに比べ、人間らしさ、安心感などの評価を上げることができたが、制作システムの処理速度、タイミング認識精度、音声認識・言語理解精度などを計画予定の状態まで構築することができず、人間と同等の評価までには至らなかった。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 3 件）

- ① 伊藤 敏彦, 北岡 教英, 西村 良太, 「音声対話における発話タイミングの影響」、電子情報通信学会, 言語理解とコミュニケーション研究会, 2008/11/10、情報通信研究機構 (NICT) 麹町会議室 (東京都千代田区)
- ② 伊藤 敏彦, 北岡 教英, 西村 良太, 「音声対話における発話タイミングの影響に関する知覚実験」、情報処理学会, 音声言語情報処理研究会, 2008/7/18-19、盛岡・つなぎ温泉ホテル紫苑 (盛岡市)
- ③ 中野 幹生, 船越 孝太郎, 伊藤 敏彦, 荒木 健二, 長谷川 雄二, 辻野 広司, 「音声対話シ

システムにおける発話の重なり情報を利用した音声認識率低下の予測」、第22回人工知能学会全国大会、2008/6/11-13、ときわ市民ホール／勤労者福祉総合センター（旭川市）

〔図書〕（計 0件）

〔産業財産権〕

○出願状況（計 0件）

○取得状況（計 0件）

6. 研究組織

(1) 研究代表者

伊藤 敏彦 (ITO TOSHIHIKO)

北海道大学・大学院情報科学研究科・准教授

研究者番号：20313926

(2) 研究分担者

なし

(3) 連携研究者

なし