論文の内容の要旨

論文題目

概念音声合成の枠組を用いた音声対話システムにおける応答生成手法の構築

氏名

八木　裕司

本文

音声は人間の最も基本的なコミュニケーション手段であり，これを計算機との情報授受に利用することの要求は高いものがある．実用上の観点から言えば，キーボードやマウスを用いるシステムに抵抗のある人にとっても，音声を用いることでシステムとの意思伝達を図ることができれば，感じられる抵抗感は軽減されると考えられる．また，公共の場等においてキーボードやマウスといったデバイスを設置するのが様々な面で問題となるような場合にも，音声を用いるのであればマイクとスピーカー（これらを内蔵型にすれば盗難の恐れもない）を用意するだけでユーザが容易に利用することが可能となる．このような観点から，音声対話システムの研究開発が盛んに行なわれるようになり，実用化されたシステムも出現している．

音声対話システム研究における技術的側面からの背景として，近年における音声認識や音声合成，自然言語処理といった音声言語情報処理技術の顕著な発展が挙げられる．音声対話システムは，様々な音声言語情報処理技術を統合して実現されるものであるため，これらの技術を統合したシステムを構築することは，各研究の方向性としても実用化の面でも重要なものと言える．

音声対話システムということを考えた時，最も重要な点は「対話調音声を取り扱う」ことである．音声言語情報処理の分野において非常に数多くの研究がなされているが，それらの研究は必ずしも「対話調」音声を取り扱う，もしくは目標としている，というものではなく，理想的な音声（適切な環境で録音された朗読調音声等）を対象としているものも多数見受けられる．しかしながら，実際に人が話す音声には様々な「対話調」音声の特徴が含まれる．例えば，言い間違いや言い澱みのようなものやフィラーなどテキストにも現れるものから，意図や感情，性別や年齢といったテキストに現れないものまで非常に様々な要因が「対話調」音声に含まれている．そのため，音声対話システム研究の多くは，「システム内のいずれかの箇所（音声認識，音声合成等）について，いかに対話調音声を取り扱うか」について研究しているものだと言い替えることができる．

音声対話システム研究の多くは，音声認識・理解に焦点を当てたものとなっている．一方，音声出力（音声合成）に関する研究は非常に少ない．特に，国内の研究ではほぼ皆無

と言っても過言ではない．実際に音声出力に焦点を当てていない研究では，音声出力にテキスト音声合成（TTS：Text-to-Speech）と呼ばれる手法を用いた既存のソフトウェアを用いている．しかしながら，この TTS システムとは，一般のテキストから所謂「朗読調」音声を生成することを目的としたものであり，高次の言語情報を反映した音声合成を想定していない，という問題点がある．音声対話システムにおいては，応答文がシステムにより生成されるため，統語構造や談話情報といった高次の言語情報を容易に得ることができるため，これらを応答音声に反映できる音声合成の枠組，すなわち概念音声合成（CTS：Concept-to-Speech）の実現が求められている．

TTS がテキストを入力とするのに対し，CTS ではシステムの内部表現（概念）から直接音声を合成するため，文の生成過程で正確な言語情報が得られ，統語構造を韻律に反映させたり，談話情報で韻律の制御を行なうといったことが容易に行なえる．また，テキスト音声合成ソフトウェアで出力される音声は，単調な朗読調であるという問題点もある．音声対話システムでは，その用途にもよるが，朗読調のみならず対話調の応答音声が求められ，またそれに発話の意図や感情を反映させることも求められる．統語構造や談話情報等の高次の言語情報，あるいは意図や感情等のパラ・非言語情報は，音声の韻律と関連する点が多く，この観点からの研究が重要であるが，実際にこのような観点から研究を行ない，音声合成システムとして構築した研究は，少なくとも国内では見受けられない．

このような背景を踏まえて，本論文では，概念音声合成の枠組を実現し，応答音声に統語構造や談話情報等の高次の言語情報を反映させる手法を構築する．また，その手法を音声対話システムに組み込み，実際のユーザにもわかりやすい音声を合成することを目指す．

音声対話システムでは，応答音声がユーザにとって「わかりやすい」ものであることが求められる．この「わかりやすい」には，応答音声自体の明瞭性等の音質に関わるものもあれば，適切な韻律制御による音声の自然性や，適切な焦点制御による意図の伝達といったものも要因として挙げられる．

本論文の構成を以下に示す．

第 1 章では，本研究の背景や目的について述べる．

第 2 章では，音声対話システムについての概略を述べた後，関連する先行研究について，着目する部分ごとに分類してまとめる．

第 3 章では，エージェント音声対話システムについて述べる．これは，仮想空間中にいるエージェントにユーザが指示することで，仮想空間中内の物体を移動させるというタスクを行なうシステムである．このシステムでは，本論文で提案する手法の基礎となっている部分が確立されている．

応答文生成手法については，生成する応答文の言語情報を常に構文木構造を保持したまま扱う，という手法を提案する．統語構造は，最終的な応答音声の韻律においては，主に文のイントネーションに深く関わってくる．そのため，正確な統語構造を保持することは非常に重要である．音声対話システムでは，自らが 1 から応答文生成を行なうため，一般

的な構文解析ツールとは異なり，始めから 100%正しい構文情報を得ることができる．そのため，システム内部情報として始めから構文木構造を保持したまま扱う手法を構築する．また，構文木構造内にタグを用いることにより，同じ属性の単語は同様に扱えるようにする等の統一的な処理を可能とする．実際の応答文生成には，適切な文テンプレートを選択し，そのテンプレート中のタグに単語を挿入することで実現する．

韻律制御手法については，上記構文木構造中のタグに「重要度」と「新規性」という 2 つのパラメータを同時に保持させ，これらを適切に応答音声の韻律に反映させることで焦点制御を行なう，という手法を提案する．談話情報は，最終的な応答音声の韻律においては，主に個々の単語のアクセントに深く関わっている．伝えるべき単語が強調されることによって，システムの意図がユーザに伝わりやすくなることが期待できる．「重要度」や「新規性」といった情報もまたシステムが 1 から作り出す情報であるため，これらの情報を応答音声に反映させないのは非常にもったいないと言える．そのため，これら談話情報を適切に設定し，また応答音声の韻律に適切に反映させる手法を構築する．

これらの手法を明瞭性・自然性の観点から聴取実験によって評価し，有効性と問題点について考察する．

第 4 章では，道案内音声対話システムについて述べる．これは，システムがユーザに指示することで目的地まで道案内する，というタスクを扱うシステムである．第 3 章で明らかになった問題点を解決するためには，より豊富な種類の応答生成が必要となることがわかったため，タスクを道案内に変更した音声対話システムを構築し，その中で提案手法の改良を行なう．

応答文生成手法については，テンプレートが文単位であったため，少しでもスタイルの異なる文章（修飾語が付く等）を生成するためにも，新たなテンプレートを用意する必要があった．そのため，タスクが拡張される等によって必要な応答文の種類が増えると，それに伴いテンプレート数を増加させる必要があった．そこで，文単位ではなく，文節単位でテンプレートを用意し，文節を適切に接続することで応答文を生成する，という手法を提案する．この手法は，文節単位のテンプレートのタグに単語だけではなく文節も挿入できるようにすることで実現される．評価実験から，従来の文テンプレートを用意する手法に比べてより少ないテンプレート数で，柔軟かつ豊富な応答生成が実現できることを示す．また，この提案手法は，タスクによらず汎用性のある応答文生成手法である．

韻律制御手法については，より自然な応答音声を目指し，新たな韻律制御規則を導入する．聴取実験から，新たな韻律制御規則の妥当性を検証する．

また，これらの提案手法による応答生成手法の評価を行なうために，さらなる聴取実験を行なう．具体的には，統語構造と談話情報の 2 項目の取扱いについて検討する．

第 5 章で本論文をまとめ，今後の展望や課題について述べる．