氏　　名	じょ　うんぴょう 徐　　雲　彪
学位(専攻分野)	博　　士（工学）
学位記番号	博甲第２８４号
学位授与の日付	平成14年11月27日
学位授与の要件	学位規程第３条第３項該当
研究科・専攻	工芸科学研究科　情報・生産科学専攻
学位論文題目	A Study on a Multilingual-Supporting Spoken Dialog System （多言語を支援する音声対話システムに関する研究）
審査委員	(主査) 教　授　黒川　隆夫 教　授　柴山　潔 教　授　辻野　嘉宏 教　授　新美　康永 助教授　荒木　雅弘

論文内容の要旨

現在我々が日常行っている情報収集活動は、文献、ラジオ、テレビ、インターネットなど多岐のメディアにわたっているが、この中で音声の占める役割は大きい。特に音声対話システムは、コンピュータやインターネットとのインタフェースとして重要な役割を果たすものと期待されている。しかし、従来研究されてきた音声対話システムは、単一の言語、単一のタスクに関するものが多かった。そこで本研究では、異なる言語間での移植性の高い音声対話システムを構成する方式について研究を行った。また、この方式に基づいて、中国語と日本語の音声対話システムを３種類のタスクについて構成し、その評価実験を行った。その結果、異なる言語およびタスクについて、容易に音声対話システムを構成できること、構成されたシステムのいずれもが良好に動作することが確認された。さらに、この研究過程で必要となった中国語の音声認識および音声合成の研究結果も含んでいる。

本論文は以下に略述する７章から構成されている。第１章では、研究の背景として従来の音声対話システムの研究を概観し、言語間で移植性の高い音声対話システムの構成方式の必要性を指摘し、本研究の目的としている。

第２章では、音声対話システムを構成するのに必要な種々の基本技術、特に音声認識、自然言語処理、応答生成について従来の研究の総括を行っている。

第３章では、音声対話システムの中国語版で必要となる中国語の音声認識システムの研究結果について報告している。まず中国語の音節構造と音声認識での基本単位との関係、認識システムの設計に使用した中国語の音声コーパスについて説明した後、音声認識システムで用いる隠れマルコフモデルの構造と音声認識率の関係についての実験結果を述べている。最後に、開発した音声認識システムを用いた小規模な文認識の実験では、94.6%の認識率が達成されたことを報告している。

第４章では、多言語を支援する音声対話システムの構成法について１つの方式を提案している。すなわち、音声対話システムの機能を言語に依存する音声インタフェース（音声認識と音声合成）と言語に依存しないで動作する対話制御、この両者を仲介する格構造変換に分割する。ある言語による音声入力は、その言語を受理する音声認識システムによって認識された後、構文および意味的に解析され、文の意味の抽象的表現である格構造に変換される。格構造変換部では、入力音声の言語によって記述された格構造を対話制御内での記述言語（これをピボット言語という）に変換する。対話制御部

では、あるタスクドメインでの対話を遂行するのに必要な知識をフレーム形式で表現しておき、ピボット言語で変換されたユーザの入力に対するシステムの動作を決定する。システムの動作がユーザへの応答である場合は、抽象的な形式の応答を生成し、これをユーザの言語に変換した後、音声合成システムを駆動する。このような方式は他の研究でも用いられているが、本研究では、言語に依存しない対話制御の機能を強化することにより、言語に依存した処理と格構造の変換操作を簡略化していること、タスクドメインごとに使用する単語全体を意味的な断層構造として与えることにより言語間の移植性を容易にしているところに特徴がある。

第５章では、中国語の音声合成方式についての研究結果を報告している。近年の音声合成方式では、多量の音声データをデータベースとして蓄積しておき、合成したい文を基本単位に分割した後、前後の音の環境やアクセントを考慮して、各基本単位に適合した音声波形をデータベースから選択し接続して合成する。これに対して本研究では、比較的少量のデータで自然な音声を合成するために、文全体のイントネーションを規則で与え、各音節の声調、時間長、振幅を決定するための規則を音声データベースから統計的に抽出した。合成音の聴取実験をインターネット経由とヘッドホンを用いて行った結果、了解性、自然性ともに実用に耐えうる音質であることが証明された。

第６章では、第４章で述べた方式に基づいて、中国語と日本語で動作する対話システムを３種類のタスクドメインについて構成し、対話実験を行った結果について報告している。中国語については、第３章と第５章で述べた音声インタフェースを、日本語については、既存の音声インタフェースを用いている。中国語、日本語ともに 15 名の学生を用いた評価実験の結果、正しく音声認識された入力の約 85%に対して対話システムとしての正しい動作を達成することができた。この結果は完全ではないが、２つの言語で異なる３種類のタスクに関する対話システムを短期間に構成でき、しかも上記の性能を得たことは、本研究が意図した音声対話システムの言語間の移植性を容易にするという目的を十分に達成できたと結論している。

第７章では、本研究の結果をまとめ、残された課題について論じている。

論文審査の結果の要旨

本研究は、従来単一言語、単一タスクに対して研究されてきた音声対話システムに関して、言語やタスクによらない統一的なシステムの構成法を提案した。すなわち、音声対話システムの機能を言語に依存する音声インタフェース（音声認識と音声合成）と言語に依存しないで動作する対話制御、この両者を仲介する格構造変換に分割する。ある言語による音声入力は、その言語を受理する音声認識システムによって認識された後、構文および意味的に解析され、文の意味の抽象的表現である格構造に変換される。格構造変換部では、入力音声の言語によって記述された格構造を対話制御内での記述言語（これをピボット言語という）に変換する。対話制御部では、あるタスクドメインでの対話を遂行するのに必要な知識をフレーム形式で表現しておき、ピボット言語で変換されたユーザの入力に対するシステムの動作を決定する。システムの動作がユーザへの応答である場合は、抽象的な形式の応答を生成し、これをユーザの言語に変換した後、音声合成システムを駆動する。この方式に基づいて、中国語と日本語で動作する対話システムを短期間に３種類のタスクドメインについて構成し、対話実験を行った結果良好に動作することが確認された。このことから本研究が意図した音

声対話システムの言語間の移植性を容易にするという目的を十分に達成できたと結論でき、音声対話システムの研究に対する貢献は大きいといえる。また、この研究の一環として中国語の認識と合成の研究を行っているが、特に合成の研究では、比較的少量のデータベースを用いても了解性、自然性ともに良好な合成音を生成することに成功しており、ポータブルな音声合成方式として、工学的価値が高い。

本研究の内容は、学術誌の論文２編（下記(1)(2)）および査読のある国際会議の論文３編（下記(3)(4)(5)）として公表されており、また現在学術誌に１編を投稿中である。

(1) Y. Xu, X. Song, M. Araki, and Y. Niimi, “A Chinese speech synthetic system based on TD-PSOLA,” Journal of Chinese Language and Computing, vol. 11, no. 1, pp. 63-79 (2001).

(2) Y. Xu, M. Araki, and Y. Niimi, “A Chinese Speech Synthetic System based on Monosyllable Concatenation,” Journal of the Acoustic Society of Japan, vol. 23, no. 3, pp.166-169 (2002).

(3) Y. Xu, M. Araki, and Y. Niimi, “A multilingual spoken dialog system,” Proc. Of the 2000 International Symposium on Chinese Spoken Language Processing, pp. 175-178 (2000)

(4) Y. Xu, M. Araki, and Y, Niimi, “A multilingual-supporting dialog system using a common dialog controller,” Proc. of the 7th European Conference on Speech Communication and Technology, pp. 1283-1286 (2001).

(5) Y. Xu, X. Song, M. Araki, and Y. Niimi, “A Chinese speech synthetic system based on TD-PSOLA,” Proc. of the International Conference on Chinese Computing 2001, pp. 171-175 (2001).