京都工芸繊維大学

氏　　名	さ しょう **左　祥**
学位(専攻分野)	博　士（工学）
学位記番号	博甲第　638　号
学位授与の日付	平成 24 年 3 月 26 日
学位授与の要件	学位規則第 4 条第 1 項該当
研究科・専攻	工芸科学研究科　設計工学専攻
学位論文題目	Two Key Technologies for a Flexible Speech Interface : From the Perspective of Human-Robot Interaction （柔軟な音声インタフェースを実現するための２つの基盤技術：ヒューマン-ロボット・インタラクションの観点から）
審査委員	(主査)教授　岡　夏樹 教授　辻野嘉宏 教授　澁谷　雄 准教授　荒木雅弘

論文内容の要旨

音声は人の一番自然なコミュニケーション手段であり、人と機械の間のインターフェースとして活用することが望まれている。しかし現状では、音声認識の性能が不十分であり、また予め登録したコマンドしか認識できないなどの問題があるため、柔軟性のある音声インターフェースの実現が難しい。そこで私は、「未知語の音韻列の学習」と「発話対象の検出」の二つの課題に注目し、柔軟性のある音声インターフェースを実現するための要素技術の開発を行った。

まず、一つ目の課題として、未知語の音韻列の学習技術を開発した。実環境における音声インターフェースにおいて未知語の音韻列の学習は大変重要である。システムは予め登録したコマンドに応じるだけでなく、オンラインで未知語の音韻列（発音）を学習できることが望ましい。例えば、システムが未知の人や未知の物体に出会ったときに、その名前の音韻列を人の発話から学習できれば、システムは自分の語彙をオンラインで拡張することができ、その後の会話の中で学習した名前を使って人とコミュニケーションすることができるようになる。

未知語の音韻列を学習するために、未知語に対して音韻認識を行えばよいが、現在の音声認識の性能は十分ではないため、音韻認識の誤りが生じる可能性が高い。従って、正確な音韻列を学習するために、音韻認識の誤りを訂正する必要がある。そこで本研究では、ユーザが未知語を繰り返すことにより、認識誤りを訂正する方法を提案した。また、訂正する際、ユーザは訂正した音韻列を確認しながら、インタラクティブに訂正することができる。提案法の特徴は次の二つである。１）ユーザは未知語をそのまま繰り返すだけではなく、未知語の音韻列の中の間違った部分だけを繰り返すこともできる。そのため、システムは認識誤りをより効率的に特定することができる。２）システムは、訂正する過程の履歴情報を用い、訂正の効率を向上させる。例えば、もし訂正後の音韻列が悪くなった場合、ユーザはその音韻列を訂正前のバージョンに戻すことができる。また、各回の訂正は、必ず違う音韻列を生成する。

訂正する際、提案法ではまず誤認識が含まれる音韻列と訂正発話の音韻列の間で DP マッチングを使ってアラインメントを行う。そして各音韻のペアからより信頼度の高い音韻を選択する。

こうしてより信頼度の高い音韻列を生成することができる。なお、音韻の信頼度としては一般化事後確率を用いる。評価実験の結果、提案法は平均 3 発話という非常に高い効率で未知語の正確な音韻列を学習できることが分かった。

次に、二つ目の課題として、発話対象の検出技術を開発した。システムとの円滑な会話を実現するためには、人の発話がシステムに向けられたものであるか否かを判断する必要がある。システムは自分に向けられた発話に対してだけ反応し、それ以外の発話に対しては反応してはならない。もしこの機能がなければ、システムは自分に向けられていない発話（例えばテレビの音や人間同士の雑談など）にも反応し、人とのコミュニケーションがうまくとれなくなり、危険な行為を起こしてしまう可能性もある。

このような問題を解決するために、私は発話対象を検出する手法を提案した。提案法は、ロボットが人の発話に従って物体を操作するタスクにおいて有効である。このタスクでは、人がロボットに対して、その時の物理環境におけるロボットが実行可能な物体操作行為を命令すると仮定する。この仮定の下では、対システム（ロボット）発話の内容と現在の物理環境とのマッチング度合いは高くなる。従って、提案法では、まず発話を現在の物理環境における実行可能な物体操作行為として解釈し、そしてその行為と物理環境とのマッチング度合いを評価することによって発話対象を検出する。

マッチング度合いの評価基準として、本研究で独自に提案したマルチモーダル・セマンティック・コンフィデンス（MSC）を使用する。MSC では、音声認識、物体認識とロボットの操作動作の生成から得られた確信度をロジスティックモデルで統合して計算する。音声確信度の計算は従来法に従うが、物体と動作の確信度の計算は本研究で新たに提案したものである。また、ロジスティックモデルのパラメータは、尤度最大化基準を用いて学習する。実験では、実機ロボットを用いて提案法を評価した。提案法では、95%以上の非常に高い精度で発話対象を検出することができた。

論文審査の結果の要旨

申請者は、柔軟な音声インタフェースの実現に向けての課題のうち、特にロボットとの音声対話の実用化を目指す上で不可欠となる、「未知語の音韻列の学習」と「発話対象の検出」の二つの課題に注目し、これらを解決するための基盤技術の開発を行った。この二つの課題はこれまであまり研究が行われてこなかったが、近い将来、ロボットが日常生活の場で活躍を始めるという期待が高まる今日、その重要性が急速に高まったものである。したがって、申請者が本研究で提案した基盤技術の意義や効果は極めて大きい。

申請者は、まず、未知語の音韻列の学習技術を開発した。ロボットが家庭環境に導入された場合は、工場等におけるロボットとは異なり、日常的に、初めての人、初めての物、初めてのできごとに遭遇することが予想される。そのため、それらを描写し、意思疎通を図るためには、未知語を獲得できる機能が必須となる。特に一般家庭での使用場面を考えると、対話的に未知語を獲得できることが望ましい。申請者は Interactive Phoneme Update 法を提案し、従来法と比べて、極めて高い効率で未知語の正しい音韻列が獲得できることを実験的に示すことに成功した。

さらに申請者は、発話対象の検出技術を開発した。日常生活場面におけるロボットは、人どう

しの会話や、テレビから聞こえてくる音声等を絶え間なく耳にしている。このような状況で、発話がロボット自身に向けられたものかどうかを判断できることは、必須の機能であると言える。もし、これができないと、ロボットは、自身に向けられたのではない発話に反応し、予想外の動きをするという大変危険な事態を招く。申請者は、Multimodal Semantic Confidence と呼ぶ尺度を考案し、これを用いて 95%以上の非常に高い精度で発話対象を検出できることを示した。

以上の結果から、本論文は、「未知語の音韻列の学習」技術と「発話対象の検出」技術を用いて、人とロボットの音声対話を円滑化することができるとともに、安全・安心なインタラクションを提供できることを多角的に示している。将来のサービスロボットの本格的な普及に必要な基礎技術を提案し、その有効性を実験的に検証したという点で、本研究内容は高く評価できる。

なお、本論文の内容は、レフェリーによる審査を経た 4 編の論文[1,2,3,4]および 1 編の審査中の論文[5]を基に構成されており、これらの論文はいずれも申請者が筆頭著者である。

[1] Xiang Zuo, Naoto Iwahashi, Ryo Taguchi, Shigeki Matsuda, Komei Sugiura, Kotaro Funakoshi, Mikio Nakano, and Natsuki Oka, ROBOT-DIRECTED SPEECH DETECTION USING MULTIMODAL SEMANTIC CONFIDENCE BASED ON SPEECH, IMAGE, AND MOTION, Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing, pp. 2458-2461, 2010.

[2] Xiang Zuo, Naoto Iwahashi, Ryo Taguchi, Kotaro Funakoshi, Mikio Nakano, Shigeki Matsuda, Komei Sugiura, and Natsuki Oka, Detecting Robot-Directed Speech by Situated Understanding in Object Manipulation Tasks, Proceedings of 19th IEEE International Symposium in Robot and Human Interactive Communication, pp. 643-648, 2010.

[3] Xiang Zuo, Naoto Iwahashi, Kotaro Funakoshi, Mikio Nakano, Ryo Taguchi, Shigeki Matsuda, Komei Sugiura, and Natsuki Oka, Detecting Robot-Directed Speech by Situated Understanding in Physical Interaction, 人工知能学会論文誌, vol. 25, no. 6, pp. 670-682, 2010.

[4] Xiang Zuo, Taisuke Sumii, Naoto Iwahashi, Kotaro Funakoshi, Mikio Nakano, and Natsuki Oka, CORRECTION OF PHONEME RECOGNITION ERRORS IN WORD LEARNING THROUGH SPEECH INTERACTION, Proceedings of IEEE Workshop on Spoken Language Technology (SLT 2010), pp. 348-353, 2010.

[5] Xiang Zuo, Taisuke Sumii, Naoto Iwahashi, Mikio Nakano, Kotaro Funakoshi, and Natsuki Oka, Correcting Phoneme Recognition Errors in Learning Unknown Words through Speech Interaction, Speech Communication, (Under review).