様式 C-19

科学研究費補助金研究成果報告書

平成２１年　６月１８日現在

研究種目：若手研究（B）
研究期間：2006～2008
課題番号：１８７００１１８
研究課題名（和文）　スイッチフリーな実環境音声言語インタフェースの研究

研究課題名（英文）　A switch-free spoken language interface for real-world speech interface

研究代表者
李　晃伸（LEE AKINOBU）
名古屋工業大学・大学院工学研究科・准教授
研究者番号：８０３３２７６６

研究成果の概要：

日常環境で音声入力の開始・終了を意識せずに誰でも自然に使うことができる音声インタフェースの実現を目指して、音響・言語情報を密に統合したロバストかつ実時間の音声区間検出の研究を行った。短時間フレームごとの GMM 尤度を用いた音声区間検出、およびフレーム単位の部分的な照合スコアから認識処理の照合度（信頼度）を判定する手法を提案した。大学案内音声対話システムを実際に構築して公共の場に設置し，データ収集と評価を行った。本研究の成果の一部は、大語彙音声認識エンジン Julius の機能として一般に公開されている。

交付額

（金額単位：円）

	直接経費	間接経費	合　計
２００６年度	1,300,000	0	1,300,000
２００７年度	1,300,000	0	1,300,000
２００８年度	800,000	240,000	1,040,000
年度			
年度			
総　計	3,400,000	240,000	3,640,000

研究分野：音声言語認識・マンマシンインタフェース
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：音声区間検出・音声インタフェース・音声認識システム・耐雑音・発話棄却

１．研究開始当初の背景

次世代マン-マシンインタフェースとして音声言語インタフェースの可能性が注目されている。機械に気軽に話しかけるだけで動作するシステムが実現できれば、日常生活におけるストレスのない機械操作が実現でき、複雑な機械操作を覚える負荷から解放され、大量の情報機器と上手に付き合える豊かな情報化社会の実現に寄与すると期待される。

音声インタフェースが使いやすく信頼できる入力手段として広く実用化されるためには、さらなる音声認識性能の向上が要求される。しかし一方で、たとえ「どんな発話でも認識・解釈できる高度な知的機械」が実現したとしても、日常環境において誤動作ばかり起こすようでは全く使い物にならない。精度の改善や発話理解の高度化はもちろんのこと、実際の人間の生活空間において避けることのできない多種多様な騒音、雑音、あるいは人のざわめき、他者の声などの音響的外乱に対して、システムが頑健に動作できることが強く求められる。

音響的外乱には、車の音やドア音などの非音声雑音といった従来の雑音のほかに、笑い声や咳、独り言、あるいはシステムとは無関係な他者との会話といった、音声であるがシステムへの明示的な発話からは区別されるべき外乱も数多く発生する。人の声から離れた雑音は、音声の音響的特徴が含まれるかどうかで識別可能であるが、システムと無関係な会話は、音響的特徴のみでは判別不可能である。

限定的な方策として、実用化されている多くの音声インタフェースでは Push-to-talk、すなわちスイッチを押してタイミングを明示的に指示して発声する方式を採用している。しかし、スイッチを押して話しかけるという行為は音声言語にとって不自然な動作であり、また手を用いずともコミュニケーションできるという音声インタフェースの良さを損ねてしまう。また、従来の研究も雑音や音楽と音声の識別、あるいはタスク外発話の検出など、限定的な対策にとどまっている。

２．研究の目的

本研究は、音声言語インタフェースの本質である「人間にとって直感的なインタラクション」の実現を大目標として、雑音や無関係な発話、笑い声や咳などの様々な音響的外乱を含む日常の音環境の中から、人間の、システムに向かって話した発話区間だけを高精度に抽出することを目的とする。

これによって、日常環境で、スイッチングを意識することなく、自分に語りかけられた発話だけに反応する、自然で信頼感のある「スイッチフリーな音声言語インタフェース」の実現を目指す。

３．研究の方法

音響レベルから言語レベル、タスクレベルまでに至る多段階の音声区間検出・発話検証手法を密統合する。アプローチは大きく分けて以下の３つの段階からなる（図１）。

1. 雑音環境からの頑健な音声区間検出（セグメンテーション、Voice Activity Detection）
2. 音響的特徴を用いた発話検証：不要音入力の棄却（Gaussian Mixture Model）
3. 言語的特徴を用いた発話検証：タスク外発話の棄却（信頼度基準、ガーベージモデル）

これらの種々のレベルの発話検証基準を密に統合することで、ユーザの明確に意図した発話のみを受理するアルゴリズムを研究する。

最初の音声区間検出では，短時間フレーム（10ms～100ms）単位で音声・非音声の判別を行う．パワーに基づく方法等を用いて，物音など音声から遠い雑音を省いて音声，あるいは音声に近い区間の開始・終了を検出する．続いての発話検証・棄却の段階では，ガウス混合分布モデル（Gaussian Mixture Model）を用いて，通常の音声発話とそれ以外の入力（笑い声・咳・マイクから離れた背景の遠隔発話など）の判定を行い，ユーザのシステムへの発声を検出する．さらに，最後の段階では，音声認識を実行し，音響尤度・言語尤度・単語信頼度といった情報から，発話がタスク内発話であるかどうかを判定する．本研究では特に，これらを密統合して，実際にリアルタイム処理を行い，システムで動作させるとともに，様々な実環境においてデータ収集および検証実験を行うことを目標とする．

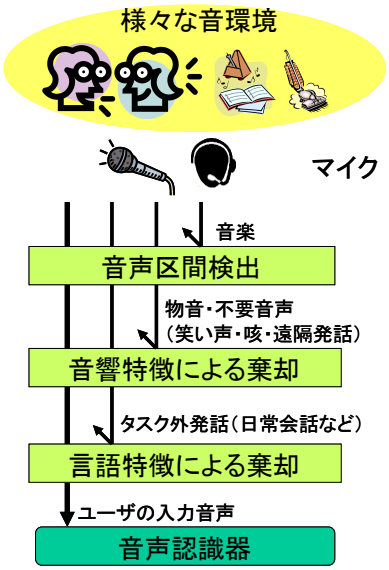

図１：提案手法の概念図

４．研究成果

本研究で解決すべき課題は、異なる性質を持つ複数の音声入力検知機構や無効入力棄却判定機構を統合することで、実環境においてユーザの意図する入力音声のみを受理するアルゴリズムを構築することである。最終的には、入力開始を意識しない「スイッチフリー」な音声言語インタフェースを実現する。最終的な目標としては、ユーザの入力音声のみを確実に取り出せる。自然で信頼感のあるスイッチフリーな実環境向け音声認識システムの構築を目指す。

これらの研究計画にしたがい、年度ごとに下記の研究成果を挙げた。

平成 18 年度は、不要音や非マッチモデルの棄却アルゴリズムについて、統合手法の検討を行った。また、研究成果を実証し、かつ

日常環境における自然なユーザの対機械発話を収集するために、大学内における学内の音声案内システムをターゲットとして、施設案内音声対話システムを試作した。具体的な研究成果を以下に示す。

(1) 音声認識におけるフレーム単位の信頼度基準に基づく動的モデル選択および棄却アルゴリズムの検討
(2) 棄却アルゴリズムの認識エンジンへの統合手法の検討と試験実装
(3) 音声対話システムの構築およびデータ収集

平成 19 年度は、不要音・不要発話の棄却アルゴリズムについて、各手法の性能検討とその統合手法の検証ならびに評価を行った、大学の学生向け案内をタスクとした音声情報案内対話システムを構築し、運用を開始すると共に、日常環境における自然なユーザの対機械発話の収集を予備的に行った。具体的な内容は以下の通りである。

(1) 雑音環境向けの GMM に基づく実時間音声区間検出プログラムを音声認識ソフトウェア上に統合した。
(2) 学内案内用音声対話システムのプロトタイプ構築および試験運用を行った。
(3) 上記システムを稼働させ、雑音や無効入力を含む全ての入力を収集した。
(4) 収集データの整備：収集したデータについて、予備的な分析，既存のデータとの比較を行った.

平成 20 年度は、昨年度までの基礎アルゴリズムの検討結果を受けて、不要音・不要発話の棄却アルゴリズムの実装と実際の音声対話システム上における動作実験、研究成果の整理と公開などを行った。またこのアルゴリズムを応用した自然な音声認識インタフェースのさらなる可能性について検討した。具体的な内容は以下の通りである。

(1) 音声認識ソフトウェア Julius における GMM に基づく実時間音声区間検出を日常環境で実践するとともに、ロボット組み込みなど特定環境への応用を図った。これは企業との音声対話ロボットに関する共同研究に繋がった。
(2) 昨年度、大学で設置した学内案内用音声対話システムの知見を活かし、マイコン向け音声認識ソフトウェアにおいて同機能を実装した質問応答システムの開発を行った。
(3) 本研究成果を広く公開すべく、音声認識ソフトウェア Julius に実装するとともに、応用を促進する目的で API や開発環境を整備した。また、その成果を論文として学会誌に寄稿した。
(4) 本研究で開発した手法を発展させた新たな研究課題として、認識中の仮説候補のスコア状態からその確定度を判断し、「しゃべり終わる前に認識する」研究の検討を開始し、その予備的結果を電子情報通信学会音声研究会にて発表した。

まとめると、本研究では以下のような研究成果を得た．まず、音響的な区間検出・入力検証については、研究開始時には区間検出後に GMM でその検出区間単位で受理・棄却を判定していたのに対して、フレームごとの GMM 尤度を前段のパワー等に基づく音声区間検出パラメータと密統合することで，GMM とパワーの 2 基準に基づくロバストな音声区間検出を実現した．また、言語的な側面については、従来照合結果の信頼度が単語の事後確率を用いて単語単位で定義されていたのに対して、認識処理の途中で得られるその時点での部分的なスコアを用いて近似的に算出する方法を提案し、その有効性を証明した。これによって、全体の認識が終わるのを待たずに入力と並行して従来に近い性能の信頼度計算が行えることが示せた。また、応用として、入力フレームごとの信頼度に基づく複数モデル並列認識法についても検討した.

大学案内をタスクとした音声対話情報案内システムを構築して公共の場に設置し，ユーザの対機械発話の収集・分析を行った．さらに、マイコン向け音声認識ソフトウェアにおける質問応答システムの開発を行った。

なお、本研究成果である音声区間検出手法の一部は、研究代表者が作成・公開しているフリーの音声認識ソフトウェア Julius の機能として実装され、広く一般に公開されている。

今後は、音響的な判定と言語的な判定の統合等による手法の高度化、および実データでの大規模な実験的評価が必要であろう。

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　２件）

① 李晃伸、河原達也、「Julius を用いた音声認識インタフェースの作成」、ヒューマンインタフェース学会誌、解説記事、2009 年 1 月、Vol.11, No.1, p.31-38.
② Tomohiro Hakamata, Akinobu Lee, Yoshihiko Nankaku, and Keiichi Tokuda、"Reducing computation on parallel decoding using frame-wise confidence scores", International Conference on

Spoken Language Processing, 査読有, 2006年, pp.1638-1641.

〔学会発表〕(計　6件)

① 小島弘, 南角吉彦, 李晃伸, 徳田恵一、「信頼度基準による解探索打ち切りに基づく超早音声認識」、電子情報通信学会音声研究会 (SP)、2009年1月29日、奈良先端科学技術大学院大学.
② 小窪浩明, 李晃伸, 河原達也, 鹿野清宏、「SHマイコン向け連続音声認識ソフトウェアを用いた質問応答システムの開発」、日本音響学会秋季研究発表会、2008年9月11日、九州大学.
③ 袴田智博, 南角吉彦, 李晃伸, 徳田恵一、「仮説の遅延確定による1パス信頼度の評価と複数デコーダ間枝刈りへの応用」、日本音響学会春季研究発表会、2008年3月17日、千葉工業大学.
④ 李晃伸、「大語彙連続音声認識エンジンJulius ver.4」、電子情報通信学会音声研究会 (SP)、2007年12月21日、NTT京阪奈ビル.
⑤ 辻洋祐, 垣鍔亮太, 小崎和正, 南角吉彦, 李晃伸, 徳田恵一、「アレイ入力と接話マイク入力のデュアルデコーディングに基づくキャンパス音声情報案内端末の構築」、日本音響学会秋季研究発表会、2007年9月21日、山梨大学
⑥ 袴田智博、南角吉彦、李晃伸、徳田恵一、「フレーム単位の信頼度を用いた並列音声認識におけるデコータ間枝刈りの検討」、情報処理学会音声言語情報処理研究会 (SIG-SLP)、2006年7月7日、ルネッサンスリゾートなると.

〔その他〕

大語彙連続音声認識エンジン Julius：
http://julius.sourceforge.jp/

研究開始時：バージョン3.5.1
9,827 views, 1,285 downloads / month
研究終了時：バージョン4.1.2
60,666 views, 3,105 downloads / month

６．研究組織

(1)研究代表者
李　晃伸 (LEE AKINOBU)
名古屋工業大学・大学院工学研究科・准教授
研究者番号：８０３３２７６６

(2)研究分担者

(3)連携研究者