氏　　名	名取　賢
博士の専攻分野の名称	博士（工学）
学 位 記 番 号	医工博甲第２９１号
学位授与年月日	平成２６年3月２０日
学位授与の要件	学位規則第４条第１項該当
専　　攻　　名	情報機能システム工学専攻
学 位 論 文 題 目	音声からキーワードを検出する技術の高度化に関する研究
論 文 審 査 委 員	主査　教　授　関　口　芳　廣
	教　授　宗　久　知　男
	教　授　福　本　文　代
	教　授　小　谷　信　司
	准教授　鈴　木　良　弥
	准教授　丹　沢　　勉

学位論文内容の要旨

[研究の位置づけ]

近年、マルチメディアデータの生成・編集環境の普及、ストレージの大容量化、ネットワークインフラの充実により、動画コンテンツに代表される音声やマルチメディアコンテンツが急激に充実してきた。また、会議や講演などにおいて音声の録音や、映像の録画を行う動きも広まってきている。これらのコンテンツはネットワークストレージや動画共有サイトなどにアクセスすることで、容易に利用することができる。

これに伴い、これらの大量のコンテンツから視聴したい場面を検索したいという要求が高まっている。しかし、多くのコンテンツは動画像と音声で構成され、テキスト情報を含んでいない。そのため、音声を含むデータに対しては、音声認識技術を適用してコンテンツを検索する方法が期待され、音声ドキュメント検索として研究が行われている。

音声中の検索語検出 STD(Spoken Term Detection)は、ある特定の検索語(1 個以上の単語からなる言葉)が、音声ドキュメント群中のどのドキュメントのどの位置に含まれているのかを特定するタスクである。従来の STD の研究の大部分は未知語と音声認識誤りの問題に焦点を合わせているが、この研究では検索精度の向上に主眼をおいている。

[新しい STD 手法の提案と検証]

この研究では、サブワードベースの CN(Confusion Network)を使用した STD 手法を提案

する。複数の音声認識システムの出力から構成された音素遷移ネットワーク PTN(Phoneme Transition Network)から検索語を検出するために、編集距離ベースの DTW(Dynamic Time Warping)フレームワークを利用している。また，音声認識システムの出力から CN を作るために PTN ベースの認識が行われている。

単一音声認識の 1 ベスト出力と CN を比較した場合、CN は豊富な情報を持っていることから、STD に対して CN の利用は有効な手法である。また、異なる言語モデルと音響モデルを利用した複数の音声認識システムとその出力を使用することは、単語抽出性能を向上させることに効果があることが知られている。この研究では、複数の音声認識システムを構築し、その出力を STD に応用したことが特徴の一つである。

具体的には、同じデコーダに基づく 12 種類の音声認識システムを構築している。認識システムで使用する音響情報と言語情報は夫々，2 種類の音響モデル(tri-phone ベースと syllableベース)と 6種類の言語モデル(単語ベースとサブワードベース)の組み合わせである。複数の音声認識システムの出力を、効果的に STD 用のインデックスとするために、CN の構造を利用したネットワーク型インデキシングを行っている。

日本語の STD テストコレクションに対し、単一の音声認識システムを利用するより、複数の音声認識システムの出力を利用することが、STD の性能を向上させることに有効であることが実験で確認された。さらに、複数の音声認識システムの出力をネットワーク型のインデックスとして利用することが STD に有効であることも確認されている。

しかし、PTN の冗長性から、多くの誤検出が発生する。複数の音声認識システムの利用は、より良好な単語抽出性能を達成することができるが、同時に多くの誤検出が発生する。この誤検出を抑制するために、複数の音声認識システムの出力を利用したネットワーク型インデックスを構築する際に得られる情報を、誤検出を抑制するパラメータとして利用した。これらの誤検出抑制パラメータを、DTW の距離計算式に導入することによって、誤検出が抑制されることが実験によってわかった。例えば、同じ音素を認識した音声認識システムの数を特徴量として導入することによって、大幅に検索性能が改善されている。

また、音素長が長い検索語は誤検出が少ないのに対し、音素長が短い検索語は検出され易く，それが誤検出である場合が多いことが判明した。そこで、検索語の音素長に着目し、音素長が短い検索語に対しては誤検出抑制パラメータの適用法を工夫した。

さらに、ネットワーク型インデックスの「複雑さ」に着目し、誤検出を抑制することが可能ではないかと考え、検索語のエントロピーを利用する方法を考案した。エントロピーを利用した手法を、日本語 STD テストセットの STD タスクと iSTD タスクで評価している。その結果、エントロピーの利用は、高 Recall 域での STD 性能の向上に有効であることがわかった。また、iSTD タスクにも有効であるということもわかった。

[提案手法の応用]

従来の STD の研究の多くは，限定された環境のデータに対するものが多く、実環境下での有効性評価の研究例は少ない。STD 技術を用いたいくつか応用分野があるものの、STD の全体的な有用性が、実際の環境で使用される実用的な検索システムで評価されたことはほとんどない。

そこで、実際に使用されている電子ノート作成支援システムでのノート見直し作業を例に、実環境下での STD 技術の有効性評価を行った。電子ノート作成支援システムに搭載されている機能で録音された音声に対し、STD 技術を利用することで記録した電子ノートから話し手の話した言葉を精度よく検索できるようになれば、書き漏らしや聞き逃しといった問題に対応できると考えられる。そこで STD 使用者と不使用者の電子ノート見直し作業にかかる時間を比較する被験者実験を行うことで、STD の有効性評価を行った。実験の結果から、STD 使用者が不使用者に比べ平均的に、試験問題に速く正答したことを確認できている。このことから、電子ノート見直し作業において、STD が有効であるということがわかった。

[結論と今後の課題]

提案手法は、STD 性能を向上させるために非常に有効であることが、実験結果から示された。しかし、実用化のためには，検索速度がまだ遅いという問題が残っている。今後、実用化のためには、DTW を使った高速検索アルゴリズムの開発等が必要である。

論文審査結果の要旨

１．博士論文につて

（１）研究の意義

このところマルチメディアデータの生成・編集技術の普及、記憶容量の大容量化、ネットワークインフラの進歩により、音声を含んだマルチメディアコンテンツ（例：動画等）が急激に増加している。また、会議や講演などにおいて音声や映像の録音・録画を行う機会も多くなっている。これに伴い、これらの大量のコンテンツから視聴したい場面を検索したいという要求が高まっている。しかし、多くのコンテンツは動画と音声は含むが、テキスト情報を含んでいない。そのためテキストによる検索は不可能で、音声を含むデータに対しては、音声を使用して検索する方法が有効であり、音声ドキュメント検索として研究が行われてきた。この中で、音声中の検索語検出 STD(Spoken Term Detection)は、ある特定の検索語が、音声ドキュメント群中のどのドキュメントのどの位置に含まれているのかを特定する問題である。従来の研究の大部分は未知語と音声認識誤りの問題等、まだ

部分的な問題に焦点を合わせているものが多いが、この研究では、STD の検索性能向上に正面から取り組んでおり、成果を上げている。また、STD の応用についても検討している。

（２）研究の内容

この研究では、検索語検出のために、複数の音声認識システムの出力から構成された音素遷移ネットワーク PTN(Phoneme Transition Network)を利用し、編集距離ベースの DTW(Dynamic Time Warping)を使って検索性能を向上させようとしている。また、PTN の性質を利用して、誤検出を抑制する方法を考案し、それが高い検出率に繋がっている。

STD の応用例として、電子ノート作成支援システムでのノート見直し作業を対象に、実環境下での STD 技術の有効性評価を行っている。被験者実験の結果から、電子ノート見直し作業等において、提案した手法による STD が有効であることが示された。

従来から有用性が指摘されていた STD であるが、これまでは音声認識率の低さ、検索の煩雑さなどから、その実用化は難しいと思われていた。しかし、この研究により、十分実用に使える検索性能を出せる STD 手法があることが示された。この研究は音声検索分野の今後の発展に大いに寄与できるものと思われる。よって、博士論文として適当と判断する。

２．研究成果の公表・貢献等について

論文提出者は、研究内容を広く公表し、この分野の発展に貢献しようという姿勢が強い。査読付き論文はいずれも英文で、論文誌や国際会議で公表している。6 編の論文の内、4 編が筆頭著者で、残りの 2 編は、後輩の研究者を指導した応用研究等である。口頭発表は 14 件ある。2010 年に発表した論文は、日本音響学会学生優秀発表賞を受賞している。また、2011 年度、2012 年度には、NTCIR の 課題タスクに対して、検索性能第 1 位を獲得しており、他の研究機関からの目標値の一つになっている。

以上、研究成果の面からみても論文提出者は博士の資格を十分備えている。

３．博士としての素養等について

音声情報処理、データ検索等に関係する専門分野の基礎知識を十分に備えている。また、研究を推進するためのプログラミング能力、データ処理能力等も十分である。エンベデッドシステムスペシャリスト等の資格もある。専門分野のみならず、周辺の知識も豊富なので、今後の発展が期待できる。さらに、企業での実用システムの開発設計に関する経験があり、今回研究開発した手法の適切な技術移転も期待できる。

このように、論文提出者は、博士としての基本的な素養を十分備えている。

以上、論文審査、最終試験の結果等から、提出された論文は博士（工学）論文として合格と判断する。