

平成 30 年 6 月 8 日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00254

研究課題名(和文) 学習・教育支援のための多様な環境で収録された授業音声の利活用に関する研究

研究課題名(英文) Studies of utilization of lecture speeches recorded in various environments for learning/educational support

研究代表者

南條 浩輝 (NANJO, Hiroaki)

京都大学・学術情報メディアセンター・准教授

研究者番号：50388162

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：学校教育の情報化が推進されている状況のもと、多様な授業に対応できる汎用的な授業音声の再利用化(学習・教育支援)システムの基盤形成のための研究を推進した。十分な録音環境が得られない教室での授業音声をはじめとする多様な授業音声の再利用化のために、雑音や児童・生徒発話などの不要音を授業音声から除去し明瞭化する技術、多様な環境の授業音声の音声認識の研究、再利用、学習・教育支援に関する検索をはじめとする諸技術の研究、を推進し重要な知見を得た。

研究成果の概要(英文)：Fundamental technologies for utilization of lectures recorded in various environments have been investigated. To deal with diverse lecture recordings including recorded materials in a class without enough recording environment, we have investigated: 1) clarification of recorded speech (e.g. enhancement of teachers' speech, noise suppression), 2) automatic speech recognition of such diverse lecture recordings, and 3) technologies including spoken document retrieval for utilization and learning/educational support. Significant knowledge in these fields has been obtained.

研究分野：音声言語情報処理

キーワード：音声言語情報処理 学習支援 教育支援 授業音声の利活用

1. 研究開始当初の背景

日本国において、学校教育の情報化が推進されている。研究開始当初において、音声言語情報処理を用いた教育支援（例：教師発話の自動字幕化）の研究は行われていたものの、基本的に十分な録音環境を備えた大学の教室での講義を対象としていた。聴き手は基本的に大人であり、授業時の騒音は少なく、教師発話も整ったものであった。しかし、現実に行われる多くの授業、特に初等中等教育授業ではこのような録音環境・発話環境は得られない。また、高等教育の授業現場でも十分な録音環境を備えた教室は多くない。

研究代表者の南條と分担者の西崎は、この研究開始までに、初等中等教育における授業の収録と、それを用いた音声認識とアーカイブの研究（再利用化）を推進してきた。この過程を通じて、授業収録には整えられた収録環境を使用できず不要音が混合すること、および、再利用時に収録音声をそのまま利用するのは計算機にとっても人間にとっても負担が大きいことに問題意識を持ち、収録された授業音声を明瞭化する必要性を強く意識していた。

我々の研究グループは、音声言語情報処理、授業支援、音声信号処理の知見を持つメンバーで構成されたグループであり、これに関連する諸技術及び知見を有していた。しかし、現実の様々な授業に対応するための音声言語処理の取り組みは不十分であり、どのような点が不十分か、どのような技術を研究・実現すべきかについての知見についても不十分であった。

2. 研究の目的

本研究は、収録環境・授業形態が異なる多様な授業を対象とした汎用的な授業音声の再利用化（学習・教育支援）システムの基盤形成を試みたものである。すなわち、不十分な録音環境で収録された種々の授業音声を明瞭化し、様々な情報を付与したうえで利用者の求めに応じて授業音声を再利用できるようにするための基盤技術の確立を目指したものである。

3. 研究の方法

汎用的な授業音声の再利用システムの基盤形成のために、以下の研究を行った。

- (1) 授業音声における不要音の解明とその除去（明瞭化）に関する研究
- (2) 多様な環境の授業音声のための音声認識に関する研究
- (3) 再利用、学習・教育支援のための諸技術に関する研究

4. 研究成果

- (1) 授業音声における不要音の解明とその除去（明瞭化）に関する研究について

授業音声を聞きやすく、または、音声言語処理を適用しやすくするための技術について研究を推進した。既存の雑音残響抑圧技術の適用可能性と問題点を検討し、必要な課題の明確化に取り組んだ。これは多様な環境で収録された音声を標準化する試みと位置付けられる。

小学校で収録された種々の授業音声について、それに重畳されている不要音（雑音や児童発話）を抑圧するための研究を行った。ここでは最も簡易的な収録環境である、1ch マイク録音音声を対象とした。はじめに、従来法（スペクトルサブトラクション法やNMF法）の適用可能性を調査した。1ch マイク録音音声中に適用できる古典的雑音抑圧手法であるスペクトルサブトラクションを適用すると、定常雑音を抑圧できるが、聴感上非定常な雑音が目立った。そこで音声情報を残し、非音声を抑圧する手法を確立することの重要性が明らかになった。そこでNMFの改良を行った。具体的には、音声スペクトル包絡形状に特化した基底ベクトルを獲得するNMFを提案した。音声スペクトル包絡の表現に広く用いられる全極モデルおよび指数スペクトルモデルを板倉斎藤ダイバージェンス基準としてNMFに統合した。提案手法による再構成パワースペクトルには、主に音声の情報を含み、非音声の情報が抑圧される仕組みを開発できた。

話者が複数人同時に話すような状況で特定の話者の音声を強調することが可能であれば、授業収録音から教員発話を強調することも可能である。これは音源分離の課題である。従来から音源分離は、空間、時間、周波数の違いを手掛かりに音源を分離することが試みられてきた。1ch マイク録音音声では、これらの手がかりは利用困難であるため、新しい音源分離の手がかりとなる情報が必要であった。音声スペクトル包絡表現をNMFに組み込んだことにより、話者間のスペクトル傾斜の違いを手掛かりとして音源を分離できる可能性を示すことができた。子供の声道は大人と比べ短いためスペクトル傾斜が急である傾向がある。この手がかりにより少なくとも母音に関して音源分離できる可能性を示唆できた。

整備された録音環境と異なり、ICレコーダー等による簡易録音で収録される

音声波形の歪は、録音音声の再利用性を低下させる要因になっている。同 SNR 条件下でも、記録された教師発話のエネルギーが不十分であると、録音機器の有効な量子化ビットの一部分によって発話が記録され、音声の明瞭度が著しく低下する。特にアクティブゲインコントロールが有効であると、児童の発話エネルギーが主になる時間区間では、教師発話は、この傾向が顕著である。IC レコーダーを一般的な設定である 16bit 量子化で使用する場合、教師発話が、背景雑音や児童発話と比べ 40dB 以上のアドバンテージがあれば、この問題の影響が十分小さいことを示した。また、40dB 未満の場合には、IC レコーダーの量子化ビット設定値にかかわらず、教師発話の量子化ビット数の実効値が低下することを示した。

(2) 多様な環境の授業音声のための音声認識に関する研究

多様な録音環境で収録された授業音声を利活用するための基礎的技術を研究した。具体的には、授業音声データベースの授業音声に様々な音声コーデックで変換を施したり、オートエンコーダを用いて変換を施したりすることで、擬似的な種々の収録環境音声を生成し、それらの扱いを検討した。また外国語を利用する授業にも対応するために、多言語音声を扱う技術の研究も行った。さらに音声認識誤りを減らすための認識訂正誤り技術の研究も推進した。

録音機では、長時間録音のために非可逆圧縮アルゴリズムを採用していることがある。圧縮方法にはさまざまな種類があるため、録音環境に頑健な音声認識を実現するために、このような圧縮音声の扱いを研究した。

録音された音声に対して音声圧縮処理を施すことで、元の音声ファイルと内容が同じであるが音声信号が異なるデータを取得することができる。様々な音声圧縮を施して、これらを音響モデルの学習データに追加することで、録音設定に頑健な音響モデルの学習を行った。収録時の音声コーデックの違いに頑健な音声認識システムを構築できること、および追加収録なしで音響モデル学習データを増やせることを示した。

また、音声圧縮による人間の聴覚に作用しないデータを除去することで音声認識が向上できる可能性を示した。

音声収録時において 16kbps に音声圧縮するのは音声認識にとって悪影響が大きいことも分かった。

変分オートエンコーダ (VAE) を用いた授業音声認識用の音響モデルの学習データを拡張する方法を研究・開発した。VAE に入力した音声から、その潜在的な意味ベクトルを抽出し、この潜在意味ベクトルから音声データを復元することで、入力音声とは内容は同じだが話者性が異なる音声を生成した。これを音響モデルの学習データに追加することで、音響モデルの学習量を擬似的に増加させた。

拡張した学習データを用いて音響モデルを学習することで、特別な音声データを追加で用意することなく、音声認識精度を改善することに成功した。

授業では英語などの外国語の授業も行われる。このような授業音声では、日本語だけではなく外国語音声も含まれることがある。複数言語の音声認識を行うためには、各言語に対応した音声認識システムを用意する必要、かつ言語識別を行うことが必要であることから手間がかかってしまう。そこで 1 つの音声認識システムだけで、同時に複数言語に対応できるような音響モデルの開発を行った。これには、例えば、日本語と英語で類似した発音の音素を共通化することで行う。複数言語同時音声認識実験の結果から、各言語の音声認識精度を改善すると同時に複数言語に同時に対応できる音声認識が実現できることを示した。

音声認識結果を利用しようとするすると音声認識誤りに対処する必要がある。そこで、深層学習を用いた音声認識誤りの訂正方法を開発した。具体的には複数の音声認識システムを用いて 1 つの音声に対して同時に音声認識を行い、それらの結果を深層学習を用いて統合する方法を開発した。実験の結果、音素正解率で約 98% を達成した。

(3) 再利用、学習・教育支援のための諸技術に関する研究について

(検索技術)

授業音声を明瞭化し、音声認識結果を含めてアーカイブ化したものを、利用者である教師や児童・生徒が学習・教育支援に利用するための要素技術を研究した。具体的には、音声の検索技術を研究した。音声の検索技術には、内容検索 (Spoken Contents Retrieval) とキーワードの検出・検索 (STD: Spoken Term Detection) 技術があり、これらに関する研究を推進した。

検索技術の共通評価基盤の整備を推進した。NTCIR-12 において、Spoken Query and Spoken Document

Retrieval ("SpokenQuery&Doc-2") 共通評価基盤を整備することで、検索技術研究の発展に寄与した。

音声の検索技術の一つである STD(音声中のキーワードの検索)の研究を推進した。

複数の音声認識システムの出力結果を利用する方法、具体的には、機械学習(CRF、深層学習)によって正解音素系列を予測する方法を提案し、検索精度が改善したことを確認した。

また、検索語が名詞、特に固有名詞となることに着目し、検索語の前または後ろに格助詞を付加して拡張し、それを用いることで誤検出を抑制する方法を提案した。これにより検索精度の向上を確認した。

アーカイブされた授業を内容から検索するとき、その授業内容を表す用語(例えば「円周率」など)が分からない場合は、求める授業をうまく検索できない。求める用語に対して思い浮かぶこと(例えば、「円のまわりの長さとか面積とかを計算するのに使うやつ。」など)を説明してもらい、そこから用語候補を出力することで支援を行えると考え、用語抽出技術の研究を推進した。この支援は、特に子供や外国人にとっては有用と考えられる。

音声ドキュメント(音声認識結果)からの用語検索において、用語の説明文と類似度の高いパッセージ(音声ドキュメントの一部)中に用語が含まれると仮定して用語を抽出する手法を検討し、その妥当性と有望性を示した。

(学習・教育支援技術)

授業を支援するためのツールを実際に開発した。

授業を支援するためのツールとして、授業中のノートテイキングで利用できる電子ノート作成支援システムの開発を行った。電子ノートシステムのユーザインタフェース上には、黒板やスクリーンに投影されているものの画像を配置でき、その上にキーボードや手書き文字によってメモを記入できる。また、講師の音声認識結果の単語をなぞることで、なぞられた単語をメモとして利用できる。

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 6 件)

Kentaro Domoto, Takehito Utsuro, Naoki Sawada and Hiromitsu Nishizaki, "Spoken Term Detection Using Spoken Document Index Based on Keyword Corrected from Automatic Speech Recognition Result," International Journal of Signal Processing Systems, Vol.4, No.4, pp.282-288, 2016

Naoki Sawada and Hiromitsu Nishizaki, "Re-Ranking Approach of Spoken Term Detection using Conditional Random Fields-based Triphone Detection," IEICE Trans. on Information & Systems, Vol.E99-D, No.10, pp.2518-2527, 2016
Kentaro Domoto, Takehito Utsuro, Naoki Sawada and Hiromitsu Nishizaki, "Spoken Term Detection using SVM-based Classifier Trained with Pre-indexed Keywords," IEICE Trans. on Information & Systems, Vol.E99-D, No.10, pp.2528-2538, 2016

高橋徹, 赤塚 俊洋, "テレビ番組音声の自発的発話音声コーパスとしての活用可能性について", 大阪産業大学論文集, 自然科学編 128, pp.21-30, 2017

森田直樹, 南條浩輝, 山本凌紀, 馬青, "音声ドキュメントを検索対象とした用語検索", 情報処理学会論文誌(テクニカルノート), Vol.58, No.3, pp.762-767, 2017

南條浩輝, 前田翔, 吉見毅彦, "音声検索語検出のための検索語拡張法", 情報処理学会論文誌, Vol.58, No.10, pp.1735-1744, 2017

[学会発表](計 24 件)

南條浩輝, 西崎博光, "初等教育における授業音声の収集と音声認識の基礎的検討", 情報処理学会研究報告 SLP-106-2, NL-221-2, 2015

Kentaro Domoto, Takehito Utsuro, Naoki Sawada and Hiromitsu Nishizaki, "Two-Step Spoken Term Detection using SVM Classifier Trained with Pre-Indexed Keywords based on ASR Result," Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH2015), pp. 834-838, 2015

澤田直輝, 西崎博光, "音素誤りパターンに基づく音声中の検索語検出の検討", 日本音響学会研究発表会講演論文集, 1-2-8, 秋季, 2015

Hiromitsu Nishizaki, Naoki Sawada, "Score Normalization using Phoneme-based Entropy for Spoken Term Detection", Proceedings of the 7th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2015, pp. 1-7, 2015
南條浩輝, 高橋徹, 西崎博光, "初等教育授業音声の利活用のためのアーカイブ技術の基礎的検討", 日本音響学会研究発表会講演論文集, 3-P-17, 春季 2016
山田一星, 西崎博光, "講義音声認識のための言語モデル学習ユーザインタフェースの設計", 日本音響学会研究発表会講演論文集, 3-P-16, 春季 2016
Tomoyosi Akiba, Hiromitsu Nishizaki, Hiroaki Nanjo and Gareth J. F. Jones, "Overview of the NTCIR-12 SpokenQuery&Doc-2 Task", In Proceedings of NTCIR-12 Workshop Meeting, pp.167-179, 2016
Naoki Sawada and Hiromitsu Nishizaki, "Recurrent Neural Network-based Phoneme Sequence Estimation using Multiple ASR Systems' Outputs for Spoken Term Detection," Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH2016), pp.3688-3692, 2016
南條浩輝, 川口達也, "検索語の説明文による音声内容検索を利用した音声検索語検出", 情報処理学会研究報告 SLP-115-5, 2017
森田直樹, 南條浩輝, 馬青, "複数の入力説明文を用いた音声ドキュメントからの用語検索", 言語処理学会第 23 回年次大会講演論文集, P1-2, pp.100--103, 2017
南條浩輝, 川口達也, "検索語説明文を利用した音声検索語検出の検討", 日本音響学会研究発表会講演論文集, 3-5-7, 春季 2017
高橋徹, "定常雑音下での AD 変換における量子化ビット数と SNR の関係", 日本音響学会研究発表会講演論文集, 1-P-31, 春季, 2017
澤田直輝, 西崎博光, "音声の中の検索語検出のための双方向回帰結合ニューラルネットワークを用いた正解音素推定", 日本音響学会研究発表会講演論文集, 2-P-20, 春季, 2017
成田陽介, 西崎博光, "電子ノート作成支援システムの利用が及ぼす学習効果の検証", 情報処理学会第 79 回全国大会講演論文集, Vol.4, 2ZE-04, pp.881-882, 2017
南條浩輝, 西崎博光, 高橋徹, "録音環境に頑健な授業音声認識のための音声

コーデックとその活用の検討", 情報処理学会研究報告 MUS-115-54, 2017

河野秀祐, 高橋徹, "人狼プレイ録音音声の NMF ベース任意話者音声強調", 電子情報通信学会技術研究報告 SP2017-35, 2017

高橋徹, 南條浩輝, 西崎博光, "教師発話から学習した音響特徴を用いた授業音声の発話強調", 日本音響学会研究発表会講演論文集, 2-Q-30, 秋季, 2017

澤田直輝, 西崎博光, "複数認識システムの重要度を用いた正解音素推定器に基づく音声の中の検索語検出の検討", 日本音響学会研究発表会講演論文集, 2-Q-10, 秋季, 2017

西崎博光, 南條浩輝, 高橋徹, "音声コーデックと変分オートエンコーダを利用した音響モデル学習データの拡張", 日本音響学会研究発表会講演論文集, 1-R-14, 秋季, 2017

Hiromitsu Nishizaki, "Data augmentation and feature extraction using variational autoencoder for acoustic modeling," Proceedings of the 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp.1222-1227, 2017

⑲ Hiromitsu Nishizaki and Yosuke Narita, "Usability and Learning Effect Evaluations of an Electrical Note-Taking Support System with Speech Processing Technologies," Proceedings of the 25th International Conference on Computers in Education 2017, pp.31--39, 2017

⑳ 丹治遥, 小嶋和徳, 李時旭, 南條浩輝, 伊藤慶明, "音声の中の検索語検出における最上位候補を含む講演及びその類似講演優先方式", 日本音響学会研究発表会講演論文集, 2-Q-7, 春季, 2018

㉑ 高橋徹, "複数スペクトルモデルの混合表現可能な NMF に基づく音声スペクトル強調", 日本音響学会研究発表会講演論文集, 1-Q-45, 春季, 2018

㉒ 原翔吾, 西崎博光, "複数言語を同時に扱える音声認識のための音響モデリングの検討", 日本音響学会研究発表会講演論文集, 1-Q-16, 春季, 2018

〔図書〕(計 2 件)

牧野浩二, 西崎博光, 「インタフェース」 CQ 出版社 2017 年 3 月号

牧野浩二, 西崎博光, 「算数&ラズパイから始めるディープ・ラーニング」 CQ 出版社 2018 年

〔産業財産権〕

出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

南條 浩輝 (NANJO, Hiroaki)
京都大学・学術情報メディアセンター・准教授
研究者番号：50388162

(2) 研究分担者

西崎 博光 (NISHIZAKI, Hiromitsu)
山梨大学・総合研究部・准教授
研究者番号：40362082

高橋 徹 (TAKAHASHI, Toru)
大阪産業大学・デザイン工学部・准教授
研究者番号：30419494

(3) 連携研究者

()

研究者番号：

(4) 研究協力者

()