

令和 4 年 6 月 13 日現在

機関番号：14701

研究種目：基盤研究(C) (一般)

研究期間：2018～2021

課題番号：18K02862

研究課題名(和文)音響信号から学修者の活性度を測るPBL指導支援システムの開発

研究課題名(英文)Development of PBL instruction support system to measure learners' activities using acoustic signals

研究代表者

西村 竜一(NISIMURA, Ryuichi)

和歌山大学・データ・インテリジェンス教育研究部門・講師

研究者番号：00379611

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、グループワークの指導者支援システムを音情報処理の技術を活用して実現するための技術開発を行った。(1)音源分離機能を評価し、学修者が装着するウェアラブルデバイスを改良した。(2)学修者の参加姿勢等を可視化するための深層学習識別アルゴリズムを検討した。(3)グループワーク記録システムとグループワーク参加情報アノテーション支援システムを開発した。(4)深層学習声質変換の応用によるグループワーク収録発話の話者匿名化手法を検討した。新型コロナウイルスの影響を受けて、当初計画を変更し、対面実験等は見送ることとなったが、オンライン教育に有益な新たな知見を得ることができた。

研究成果の学術的意義や社会的意義

大学においてもアクティブラーニングやオンライン教育等の教育手法の導入が求められており、指導者の負担は増加傾向にある。指導者を技術的に支援するシステムの開発は必要である。現状では、実用には至っていないが、必要な要素技術の整理と開発はできており、コロナ禍後の社会や教育現場の動向にも注意しながら、研究を継続する必要があると考える。なお、本研究の成果に対して、筆頭発表の学生が学会賞(DICOM02020ヤングリサーチ賞)を受賞している。

研究成果の概要(英文)：In this study, we developed a technology to realize an instructor support system for group work by applying sound information processing technology. (1) Wearable devices worn by learners were improved by evaluating sound source separation features. (2) Deep learning identification algorithms were developed to visualize the participation attitudes of learners. (3) We developed a group work logging system and a support system for annotating group work participation information. (4) We developed a method for speaker anonymization of recorded group work speech by applying deep learning voice transformation. Due to the impact of the new coronavirus, we had to change our original plan and decided not to continue the face-to-face experiments, but we were able to obtain new knowledge that is useful for online education.

研究分野：音情報処理

キーワード：グループワーク PBL 深層学習 オンライン授業 敵対的生成ネットワーク 音源分離 話者匿名化

1. 研究開始当初の背景

本研究では、研究代表者及び研究分担者が専門とする音情報処理を応用した PBL (Project-Based Learning; 課題解決型・プロジェクト型学習) の指導者支援システムの開発を目指す。

現在、大学等の高等教育機関では、アクティブラーニング(学修者の能動的な学修への参加を取り入れた教授・学習法)の導入が進む。中でも、PBL は、学生自らがチームを組織し、課題解決のプロジェクトを遂行する実践的な教育プログラムである。教育学等の領域では、学生の主体性を育てる PBL の教育効果が報告されている。一方で、教育の現場では、さまざまな教育方法が混在し、新しい取り組みには慣れることができない指導者の間で混乱が起こっている。特に、グループワークを基本とする PBL では、活動に参加できない学修者を見つけて適切なタイミングでサポートすることや、学修者を個人評価することの難しさが指摘されている。

2. 研究の目的

本研究の目的は、PBL を実施する指導者を支援するための ICT (情報通信技術) システムの開発を目指すことである。本システムでは、PBL に参加する学修者が発生した音(活動音)を収録し、音声認識を応用した機械学習アルゴリズムによって分析する。その結果から、学修者の活性度を測り、音と同時に記録した映像とともに提示する。指導者は、その提示結果を参照することで、PBL の結果を採点する際の参考にすることができる。また、PBL を実施している最中に、活性度が低い学修者を検出する機能を提供する。指導者は、作業に参加できていない学修者を知ることができるようになり、その学修者を適切なタイミングでサポートすることが可能になる。このような支援を充実させることで、指導者の不安を抑えることができるようになる。

本研究は、PBL の教育方法を確立する一助になるものである。システムの実現によって、PBL の成績評価及び単位認定の客観性・信頼性を向上させることができる。結果として、学生による主体的な学びの教育効果の向上に寄与できると考える。以前からの一般的な教授法では、単位認定に関わる学修者の個人評価は筆記試験やレポート課題に基づいていた。一方、PBL では、グループワークを基本とするため、教員が明確に観察できるのは、成果物や成果報告書、プレゼンテーションである。これらの材料からは、個人がどの程度、課題解決に貢献できたかを把握することが難しい。例えば、作業中のアクティビティが低い個人に対して、高い成績評価を指導者が与えてしまうようなミスを防ぐ対策が必要である。このような学修者の間に生じる不公平は、提案システムによって解消することができる。

音情報処理を代表する音声認識(音声テキストに変換する技術)は、急速な実用化が進んでいる。これまでの音声認識研究では、音声以外の音は、雑音、不要な信号としてきた。雑音除去や抑圧の技術開発は、いまでも世界の音情報処理研究の主流である。しかしながら、音声以外の音にも価値がある。すべての音には発生した理由があり、その音源を知るとは、音が発生した状況を把握するための手がかりになる。本提案は、音情報処理の有力なアプリケーションの開発につながると思う。

3. 研究の方法

本研究では、事業開始後に生じた新型コロナウイルス感染症の影響を受けて、本研究が当初想定していた対面による PBL やグループワークを実施することが困難になった。特に、グループワークでは、対面で人と人が積極的に会話・議論をすることが求められるため、いわゆる三密の状態になりやすい。感染症対策を講じて、従来と同じ実施形態のグループワークでは、感染予防は難しく、リスクを伴うため、研究内容を変更する必要性が生じた。このような状況において、本研究の目的の本質を整理して、実施内容の見直しを行っている。

具体的には、実験協力者を募った実験等は見送ることになった。その一方で、コロナ禍の教育現場で急速に普及したオンライン技術を使った PBL を想定して、オンライン(同時双方向型)のビデオ会議システムを用いたグループワーク等を対象とした新たな研究開発をはじめることができた。今では、さまざまな理由で、コロナ禍後であっても、教育現場でのオンライン技術の活用は必要不可欠であると考えられている。また、オンラインでは、教員と学修者の距離が広がる傾向にあり、目的でも述べた学修者のサポートがますます重要となる。外的要因により、研究の方法は当初計画からの変更が生じることになったが、結果として、想定していなかった新しい知見を得ることに成功しており、これからの時代に適した研究開発を実行することができたと考える。

本研究の4年間(新型コロナウイルス感染症の影響により、本研究の事業期間は、当初計画の3年間から1年延長している)の主な実施項目は、次の通りである。

- (1) 音源分離機能の実装による、学修者が装着するウェアラブルデバイスの改良
- (2) 学修者の活性度、発話意図、参加姿勢等を可視化するための深層学習識別アルゴリズムの検討
- (3) 360度カメラ等を用いたグループワーク記録システムの開発と、指導者が操作するグループ

ワーク参加情報アノテーション支援システムの開発

(4) 深層学習声質変換の応用によるグループワーク収録発話の話者匿名化手法の検討

なお、このうち、(4)は、個人情報の慎重な取り扱いが求められるようになったことから追加した実施項目であり、音声から個人が特定できないようにする加工処理が該当する。これは、プライバシー保護の観点からも必要な項目であるが、本研究で実施することになったのには、もう一つの理由がある。研究を進める過程で、本質的に収録音声から個人性を排除する必要性に迫られた。一般に、グループワーク中の発話には、議論の高まり等により、その話者の個人性が強く含まれるようになる。収録データに対するアノテーションや可視化の際には、結果が、その個人性に強く依存することになった。感情表現やニュアンスは保持しながらも、個人性を抑える必要がある場面があったことが(4)を実施した理由である。

4. 研究成果

- (1) シングルボード小型コンピュータ (Raspberry Pi) を用いて、学修者が装着するウェアラブルデバイスの開発を行った。本研究では、音源分離機能を実現するために、同デバイスに対して、複数のマイクロフォンの実装を検討した。また、学修者が服の上に直接装着するためのハードウェアのデザインを検討した。

音源分離機能に関しては、収録信号に含まれる音源が異なる信号 (混合音) のブラインド分離手法の適用を試みた。音による状況察知や周辺環境理解のためには、収録信号から環境音等を抽出する必要がある。しかし、通常、日常生活下での収録信号には、人間の発話が含まれており、収録信号から音声以外の信号を抽出する処理が必要となる。本研究では、従来は雑音や不要な信号とされてきた音声以外の信号の可視化が必要であるため、人間の発話信号を抑圧し、環境音を抽出するための手法を検討した。実験では、独立成分分析 (ICA)・独立ベクトル分析 (IVA)・独立ベクトル分析 + IP 法による最適化 (IVA + IP) の各手法を比較した。実環境で収録した信号を分離し、SNR と処理時間を評価した。実験結果を図 1 及び図 2 に示す。Raspberry Pi を用いて、混合音を収録し、分離処理を実行できることは確認したが、実用には、処理速度の改良が必要であることを確認した。

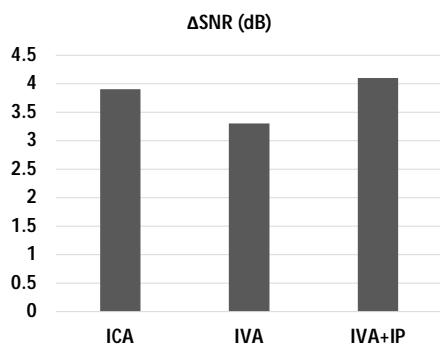


図 1 分離手法比較結果 (SNR)

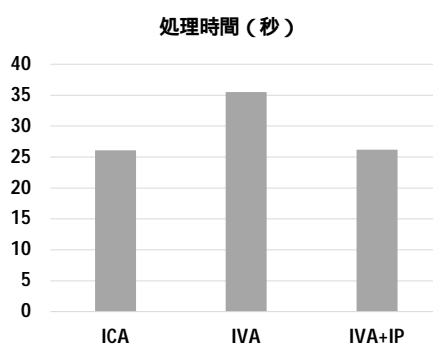


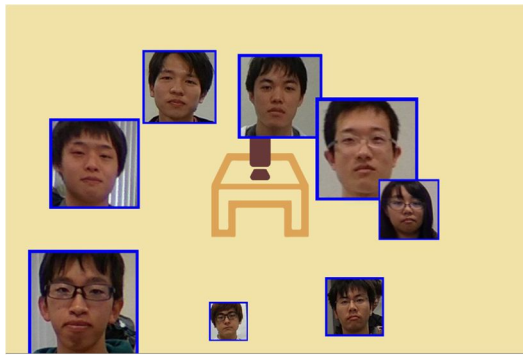
図 2 分離手法比較結果 (処理時間)

- (2) 敵対的生成ネットワーク (GAN) を用いた音響信号の識別アルゴリズムの検討を行った。その際、比較的量が少ない訓練データを用いた統計モデルの学習手法を検討した。音響信号から抽出した特徴量を訓練データに用いてモデルを学習する。そのモデルから出力した疑似的な生成特徴量を実特徴量に加える。学習モデルを用いた深層学習型ニューラルネットワークによって自動推定する。以上の方法によって推定正解率の向上を確認した。

また、学修者の活性化抽出アルゴリズムの一部として、CNN Autoencoder による音分類方式の検討を進めた。特に、収録環境の異なる 2 つのコーパスを用いた実験により、単に両者を混ぜたコーパスからモデル学習をするのではなく、Autoencoder による特徴量抽出を介することで、より高い分類性能を持つモデルを学習できることを示した [1]。画像と音声を含む会議データを収録し、その画像としての特徴と音声としての特徴を併用することで、会議参加者の発話意図の推定や参加姿勢の良し悪しを推定する検討を行った。

学修者の活性化を可視化するために、機械学習のアプローチを用いて、音声と映像を併用した議論への関与姿勢の分類タスクを検証した [2]。音だけでなく画像を併用することで、識別率の改善の傾向を確認した [3]。しかし、画像を用いたわずかな性能向上の代償として、プライバシーの問題やデータ量の問題などが発生することがわかった。

- (3) グループワークを 360 度カメラによって撮影し、記録動画と音声を用いた学修者の評価方法を検討した。PBL の記録及び画像データの抽出に 360 度画像を利用することで、PBL のような複雑な状況下においても、音響信号と画像を複合的に利用することが可能になった。記録データに学修者の参加度や貢献度等のメタ情報を付与するためのアノテーション支援



タッチ操作で記録できるグループワーク情報（例）





-  **シングルタップ：**
顔画像の位置を移動して参加者の立ち位置を記録
-  **ダブルタップ：**
相槌や短い発言の回数を記録
-  **長押し：**
参加者の興味深い発言の回数を記録
-  **スケーリング(拡大・縮小)：**
参加者の貢献度を記録
(貢献が大きい人ほど顔を大きく表示)

図3 システム利用例（タッチ操作によるグループワーク情報記録システム）

システムとして、グループワーク参加学生個々の参加情報記録システムを開発した。図3にシステムの利用例を示す。本システムでは、指導者がタブレット端末に提示された学生の顔画像をタッチ操作することで、その学生のグループワークへの参加状況を記録することが可能である。本研究では、発言や相槌の回数、グループワークに対する貢献度を記録できるインタフェースを設計した。システムで提示する顔画像は、360度全方位の撮影が可能なビデオカメラで撮影したパノラマ画像から抽出する。実験では、協力者に依頼し、実際のグループワークを撮影した。

2年目以降は、コロナ禍で、対面での実施が不可能となったため、オンライン（同時双方向型）のビデオ会議システムを用いたグループワーク実験を行い、同システムの改良・評価を進めた。提案システムでは、指導者がタブレット端末に提示された学生の顔画像をタッチ操作することで、その学生のグループワークへの参加状況を記録することが可能である。改良システムでは、(a) グループ内での役割や要注意行動等を定義し、それらを記録できるマーク機能を実装した。(b) 同時並行的に実施される複数のグループワークの参加情報を一括して記録できるタブ機能を実装した。

- (4) 発話データの利用を円滑にするため、深層学習を用いて音声の話者を匿名化する声質変換手法の開発を行った[4]。開発したプロトタイプシステムは、(a) マイクロホンアレイを用いた対話音声の収録 (b) 各話者の音声に含まれる特徴量の抽出 (c) 深層学習による音声の声質変換を実現する。

パラメータの変換に用いる敵対的生成ネットワーク（GAN）のアルゴリズムを検討するとともに、パラメータを構成する音響特徴量の検討を行った[5, 6]。特に、話者の特定に寄与する音響特徴量を実験で検証した。先行研究ではパラメータ変換の適用外とされてきた非周期性指標が話者特定に与える影響を確認し、スペクトル包絡と比較して影響が少ないことを確認した。未知話者音声を入力に用いた場合に不自然な変換結果が得られる問題について、特徴量の変換式を導入した。

<引用文献>

- [1] 松原拓未, 原直, 阿部匡伸, CNN Autoencoder から抽出したボトルネック特徴量を用いた環境音分類, マルチメディア, 分散, 協調とモバイルシンポジウム(DICOMO)2019論文集, pp.339-346, 2019.
- [2] 金岡翼, 上原佑太郎, 原直, 阿部匡伸, 映像と音声を用いた議論への関与姿勢や肯定的・否定的態度の推定方式の検討, マルチメディア, 分散, 協調とモバイルシンポジウム(DICOMO)2020論文集, pp.1422-1429, 2020.
- [3] 金岡翼, 原直, 阿部匡伸, 音声と映像から議論への関与姿勢を推定するための特徴量の検討, 電子情報通信学会技術研究報告, vol.121, no.401, L01S2021-49, pp.57-62, 2022.
- [4] 近藤伊佐直, 西村竜一, CycleGAN-VC 単一モデルによる音声匿名化変換の検討, 第19回情報科学技術フォーラム(FIT2020)講演論文集, vol.2, pp.17-20, 2020.
- [5] 近藤伊佐直, 西村竜一, CycleGANを用いた音声匿名加工に必要なWORLD音響特徴の構成, 情報処理学会第83回全国大会講演論文集, pp.237-238, 2021.
- [6] 近藤伊佐直, 西村竜一, 音声匿名化変換に用いる音響特徴構成の評価, 日本音響学会2021年秋季研究発表会講演論文集, pp.1015-1016, 2021.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計14件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 近藤伊佐直, 西村竜一
2. 発表標題 音声匿名化変換に用いる音響特徴構成の評価
3. 学会等名 日本音響学会2021年秋季研究発表会
4. 発表年 2021年

1. 発表者名 池川心, 西村竜一
2. 発表標題 音声認識を用いた講義音声の早口分析
3. 学会等名 情報処理学会第84回全国大会
4. 発表年 2022年

1. 発表者名 奥本佑哉, 西村竜一
2. 発表標題 クラウドソーシング収集発話を用いた深層学習若年話者判別
3. 学会等名 情報処理学会第84回全国大会
4. 発表年 2022年

1. 発表者名 金岡翼, 原直, 阿部匡伸
2. 発表標題 音声と映像から議論への関与姿勢を推定するための特徴量の検討
3. 学会等名 電子情報通信学会ライフインテリジェンスとオフィス情報システム研究会
4. 発表年 2022年

1. 発表者名 近藤伊佐直, 西村竜一
2. 発表標題 CycleGAN-VC 単一モデルによる音声匿名化変換の検討
3. 学会等名 第19回情報科学技術フォーラム (FIT2020)
4. 発表年 2020年

1. 発表者名 近藤伊佐直, 西村竜一
2. 発表標題 Cycle GANを用いた音声匿名加工に必要なWORLD音響特徴の構成
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 西村竜一, 井ノ上彩海, 曾我真人, 腰塚尚志, 細川瑞彦
2. 発表標題 インタラクティブなコミュニケーションを可能とした 分散協働型の遠隔講義
3. 学会等名 大学eラーニング協議会 / 日本リメディアル教育学会 合同フォーラム 2020
4. 発表年 2021年

1. 発表者名 金岡翼, 上原佑太郎, 原直, 阿部匡伸
2. 発表標題 映像と音声を用いた議論への関与姿勢や肯定的・否定的態度の推定方式の検討
3. 学会等名 マルチメディア, 分散, 協調とモバイルシンポジウム (DICOM02020)
4. 発表年 2020年

1. 発表者名 西村竜一, 曾我真人, 石塚互, 腰塚尚志, 細川瑞彦
2. 発表標題 クラウド型会議サービスを用いた遠隔講義を活用する大人数授業の取り組み
3. 学会等名 大学eラーニング協議会 / 日本リメディアル教育学会 合同フォーラム 2019
4. 発表年 2020年

1. 発表者名 松原拓未, 原直, 阿部匡伸
2. 発表標題 CNN Autoencoderから抽出したボトルネック特徴量を用いた環境音分類
3. 学会等名 マルチメディア, 分散, 協調とモバイル (DICOM2019) シンポジウム
4. 発表年 2019年

1. 発表者名 呉セン陽, 朝田興平, 原直, 阿部匡伸
2. 発表標題 i-vectorに基づく賑わい音の推定方式の検討
3. 学会等名 情報処理学会 音学シンポジウム
4. 発表年 2019年

1. 発表者名 納庄貴大, 西村竜一, 入野俊夫
2. 発表標題 敵対的生成ネットワークを用いた楽曲の自動コード推定法の検討
3. 学会等名 情報処理学会第120回音楽情報科学研究会
4. 発表年 2018年

1. 発表者名 曾我真人, 西村竜一, 西川明宏, 尾久土正己
2. 発表標題 アクティブラーニングをとりいれた英語学習の実践
3. 学会等名 大学eラーニング協議会 総会・フォーラム2018
4. 発表年 2019年

1. 発表者名 原直, 阿部匡伸
2. 発表標題 クラウドソーシングによる環境音マップ構築のための主観的な騒々しさ推定方式の検討
3. 学会等名 第17回情報科学技術フォーラム (FIT2018)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	原 直 (HARA Sunao) (50402467)	岡山大学・ヘルスシステム統合科学学域・助教 (15301)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------