

平成 30 年 6 月 5 日現在

機関番号：14701

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K01069

研究課題名(和文) PBLの個人評価を目的とする活動音の可視化技術

研究課題名(英文) Development of activity sound visualization method for personal evaluation of PBL

研究代表者

西村 竜一 (NISIMURA, Ryuichi)

和歌山大学・システム工学部・助教

研究者番号：00379611

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)：本研究では、PBL (Project-Based Learning; プロジェクト型学習) に参加する学修者の評価を支援する手法を音情報の可視化を通じて実現するための技術開発を行った。対話音声を用いたグループワークにおけるアクティビティの自動検出法を検討した。その結果を応用し、ウェアラブル音声収録端末を用いたグループワーク状況提示システムを試作した。加えて、深層学習型ニューラルネットワークに基づく音源情報可視化手法を検討した。

研究成果の概要(英文)：In this study, we have developed methods for realizing to support evaluations of students participating in PBL (Project-Based Learning) on the basis of visualization technologies of sound information. A method of detection of activated communication in group work from dialogue voice was examined. We developed the prototype system for presenting a whole condition of a group work using wearable voice recording terminals. In addition, sound source information visualization methods based on deep learning neural networks have been investigated.

研究分野：音情報処理

キーワード：PBL 音声情報処理 笑い声 環境音 ウェアラブル端末 ニューラルネットワーク 可視化

1. 研究開始当初の背景

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。そのなかでも、PBL(Project-Based Learning; プロジェクト型学習)は、学生らがチームを組み、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。PBLは、学生の主体性を育み、教育工学等の学術分野においても効果の高さの報告が続いている。

しかしながら、PBLの指導者からは、学修者の評価の難しさに対する指摘がなされている。グループワークを基本とするため、教員・指導者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。

2. 研究の目的

本研究の目的は、PBLに参加する学修者の評価を支援する手法を音情報の可視化を通じて実現することである。具体的には、PBLに参加している学修者が発生したあらゆる音を収録し、音声認識を応用した機械学習アルゴリズムによって可視化する。加えて、可視化した情報を学修者個人の定量的な評価材料として利用するための方法を検討する。

ただし、コンピュータによる自動化に基づいた学修者評価を教育の現場に持ち込むことには慎重な検討が必要である。よって、本研究では、要素技術の開発及びデータの取得法の確立を重点的な実施事項としている。

中長期的な目標としては、本研究の成果を拡充し、PBLの指導者をサポートするためのシステムとして実用化を目指す。

3. 研究の方法

(1)対話音声を用いたグループワークにおけるアクティビティ自動検出法の検討

グループワークの指導・採点を行なう者(評定者)がグループワークの中で生じるインタラクションの一つとして注視することができ、指導や採点の参考にできる指標を盛り上がりとして、これを自動検出するための技術開発を行った。

具体的には、盛り上がりの区間を音響信号から検出することを目指し、グループワークデータの収集、評定者が盛り上がりの区間を定義するためのアノテーション方法、収録音声から特徴ベクトルの抽出方法、機械学習による自動検出の検討を行った。

グループワークデータを収集する際には、協力者4名によるグループワークを音声及び動画で記録した。練習10分、本題20分、まとめ10分の計40分にわたるグループワーク課題を協力者(参加者)に与え、図1に示す実験環境において、実際にグループワークを実施して記録を行った。

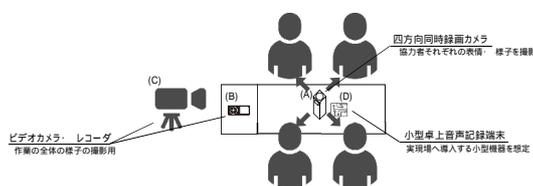


図1 グループワークの収録環境



図2 マルチタッチパネルを用いて実装したアノテーション支援システム

では、収録したグループワークデータに対して、評定者が考える盛り上がりの区間を定義するため、時間軸上の注釈情報を付与するための手法を検討した。その過程では、アノテーション作業の効率化を支援するシステムが必要であった。そこで、マルチポイントに対応したタッチパネルを利用したユーザインタフェースを有する支援システムを開発した(図2)。実装の結果、参加者4人の発話に対し、同時かつ効率的に、盛り上がりを示す区間ラベルを付与できるようになった。

評定者2名がアノテーションを行った結果、評定者Aは収録データ全区間の8.1%、評定者Bは33.8%を盛り上がりの区間とした。評定者A及びBの間には、注釈を付与する頻度に違いはあるが、A、Bともに積極性の低い同じ協力者に対しては、盛り上がりの区間ラベルを少なく付与していることがわかった。

グループワークの収録音声に対する音声特徴量の抽出と機械学習を用いた自動検出を試みた。F0(基本周波数)やパワーなどに由来する音声特徴を合計384次元抽出してベクトルを構成した。そして、の支援システムで定義した盛り上がりの区間ラベルと合わせて、機械学習に用いるデータセットとした。実験では、二値分類器であるサポートベクターマシンで識別器を構築し、性能を検証した。

(2) ウェアラブル音声収録端末を用いたグループワーク状況提示システムの開発

グループワークのとき発生する音響信号を収録する際、収録機器が本来の作業や対話を妨げることが無いようにする必要がある。加えて、グループワーク中の個人の発話を録音することを目的として、本研究では、ウェアラブルで使用できる音声収録端末を開発した。図3に示すように、同端末は、小型ボードコンピュータ Raspberry Pi Zero Wを用

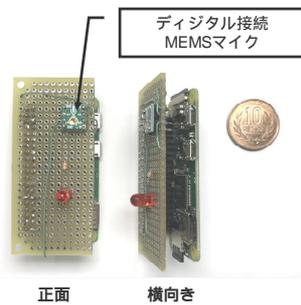


図3 ウェアラブル収録端末（プロトタイプ）

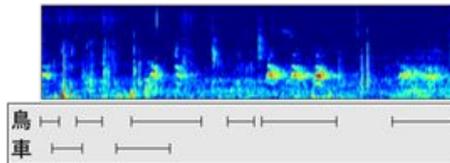


図4 環境音信号と音源ラベル

いて実装されており、デジタル接続の MEMS (Micro Electro Mechanical Systems) マイクを搭載する。

本研究で開発したグループワーク状況提示システムでは、上記のウェアラブル端末で収録した信号を自動分類し、その結果を無線 LAN 経由でサーバに送信する。また、サーバ側に集約された情報を指導者にわかりやすく提示するためのインタフェースを有する。指導者は、提示された情報を確認することで、グループワークに参加できていない学修者等に気付き、適切な指導を与えることができると考える。

開発システムは、Python を実装言語とした。ウェアラブル端末とサーバ間の通信には、HTTP (Hypertext Transfer Protocol) リクエストの protocols を利用し、通信負荷を抑えた実装とした。交換する情報には、アクティビティを知るためのファクタとなる学修者の話し声・笑い声の検出回数・検出時間が含まれる。

(3) 深層学習型ニューラルネットワークに基づく音源情報可視化手法の検討

同一空間内に多数存在する音源情報を可視化するため、収録したパラメタから音源情報に変換するための認識器およびその特徴量について検討を加えた。深層学習型ニューラルネットワークに基づくアルゴリズムである LSTM (Long Short-Term Memory)、CNN (Convolutional Neural Network) 及び CNN-LSTM を比較した。

特徴量とネットワーク構成の異なる各手法について、10-fold cross validation によって評価した。入力信号と出力すべき音源ラベルの例を図4に示す。CNN の入力とする特徴量には 25 次元のメルフィルタバンクスペクトログラムを用いた。それに加えて、動的特徴である特徴量を追加で入力に用いる場合と用いない場合での比較を行った。

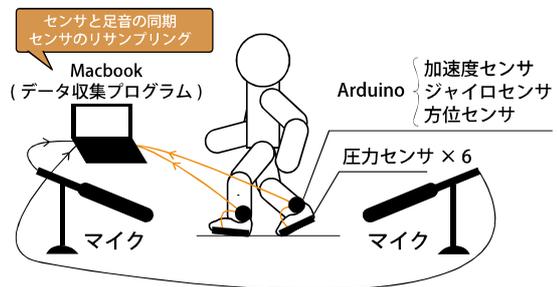


図5 歩行データ収集システム概略



図6 足に装着するセンサ

また、ネットワーク構成には CNN (2 層の畳み込み+プーリング層 2 層の全結合層) と CNN-LSTM (2 層の畳み込み+プーリング層 2 層の LSTM 層 1 層の全結合層) の 2 種類を用いて、ネットワークの構成の違いによる比較も行った。

(4) 人間の歩行動作のセンシングデータを入力にした足音信号の合成手法の提案

本研究では、複数のフレームのセンサデータを連結して変換モデルを学習することで、足音が発生する直前の歩行動作を考慮することができる足音信号合成手法を提案した。

学習に用いるデータセットとして、時間同期した歩行動作のセンシングデータと足音を収集した。そのために、加速度、ジャイロ、地磁気や圧力のセンシングデータと時間同期した足音の音響信号の収録システムを開発した(図5, 図6)。本学で学生プロジェクト活動に参加する学生 7 名の協力を得て、7 時間分の歩行データ(音響信号、センシングデータ)を収集した。

4. 研究成果

(1) で構築した盛り上がり区間の自動検出器を用いて、精度を評価した実験の結果を図7に示す。この実験では、従来は固定長であった音声特徴の分析単位を 1 秒から 10 秒まで拡張し、10-fold cross validation で精度を確認した。実験の結果、分析単位 1 秒のときは、0.54 だった F-measure が、分析単位を長くすることで 0.7 程度まで向上することを確認した。

比較のため、分析区間長を 1 秒に固定し、盛り上がりの区間ラベルを時間軸上でずらしたデータセットを作成して、前述と同条件

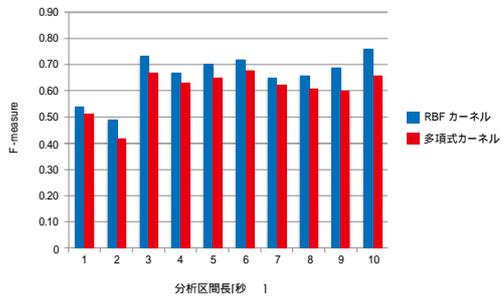


図7 長時間区間分析導入後の平均F値

の実験を行なった。その結果からは、分析単位に長時間区間を導入した提案法と同程度の性能を確認することはできなかった。

実際の盛り上がり区間と、評定者によって付与された区間ラベルの間には、時間上のズレが生じることが予想される。このズレは評定者ごとに異なるため、一意に決定して処理することは困難である。長時間区間分析を導入することで、そのズレを吸収しながら、自動検出の精度向上を得ることができ、提案法の有効性を示すことができた。

(2) グループワーク状況提示システムを用いた実験では、グループワークの記録映像とシステムが録音した音声のログを比較した。

各協力者のグループワークに対する積極性とシステムが収集した音声の検出回数を比較した。グループワークに対する積極性の評価は、3分ごとに行い、グループワークに参加できているか指導の必要性を感じるかの基準で各協力者に（マル）または×（バツ）の2段階の情報を評定者が付与した。図8、図9に2つのグループの評定の結果とシステムが記録した音声の検出回数を示す。

この事例では、図8のグループ1は指導の必要はないと判断され、図9のグループ2は指導が必要と判断された。システムが記録したログでは、グループ1全体での音声検出回数は平均294回、グループ2の音声検出回数は平均209回であり、音声の検出回数に差が生じていた。さらに、指導者にサポートが必要と評定されることが多かったグループ2の協力者Gの検出回数は最も少なく163回であった。この結果から、提案システムで記録した情報からグループ全体の状況、グループ内での個人の状況について知ることができることを確認した。

記録した検出音声の中に装着者本人の音声のみが含まれる割合は、60%から90%と協力者によって大きく差があった。これは声が大きい他の協力者の音声を端末が検出してしまったことが原因である。

(3) 手法ごとの評価結果を表1に示す。CNNに特徴量を用いることでF-score、Error Rateともに性能が向上している。さらに、CNNの代わりにCNN-LSTMを用いることでCNN+

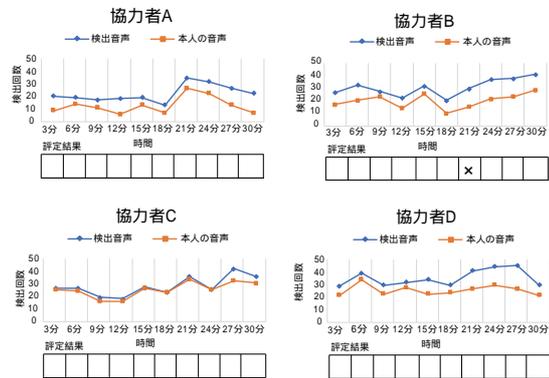


図8 音声の検出回数と積極性の評定結果（グループ1）

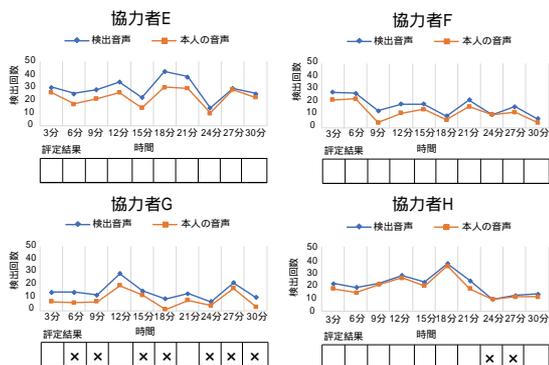


図9 音声の検出回数と積極性の評定結果（グループ2）

よりも良い性能が得られており、CNN-LSTMを用いた場合、短時間分類において最も優秀な性能が得られた。よって、総合的な結果としては、CNNにLSTMと特徴量の両方を用いることが有効であることが確認できた。

各手法のラベルごとのF-scoreを表2に示す。「人」「信号機」では、LSTMを用いた場合、そうでない場合に比べF-scoreが上昇しており（人の場合、CNN CNN-LSTM:0.469 0.577、CNN+ CNN-LSTM+ :0.493 0.576）これらは連続的な長い音もしくは断続的に同じ音が続くためLSTMが有効に働いたと考えられる。「鳥」「信号機」「タップ音」は、特徴量を用いた場合、そうでない場合に比べF-scoreが大きく上昇しており（鳥の場合、CNN CNN+ :0.274 0.328、CNN-LSTM CNN-LSTM+ :0.284 0.355）これらは、離散的で短い音であるため、時間変化が確認できる特徴量が有効に働いたと考えられる。

(4) 収集したデータセットを用いて、足音の音響信号合成のための予備実験を行なった。音響特徴量の比較、入力データの抽象化による精度への影響などを実験で比較をした。その結果、オクターブフィルタバンクを用いた入力データの抽象化や、フレーム連結による精度向上を確認した。また、深層学習型ニューラルネットワークの出力を抽象化し、次元数を減らした音響特徴量とした際に、教師データとのケプストラム距離は大きく下がり、

元の波形と近い波形を合成できることを確認した。

表1 分類精度 (10-fold CV)

Method	F-score	Error Rate	Precision	Recall
CNN	0.619	1.003	0.645	0.595
CNN+	0.639	0.963	0.657	0.622
CNN-LSTM	0.645	0.951	0.660	0.630
CNN-LSTM+	0.650	0.943	0.661	0.639

表2 音源ラベルごとの F-score

	CNN	CNN+	CNN-LSTM	CNN-LSTM+
人	0.469	0.493	0.577	0.576
鳥	0.274	0.328	0.284	0.355
虫	0.470	0.542	0.454	0.434
車	0.733	0.738	0.745	0.735
風	0.652	0.679	0.695	0.690
バイク	0.238	0.244	0.248	0.219
踏切	0.652	0.738	0.761	0.616
電車	0.321	0.420	0.400	0.418
サイレン	0.281	0.302	0.183	0.195
信号機	0.483	0.540	0.554	0.593
動物	0.318	0.312	0.278	0.227
音楽	0.315	0.326	0.312	0.302
タップ音	0.637	0.705	0.637	0.681
足音	0.179	0.197	0.292	0.252
ノイズ	0.699	0.722	0.747	0.757

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)

[学会発表](計17件)

西村竜一, 武田悠, 三上菜穂, 入野俊夫, 小型ボードコンピュータ Raspberry Pi を用いた笑い声の収集, 日本音響学会 2018 年春季研究発表会, 2018.

Sunao Hara, Asako Hatakeyama, Shota Kobayashi, Masanobu Abe, Sound sensing using smartphones as a crowdsourcing approach, APSIPA Annual Summit and Conference, 2017.

三上菜穂, 西村竜一, 入野俊夫, グループワーク対話の分析を通じた盛り上げりの定量化の検討, 日本音響学会 2017 年

秋季研究発表会, 2017.

小林将大, 原直, 阿部匡伸, DNN による人間の感覚を考慮した騒々しさ推定方式に基づく騒音マップの作成, 日本音響学会 2017 年秋季研究発表会, 2017.

Shota Kobayashi, Sunao Hara, Masanobu Abe, Prediction of subjective assessments for a noise map using deep neural networks, UbiComp/ISWC, 2017.

鳥羽隼司, 原直, 阿部匡伸, スマートフォンで収録した環境音データベースを用いた CNN による環境音分類, 日本音響学会 2017 年春季研究発表会, 2017.

吉田起, 西村竜一, 入野俊夫, 歩行動作のセンシングデータを入力とした足音合成, 情報処理学会インタラクシオン 2017, 2017.

遠山智明, 西村竜一, 入野俊夫, ユーザ訂正情報を用いた音声認識 API のカスタマイズ手法, 日本音響学会 2016 年秋季研究発表会, 2016.

原直, 田中智康, 阿部匡伸, 重複音を含む環境音データベースにおける環境音検出のための特徴量の基本検討, 日本音響学会 2016 年秋季研究発表会, 2016.

鳥羽隼司, 原直, 阿部匡伸, RNN による実環境データからのマルチ音響イベント検出, 日本音響学会 2016 年秋季研究発表会, 2016.

三上菜穂, 西村竜一, 入野俊夫, RaspberryPi を用いた笑い声検知システムの提案, 日本音響学会 2015 年秋季研究発表会, 2015.

[図書](計1件)

西村竜一, “Web 音声インタフェース”, 日本音響学会編, コロナ社, 音響キーワードブック, pp. 458-459, 2016.

[その他]

ホームページ等
<https://w3voice.jp/>

6. 研究組織

(1) 研究代表者

西村 竜一 (NISIMURA, Ryuichi)
 和歌山大学・システム工学部・助教
 研究者番号: 00379611

(2) 研究分担者

原 直 (HARA, Sunao)
 岡山大学・自然科学研究科・助教
 研究者番号: 50402467