

6ZA-05

協調学習のマルチモーダル分析に向けた IoT センサの時刻同期精度に関する基礎評価

山口 大地[†] 山口 隼平[†] 大島 律子[‡] 大島 純[‡] 椎名 亮太^{‡†}
 藤橋 卓也[†] 猿渡 俊介[†] 渡辺 尚[†]
[†] 大阪大学大学院情報科学研究科 [‡] 静岡大学大学院総合科学技術研究所
^{‡†} 日本電信電話株式会社 NTT アクセスサービスシステム研究所

1 はじめに

協調学習とは、他者と協調しながら創造的な問題解決に取り組む学習形態である。学習科学の分野でなされている協調学習のマルチモーダル分析をサポートする IoT システムとして、文献 [1, 2] ではセンサ間で誤差 $\pm 30 \mu\text{s}$ 以内の時刻同期を取りながら学習者の活動をセンシングする名刺型センサとセンサデータを統合して学習活動をマルチモーダルに可視化する学習分析アルゴリズムを提案している。本稿では、SRP を用いた協調学習のマルチモーダル分析において名刺型センサに求められる時刻同期精度を検証するために、学習分析アルゴリズムの 1 つである話者検出技術の精度と名刺型センサの時刻同期精度との関係を探る。

2 協調学習の場に対して定量分析ツールを導入するときの課題

協調学習において個人に合った学習活動を提供するためには個々の学習の様子を定性分析する必要があることから分析に長けた人物の存在や分析自体の労力を要する。協調学習の場に対して定量分析ツールを導入することで従来の定性分析に即時性を持たせることができる。導入する定量分析ツールは以下の 3 つの要件を満たす必要がある。

1 つ目は、「学習環境に簡単に導入できること」である。学習環境に簡単に導入できることは定量分析ツールを広く展開するために必要である。学習環境に対して設置するコストが高い技術を利用することは現実的ではない。本稿では、学習環境に簡単に導入することを目的として学習者が胸元に装着する名刺型センサの導入を考えている。

2 つ目は、「学習活動を自動で可視化すること」である。学習活動の自動可視化は、協調学習における定性分析と定量分析を接続するうえで必要である。理想的には、現在の学習科学の研究者が行っている定性分析と同じ分析結果を全て自動抽出できることが望ましい。しかしながら、現状では機械と人間が取得・処理できる情報には大きな乖離があるため、定性分析で行っていることを定量分析で完全に置き換えるのは難しい。本稿では、定性分析において分析者が注目している学習者同士の対面、学習フェーズ、学習者の発話、学習者の活動量に対象を絞る。

3 つ目は、「データの時刻同期が取れていること」である。データの時刻同期は、各個人ではなく人と人との協調を対象とするうえで必要となる。本稿では、学習者や学習環境などの様々な対象にセンサが設置されており、かつ多様なセンサからのデータの整合性を確認しながら分析する。時刻同期が取れていないデータ同士を結合して分析した場合、実際に起こった事象の時系列的な順序が崩れるため分析結果が意味をなさないものになってしまう。ユーザー同士の協調を抽出するためのセンサとして、日立のビジネス顕微鏡 [4] やマサチューセッツ工科大学 (MIT) の Rhythm Badge [5] が存在するが、各センサ間で高精度な時刻同期を実現する仕組みは提供されていない。本稿では、学習活動を可視化する手法の 1 つとして実現されている話者検出技術を対象として精度を担保するうえで求められるセンサデータの時刻同期精度を評価する。

3 定量分析を行う IoT システム

2 節で述べた要件を同時に実現するための手法として、文献 [1, 2] では Sensor-based Regulation Profiler (SRP) を提案している。SRP は、センサ間で時刻同期を取りながら学習者の活動をセンシングする名刺型センサとセンサデータを統合して学習活動をマルチモーダルに可視化する学習分析アルゴリズムを提案している。SRP を用いることによって、人手による定性分析に起因して生じていた学習分析に要する労力を削減する。

3.1 名刺型センサ

学習者からデータを収集するための名刺型センサとして、文献 [1] では Sensor-based Regulation Profiler Badge (SRP Badge) を提案している。図 1 (a), (b) に SRP Badge の概観、ブロック図を示す。SRP Badge は学習者の胸元に装着することを想定した名刺型のセンサとなっている。SRP Badge は、電源制御部・CPU センサ部・無線部からなる。電源制御部では、SRP Badge を動かすためにリチウムイオン電池が搭載されている。リチウムイオン電池から電源スイッチと Micro Controller Unit (MCU) に電源が供給されている。SRP Badge の連続稼働時間は 24 時間である。

CPU センサ部では、STMicroelectronics 社の STM32L476RGT6 を MCU として、ANALOG DEVICES 社の ADXL362 の加速度センサ、OptoSupply 社の OS15LAS1C1A の赤外線 LED、KODENSHI CORP. の PIC79603 の赤外線受光器、TDK 社の INMP510 のアナログマイクروفオンが搭載されている。加速度センサは 100 Hz 3 軸 12 bit、音圧センサは 100 Hz 12 bit でサンプリングする。記憶装置として、Hirose Electric 社の DM3AT-SF-PEJM5 の microSD カードコネクタを具備している。加速度データ・赤外線データ・音圧データは microSD カードによって保存できる。FAT32 形式でフォーマットした 32 GB 以下の microSD カードを SRP Badge のスロットに差し込むことでセンサ値のバイナリデータが保存される。

無線部では、無線同期モジュールが搭載されている。無線同期モジュールは、SRP Badge の親機から 10 ms ごとに送信される同期信号を他の SRP Badge に転送して SRP Badge 間の時刻同期を実現する。図 1 (b) に示した RF モジュールである Texus Instruments 社の CC2650 上では、SRP Badge 間の高精度な時刻同期を実現する同時送信型フラッドイングプラットフォーム「UNISONet (別名 Choco)」が動作している。同時送信型フラッドイングでは、任意のセンサノードが近隣のセンサノードに対して時刻同期パケットを転送することで宛先センサノードに時刻同期パケットを伝搬する。センサノードは新たに時刻同期パケットを受信すると即座に他のセンサノードに対して時刻同期パケットを転送する。フラッドイングによって各センサノードが時刻同期パケットの受信・転送を繰り返すことで、時刻同期パケットをセンサノード全体に高速で伝搬することができる。

3.2 話者検出アルゴリズム

3.1 節で述べた名刺型センサで収集した 100 Hz の音圧データを用いて高精度に話者を検出する話者検出アルゴリズムを提案している [3]。図 2 に話者検出アルゴリズムの全体像を示す。提案アルゴリズムは、非話者の名刺型センサにノイズとして乗ずる話者の声を検出して除去することによって種々の条件下で高精度な話者検出を実現している。本アルゴリズムは、1) データの前処理、2) 発話区間検出、3) 発話者特定の 3 ステップで話者を検出する。

1) データの前処理: データの前処理では、名刺型センサを用いて収集した音圧データの補正および立ち上がり検知を行う。収集した音圧データの最小値を基準としてユーザごとにゼロ点補正を行う。補正データに対して、ウィンドウサイズ 2s、スライド幅 1s のスライディングウィンドウを適用しながら、

- ウィンドウ内において音圧検出の有無を判定する閾値 (閾値 1)
- 音圧検出ウィンドウ内で音圧の検出・非検出を 2 値化する閾値 (閾値 2)

を用いて音圧データの立ち上がりを検知する。閾値 1 をユーザごとに

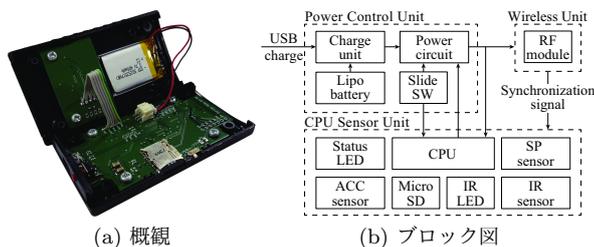


図 1: 名刺型センサ

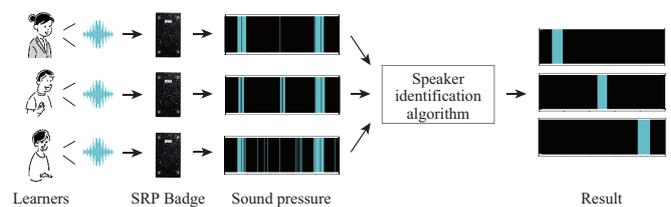


図 2: 話者検出アルゴリズム

表 1: 各時刻同期誤差における話者検出精度の混同行列

	0 ms		0.1 ms		0.3 ms		1 ms		10 ms		50 ms	
	P	N	P	N	P	N	P	N	P	N	P	N
T	440	10	440	10	440	10	440	10	440	10	439	10
F	0	450	0	450	81	369	98	352	192	258	204	247

定めて各ウィンドウ内で閾値 1 を超える音圧が検出されるか判定する。ウィンドウ内で閾値 1 を超える音圧が観測された場合、閾値 2 を用いてウィンドウ内の音圧検出区間をより詳細に抽出する。閾値 2 を超える音圧にラベル 1, 超えない音圧にラベル 0 を付与してユーザごとの音圧検出データ「各ユーザの 1-0 データ」を抽出する。

2) 発話区間抽出: 発話区間抽出では、各ユーザの 1-0 データからユーザの発話の有無を抽出する。各ユーザの 1-0 データにおいてラベル 1 で囲まれた 90 ms 以内のラベル 0 区間を 1 発話の間とみなしてラベル 1 で補完する。各ユーザの 1-0 データにおいてラベル 0 で囲まれた 150 ms 以内のラベル 1 区間をパルスノイズとみなしてラベル 0 で上書きする。得られたデータにおいて時間軸に沿って全ユーザの論理和をとることで「発話区間データ」を抽出する。

3) 発話者特定: 発話者特定では、各ユーザの 1-0 データおよび発話区間データを用いて発話区間ごとに発話者を推定する。発話区間データにおける各区間に相当する各ユーザの 1-0 データを参照する。区間内で最もラベル 1 が多く存在するユーザを発話者として判定することで発話者特定を行う。

4 評価

SRP Badge の時刻同期精度と話者検出アルゴリズムの精度との関係の評価するために SRP Badge を用いた評価実験を行った。実験は、20 代前半の男子大学生 3 人を被験者として 10.6 m × 7.05 m × 2.65 m からなる部屋で行った。被験者は胸部に SRP Badge を装着して、机を中心にして隣接する被験者から 1.5 m ずつ離れた椅子に着座した。机の中央に同期装置を設置して SRP Badge 間の時刻同期を行った。各被験者は文献 [3] における英文のスキriptに従って 1 人ずつ一定の時間間隔を空けて発話を行った。

本評価にあたって、SRP Badge で取得した音圧データのタイムスタンプに乱数を加えることで時刻同期精度の異なるデータを擬似的に作成した。図 3 に SRP Badge および親機との間で実測される時刻同期誤差を示す [1]。図 3 の横軸は、SRP Badge および親機との間で計測された時刻同期誤差 (μs)、縦軸は度数を示す。図 3 における時刻同期誤差の分布が正規分布に近いと認められることから、本評価では音圧データのタイムスタンプに加える乱数として正規分布に従うものを用いた。正規分布の標準偏差を調節することで乱数の範囲を指定した。たとえば標準偏差 $\sigma = 0.5$ ms の場合、 $\pm 2\sigma = \pm 1.0$ ms に全体の約 95% が収まるように乱数が生成される。本評価では、 $\pm 2\sigma$ を時刻同期誤差として扱うこととした。日立のビジネス顕微鏡が名刺型センサ間で誤差 100 ms の時刻同期を実現していること [4]、MIT の Rhythm Badge が名刺型センサ間で誤差 10 ms の時刻同期を実現していること [5] を考慮して、評価実験で得られた音圧データのタイムスタンプに対して時刻同期誤差 $2\sigma = 0, 0.1, 0.3, 1.0, 10, 50$ ms の正規分布に従う乱数をそれぞれ加えた。SRP Badge で取得した音圧データを 3 次曲線で補間し、その補間関数から、タイムスタンプのずれた音圧データを取得することで、擬似的に時刻同期誤差を加えた音圧データを生成した。乱数による偏りを考慮するために時刻同期誤差ごとに 10 回ずつ乱数を生成して話者検出を行い、F 尺度を用いた話者検出精度の平均値を算出した。

図 4 に各標準偏差における話者検出精度の F 値を示す。図 4 において、横軸は音圧データのタイムスタンプに加えた時刻同期誤差 (ms) を、縦軸は話者検出精度の F 値を示す。時刻同期誤差 (ms) が 0, 0.1, 0.3, 1.0, 10, 50 の各場合において、ユーザ 1 の F 値は 1, 1, 0.916, 0.895, 0.792, 0.780, ユーザ 2 の F 値は 1, 1, 0.884, 0.858, 0.746, 0.735, ユーザ 3 の F 値は 0.966, 0.966, 0.966, 0.966, 0.966, 0.966 となった。図 4

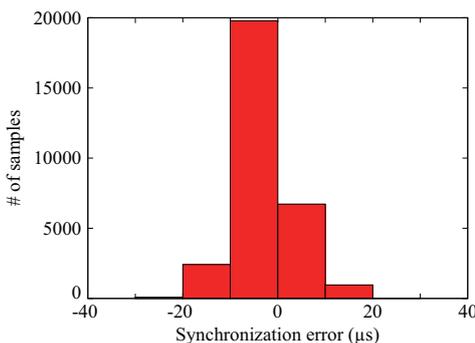


図 3: 名刺型センサの時刻同期精度

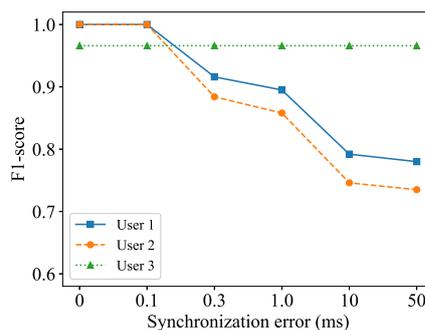


図 4: 話者検出精度の F 値

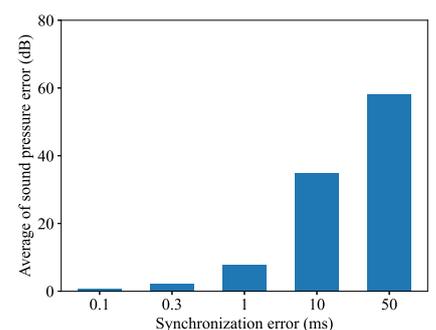


図 5: 時刻同期誤差ごとの平均音圧誤差

に示すとおり、時刻同期誤差が大きくなるにつれて話者検出精度の低下が見られた。

表 1 に各時刻同期誤差における話者検出精度の混同行列を示す。表 1 から、時刻同期誤差が大きくなるにつれて非発話区間における発話の誤検出が増えることがわかる。

図 5 に各時刻同期誤差の平均音圧誤差を示す。図 5 において、横軸は音圧データのタイムスタンプに加えた時刻同期誤差 (ms) を、縦軸は音圧誤差の平均値 (dB) を示す。図 5 から、時刻同期誤差が大きくなるにつれて音圧誤差も大きくなることわかる。本音圧誤差が、図 4 に示す話者検出精度の低下を引き起こす要因の 1 つになりうると考えられる。

本結果における有意差を検証するために各時刻同期誤差における平均の差の検定 (t 検定) を行った。標本数 10 であることから自由度 18、有意水準 5% として検定したところ、時刻同期誤差 0.3 ms および 1 ms 間に有意な差は見られなかったが、時刻同期誤差 0.3 ms および 10 ms, 50 ms 間にはそれぞれ有意な差が見られた。本検定結果より、SRP Badge の高精度な話者認識精度を担保するうえで誤差 10 ms 以内の時刻同期が SRP Badge 間において必要であると考えられる。

5 おわりに

本稿では、協調学習を定量的に分析する Sensor-based Regulation Profiler (SRP) における名刺型センサの時刻同期精度と話者検出アルゴリズムの精度との関係について述べた。本評価によって、話者検出アルゴリズムの精度を担保するうえで名刺型センサの時刻同期精度の重要性を示すことができた。今後の展望として、学習活動のマルチモーダル分析に用いる学習者同士の対面や学習フェーズ、学習者の活動量を抽出するアルゴリズムの精度と名刺型センサの時刻同期精度との関係の評価して、時刻同期精度を担保するうえで必要とされる名刺型センサのハードウェア性能を明らかにすると同時にハードウェアの低コスト化を図る。

謝辞

本研究は JSPS 科研費 19H01714 及び NTT アクセスサービスシステム研究所の支援の下で行った。

参考文献

- [1] Shunpei Yamaguchi, Shusuke Ohtawa, Ritsuko Oshima, Jun Oshima, Takuya Fujihashi, Shunsuke Saruwatari, and Takashi Watanabe, “Collaborative Learning Analysis Using Business Card-Type Sensors,” in *International Conference on Quantitative Ethnography*, pp. 319–333, 2021.
- [2] Daichi Yamaguchi, Shunpei Yamaguchi, Ritsuko Oshima, Jun Oshima, Takuya Fujihashi, Shunsuke Saruwatari, and Takashi Watanabe, “A Web Application with Business Card-Type Sensors for Collaborative Learning Analysis,” in *IEEE Global Conference on Consumer Electronics*, pp. 748–749, 2021.
- [3] Shunpei Yamaguchi, Ritsuko Oshima, Jun Oshima, Ryota Shiina, Takuya Fujihashi, Shunsuke Saruwatari, and Takashi Watanabe, “Speaker Identification for Business-Card-Type Sensors,” *IEEE Open Journal of the Computer Society*, pp. 216–226, 2021.
- [4] 辻 聡美, 佐藤 信夫, 上垣 映理子, 佐々木 真美, 賀 曉琳, 矢野 和男: 行動センシングによる働き方パーソナルアドバイザーの設計と試行, 情報処理学会デジタルプラクティス, pp. 267–282, 2019.
- [5] Oren Lederman, Akshay Mohan, Dan Calacci, and Alex Sandy Pentland, “Rhythm: A Unified Measurement Platform for Human Organizations,” *IEEE MultiMedia*, vol. 25, no. 1, pp. 26–38, 2018.