7ZJ-02

ハイブリッド環境下のコラボレーションを抽出するIoTバッジを用いた話者検出に関する一検討

藤原 拓也 † 山口 隼平 ‡ 大島 律子 ‡† 大島 純 ‡† 藤橋 卓也 ‡ 猿渡 俊介 ‡ 渡辺 尚 ‡
† 大阪大学工学部 ‡ 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科

1 はじめに

ハイブリッド協調学習とは，学習者らがハイブリッド環境下で協調 (コラボレーション) しながら問題解決に取り組むアフターコロナの時代に向けた学習形態である [1]．ハイブリッド協調学習の分析をサポートする一手法として IoT バッジを用いた自動話者検出技術があげられる．しかしながら，ハイブリッド環境下ではスピーカから流れてくる話者の声がノイズとなって正確な話者検出を妨げてしまう． 本稿では，ハイブリッド学習環境に対応する IoT バッジを用いた話者検出技術を提案することでハイブリッド環境下での高精度話者検出が実現できることを示した．

2 ハイブリッド環境下での課題

IoT バッジを用いた話者検出をハイブリッド環境に適用する場合，スピーカから流れてくる話者の声が非話者のバッジにノイズとして乗るため検出精度の低下を招いてしまう．具体的には，スピーカの音量設定上もとの発話の音圧がスピーカを通じてより大きく IoT バッジに検知される場合にスピーカノイズを発話と誤検出するケースが生じてしまう．

本課題を解決しうる IoT バッジを用いた話者検出に関連する研究として以下のものがあげられる．日立のビジネス顕微鏡 [2] およびマサチューセッツ工科大学 (MIT) の Rhythm [3] では，組織内のコラボレーション抽出を目的として音圧センサを搭載した IoT バッジおよび話者検出アルゴリズムを提案している．しかしながら，ビジネス顕微鏡は 8 kHz の高サンプリングレートで音圧を収集・処理することからハードウェアコストの観点で学習現場への導入が難しい．Rhythm は 700 Hz の低サンプリングレートで音圧を収集・処理するが，バッジ間での高精度時刻同期を実現していないためデータ精度の観点で学習現場への導入が難しい．

オフライン協調学習の分析をサポートする研究 [4–6] では，Sensor-based Regulation Profiler Badge (SRP Badge) と呼ばれる IoT バッジを用いた話者検出技術を提案している．本手法は，学習現場への導入を目的として低コストかつ高精度で音圧情報を収集する SRP Badge と，SRP Badge から得られる音圧データを用いて高精度な話者検出を実現するアルゴリズムから構成される．SRP Badge は，音圧を 100 Hz でサンプリングすることで低ハードウェアコストを実現しながら発話を高精度に収集するためにデバイス間で ±30 µs 以内の高精度時刻同期を実現している．話者検出アルゴリズムは，100 Hz の音圧情報をもとに高精度な話者検出を実現している．しかしながら，本手法をハイブリッド環境に適用した場合はスピーカノイズに起因して話者検出精度が低下してしまう．具体的には，ハイブリッド環境下でもとの発話の音圧がスピーカを通じてより大きく SRP Badge に検知される場合，スピーカノイズを発話と誤検出するケースが生じてしまう．

3 提案手法

2 節で述べた課題を解決するための手法として，SRP Badge を用いたハイブリッド環境に対応する話者検出技術を提案する．提案手法の全体像を図 1 に示す．提案手法は，学習者らの発話を音圧として正確に収集・記録する SRP Badge と収集した音圧データからハイブリッド環境下での発話を正確に判定する話者検出アルゴリズムからなる．

3.1 Sensor-based Regulation Profiler Badge

SRP Badge は，各学習者の胸元に装着することを想定した名刺型のバッジとなっていて赤外線・音圧・加速度センサを具備する．本バッジはリチウムイオン電池を搭載して 24 時間連続で稼働する．SRP Badge の音圧センサとして，TDK 社の INMP510 のアナログマイクロフォンが搭載されている．INMP510 の周波数特性に従って，60 Hz から 20 000 Hz の周波数をもつ音声が収集される．収集された音声を，SRP Badge の MCU

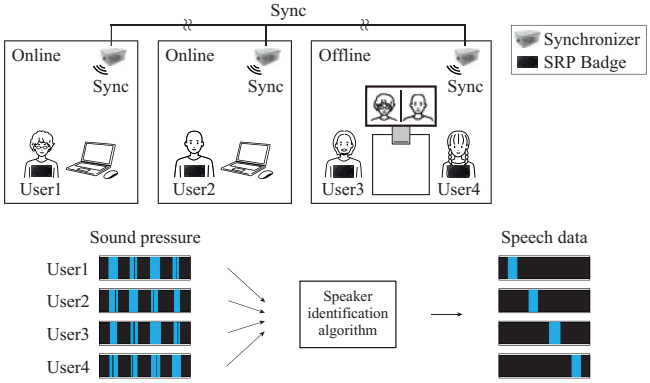


図 1: 提案手法の全体像

として搭載されている STMicroelectronics 社の STM32L476RGT6 が 100 Hz でサンプリングする．サンプリングで得られた音圧データは，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 カードをセンサのスロットに差し込むことでデータが保存される．また，SRP Badge には無線同期モジュールとして Texas Instruments 製 CC2650 が搭載されている．無線同期モジュールは，図 1 に示す SRP Badge の親機 (Synchronizer) から同期信号を受信・転送することによってバッジ間誤差 ±30 µs の高精度時刻同期を維持しながら音圧を正確に収集・記録する．

3.2 話者検出アルゴリズム

図 2 に話者検出アルゴリズムの全体像を示す．話者検出アルゴリズムは，SRP Badge で得られた音圧データの前処理，発話区間推定，話者候補抽出，発話者特定の 4 ステップから構成される．

音圧データの前処理

前処理では，音圧データのゼロ点補正およびユーザごとの音圧立ち上がり検出を行う．各ユーザの音圧データから全区間における音圧の最小値を差し引くことでゼロ点補正を行う．補正データに対して，ウィンドウサイズ 2 s，スライド幅 1 s のスライディングウィンドウを適用しながら各ウィンドウ内で 2 つの閾値を用いて音圧の立ち上がりを詳細にラベル付けする．全ウィンドウ共通の閾値 (閾値 1) を設定してウィンドウごとに音圧の立ち上がりの有無を決定する．立ち上がりが検出されたウィンドウに対しては，ウィンドウごとに閾値 (閾値 2) を定めて閾値を超えた区間を 1 とラベリングする．上記の過程でラベル 1 が付与されない区間にはラベル 0 を付与することで音圧の立ち上がりの有無を詳細に 2 値化する．本ステップで得られる音圧情報を「各ユーザの 1–0 データ」と呼ぶこととする．

発話区間推定

発話区間推定では，音圧データの前処理で得られる各ユーザの 1–0 データを用いてハイブリッド空間内での発話有無を推定する． 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 区間の間とみなしてラベル 1 で上書きする．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．上記プロセスを通じて得られるユーザごとのラベルの論理和をとることで空間内の発話有無を表す「発話区間データ」を抽出する．

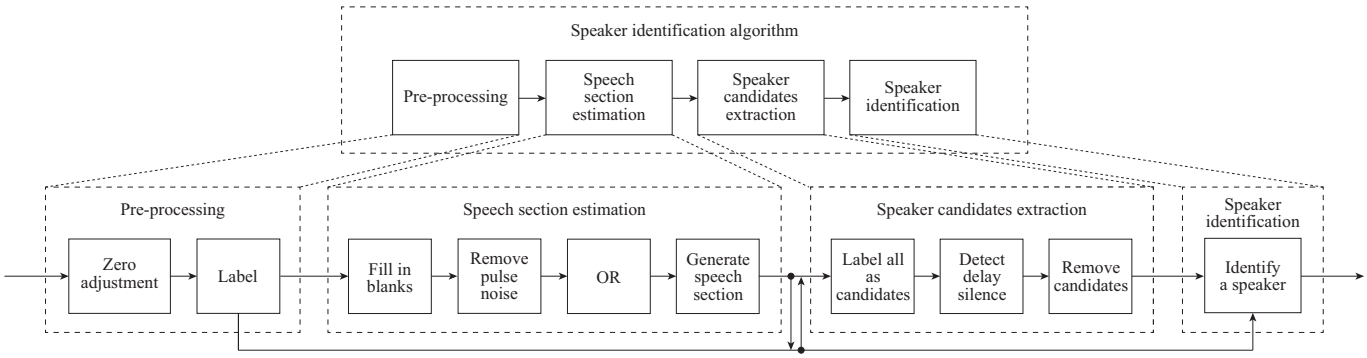


図 2: 提案アルゴリズム

Copyright ©2023 Information Processing Society of Japan.
All Rights Reserved.

話者候補抽出

話者候補抽出では，発話区間データの各区間においてスピーカノイズを有するユーザを除外することで各発話区間の話者候補を抽出する．はじめは各発話区間において全ユーザを話者候補とする．一方の部屋で生じた発話が通信遅延の影響で他方の部屋のスピーカから遅れて聞こえることを利用してスピーカノイズを拾うユーザを話者候補から除外する．具体的には，一定の遅延時間を仮定して各発話区間の先頭から遅延時間内で音圧の立ち上がりが一定割合以上検出されないユーザを話者候補から除外する．本ステップを通じて得られる発話区間ごとの話者候補を「話者候補データ」と呼ぶこととする．

発話者特定

発話者特定では，各ユーザの 1–0 データ，発話区間データ，および各発話区間の話者候補データを用いて発話区間ごとに発話者を特定する．発話区間データにおける各区間に対応する話者候補データを参照してスピーカノイズを拾っているユーザを除外する．残った話者候補から，各区間に対応する各ユーザの 1–0 データを参照して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．

4 評価

ハイブリッド環境を対象とした話者検出アルゴリズムの精度を評価するために SRP Badge を用いた話者検出実験を行った．被験者は，20 代前半の男子大学生 4 名とした． ハイブリッド空間を想定して，オフラインおよびオンライン環境双方からのノイズが入らないように密閉された隣接する 2 部屋 (部屋 A, B) を用いて実験を行った． 部屋 A では，オンライン環境を想定して 1 名の被験者と 1 台の PC (Apple MacBook Pro, 14 inch, 2021) を配置した． 部屋 B では，オフライン環境を想定して残りの被験者と 1 台の PC (FUJITSU LIFEBOOK WU2/D2) を配置した． 各部屋は，弊学が所属学生に提供している Wi-Fi のアクセスポイント (SSID: odins-1x) 下で Web 会議システム Zoom を用いて無線接続した．各被験者は胸部に SRP Badge を 1 台装着した． 2 部屋の各 SRP Badge が同期信号を受信できるよう全バッジの通信可能範囲である部屋 B の隅に SRP Badge の親機を配置した．各被験者は，文献 [5] の発話実験で用いられた 15 文からなる英文スクリプトに従って 10 s のインターバルを設けながら発話を行った． 各被験者の発話を他方の部屋に伝達するにあたっては PC のスピーカを用いて最大音量で発話を流した． 部屋 A では，被験者と対面するように PC を設置した．部屋 B では，被験者同士が PC を囲むように 1 m の間隔をあけて着座した．提案アルゴリズムの前処理で用いる閾値 1 は文献 [5] に基づいて 75 dB とした． 閾値 2 は，無音区間の音圧が閾値 1 を上回らないようウィンドウ内の最大音圧値の 0.15 倍とした．提案アルゴリズムの話者候補抽出で用いる通信遅延時間および話者候補の除外に用いる音圧の立ち上がり割合は同環境の簡易遅延評価に基づいてそれぞれ 0.25 s，0.8 とした．話者検出精度の評価指標には F 値を用いた．比較手法として，文献 [5] で提案されている話者検出アルゴリズムを用いた．

4.1 多様な学習環境を想定した評価

ハイブリッド協調学習の多様な参加形態を考慮して 4 環境で実験を行って話者検出精度を評価した．本実験では，オフライン側に 2 名，オンライン側に 1 名の被験者を配置して計 3 名で発話を行った．表 1 に実験環境の設定を示す．本稿では，表 1 の env 4 に相当する環境下で実験を行ったのち，得られた実験データを編集することで env 1, 2, 3 に相当する環境を想定した．env 1, 2, 3 を想定した実験データは，被験者がイヤホンを装着している場合は SRP Badge にスピーカノイズが乗らないことを用いて env 4 の実験データ上に乗っているスピーカノイズを無音データで上書きすることによって生成した． 無音データは，いずれの被験者も発話していない区間の中から音圧データを数百サンプル抽出して得られた平均および標準偏差をもつ正規分布に従う乱数から生成した．

図 3 に実験結果の F 値を示す．図 3 の横軸は想定環境を，縦軸は F 値を，凡例は既存手法 [5] (existing) および提案手法 (proposed) を表している．env 1, 2, 3, 4 におけるオフライン側の全ユーザの F 値は，既存手法ではいずれの環境でも 0.967 となり，提案手法ではそれぞれ 0.967, 0.982, 0.967, 0.982 となった．env 1, 2, 3, 4 におけるオンライン側のユーザの F 値は，既存手法では 0.965, 0.483, 0.965, 0.508 となり，提案手法で 1.000, 0.938, 1.000, 0.938 となった．いずれの環境においても提案手法の精度が既存手法の精度と同等以上であることが示された．とりわけオンライン側でスピーカを用いた env 2 および 4 においては，既存手法を用いた場合はスピーカノイズの影響で F 値の精度低下が見られたが，提案手法を用いた場合はスピーカノイズを有するユーザを話者候補から除外することで高精度な話者検出を維持することがわかった．

		Online	
		earphone	speaker
Offline	earphone	env 1	env 2
	speaker	env 3	env 4

表 1: 本稿で想定した学習環境

4.2 参加人数による評価

ハイブリッド環境におけるオフライン側の参加人数による話者検出精度を評価した．本実験では，オンライン側に 1 名，オフライン側に 1, 2, 3 名を配置して計 2, 3, 4 名の被験者でそれぞれ発話を行った． 図 4 に実験結果の F 値を示す． 図 3 の横軸はハイブリッド環境下の全ユーザ数を，縦軸は F 値を，凡例は既存手法 [5] (existing) および提案手法 (proposed) を表している．オフライン環境の全ユーザの F 値は，オフライン側のユーザが 1, 2, 3 名のときに既存手法では 1.000, 0.967, 0.944 となり，提案手法では 1.000, 0.982, 0.952 となった． オンライン環境のユーザの F 値は，オフライン側のユーザが 1, 2, 3 名のときに既存手法では 0.683, 0.508, 0.462 となり，提案手法では 1.000, 0.938, 0.737 となった．ユーザ数が増加するに従って両環境におけるユーザの F 値の低下がみられた．とりわけ被験者 4 名におけるオンライン側のユーザの F 値が著しく低下したが，既存手法の F 値が 0.462 であったのに対して提案手法の F 値は 0.737 であったことから，提案手法を用いることでオンライン側の話者検出精度の低下を抑えられることがわかった．両手法における精度低下の原因として，被験者 4 名のうちオフライン側の 1 名の声量が小さく発話を非発話と誤判定するケースが増えたことが考えられる．

5 おわりに

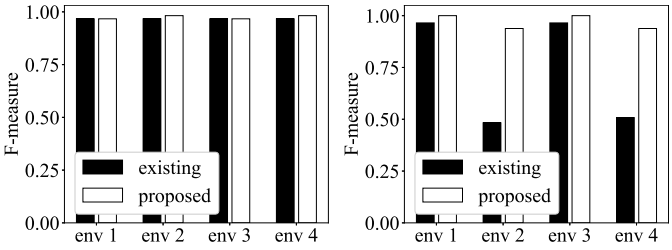
本稿では，ハイブリッド協調学習に対応する SRP Badge を用いた話者検出技術を提案して様々な環境下で提案手法が高精度話者検出を実現することを示した．今後の取り組みとして，協調学習で用いられる動画教材から流れてくる多様なノイズの影響，オンライン側のユーザ増加に伴ってより多くの部屋をまたいだ場合の影響を評価することがあげられる．加えて，図 1 に示す SRP Badge の親機間で高精度時刻同期をとる通信手法を実現することで遠く離れた部屋同士でも本提案手法が適用できるよう研究を進めている．

謝辞

本研究は，JSPS 科研費 (19H01714, 22J20391) および JST さきがけ (課題番号: JPMJPR2032) の支援の下で行った．

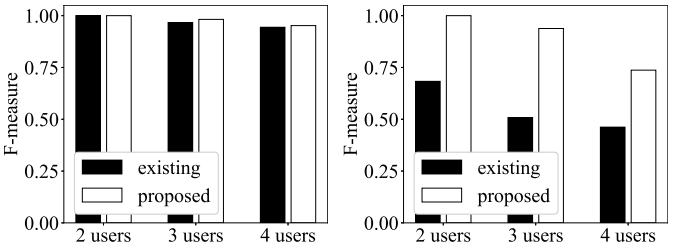
参考文献

[1] Sandra Phek-Lin Sim et al., “Online Learning: A Post Covid-19 Alternative Pedagogy For University Students,” *Asian Journal of University Education*, vol. 16, no. 4, pp. 137–151, 2020.
[2] 辻 聡美 ほか：行動センシングによる働き方パーソナルアドバイザの設計と試行，情報処理学会デジタルプラクティス，pp. 267–282，2019.
[3] O. Lederman et al., “Rhythm: A Unified Measurement Platform for Human Organizations,” *IEEE MultiMedia*, vol. 25, no. 1, pp. 26–38, 2018.
[4] S. Yamaguchi et al., “Collaborative Learning Analysis Using Business Card-Type Sensors,” in *International Conference on Quantitative Ethnography*, pp. 319–333, 2021.
[5] S. Yamaguchi et al., “Speaker Identification for Business-Card-Type Sensors,” *IEEE Open Journal of the Computer Society*, vol. 2, pp. 216–226, 2021.
[6] S. Yamaguchi et al., “Web Services for Collaboration Analysis With IoT Badges,” *IEEE Access*, vol. 10, pp. 121318–121328, 2022.



(a) オフライン　(b) オンライン

図 3: 環境によるユーザの F 値



(a) オフライン　(b) オンライン

図 4: 人数によるユーザの F 値

Copyright ©2023 Information Processing Society of Japan.
All Rights Reserved.