

# 自立生活支援のための音響イベント検出の連合学習

---

竹本志恩

July 18, 2025

# 研究テーマ

---

# テーマ概要

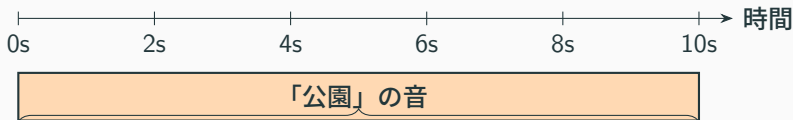
- 少子高齢化により医療と介護の負担は増大
- AAL (Ambient Assisted Living) [1, 2]
  - 情報技術で自立的な生活を支援し, 在宅介護で問題解決を目指す
- 従来手法の課題
  - 主にカメラを用いるが, 高価でプライバシー受容性に難点
  - ウェアラブルは充電や装着し忘れ, 侵襲性の問題
- 音響イベント検出 (Sound Event Detection, SED) で解決
  - 音による行動認識で, 異常検知や健康状態の把握を目指す
  - カメラより安価で, 音特有の異常兆候を検出
  - 機械学習で多様な環境や対象に柔軟に対応
  - SED で「いつ, どんな行動があったか」を理解し, 説明性のある異常検知モデル構築を目指す

# 音響イベント検出とは

- 与えられた音から、イベントを検出するタスク
- イベントの開始と終了，ラベルを予測
- 音響シーン分類との違い
  - 音響シーン分類は一定時間の音のラベルのみ予測
  - SED は音の発生時間や持続時間も考慮
- Acoustic と Sound の違いは未調査

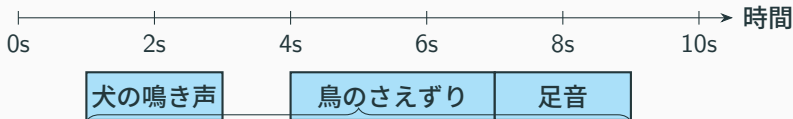


## 音響シーン分類 (Acoustic Scene Classification)



音声全体に単一のラベルを予測

## 音響イベント検出 (Acoustic Event Detection)



個別のイベントとその時間を予測

# 研究課題

---

# テーマで取り組む課題

- 総合的な課題: 連合学習によるプライバシー問題の解決
  - AAL はプライバシー性の高いデータを扱う
  - 従来の中央集権的な機械学習はプライバシーに課題
  - 本研究では, この課題を連合学習 (Federated Learning, FL) で解決
  - エッジデバイスの先行研究もある [3] が, より柔軟な FL に着目
- 本研究における問い
  1. 中央集権的手法と比較し, 連合学習でどれだけ精度を維持できるか?
  2. どの連合学習アルゴリズムが家庭内環境に対して適切か?
- 展望
  - SED を手がかりに異常検知

# 実験計画

---



# 実験評価計画 (1/2)

- 7月
  - 研究計画書を作成
  - 各評価指標の意味を理解
  - 具体的な目標精度を決定
- 7-8月
  - 適切なモデルアーキテクチャの比較・検討 [4, 5]
    - 事前学習済みモデルの比較
    - 適切な CNN, RNN の構成を模索
    - 事前学習済みモデル+CNN+RNN の構成を模索
  - 適切な学習戦略を決定
    - 前処理/後処理の方法, アンサンブルの有無など
    - 特に半教師あり学習の手法を検討 (Mean-Teacher か FixMatch を想定)
    - DCASE 2024 を参照

# 実験評価計画 (2/2)

- 9月
  - 前処理, 後処理, モデル構成要素の精度への影響を調査
- 9-10月
  - 連合学習モデルの精度を従来手法と比較
    - ベースライン: FedAVG [6]
    - 比較対象: FedProx [7], SCAFFOLD [8]などを予定
    - 連合学習のハイパーパラメータを調整
- 11月
  - 考察

# 評価方法

---

# 評価方法

- モデル共通の前提
  - 基本モデル: DCASE 2024 のベースライン
  - データセット: DESED, MAESTRO
- モデルの比較対象
  - ベースラインや SOTA(State-of-the-Art) と比較
- 評価指標
  - DCASE 2024 の Supplementary metrics を参照
  - 各種 F1 スコア+PSDS (Polyphonic Sound Detection Score) 1, 2 を使用
- 精度の基準
  - 連合学習を適用した場合の精度を中央集権的な手法と比較
  - 従来手法から少し劣る精度が目標

進捗



- ✓ 各種サーベイを実施
  - 技術的な理解はこれから
  - 連合学習/音響イベント検出/半教師あり学習の調査
  - 自立生活支援の研究目的を理解
  - DCASE task4 2018-2024 の大まかな内容を把握
- 実験評価計画タスク
  - 研究計画書: ほぼ完了
  - 各評価指標, 目標精度: 指標は把握
  - モデルアーキテクチャ: 使用するモデルは概ね把握
  - 学習戦略を決定: 半教師あり学習手法は把握, 全体は論文を参考に検討中
  - 精度影響の調査: 未着手
  - 連合学習: FedAVG, FedProx は追試済み, 他は未調査

# References

---

- [1] S. Blackman et al. **“Ambient assisted living technologies for aging well: a scoping review”**. In: *Journal of Intelligent Systems* 25.1 (2016), pp. 55–69.
- [2] R. Stodczyk and F.-H. Uhp. **“Ambient assisted living an overview of current applications, end-users and acceptance”**. In: *Biomedical Journal of Scientific & Technical Research* 30.3 (2020), pp. 23374–23384.
- [3] R. M. Alsina-Pagès et al. **“homesound: Real-time audio event detection based on high performance computing for behaviour and surveillance remote monitoring”**. In: *Sensors* 17.4 (2017), p. 854.

- [4] Y. Li et al. **“A hybrid system of sound event detection transformer and frame-wise model for dcase 2022 task 4”**. In: *arXiv preprint arXiv:2210.09529* (2022).
- [5] F. Schmid et al. **“Multi-iteration multi-stage fine-tuning of transformers for sound event detection with heterogeneous datasets”**. In: *arXiv preprint arXiv:2407.12997* (2024).
- [6] B. McMahan et al. **“Communication-efficient learning of deep networks from decentralized data”**. In: *Artificial intelligence and statistics*. PMLR. 2017, pp. 1273–1282.
- [7] T. Li et al. **“Federated optimization in heterogeneous networks”**. In: *Proceedings of Machine learning and systems 2* (2020), pp. 429–450.



- [8] S. P. Karimireddy et al. **“Scaffold: Stochastic controlled averaging for federated learning”**. In: *International conference on machine learning*. PMLR. 2020, pp. 5132–5143.