

日本－フランス 国際共同研究 「エッジ AI」 2023 年度 年次報告書	
研究課題名（和文）	組み込み機器による音響信号の意味論的理解と分離
研究課題名（英文）	Semantic Segmentation of Complex Sound Scenes on Edge Devices (CONFLUENCE)
日本側研究代表者氏名	原田 登
所属・役職	日本電信電話株式会社 コミュニケーション科学基礎研究所 メディア情報研究部・上席特別研究員
研究期間	2023 年 12 月 1 日 ～ 2027 年 11 月 30 日

1. 日本側の研究実施体制

氏名	所属機関・部局・役職	役割
原田 登	日本電信電話株式会社・コミュニケーション科学基礎研究所・上席特別研究員	研究代表者・全体統括・ワークパッケージ No. 1, 2, 4, 5
小野 順貴	東京都立大学・システムデザイン学部・教授	研究代表者（学）・ワークパッケージ No. 1, 2, 3, 5

2. 日本側研究チームの研究目標及び計画概要

本研究は、我々を取り巻く様々な音響イベントを認識し意味的な音の塊を分離する AI 技術の開発、及び通信サービスへの応用を目的とし、没入型コミュニケーションとホームアシスト（見守り）の2つのユースケースを例として、それらを実装できる組み込みシステムの設計を行う。具体的には、様々な音を分析して分離しメタデータを付与することで、状況に応じて必要な音だけを伝送し、空間的に再構成して再生できるようにする。

日本側チームでは音源の分離と抽出及び没入型通信(Immersive communication)への応用に関する開発を行う。フランス側チームでは、エッジ AI デバイスへの実装と没入型コミュ

ニケーション・ホームモニタリングへの応用検討を行う。

開発した技術は、移動体通信システム国際標準化（3GPP）への提案も行い、通信サービスでの標準化もめざす。本技術により、例えば、生活音を遮断しビジネス対話に必要な音声のみを通す Web 会議システムや、ホームアシスタントなどプライバシーに配慮した通信への応用、国際会議の現地参加者とリモート参加者が体験を共有できる新しいインクルーシブコミュニケーションなどへの応用が期待される。

3. 日本側研究チームの実施概要

初回である 2023 年度の年次報告書では、2023 年 12 月 1 日より 2024 年 3 月 31 日までの 4 ヶ月間に実施した内容について報告する。

（1）全体 Kick-off（ワークパッケージ 5）

日仏各研究機関の主要メンバーが 2024 年 3 月に開催された Kick-off 会合に参加した。Kick-off 会合では、プロジェクト概要についてプレゼンテーションを行ったほか、NTT より 2 件の関連既存技術の実演デモンストレーション展示を行った。

併せてフランス側の研究機関 3 拠点を訪問しプロジェクトミーティングを開催した。

（2）ベースラインの検討（ワークパッケージ 1 および 2）

音響信号の意味論的理解と分離に関わる市中技術の調査を実施し、音響イベント検知およびシーン分類に関する国際会議である Detection and Classification of Acoustic Scenes and Events の Challenge Task 9 などの baseline 実装の評価等を行った。

（3）技術評価と有効性検討（ワークパッケージ 3）

音光変換の音響シーン認識応用を探るため、occlusion（遮蔽）を含むシミュレーション環境を設定して、音光変換、ならびにカメラで信号取得後の認識器の end-to-end 学習、ならびにその評価を行った。

また、音光変換デバイスの分散配置に基づく発話検出のシミュレーション評価を行い、分散配置の有効性を検討した。

（4）Use case の調査（ワークパッケージ 4）

3GPP IVAS 標準化の議論に参加し Use case シナリオの調査を行った。モデル学習・評価に用いるデータを収録するための特注品マイクロホン等を設計し調達した。