

# The Sounds of Home: A Speech-Removed Residential Audio Dataset for Sound Event Detection

GABRIEL BIBBÓ<sup>1</sup>, THOMAS DEACON, ARSHDEEP SINGH, MARK D. PLUMBLEY

CENTRE FOR VISION, SPEECH AND SIGNAL PROCESSING (CVSSP), UNIVERSITY OF SURREY, UK

<sup>1</sup>Email: g.bibbo@surrey.ac.uk, ORCID: <https://orcid.org/0009-0009-6262-5907>

## INTRODUCTION

1

- Smart home systems offer opportunities to enhance older adults' wellbeing through audio machine learning, but require substantial labeled data capturing Activities of Daily Living (ADL).
- We present **The Sounds of Home**, a comprehensive residential audio dataset for sound event detection research in smart home applications, in collaboration with LiCalab as part of the Vitalise project <https://vitalise-project.eu/>
- Our dataset of 1342 hours length was constructed by deploying audio recording systems in 8 participants' homes (ages 55-80) for 7 days, with a novel automated speech removal pipeline to address privacy concerns.
- The resulting privacy-compliant dataset accurately captures residential soundscapes, enabling development and benchmarking of sound event detection models for in-home applications.

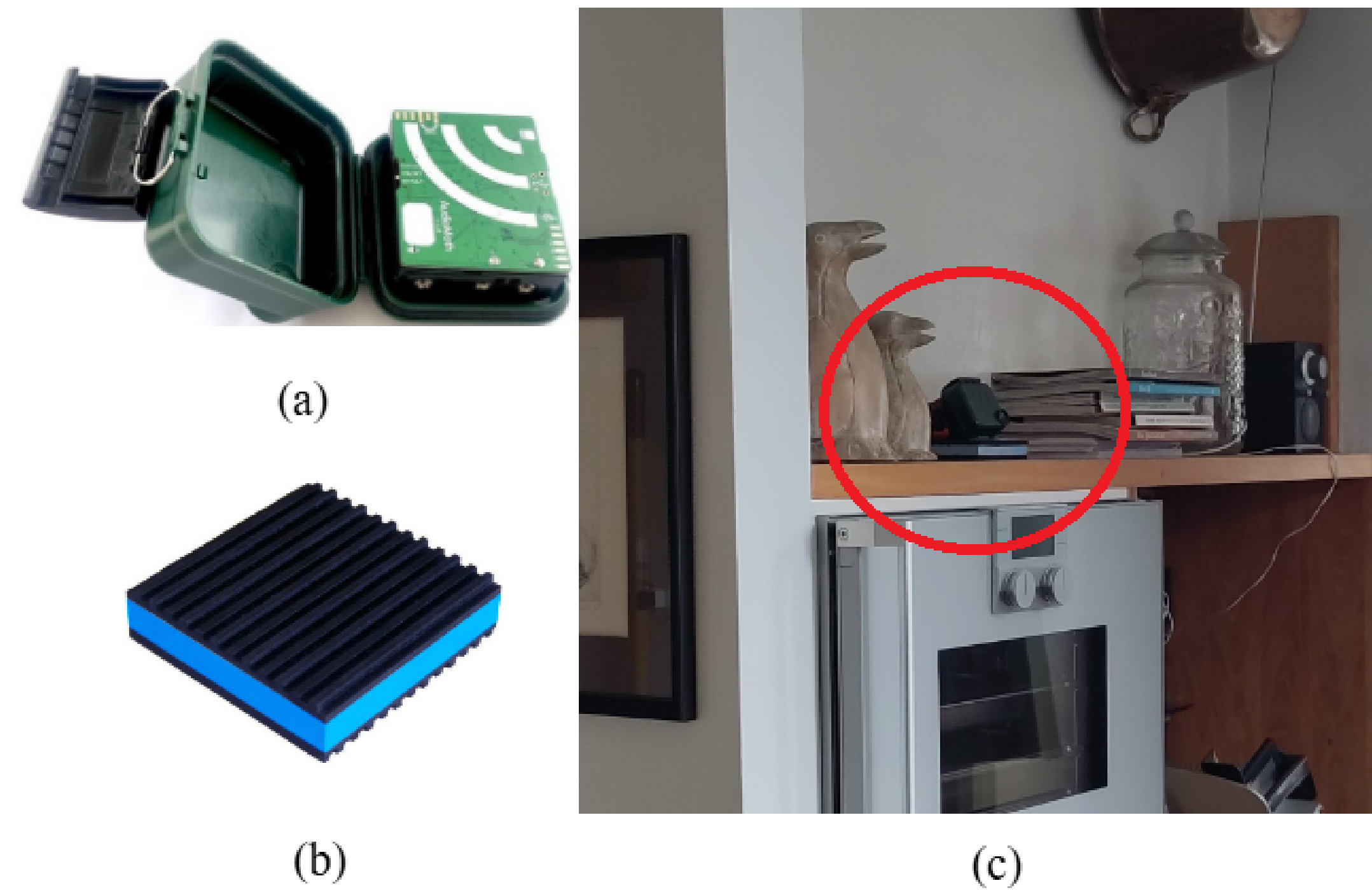
## OBJECTIVES

2

- Create a large-scale, privacy-compliant audio dataset capturing real-world domestic soundscapes of older adults for sound event detection research.
- Develop and implement an automated speech removal pipeline to ensure participant privacy while preserving non-speech audio events.

## DATASET COLLECTION METHODOLOGY

3



- Deployed two Audio Recording Systems (ARS) in each of 8 participants' homes for 7 days, using AudioMoth devices enclosed in IPX7 Waterproof Cases to ensure durability and high-quality audio capture.
- Configured devices to record WAV audio at 48 kHz sample rate in 1-hour segments from 8:00 AM to 9:00 PM daily, capturing participants' most active periods while minimizing interference with daily routines.
- Placed devices strategically in living rooms and kitchens at 1-2 meters height, away from windows and obvious noise sources, with vibration pads to reduce ground-borne vibration impact.
- Documented acoustic characteristics through detailed floor plans, construction material information, and placement photographs to enable future replication of recording environments.

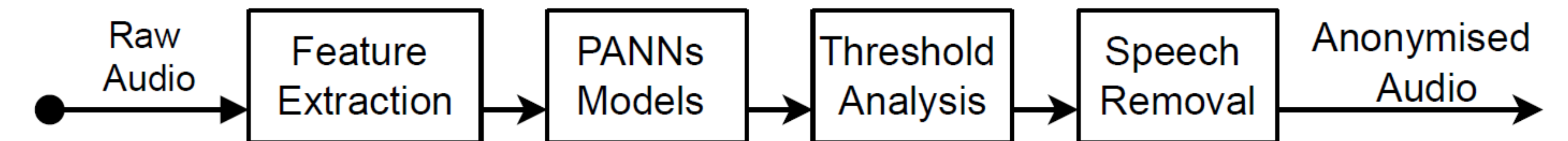
## ACKNOWLEDGEMENT

6

This work was supported by Engineering and Physical Sciences Research Council (EPSRC) Grant EP/T019751/1 AI for Sound. AI4S project link: <https://ai4s.surrey.ac.uk/index>.

## DATASET ANONYMISATION METHODOLOGY

4



### 1. AI Models for Inference

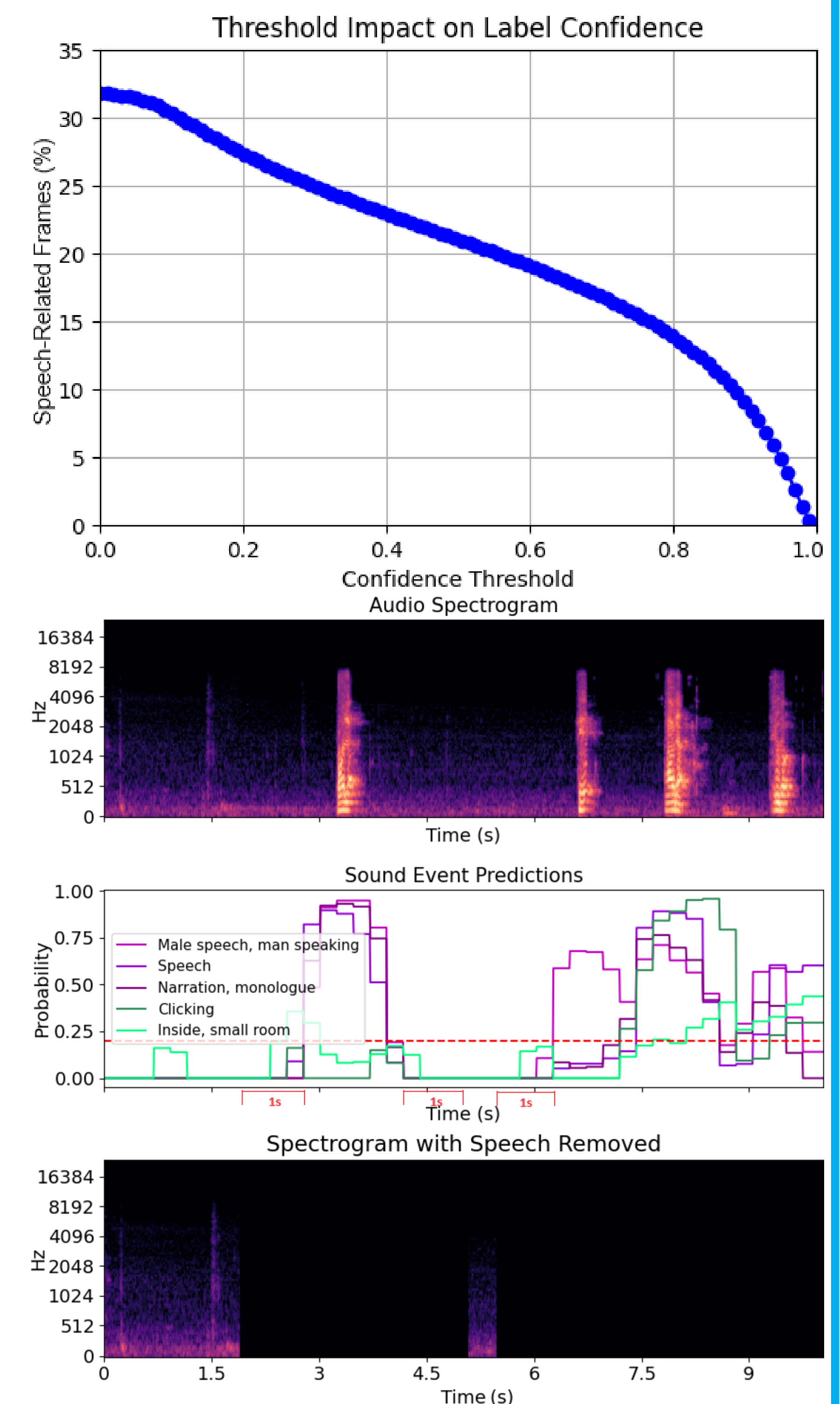
- Employed Pretrained Audio Neural Networks (PANNs)
- Used two specific models: Cnn14\_DecisionLevelAtt and Cnn14\_DecisionLevelMax
- Models pre-trained on AudioSet with 48% speech-labeled data

### 2. Feature Extraction and Analysis

- Extracted log-mel spectrogram features (frame size: 1024 samples, hop size: 512 samples)
- Analyzed 32 frames at a time for inference
- Saved the 7 detection results with most confidence in JSON format, including frame indices and probabilities

### 3. Threshold Analysis and Speech Removal

- Focused on distribution of predefined vocal-related labels
- Set confidence threshold at 0.2 based on statistical analysis and manual verification
- Removed segments containing speech-related labels above threshold
- Extended removal intervals by one second before and after detected speech
- Speech-related labels considered: Speech, Singing, Male singing, Female singing, Child singing, Male speech, man speaking, Female speech, woman speaking, Conversation, Narration, monologue, Music
- Replacement of identified speech segments with low-amplitude noise (1e-10).



## ACCESS TO DATASET AND ANONYMIZATION SOFTWARE

In the following link you will find: 1. The "The Sounds of Home" dataset, 2. The software used to remove the sections containing spoken voice and 3. A Jupiter Notebook uploaded to Google Colab where you can analyze your own audio files using our software. Just upload the file and you're done!

[https://github.com/gbibbo/voice\\_anonymization](https://github.com/gbibbo/voice_anonymization)

