Acoustics ML Voice Screening

## Problem Statement

Machine learning is increasingly used in acoustics to identify active noise sources. Such classification models have a wide range of applications, including:

* Voice recognition - understanding, interpreting and acting on spoken language
* Bioacoustics - surveying populations of noise-making animals (such as birds, frogs, or bats)
* Industrial noise assessment - monitoring noise emissions from industrial sites

Environmental applications (eg bioacoustics and industrial assessment) typically involve the installation of audio recording equipment for long-term monitoring. The equipment, by necessity, records high-quality audio data in order for the ML model to identify the noise sources.

However, this may result in undesirable data being captured. In particular, the voices of anyone in the vicinity will be recorded, potentially without their knowledge or consent. This is an ethical issue for the acoustic ML practitioners.

Our proposed solution is to apply a pre-screening stage to recorded data. Our system would identify sections of data that represent voices, prevent them from being stored, and exclude them from further analysis. In this way, the privacy of the public is protected and acoustic ML practitioners do not need to consider potential privacy impacts in the design and operation of their systems

## Stakeholder Mapping

| Stakeholder | Level of effect | Impacts |
| --- | --- | --- |
| Acoustic ML developers |  |  |
| Acoustic ML customers |  |  |
| Public users of monitored spaces |  |  |
| Government |  |  |
| Hardware producers |  |  |
| Data customers |  |  |

Required characteristics:

* Successfully able to distinguish between voice and non-voice audio
  + Should work for all voice types and languages (or as many as possible)
* Does not compromise later assessment
* Records discarded data timestamps for traceability

Desirable characteristics:

* Minimal computation power/time required (save battery)
* Adaptable to multiple input formats (time steps, frequency ranges) to minimise data reprocessing