Synthesising underwater acoustic datasets for noise robust sound event detection

Cuyx, B; Parcerisas, C; Debusschere, E; and Dietzen, T

Bram Cuyx
bram.cuyx@vliz.be
Flanders Marine Institute (VLIZ) / KU Leuven, Belgium

Underwater recordings in urbanised marine environments are characterised by the presence of high levels of continuous background noise. This noise can originate from human activity, such as wind farms, shipping, fishing activities, from natural sources like currents over hard structures, wind, waves or rainfall, or can come from the measuring instruments themselves. When analysing these soundscapes for distinct sound events (e.g. biological activity, ...), the noise presence negatively affects the detection ability and range. However, dealing with this noise is not trivial and care needs to be taken in its implementation. Pre-processing methods can shift the input data distribution, if preprocessing is not consistently applied in training and inference, the model may receive inputs outside of its trained data distribution, possibly degrading its performance. If one aims to evaluate the performance of sound event detectors including a pre-processing step, having a synthetic dataset with known event presence and different background noise levels is essential. In this work, we introduce an open toolbox that enables users to synthesise recordings made in different noise scenarios, controlling the noise origin and signal-to-noise ratio (SNR). It allows users to synthesise their own dataset of new recordings mixed from existing events and background noises. Using the toolbox, one has control over the number of events in each generated recording, the SNR range, and the sample rate. To isolate events for mixing, we introduce the use of event presence masks. For every event, a binary two-dimensional mask is first calculated, indicating where the event energy is located in a time-frequency representation of the audio. When generating new files using these isolated events, the masks are aggregated, indicating where event energy is present in the synthesised recording. The generated dataset is accompanied by metadata on aggregated masks, SNR, and events classes for each synthesised recording. Furthermore, evaluations of noise reduction in terms of SNR improvement and probability of misdetection are implemented. This allows one to evaluate the noise reduction separately from the event detector, optimising this step in isolation first.