Addressing the Bioacoustic Cocktail Party Problem for Wolf Population Estimates in the Greater Yellowstone

Jeff Reed, Peter Bermant, Cody Goldhahn

Recent advances in machine learning (ML) have revolutionized methodologies for acoustic wildlife monitoring by providing tools for processing and analyzing large quantities of real-world audio data. The improved computational techniques provide important insight into wildlife populations, interactions with humans and prey, and migratory patterns and facilitate the development of enhanced conservation and management strategies.

To date, however, the state-of-the-art ML-based acoustic detectors (such as Google's convolutional neural network (CNN)-based humpback detector) treat detection as a binary classification (or presence indication) problem. With this approach, the downstream conservation implications are limited since these detectors effectively ask only if one or more individuals are simply present at a given time.

One of the key challenges in expanding beyond presence indication is the bioacoustic cocktail party problem (CPP), a phenomenon in which conspecific and heterospecific individuals vocalize concurrently, often in the presence of biotic and abiotic noises. Recently, there have been attempts to apply deep learning (DL) techniques to address the CPP. Bermant, 2021 provided a framework for separating overlapping mixtures of calls produced by macaques, Egyptian fruit bats, and bottlenose dolphins; this study also represents the first attempt at solving the CPP with canids (at timestamp 2:40). Further, there have been additional studies exploring unsupervised source separation in birdsong, and there is a recent toolkit for overcoming the CPP with orcas.

Solving the bioacoustic CPP can have significant implications for wildlife monitoring and conservation. In particular, by providing information regarding the number of individuals present (as opposed to presence indication of one or more individuals), solutions to the CPP can enhance acoustic-based censusing pipelines. To that end, we propose a system for addressing the bioacoustic cocktail party with wolves in the Greater Yellowstone. We can take inspiration from advances in human, music, and bioacoustic source separation, and we can leverage algorithms such as MixIT to provide for source separation while eliminating the need for large-scale manual annotation. It is important to note that there remain computational, technical, and practical challenges to solving the CPP with real-world acoustic data, and we propose additional proxy tasks to simplify the problem while still answering key questions pertaining to wolf populations. For instance, instead of treating the CPP as a degenerate regression problem with the aim of separating mixtures of calls into predictions for the individual constituent calls, we can treat the problem as a multi-class counting problem with the aim of predicting the number of overlapping wolves in a chorus. With this approach, we can push the state-of-the-art beyond presence indication and provide real-world conservation impact by establishing minimum bounds on the number of wolves in a population at a given time.