
Detecting human and non-human vocal productions in large scale audio recordings

Guillem Bonafos^{*1}, Pierre Pudlo , Jean-Marc Freyermuth , Thierry Legou , Joel Fagot ,
Samuel Tronçon , and Arnaud Rey

¹Institut de Mathématiques de Marseille – Aix Marseille Université, Ecole Centrale de Marseille, Centre
National de la Recherche Scientifique – France

Abstract

We propose an automatic data processing pipeline to extract vocal productions from large-scale natural audio recordings. Through a series of computational steps (windowing, creation of a noise class, data augmentation, re-sampling, transfer learning, Bayesian optimisation), it automatically trains a neural network for detecting various types of natural vocal productions in a noisy data stream without requiring a large sample of labeled data. We test it on two different data sets, one from a group of Guinea baboons recorded from a primate research center and one from human babies recorded at home. The pipeline trains a model on 72 and 77 minutes of labeled audio recordings, with an accuracy of 94.58

Keywords: detection, classification, neural network, automatic pipeline, natural environment, baboon, baby, vocalization

^{*}Speaker