Using audio content and emotional response to predict soundscape perception through machine learning FREE

Volkan Acun; Semiha Yilmazer

Check for updates

This study is concerned with utilizing machine learning techniques for predicting soundscape perception by identifying the audio content of soundscapes and linking it with people's reported emotional responses. This research goal required developing an environmental sound classification model; however, the capabilities of these algorithms have some significant drawbacks. Supervised learning algorithms need a large number of labelled audio samples for each sound category. Given that a model for classifying environmental sound must be trained using a wide range of sound sources, this presents a substantial problem for developing a robust model that generalizes well to different environments. We prepared a convolutional neural network (CNN) based classifier; however, to tackle the limitations, we used musical instruments for the training dataset rather than environmental sound sources and optimized the neural network for this task. Based on how closely the soundscapes' audio content resembled the musical instruments in the dataset, CNN classified the soundscapes' audio content. We then conducted an online soundscape perception survey to evaluate participants' emotional responses to numerous soundscape clips. We prepared a feedforward neural network, which used the sound classification model's audio content output with the survey data to create a model for predicting people's responses to different soundscapes.