

From light to sound

BY S. CHANDRASEKARAN

September 9, 2024

September 9, 2024

Email: shiv@ucsb.edu

The aim of this project is to develop a system to aid visually impaired people to understand their surroundings with the help of a single camera (like the one on a phone or a laptop) and a stereo audio system (headphones attached to a phone or laptop).

When the camera is pointed at a scene the person should be able to hear words describing the objects in the scene (car, truck, building, dog, etc.), and the word for an object should appear to come from the location (direction) of the object in the scene. The amplitude (loudness) of the word should be inversely proportional to the distance of the object from the camera, while the pitch (frequency) of the word should be related to the radial velocity of the object (so objects moving towards the camera will have higher pitches than objects moving away from the camera). The idea is to give the visually impaired person an audio description of the scene.

It is anticipated that the project will leverage modern neural networks like **mast3r** or **DepthAnything**, to help with acquiring the 3D structures of the scene from a single camera. Similarly neural networks like **EVA** or **Yolo** could be leveraged to find and identify the objects of interest in the camera.

Converting this information into stereo audio signals can be done using traditional signal processing techniques.