Effect of Distance and Direction on Distress Keyword Recognition using Ensembled Bagged Trees with a Ceiling-Mounted Omnidirectional Microphone

ABSTRACT

Audio surveillance can provide an effective alternative to video surveillance in situations where the latter is impractical. Nevertheless, it is essential to note that audio recording raises privacy and legal concerns that require unambiguous consent from all parties involved. By utilizing keyword recognition, audio recordings can be filtered, allowing for the creation of a surveillance system that is activated by distress keywords. This paper investigates the performance of the Ensemble Bagged Trees (EBT) classifier in recognizing the distress keyword "Please" captured by a ceiling-mounted omnidirectional microphone in a room measuring 4.064m (length) x 2.54m (width) x 2.794m (height). The study analyzes the impact of different distances (0m, 1m, and 2m) and two directions (facing towards and away from the microphone) on recognition performance. Results indicate that the system is more sensitive and better able to identify targeted signals when they are farther away and facing toward the microphone. The validation process demonstrates excellent accuracy, precision, and recall values exceeding 98%. In testing, the EBT achieved a satisfactory recall rate of 86.7%, indicating moderate sensitivity, and a precision of 97.7%, implying less susceptibility to false alarms, a crucial feature of any reliable surveillance system. Overall, the findings suggest that a single omnidirectional microphone equipped with an EBT classifier is capable of detecting distress keywords in a low-noise enclosed room measuring up to 4.0 meters in length, 4.0 meters in width, and 2.794 meters in height. This study highlights the potential of employing an omnidirectional microphone and EBT classifier as an edge audio surveillance system for indoor environments.