A ROS-Based Audio Platform for Human-Robot Interaction

Chandra Suryadevara[1,3], Ramiz Hussain[2], Lukas Grasse[1,3] and Matthew Tata[1,3]
[1] Department of Neuroscience, University of Lethbridge, Lethbridge, AB, Canada.
[2] Department of Computing Science, University of Alberta, Edmonton, AB, Canada.
[3] Reverb Robotics Inc., Lethbridge, AB, Canada.
Email: matthew.tata@uleth.ca

INTRODUCTION

The alignment problem of robotics refers to the need to ensure that the goals and intentions of an autonomous system align with those of its human users. To ensure alignment and seamless human-robot interaction it is necessary to communicate our goals and intentions to robots. The most intuitive way to do this is through speech interaction. However, auditory perception and audio programming present unique hardware and software challenges in robotics. We set out to build an easy to use, ROS-ready audio system that can be conveniently integrated into existing and future robots. We demonstrate a prototype system that is lightweight, and can be quickly adapted to perform various audio tasks. As an example we demonstrated voice activity detection, speaker localization, and control of a Turtlebot2 mobile base.

MATERIALS AND METHODS

We configured Reverb Robotics audio hardware and software to include a ROS2 package for a mobile robot base (Turtlebot2). The Reverb Robotics system consists of a low-latency audio DSP and a Raspberry Pi compute module on a customized PCB with up to four far-field digital MEMS microphones. ReRo Server provides light-weight low-latency access to microphone outputs via the Google Remote Procedure Calls (gRPC) framework. We created a ROS2 package that captures the microphone data and publishes a multichannel audio stream to ROS for other nodes to process and use.

We then built a Direction-of-Arrival (DOA) finder incorporating speech onset detection with the SileroVAD [2] neural network. This node publishes a present/absent boolean for voice activity. We then used the GCC-Phat algorithm [1] to find the azimuth of the speech signal. Finally, we published the talker angle to a ROS topic which is interpreted by the TurtleBot control. To evaluate, we presented 20 trials of speech onset to the TurtleBot with the task of orienting towards the talker. We considered <20 deg error to be success.

RESULTS AND DISCUSSION

We built a ROS2 package that provides an easy interface with the capabilities of the Reverb Robotics audio hardware and software. We demonstrated the example use case of detecting and localizing a talker in the audio environment. This task was accomplished with a set of modular ROS nodes that conveniently expose this information to other nodes on the ROS network, such as control of a mobile robot base. Localization success rate is reported in figure 1. We found the 4-microphone array to be both low-latency and accurate even using lightweight voice detection and localization algorithms.

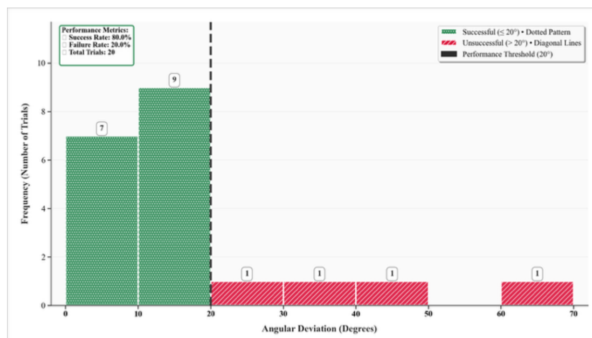

Figure 1. Speaker direction estimation. System accurately finds the talker to within 20 degrees, 80% of the time.

CONCLUSIONS

The developed system is easy to integrate into existing ROS-based robotics and provides an opportunity to build voice control and audio awareness into autonomous systems, especially for human-centric applications such as health care.

REFERENCES

[1] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* p. 320–327, 1976.

[2] S. Team, "Silero Voice Activity Detector," 2021. [Online]. Available: https://github.com/snakers4/silero-vad.