Pythia: Speech Interface for UNR's Autonomous Car

Sybille Horcholle, Bryson Lingenfelter, Nathan Thom
Advisor: Dr. Richard Kelley Instructors: Dr. Sergiu Dascalu, Devrin Lee
University of Nevada, Reno

CSE

Abstract

Pythia is a speech interface for UNR's self-driving car that allows intuitive two-way communication between driver and vehicle. Users can obtain real-time information about obstacle detection and vehicle status simply by asking - at no point needing to take their eyes off the road. The speech interface allows the car to make announcements when critical events occur, such as pedestrians entering the vehicle's path. By providing a simple and easily extensible interface to the vehicle, Pythia allows drivers to safely and quickly check that the vehicle is operating correctly.

Hardware

The speech interface has been developed specifically for UNR's self-driving car, a Lincoln MKZ shown in Figure 1. It is equipped with an on-board computer running Ubuntu that can access the vehicle CANBUS (controller area network).

The microphone chosen to capture voice for the interface is ReSpeaker Mic Array v2.0, shown in Figure 2. It is small enough to fit easily on the car dashboard and is able to detect the direction of the loudest audio input, helping reduce noise interference.

Figure 1. NAASIC's autonomous vehicle

Figure 2. ReSpeaker Mic Array

Interface Functionality

The primary functionality of the interface is responding to questions regarding the status of the car and its neighboring environment. Users can ask questions on obstacle detection, vehicle status, and odometry data. The GUI, shown in Figure 3, displays what the car most recently heard and said.

The interface is designed with flexibility in mind. The text size of the interface can be changed using a slider. The voice of the interface can be changed using a drop down menu. The wake word (used to activate the interface) can be set to any valid text.

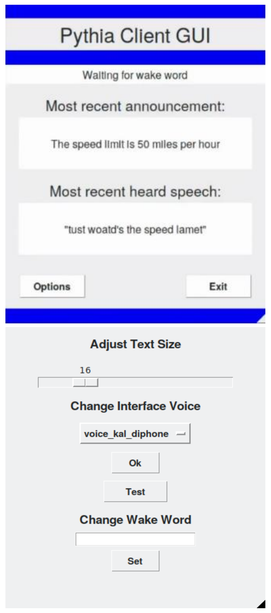

Figure 3. Interface main menu (top) and settings (bottom)

Event Driven Alerts

The interface provides announcements for situations where the driver should be immediately notified, such as a pedestrian entering the vehicle's path. Figure 4 shows an example where a pedestrian enters a crosswalk in front of the vehicle, triggering an announcement from the interface.

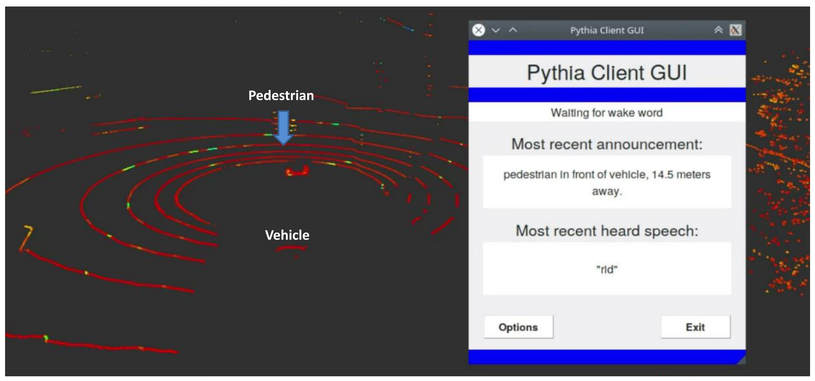

Figure 4. Interface responding to a pedestrian entering vehicle path

Open Source

As part of the Pythia project, an open source speech recognition package for ROS, the Robot Operating System, was developed and made publicly available. The package, shown in Figure 5, is now being used by several roboticists outside of UNR.

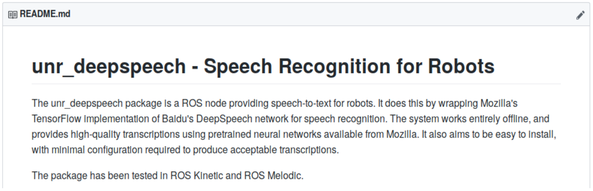

Figure 5. Github page for the unr_deepspeech speech recognition package

Future Development

Currently, Pythia receives status information from the vehicle and camera, and communicates it to the user. This provides an accurate representation of what the car knows, but it does not explain how the car is making decisions based on the data. To fulfill the project goal of allowing the driver to understand the car's intentions, the interface will need to be integrated with the car's planner.

Additionally, the interface needs to be optimized for a faster response time. Delays between questions and answers can reach a maximum of 8 seconds, which is too large for time-sensitive information.

Acknowledgements

We thank our advisor Dr. Richard Kelley, our instructors Dr. Sergiu Dascalu and Devrin Lee, and the CSE Department of the University of Nevada, Reno.