Robot Communication through Text-to-Speech
By Christian Onuogu and Jordan Taylor

Problem

When the segway robot has been given a task and is on the move, usually the only person who knows what the bot is doing is the task-giver. It is difficult for a casual observer to understand the state of the robot, especially if when process has an anomaly or was interrupted. If the robot ended up in the wrong place or lost, spinning helplessly while not accompanied by the task-giver, how would anyone unaffiliated with the robotics program know that it isn't functioning properly? This lack of understanding raises not only concerns of inefficiency and inconvenience, but also safety. There needs to be a way for the robot to communicate its actions and intentions more liberally with the users around it.

Task

To solve this issue, we intend to have the robot verbally communicate more of its intentions through text-to-speech. Such an interface would allow people in the vicinity of the robot to confirm their orders, understand the current objective of the robot, move out of its way, or decide if it needs assistance. For example, when the robot is enroute to a location, it should announce that it is heading to that room. When it reaches that location, it should say something along the lines of "I have arrived." If it gets lost or fails to localize, instead of just spinning helplessly, it should announce that it is lost and perhaps ask those around it to get someone from the lab. This could be integrated with the BWI robot planner, so that it could announce the actual task. We've even considered having the robot in a state in which it announces the room it just entered, conveying this information to anyone nearby who may need to adjust their position, or to someone who might not know a lot about the GDC.

Previous Experience

We've gained experience in ROS and C++ during the semester through the homework assignments and online tutorials. Perhaps some of the most pertinent experience was gained through the first part of homework 4 in which we delocalized the robot and moved it using teleoperation and navigation goals. This gave us a better understanding of how the robot moves and how the floors of the GDC are mapped in RViz, which may help us in determining what constitutes a room/location. Additionally, the lectures on planning and computer vision will

allow us to better understand how the robot may perceive and act upon stimuli, which will be useful in choosing what and when to vocalize.

Approach

The implementation of these features is a several step process. Firstly, we will study the *sound_play* package, a common choice for text to speech in ros that makes use of The University of Edinburgh's *Festival Speech Synthesis System*. Additionally, we will familiarize ourselves with the BWI planning and environmental sensors for the robots. Once familiar with both systems, the next step will be to write handlers to recognize common states of the robot that we would like to vocalize, such as 'delocalized', 'in transit' or 'entering a room'. Finally, each handler will call a speech function appropriate for the current state.

Writing handlers to recognize these desired states will likely be the bulk of the process, aside from refining the quality and accuracy of the speech. It is possible that the current system either does not recognize or ignores events that are of interest to this project, which would necessitate code that taps into the robot's sensory data to trigger those events manually. The difficulty of recognizing each event will likely determine which events will successfully have speech attached to them. However, as a loose guideline, we will seek to accomplish speech to text for as many of the following triggers as possible, listed in no particular order:

- Accepted a movement instruction.
- Obstructed and needs to reroute significantly.
- Entered a new room.
- Is delocalized.
- Successfully completed an instruction.

Evaluation of Success

Success hinges on the number of above proposed or additional events that have speech properly associated with them. To satisfy the condition for a successful speech association, the speech following an event must be: audible and understandable from a distance of at least a few meters away, given with correct and consistent timings, and non-destructive to the normal functionality of the robot. To these ends, we will test the program with several unassociated participants. Firstly with a listening test to ensure that the speech of the robot is clear and understandable. Next, we will simulate each event rigorously to guarantee that the correct announcements are given at the right time. This will be done, again, with unaffiliated test subjects and commands such as "global_localization", previously used to forcibly delocalize the robot. Finally, we will compare the results of the robot running with speech active and without

speech active, to certify that the speech feature minimally impedes the efficiency of the robot and runs smoothly. Given that there is little in house documentation of this feature, we will consider the project successful and extendable for future streams if we successfully associate speech with at least 2 of the previously mentioned events.

Expected Result

Given that the project is not too ambitious in scope, we expect to be able to accomplish all of the goals mentioned in the task section for at least 2 events. If there is room for failure, it will come from attempting to verbalize an event that is simply too difficult for the robot to recognize, such as announcing the room it just entered. For example, this would require spending quite a bit of time mapping each room to a range of coordinates, and even longer to ensure the process is always accurate. With extra time, we could perhaps get the robot to respond verbally to even more stimuli, such as that announcing a low battery, or that it needs help opening a door.

Related Work:

In terms of personal experience, many of the articles we read at the beginning semester relate to increasing communication between humans and robots, more specifically, Rosenthal, Biswas, Coltin, and Veloso's *An Effective Personal Mobile Robot Agent Through Symbiotic Human-Robot Interaction*, and *CoBots: Robust Symbiotic Autonomous Mobile Service Robots*. Both articles discuss the idea of a symbiotic relationship between humans and robots that is achieved by the robots being able to verbally communicate. We hope to strengthen this symbiotic connection through our work and to improve the perceived relationship between the robots and their users.

Many Autonomous Robotics FRI groups of past and present have, and are focused on the reverse, speech-to-text, but none on text-to-speech, meaning we have little local content to expand upon or compete with in terms of other projects. However, there is a plethora of online resources concerning text-to-speech using The University of Edinburgh's *Festival Speech Synthesis System*.

Timeline

A cautious proposed timeline is described below, and is subject to change. Due to the modular nature of the event handlers, it is less important the order of their implementation than that as many are implemented correctly as possible.

Date	**Milestone**
April 3	Turn in project proposal, begin first event handler.

April 13	Finish the study of our chosen tools and wrap up the implementation of the first event handler. (Likely for ‘Accepted an instruction’)
April 20	Partially complete another event handler, test the first..
April 27	Test and complete the second event handler, begin any attainable extra features.
May 4	Improve the quality of speech and finish up any extra features, while testing.
May 11	Turn in final report.
May 13	Present.