Project Lit Review

Group4

9212420231 Will Webster

2018007938MinKwan Kim(김민관)

2021029443 Yeeun Jeon(전예은)

[Articles]

1. Localization of sound sources in robotics: A review

https://www.sciencedirect.com/science/article/pii/S0921889016304742#sec4

To perform sound localization using sound, two or more microphone sensors are utilized. One prominent example of this approach is TDOA (Time Difference of Arrival). This method relies on the time difference between signals captured by two microphones. To calculate this time difference, a cross-correlation function is employed. This function compares the similarity of the two signals over time. Using this, the most similar signal between the two sensor inputs is identified, and the time it arrived is calculated to estimate the location.

2. Sound direction estimation using an artificial ear for robots (robot 1)

https://www.sciencedirect.com/science/article/pii/S0921889010002022?casa_token=YWH7yWFla3QAAAAA:ag9q59Cs_tc7128ts4dk6uQmUUmLU1K2_NW4wvAVMSfp1YMTwmnvplaijDwK85ZpBCPviXS74Wmu

Humans use both ears to estimate the direction of sound. Similarly, artificial ears are created to mimic this process and infer the direction of sound. This is achieved by utilizing two methods: ICTD (Inter-Channel Time Difference) and ICLD (Inter-Channel Level Difference), which replicate how humans use both ears to perceive the direction of sound. ICTD is defined as the difference in the arrival times of a sound wave-front between two microphones. Likewise, ICLD is defined as the difference in sound pressure levels between two microphones.

3. Kinematic model of three wheeled mobile robot Dayal R. Parhi* and B. B. V. L. Deepak

http://200.19.248.10:8002/professores/stefanie/2011/engenharia/Parhi%2520and%2520Deepak.pdf

We are planning to create a moving robot using two fixed wheels. Therefore, we need to calculate how to control these two wheels in a way that allows the robot to move in the desired direction and at the desired speed. The fixed standard wheel used in this context lacks a vertical axis for steering, resulting in a fixed angle relative to the chassis. Its movement is constrained to back-and-forth motion within the wheel plane and rotation around its point of contact with the ground.

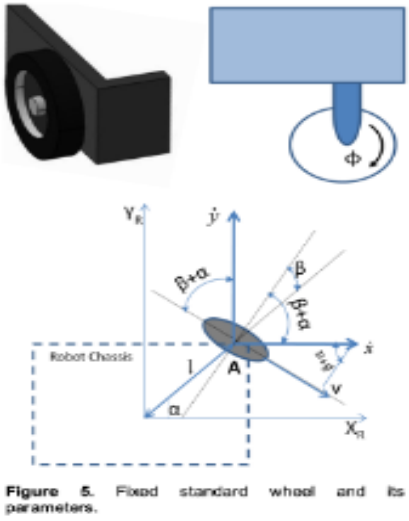


Figure 5. Fixed standard wheel and its parameters.

4. Portable Keyword Spotting and Sound Source Detection System Design on Mobile Robot with Mini Microphone Array (robot 2)

https://ieeexplore.ieee.org/abstract/document/9108086

This paper details experiments conducted using a robot that has a 6 microphone array embedded on the body of a robot. The experiment in the article uses the SMART-4WD mobile robot platform that is controlled by an Arduino UNO. The goals of this experiment are both sound source detection and keyword spotting. They use the Frequency Domain Binaural Model (FDBM) to detect the Direction of Arrival (DOA) of the sound source. To estimate the distance using triangulation as the robot moves.

This can be helpful in our project because we also plan on doing sound source detection in our project. The accuracy for spotting DOA in the article's experiment is 90%; however, error when estimating distances was around 30%. We are also using an Arduino UNO to control our robot. This

would be a good starting point for the library "shusher" idea.

5. A Sound Source Tracking Humanoid Robotic Platform: Design and Architecture

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&as_ylo=2019&q=robot+humanoid+%22ears%22&btnG=

This article proposes a design for a sound source tracking system that can be implemented in a humanoid robot. The design closely resembles a human head; however, it uses 4 microphones instead of 2 in order to better detect in 3D space as opposed to 2D. This design also includes 2 axes of rotation: 90 degrees to left and 90 degrees to right as well as 45 degrees up and 45 down. The main objective of this design is to allow the robot to detect the direction of the noise coming from the subject in a conversation and orient the head towards the subject.

A robot orienting its head towards a human in conversation is one piece of the large puzzle of an anthropomorphic human-robot interaction. This piece could be feasible in one semester. We want to have a robot that will turn towards a human when the human is speaking to the robot.

6. Local LLM

https://www.infoworld.com/article/3705035/5-easy-ways-to-run-an-llm-locally.html

This would give the 4-eared head a voice. A head that can orient itself towards a user is the first step, but there needs to be some sort of chat feature. Modern language models such as Chat-GPT offer a better conversation for people than virtual assistants such as Siri, Google Assistant, and Alexa; however, there are two major drawbacks: 1. This would require constant WiFi connection. 2. There is no privacy in using it whatsoever. With privacy being important to people, many will not want a chat bot in their homes that collects their data and sends it off.

GPT4All is a locally run Large Language Model (LLM) who's desktop client is simple to use and has a fast response time. It requires Python to run. We can look into how we use Python to automatically send a message in the desktop client and extract.

7. You just do not understand me! Speech Recognition in Human Robot Interaction

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1ac19e217b1e0e28e396e02c6e8ede84752c82fe

In this paper, authors implement many tests on speech recognition. They compared speech recognition friendly artificial language(ROILA) and English and found out that there was no

significant difference between them. They also found out that the type of microphone and robot's head movement had a significant effect on speech recognition. In conclusion, they suggest implications for Human Robot (Speech) Interaction. For building our robot, we have to use speech recognition technologies to understand people's words. This paper recommends ways to improve speech recognition rate such as choice of the microphone, so we can reference this paper in terms of speech recognition.

8. ASKA: receptionist robot with speech dialogue system

https://ieeexplore.ieee.org/abstract/document/1043936

This paper is about a humanoid robot, ASKA, a receptionist robot with speech dialogue system. ASKA can recognize a user's question utterance, and answer the user's question by its text-to-speech voice, hand gesture and head movement. This paper describes the speech related parts of ASKA. ASKA can deal with a wide task domain of 20k large vocabulary using a word trigram model and an elaborated speaker-independent acoustic model. ASKA can also make a response with keyword and key-phrase detection in the N-best speech recognition results. Our robot has to understand user's questions and answer to the user by text-to-speech technology, so we can reference this robot and we can use some technologies we need for our robot.

9. Robots that Use Language

https://www.annualreviews.org/doi/abs/10.1146/annurev-control-101119-071628

This article surveys the use of natural language in robotics from a robotics point of view. To use human language, robots must map words to aspects of the physical world, mediated by the robot's sensors and actuators. This paper describes central aspects of language use by robots, including understanding natural language requests, using language to drive learning about the physical world, and engaging in collaborative dialogue with a human partner. Also this paper describes common approaches, roughly divided into learning methods, logic-based methods, and methods that focus on questions of human–robot interaction. Our robot has to do various language related tasks such as understanding human speech and making text into speech, so this paper works as a guide to learn about language in robotics.

<Overview>

These days, Chat GPT is a widely used and beloved tool among young people; however, GPT does have several drawbacks. Among them, one is the necessity for a device to run GPT and a constant internet connection. Additionally, individuals with visual or physical impairments, as well as those who are not familiar with such devices, may find it challenging to use GPT; therefore, we aim to create a LLM that can be used just like conversing with a human. A benefit of running our own LLM, as opposed to using Chat GPT, is we can run it locally. This means we will not need a wifi connection, and we can assure privacy for the user. Our robot will use this LLM to respond to people's voices, locate the source of sound, and provide answers to their questions, much like asking an information assistant. The nine articles above are related to creating a robot that operates by detecting sound direction, recognizing voices, and responding to them.

ex) we can use our robot to automate the role of a receptionist at an information desk to answer questions visitors may have about the building.

This is the rough idea for our robot: Our plan for our final product is our robot will remain idle until a keyword is spoken such as the robot's name. When someone says its name, the robot will turn to the direction its name was spoken to indicate to the user it is listening. The user will then ask the robot a question. The robot will send the message to a computer via bluetooth or wifi to the computer that is running the LLM. The LLM will produce the response and send it back to the robot. The robot will either display the message as text or convert it to an audio output.

[3 Robots]

1. Robot Dog - Spot

https://www.youtube.com/watch?v=Y1-s37zrm1M

Spot is a compact, nimble four-legged robot that can trot around office, home or outdoors. It can map its environment, sense and avoid obstacles, climb stairs, and open doors. Spot is equipped with ChatGPT, so if people ask questions it can give answers. Inside the robot, JSON is given to ChatGPT and explain what the structure is and how to read that JSON, and ChatGPT can answer questions about that JSON. Spot uses Google Text to Speech to speak.

2. Robot Ear

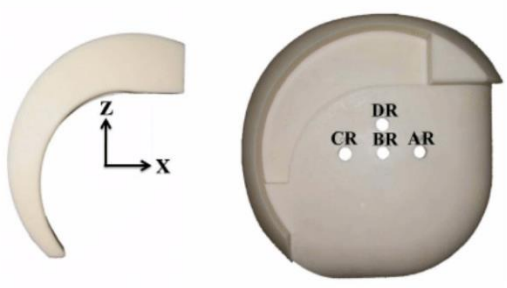


Fig. 1 Designed ear and microphone positions.

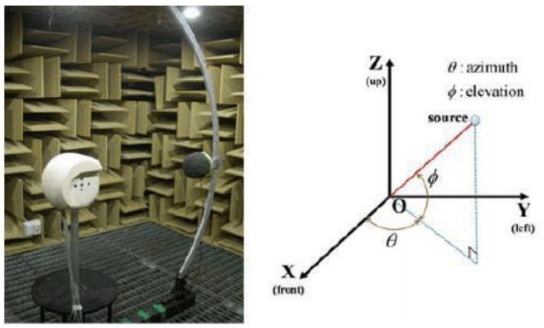
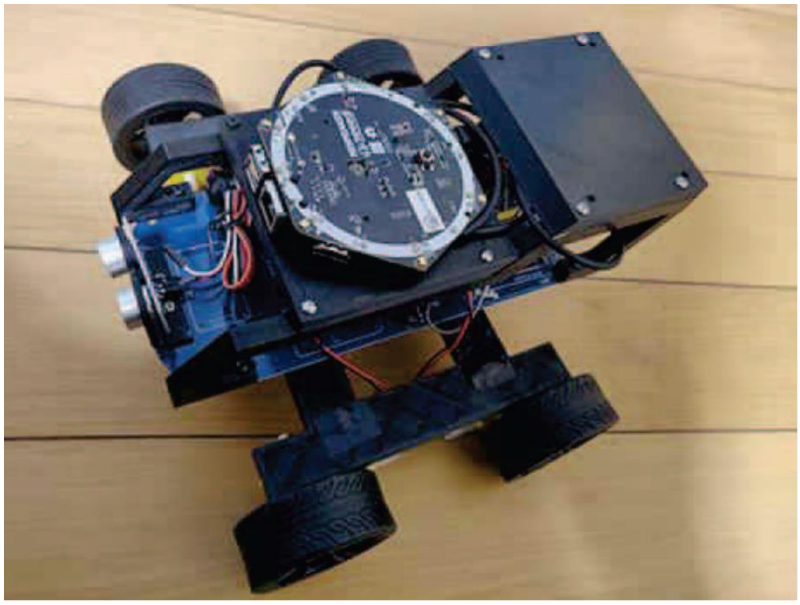


Fig. 2 HRTFs measurement setup (left) and vertical-polar coordinates (right).

Robot ears hear sound and tell the direction of the sound.

3. Robot using 6 microphone array controlled with Arduino



This robot uses the 6 microphone array to detect the direction and estimate the distance to a sound source as well as keywords