October 30, 2017

Visual Matching System

6.111 Project Proposal

Ertem Nusret Tas and Faysal Shair

1) Overview

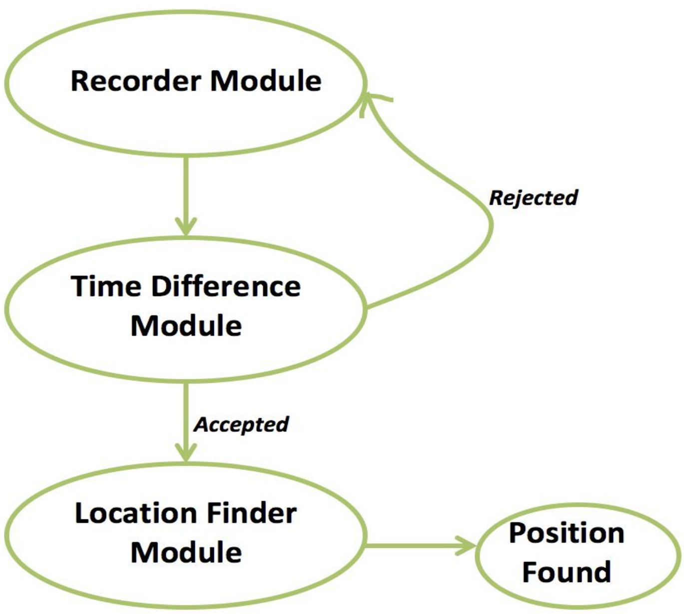
There are many different ways to detect the position of an object. For years this has been done in all sorts of industries, using a variety of different techniques. Some tracking methods rely on the physical properties of the objects, while others rely on digital processing and machine learning, such as image recognition.

Our goal in this lab project is to be able to detect the position of an object based on its sound properties. In the real world, there are sound disturbances in the background, which need to be ignored. That said, our system will be able to ignore invalid sounds such as white noises or sine waves that do not change over time.

Once the object's position has been correctly identified on a given plane, we will have a camera point in the direction of the object. This could be useful in presentations, where the camera can track the presenter based on voice. It could also be useful in a security surveillance setting, where the CCTV camera faces the direction of the sound.

2) Design

The design consists of three main modules: Microphone module, Time Difference module and Location Finder module. The flow chart below briefly shows the sequence of processes that the sound signals go through.



I will briefly introduce the three main modules here. They will be discussed in depth below.

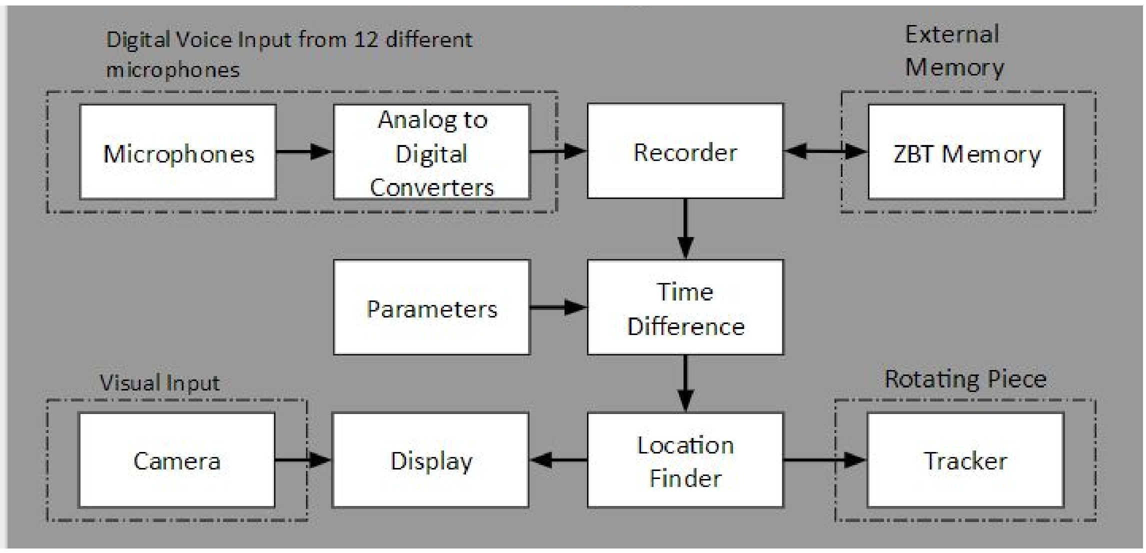
1. Recorder Module: Each microphone will have its own module. They basically receive the sound signal, and send it to the analog-to-digital converters. After the signal is changed to digital form, it is stored in a ZBT memory.

2. Time Difference Module: In this module, we find the difference between the arrival times of a sound wave at two microphones positioned on the same axis. Since there will be four microphones, two positioned on the x-axis and two positioned on the y-axis, this module will output two time delay values corresponding to the x and y coordinates of the sound source.

3. Location Finder Module: After the time differences are calculated, we move to the location finder module. Here, we basically detect the location of the sound source on a given plane, namely its x and y coordinates, based on the time differences between the microphones.

3) Implementation and Testing

3.1) Block Diagram



3.2) External Components

Four external components are needed for the system: 4 microphones, 4 ADCs, a camera and a servo to rotate the camera. We will place the microphones on the same plane, namely the **microphone plane**, such that two of the microphones lie on the x-axis and two on the y axis. However, their distance from each other will be determined by trial-and-error until we are satisfied with the level of precision. For instance, if sound waves are observed to hit the microphones within very small numbers of clock cycles, then, microphones will be placed further apart. On the other hand, since we would like to identify the same sound across the all 4 microphones, they will have to be within a certain range of each other. Hence, influence of these factors will have to be be examined prior to the microphones' placement.

Secondly, the camera will be attached to a servo that will be placed approximately in the center of a predetermined microphone plane. Since our system aims to detect the source location in 2 dimensions, this source will be assumed to move on a plane parallel to the microphone plane, namely, the **sound plane**. The servos will rotate until the camera points toward the sound source. Finally, for the case where the sound and microphone planes are different, we might need to add more servos to give the camera more degrees of freedom.

3.3) Verilog Modules

3.3.1) Recorder Module (Faysal)

This module will receive the sampled sound signals from the ADCs as its input and store them inside the external ZBT memory. It will also be used to access the sound samples at the ZBT memory addresses indicated by the time difference module. Four instances of this module will be used in our system corresponding to the four microphones.

The recorder module will be based on the design provided in the lab 5 with only minor modifications if deemed necessary.

3.3.2) Time Difference Module (Nusret)

This module periodically retrieves the sound samples recorded by the 4 microphones and calculates the difference between the arrival times of a sound wave at the two microphones on the same axis. In other words, when a sound signal is incident on two microphones on the same axis, one of the microphones (the one closer to the sound source) will receive the sound before the other microphone. To calculate how long it took for the sound wave to hit the second microphone after hitting the first, we shift the memory window of the second sample over the first sample and find the difference between the two samples. This difference is called the **intensity difference**, and it will be minimized when we have shifted one of the signals by approximately the time difference.

The intensity difference is calculated as the sum of the squared differences between the intensities at the memory addresses referred during the calculation. As the module can sum over different number of addresses over the time, intensity difference is normalized by the number of addresses that were considered. Once the amount of shift in memory addresses is determined for the two sample sets, time difference is derived by multiplying this shift with the sampling rate of the microphones. Finally, the same procedure is applied for the top and bottom microphones on the y-axis to find the delay between these two microphones.

There are three parameters that have to be tuned to enhance this module's performance. These parameters allow us to filter out any background sounds. Therefore, we only consider valid sound signals, and not every sound signal, because this is would computationally expensive and would violate the system's time constraints.

The memory window is the set of addresses that we consider. If we choose to use a very small memory window (only a very few samples), then the tracking system may arrive to incorrect results. That said, it is important to choose a memory window that is not too big, because that would slow down our intensity difference calculation, and not too small.

However, if the size of the **memory window**, which refers to the set of considered addresses, is very small, intensity difference might prompt incorrect responses by the tracking system. Hence, it is important to find the right balance for the memory window size.

Second, it is possible that the two sets of samples combine different sounds coming from different directions. In this case, the intensity difference might never be low enough to predict the location of one sound source. Hence, we should set an **intensity threshold** value such that if the minimum intensity difference is above the threshold, the system will not make a prediction for the time delay. Again, the value of this threshold should reflect the right balance between the system's fault-tolerance and its response rate.

Finally, **calculation frequency of the time difference** should be adjusted to maximize the accuracy and the response rate of the system. If the size of the memory window is kept too low, corresponding to a high frequency of calculation, the small size might prevent the system from identifying the correct time difference between the two microphones on any axis. On the other hand, if the frequency of calculation is too low, then the camera angle would not be updated soon enough for the human eye to perceive a continuous tracking of the sound source. Consequently, this frequency should be high enough to render the tracking continuous from a human's perspective.

3.3.3) Location Finder Module (Faysal)

This module receives the two time differences between the upper-lower and left-right microphones and calculates the location of the sound source with respect to the location of the microphones (or the camera). It provides this location to the display module so that the camera axis can be updated accordingly.

Basic trigonometry will be used to estimate the location of the sound source. The main challenge in the design of this module will be implementing certain trigonometric functions such as arctan etc. in Verilog. Moreover, this module will have to solve two equations corresponding to the coordinates at two axis, together to find the correct location in 2 dimensional tracking. Consequently, it might require multiple clock cycles for calculating the location.

3.3.4) Tracker Module (Nusret)

This module will take the location estimate from the location finder module as its input and rotate the camera so that it faces the location of the sound source. In the case that there is a hardcoded offset between the microphone and sound planes, the camera will be rotated until its axis cuts the sound plane at the point where the sound source is located. The outputs of this module will be the control signals for rotating the servos on which the camera is attached.

3.3.5) Display Module (Faysal)

This module will take the camera image as its input and display it on the screen. It will also display the sound plane and the location of the sound source on this plane. Finally, if there is time left for our project, we will enable it to zoom in or out of the displayed image depending on a user input.

3.3.6) Parameters Module (Nusret)

This module will enable the users to modify certain parameters related to the system such as the memory window size, intensity threshold and the calculation frequency of the time delay. We should note that each input signal will be debounced prior to its use in the system.

4) Goal and Objective

In order for the end goal of this project to be achieved, each stage needs to be completed successfully. That said, our base goal is essential for our expected goal and reach goal.

Base Goal: Our base goal involves detecting the location of a sound source on one axis whose y (or x) coordinate is fixed on the sound plane. By detecting this location over time, the camera will be able to track the sound source as it moves on the given axis.

Expected Goal: After achieving the base goals, we plan to detect the source location in two dimensions on the sound plane. Although we will implement the functionality in the location finder module to track the source on a sound plane parallel to but different than the microphone plane, demonstration of this capability is restricted by the capabilities of the servos on which the camera will be attached. Hence, we will restrict our demonstration to the case where image and the microphone planes overlap with each other.

Reach Goal: Our reach goal involves demonstrating the functionality of the location finder module to track the sound source on a sound plane that is different from the microphone plane. Finally, we can try to extend the system into three dimensional detection of the source location as the ultimate case, with 6 microphones and a camera that can rotate to face any point in the space.

5) Project Timeline

	Implementation	Verification
Week of November 6	Buy microphones and the ADCs. (Nusret and Faysal) Modify the recorder module from lab 5 with ZBT memory implementation. (Faysal)	Test the microphone, ADC unit. (Nusret and Faysal) Test the recorder module and its memory. (Faysal)
Week of November 13	Start working on the time difference (Nusret) and location finder (Faysal) modules.	Test the time difference and location finder modules. (Faysal and Nusret)
Week of November 20	Complete the time difference and location finder modules. Combine the modules. (Faysal and Nusret)	Test the combined system. (Faysal and Nusret)
Week of November 27	Implement the display (Faysal) and tracker (Nusret) modules. Implement the Parameters module (Nusret).	Test the combined system again with the added modules from this part. (Faysal and Nusret)
Week of December 4	Debugging	Debugging
Week of December 11	Debugging (if necessary)	Debugging (if necessary)