Localization of Natural Disaster Survivors Through Drone-based Sound Source Localization

Abstract

Most natural disaster casualties arise from entrapment underneath collapsed structures, as disaster-relief efforts can be hindered by the danger and inaccessibility. Consequently, unmanned aerial vehicles (UAVs) can be used for survivor detection. Some UAVs are mounted with microphones so that shouts or whistles can be detected in low-visibility conditions. In order to determine the location of a sound of interest, UAVs use sound source localization (SSL), but current methods can be inefficient or inaccurate. A novel method of SSL was designed using the volume ratios of the left and right channels of a stereo microphone to calculate the horizontal and vertical angles of origin of an incoming sound within a given frequency range.

A 1,760 Hz sound was played at seven locations, each a multiple of 30 degrees 20cm away from a stereo microphone. The Wolfram Mathematica programming language was used to compute seven left/right ratios that each corresponded to an angle. These served as the groundwork for an exponential model that could output an angle of sound origin when given a volume ratio. A band-pass filter was applied so that a specific range of frequencies could be analyzed. The system calculated vertical and horizontal angles of origin with an average of 4.9 degrees off from the actual position in quiet conditions, and 19.2 degrees in loud conditions. Calculations took around 0.03 seconds. The accuracy and speed of the system show the viability of this form of SSL, and its potential in life-saving applications.

Contents

Hypothesis

A system that calculates the horizontal and vertical angle of origin of a specified incoming frequency range can be designed using the volume ratios of the left and right channels of two stereo microphones. The system would offer a more efficient and effective option for drones equipped with microphones and optimized for disaster relief.

Background Research

1. Usage of Drones in Disaster-Relief Roles

In the wake of a large-scale natural disaster, short-term response relief efforts typically span from 0 to 30 days after the initial event. During this phase, evacuations, supply delivery, and medical care are the most critical tasks (Labrador & Cheatham, 2020). Robots have been used around the world to survey the vicinity, deliver supplies, and locate survivors for several years, with the first recorded usage originating from the September 2001 terrorist attacks on the World Trade Center in New York City. In addition, unmanned aerial vehicles (UAVs) work to facilitate survival rates through real-time mapping from the sky as well as the detection of survivors (Feuilherade, 2017). Most UAVs are operated autonomously, and a wide array of sensors ranging from infrared cameras to wifi antennae allow for the best chance of survivor detection.

2. Current Methods of Sound Source Localization

There are currently several sound source localization (SSL) systems optimized for drone usage, but many challenges hinder their effectiveness and time-efficiency. Wind noise and drone motor noises can disrupt microphone readings and obscure sounds of interest, and the nature of search-and-rescue requires very fast turnaround times. Some forms of SSL use machine-learning algorithms, but these require a large number of complex datasets for training, which can be difficult to obtain, leading to the usage of simulated data (Deleforte et al., 2019). Simulated data is not optimal as it can reduce the accuracy of a machine-learning model. Other systems analyze the slight time difference of sound arrival between left and right-facing microphones.

Materials List

- Zoom H1 Handy Recorder (1x)
- Zoom H1N Handy Recorder (1x)
- Wolfram Mathematica 12.1 Computer Language (1x)
- Intel® Core™ 10700K CPU @ 3.80GHz Desktop Computer (1x)
- iPhone 7 Plus (1x) (for testing and tone generation)
- *Sonic* application (1x) (for testing and tone generation)
- 3,300Hz Rescue whistle (1x)

Variables

1. Independent

- The location of a whistle or speaker emitting noise that lies within 300-8,000Hz
- Levels of background noise

2. Dependent

- The horizontal angle of origin of the sound of interest, where zero degrees is to the immediate left of the microphone and 180 degrees is to the immediate right
- The vertical angle of origin of the sound of interest, where zero degrees is immediately below the microphone while 180 degrees represents the location directly above
- Time taken by the program to calculate results

Experimental Procedure

1. Measurement of Microphone Response

One hand-held stereo microphone was placed on a carpeted surface in a room that minimized sound reverberation. Pieces of tape were situated at seven different locations, all 30 centimeters away from the microphone and spaced 30 degrees apart, as shown in Figure 1.

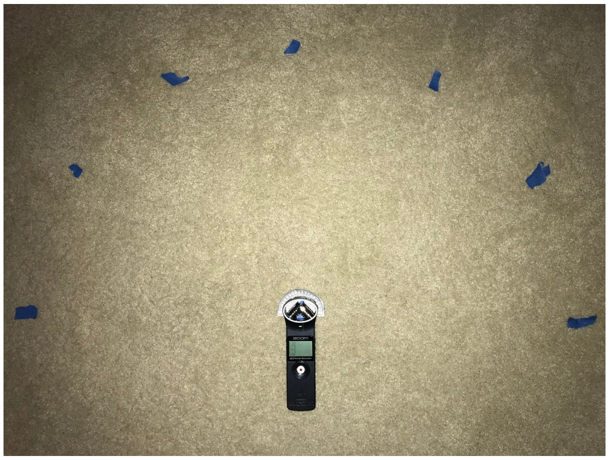


Figure 1

Then, a 1,760Hz sound was played from an iPhone speaker at each location for three seconds, with three seconds of silence in between. While 1,760Hz does not fall within the comfortable human vocal range, it would likely be the frequency of a rescue-whistle. Frequency does not affect the system's ability to localize sound, for the system of SSL designed relies on relative volume instead. The process was repeated so that both microphones (Zoom H1 Handy Recorder and Zoom H1N Handy Recorder) had one audio file. Two microphones were used so that both vertical and horizontal directions could be localized in the future.

2. Analysis of Preliminary Trials

After recordings were obtained, the audio data was analyzed so that the measured volume of the left channel could be divided by that of the right-channel so that each of the seven angles (0, 30, 60, 90, 120, 150, and 180 degrees) could be paired with its corresponding volume ratio. First, the

RMS amplitudes of the left and right channels were calculated using the AudioLocalMeasurements command in the Wolfram Mathematica language. The code below outlines how the command was used.

$$leftVolume\ =\ AudioLocalMeasurements[Audio[leftChannel],\ "RMSAmplitude"]["Path"];$$

$$rightVolume\ =\ AudioLocalMeasurements[Audio[rightChannel],\ "RMSAmplitude"]["Path"];$$

Calculating the volume allowed the data to be converted to a list of numbers, with each number representing a volume level at a specific point in time. Figure 2 shows the data of the left channel's volume levels as a function of relative time. 7 distinct blocks are clearly visible, representing the three second tones.

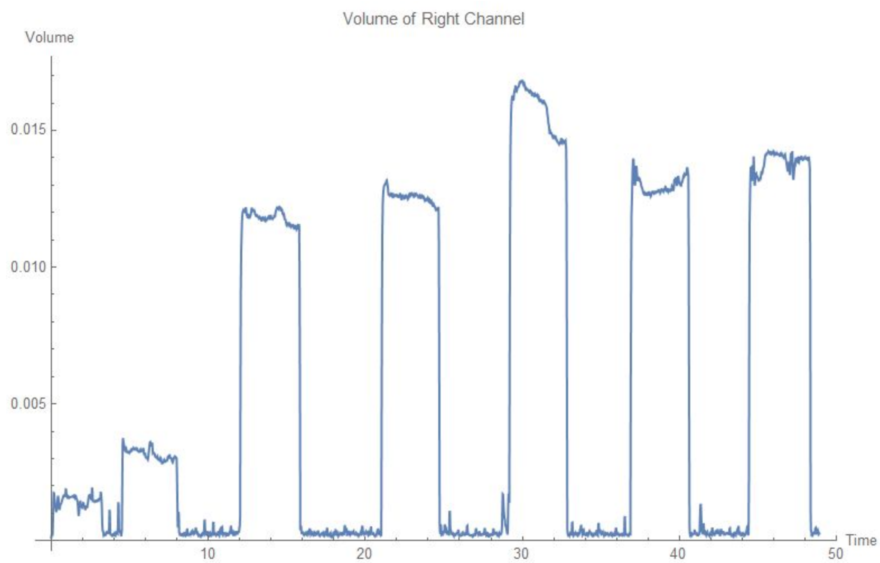


Figure 2

An automatic segmentation algorithm was written so that only the seven chunks data chunks of interest could be analyzed. Any volumes below a threshold of 0.001 RMS Amplitude units were equated to 0, and the SequenceSplit command was used in Mathematica to extract all relatively loud and continuous sounds. In addition, only segments above a length of 80 time units were considered, as three seconds of recording translated to approximately 100 time units. This step ensured that the seven tones would be the only blocks of sound analyzed, and any other loud occurrences would be filtered out. With a total of 14 sets of volume data (seven for the left and right channels), the average levels were calculated and divided by each other to produce a one-dimensional list of ratios. The list of averages was plotted on the XY-plane with the seven angles on the X-axis and the volume ratios represented on the Y-axis. Each input corresponded to exactly one input, meaning that if a ratio was supplied with an unknown angle, that angle could be determined based on the data. A mathematical model was then created to accurately represent the seven ratio data points. The FindFit command was used to estimate several coefficients of a logarithmic function, as shown in the code below:

$$FindFit[ratios,\ \{a * Log[x/b],\ \{a < 0,\ 8 < b < 9\}\},\ \{\{a,\ -70\},\ \{b,\ 9\}\},\ \ x]$$

The command generated two coefficients that could then be used in the following function:

$$-43.154321712525416\ Log[0.12974412755601705\ ratio]$$

The resulting model could return an angle within the range of 0 and 180 degrees based on the left and right volume ratios, while still adhering to the initial seven reference points, as shown in Figure 3, where the dotted red line represents the model and the green points represent the original ratios.

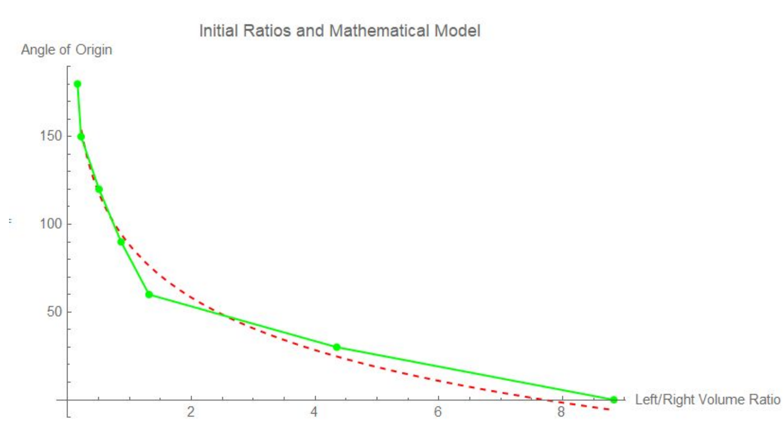


Figure 3

3. Selection of Specified Frequency Range

SSL relies on a substantial input of audio of interest, meaning that background noise can hinder localization performance. Most background noise, such as wind or crashing sounds, occupy low frequencies, so code that applies a band-pass filter to inputted audio was written so that only frequencies between 300Hz and 8,000Hz would be analyzed. The code below shows how the BandpassFilter command was used.

$$BandpassFilter[\ audio,\ \{Quantity[300,\ "Hertz"],\ Quantity[8000,\ "Hertz"]\}]$$

Figures 4 and 5 show an example of one test trial before and after a filter was applied; the lower frequencies are noticeably weaker once filtered.

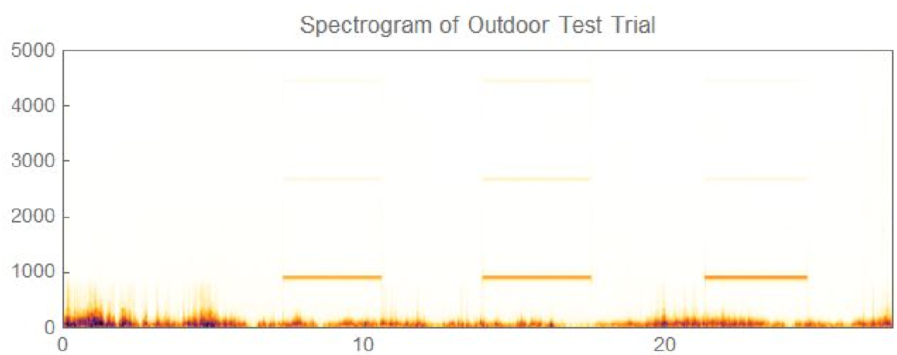


Figure 4

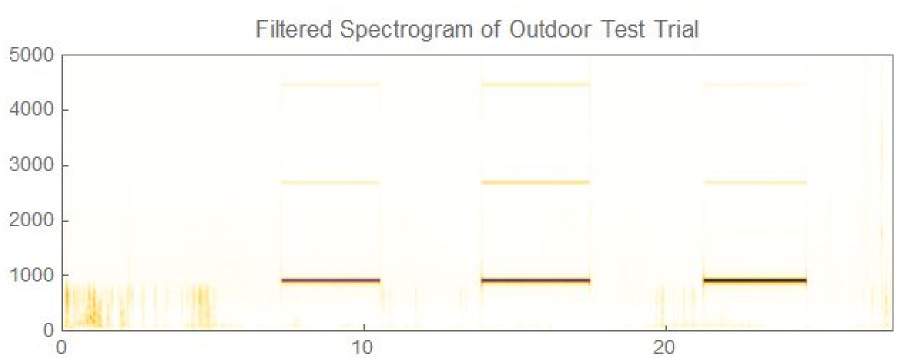


Figure 5

All the necessary components and functions were compiled with some intermediate code to smoothen the transitions between them. The final algorithm could receive an audio input and output the angle of direction of a scream for help or a whistle sound.

4. Accuracy Testing

Once the program had been written, several trials were conducted to assess the accuracy of the design as well as its ability to hold up in environments with high levels of background noise. To test the latter, an experiment was set up where both horizontal and vertical positions were tested at 45, 90, and 120 degrees. These angles were chosen to both assess positions previously calculated (in the calibration trials) as well as new angles that had not been previously calculated. One microphone was placed on a short ledge outdoors in high-wind conditions as shown in Figure 6, and both vertical and horizontal angles were tested by first recording when the microphone was laying flat and then sideways. The goal of these trials was to see how effective the system was in windy conditions.

Figure 6

Next, test trials were conducted indoors to assess the overall error range of the system when a significant level of background noise was not present. This time, the vertical and horizontal positions were calculated simultaneously, and a 3,300Hz whistle was blown five times in each area to create replicates. The whistle was positioned at two positions: 90 degrees horizontal/110 degrees vertical, and 70 degrees horizontal/100 degrees vertical. While varying angles were used, this time only obtuse vertical angles were chosen because survivors would most likely be situated underneath a UAV. Figure 7 shows how the two microphones were positioned so that both directions were detected.

Figure 7

Data Analysis and Discussion

1. How the Preliminary Trials Changed

Five trials were conducted before an adequate series of ratios were acquired. The first four trials used 13 distinct ratios by recording the controlled trial with angle multiples of 15 degrees rather than 30. In addition, three out of the five trials were conducted in large environments on hardwood surfaces, which introduced unwanted variables including reverberation. When the preceding trials were analyzed, their angle/ratio pairs were not suitable as some outputs (ratios) corresponded with multiple inputs (angles). Figure 8 shows an example of one such unsuitable set of data, where the graph does not pass the "horizontal line test." The green dotted line represents one area where one ratio corresponded with multiple distinct angles. If the data were to be used as a basis for the design, a microphone measuring a ratio of 0.5 could result in angles of approximately 105, 149, 160, and 170 degrees.

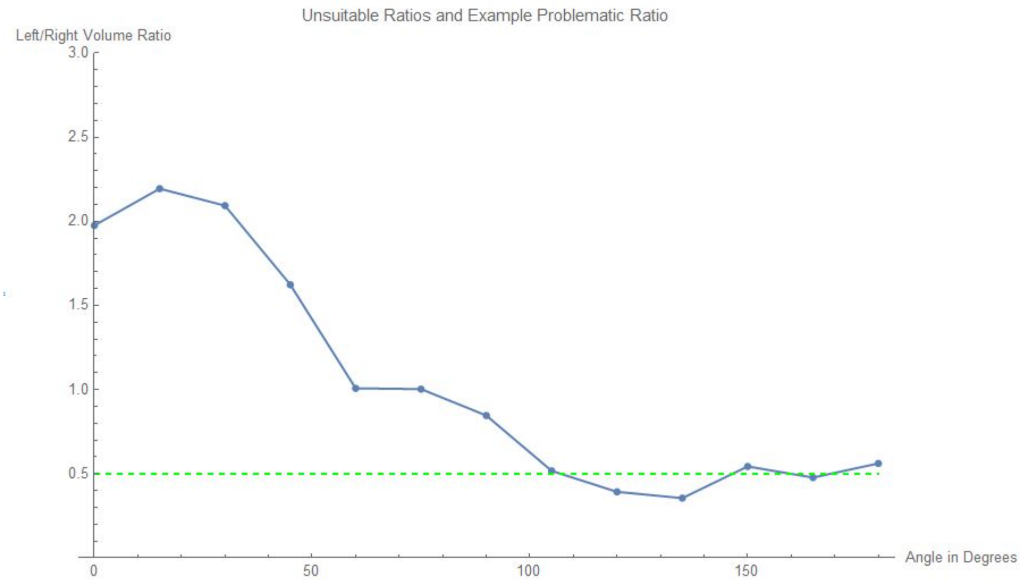


Figure 8

2. Diagnosis of the Problem

In order to isolate the source of this repeated scenario where "bumps" would appear in the ratio-angle graphs, the effectiveness of the microphones were assessed. The microphones could have been recording in inconsistent patterns where sounds originating from certain locations were measured in different ways. A 13 angle trial was conducted to test the possibility, but this time one microphone was placed upside-down. As shown in Figure 9, the ratio graph of the result showed two trends that mirrored each other along the x-axis. The mirror-image pattern showed that the microphones recorded in an unbiased way, meaning that irregularities could be originating from unwanted variables such as reverberation and the orientation of the speaker. If the trends had appeared to be reflecting across the line x=90, it would have been highly likely that some angles were being recorded by the microphones in a non-cardioid manner.

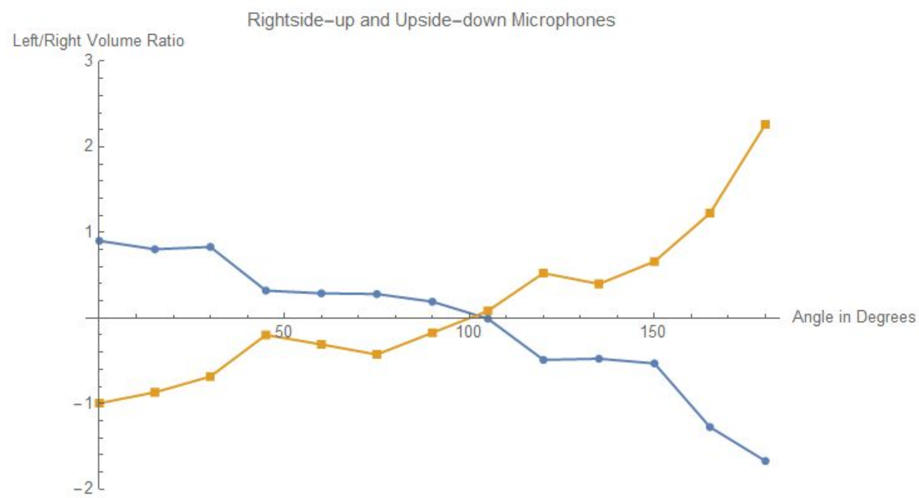


Figure 9

After the findings ruled out the initial hypothesis of microphone irregularity, several changes were made to the trials to minimize external factors. Firstly, the recording location was moved from a hardwood floor surface to a carpeted one to reduce echo and reverberation. Also, the initial 13 angles were cut down to only 7, each being 30 degrees apart. A large amount of angles were not needed for the ratio calculations because it would only serve as the basis for a seamless mathematical model. Thus, by reducing the amount of recording instances, chances of variability were lowered while maintaining accuracy. Finally, the direction of the speaker was emphasized so that it was always pointing towards the microphone's center along a flat line. The changes allowed for a more reliable set of ratios to be obtained, as shown in Figure 3.

3. Analysis of Results

Both horizontal and vertical positions were calculated in the accuracy testing trials, but no significant differences were found between the two. Trials with 45 degrees, 90 degrees, and 120 degrees in both horizontal and vertical positions were analyzed in an outdoor environment with a large amount of wind as well as an indoor environment. The outdoor trials were on average 14.3 degrees more inaccurate than the results of the indoor trials, which on average was 4.9 degrees from the actual angle of origin. The results showed that wind could heavily distort the system's readings, and in turn its accuracy. In addition, another set of trials were conducted inside where the locations of 90 degrees horizontal/110 degrees vertical and 70 degrees horizontal/100 degrees vertical were tested with five rapid whistle bursts. Angle locations that best represented a disaster survivor scenario were chosen (the vertical angles were larger than 90 degrees because the survivors would be more likely to be under the drone). In the whistle trial, vertical (6.99 degrees on average) calculations were slightly more inaccurate than the horizontal calculations (4.53 degrees), but the difference was not very large. Overall, across all 4 trials, horizontal calculations were on average 7.06 degrees off from the actual angle while vertical calculations were about 9.94 degrees off. A windy environment disrupted angle calculations by around 19.2 degrees, while an indoor environment had an inaccuracy of only 4.9 degrees. Calculation time was also recorded and 0.03 seconds was found to be the average time taken.

Conclusion

An accurate, fast, and novel system of SSL that used the ratio of a left and right channel's volumes was created and optimized for search and rescue. A mathematical function was able to model the angle of origin as a function of left/right volume ratios, and a bandpass filter could specify the frequency ranges of the human voice or a survival whistle to an extent. Since the dynamic nature of UAV flight demands the continuous running of an SSL program, certain levels of inaccuracy can be mitigated as the drone could potentially be receiving 33 new angle measurements per second. Looking forward, the performance of the system in high-wind conditions can be improved and the effect of distance could be explored.

Ideas for Future Research

Machine-learning algorithms could be explored to facilitate signal identification, while testing could be conducted on an actual UAV. The system of frequency extraction could be improved so that a multi-layer algorithm can select the necessary frequency range rather than one bandpass filter. The topic of rescue UAVs could be further explored with a system that calculates the fastest route towards a trapped survivor in harsh conditions or a system that maps out all survivors in a given array and computes the most efficient way of evacuation.

Acknowledgments

Dr. Choi, Mrs. Lee, Dr. Bae

Bibliography

Feuilherade, P. (2017, July 19). ≡. Retrieved December 30, 2020, from https://robohub.org/robots-to-the-rescue/

Cara Labrador, R., & Cheatham, A. (2020, August 24). U.S. Disaster Relief at Home and Abroad. Retrieved December 30, 2020, from https://www.cfr.org/backgrounder/us-disaster-relief-home-and-abroad

Madisetti, V. K. (2010). The Digital Signal Processing Handbook. In 1225997811 911504594 R. C. Dorf (Ed.), *The Digital Signal Processing Handbook* (2nd ed.). Boca Raton: CRC Press.

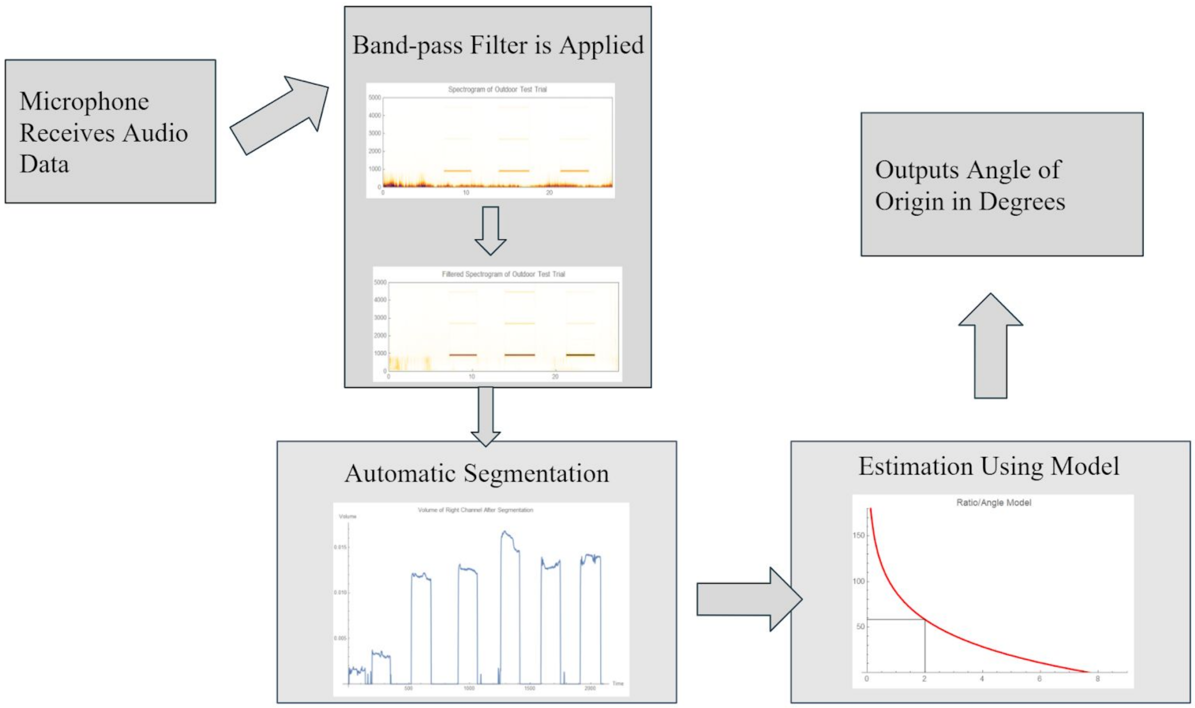
Valin, J., Michaud, F., Rouat, J., & L´etourneau, D. (2016). Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. *LABORIUS - Research Laboratory on Mobile Robotics and Intelligent Systems Department of Electrical Engineering and Computer Engineering*, 1. https://arxiv.org/pdf/1602.08213.pdf

Deleforge, A., Di Carlo, D., Strauss, M., Serizel, R., & Marcenaro, L. (2019). Audio-Based Search and Rescue with a Drone: Highlights from the IEEE Signal Processing Cup 2019 Student Competition. *IEEE Signal Processing Magazine,* 1.

Ritchie, H. (2014, June 3). Natural Disasters. Our World in Data. https://ourworldindata.org/natural-disasters

Appendix

1. System Flowchart



2. Code Written:

```
FrequencyExtractor[audio_] :=

 BandpassFilter[

  audio, {Quantity[300, "Hertz"],

   Quantity[8000,

    "Hertz"]}];

AngleCalculator[

  ratio_] := -43.154321712525416` Log[0.12974412755601705` ratio];

sound = FrequencyExtractor[Audio[Input Sound]];
left = AudioData[sound][[1]];
right = AudioData[sound][[2]];
```

```
left = AudioLocalMeasurements[Audio[left], "RMSAmplitude"]["Path"];
right = AudioLocalMeasurements[Audio[right], "RMSAmplitude"]["Path"];
left = Map[If[#[[2]] < 0.005, 0, #[[2]]] &, left];
right = Map[If[#[[2]] < 0.002, 0, #[[2]]] &, right];
left = SequenceSplit[left, {0}];
right = SequenceSplit[right, {0}];

lengthsLeft = Map[Length[left[[#]]] &, Range[Length[left]]];
leftindex = Map[
  If[Max[lengthsLeft] == lengthsLeft[[#]], #] &,
  Range[Length[lengthsLeft]]
  ];
leftindex = Flatten[Select[leftindex, IntegerQ[#] &]];
left = left[[leftindex]];
lengthsRight = Map[Length[right[[#]]] &, Range[Length[right]]];
rightindex = Map[
  If[Max[lengthsRight] == lengthsRight[[#]], #] &,
  Range[Length[lengthsRight]]
  ];
rightindex = Flatten[Select[rightindex, IntegerQ[#] &]];
right = right[[rightindex]];
left = Map[
  #[[Floor[Length[left[[1]]]/10] ;;
     Floor[Length[left[[1]]]*0.8]]] &, left];
right = Map[
  #[[Floor[Length[right[[1]]]/10] ;;
     Floor[Length[right[[1]]]*0.8]]] &, right];
minL = Min[{Length[left[[1]]], Length[right[[1]]]}];
```

```
left = left[[1]][[;; minL]];
right = right[[1]][[;; minL]];
quotients = Map[left[[#]]/right[[#]] &, Range[minL]];
average = Mean[quotients];
angle = AngleCalculator[average]
```

3. Full Set of Test Calculations and Accuracy Assessment Data:

Condition Testing						
Actual Angle (degrees)	45	90	120	Average Calculation Time	Condition	Average Difference (degrees)
Horizontal Calculation	54.2929	81.5178	85.6215	0.051 seconds	High wind	17.3845
Vertical Calculation	46.28	79.3188	68.99	0.05 seconds	High wind	20.9904
Horizontal Calculation	54.8665	92.1806	113	0.05 seconds	No wind	6.349
Vertical Calculation	43.6809	91.8793	108.818	0.051 seconds	No wind	4.7935

Horizontal/Vertical Testing				
	Lateral	Vertical		Average Calculation Time
Actual Angle (degrees)	90	110		
Trial 1	92.5886	104.643		0.02767 seconds
Trial 2	92.0685	107.521		0.02307 seconds
Trial 3	93.2119	108.355		0.02292 seconds
Trial 4	95.7865	75.198		0.03067 seconds
Trial 5	93.5289	106.242		0.04286 seconds
Average Inaccuracy	3.43888	9.6082		

Actual Angle (degrees)	70	100		
Trial 1	70.25	103.025		0.02071 seconds
Trial 2	71.3979	102.192		0.0269 seconds
Trial 3	70.1597	102.071		0.0163 seconds
Trial 4	66.5222	103.088		0.02529 seconds
Trial 5	69.8221	88.5032		0.02021 seconds
Average Inaccuracy	1.09266	4.37456		