Estimation of Scuba Diver Location from Hydrophone Data using Acoustical Direction of Arrival Methods

Pranav Barot[1] **and Benjamin Masters**[1]

[1]**University of Waterloo, Department of Systems Design Engineering**

ABSTRACT

As a submission for the International Student Challenge in Acoustical Signal Processing, we applied traditional acoustic direction of arrival methods to estimate the location of a scuba diver passing a hydrophone array, and propose a simple mathematical model for their motion. This manuscript outlines the chosen techniques, signal processing implementation, design decisions and final results for the challenge.

INTRODUCTION

The international student challenge in acoustical signal processing presents a scenario in which a scuba diver is swimming past a transducer array consisting of three hydrophones, and provides recordings from each hydrophone. Participants are tasked with analysing the spectral content of the hydrophone signals and determining characteristics of the diver's motion.

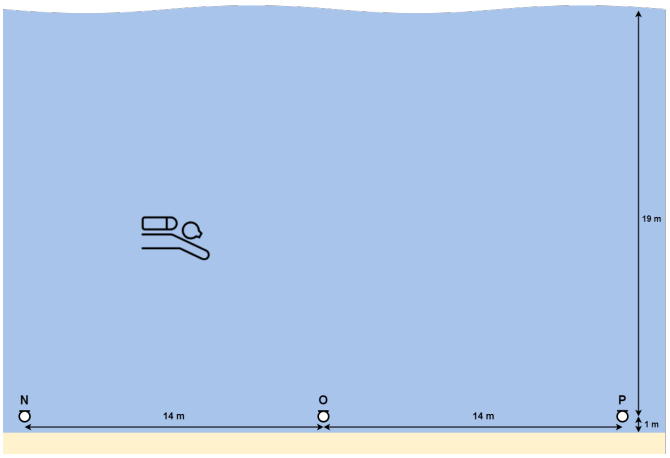

Figure 1. A diagram of the scuba diver and hydrophone array set up.

We apply spectral and time domain signal processing techniques to estimate the diver's breathing rate, time of closest approach, altitude, and swim speed for the duration of the hydrophone recordings. Tasks are reported on in order by section, with final challenge results presented at the end of this manuscript.

TASK 1

Part A: Hydrophone Spectrogram and Characteristics

A spectrogram can be produced to allow for time-frequency analysis of the signals. This is done using the Python Librosa library, and the output spectrogram of a slice of the signal from hydrophone N is shown in figure 2.

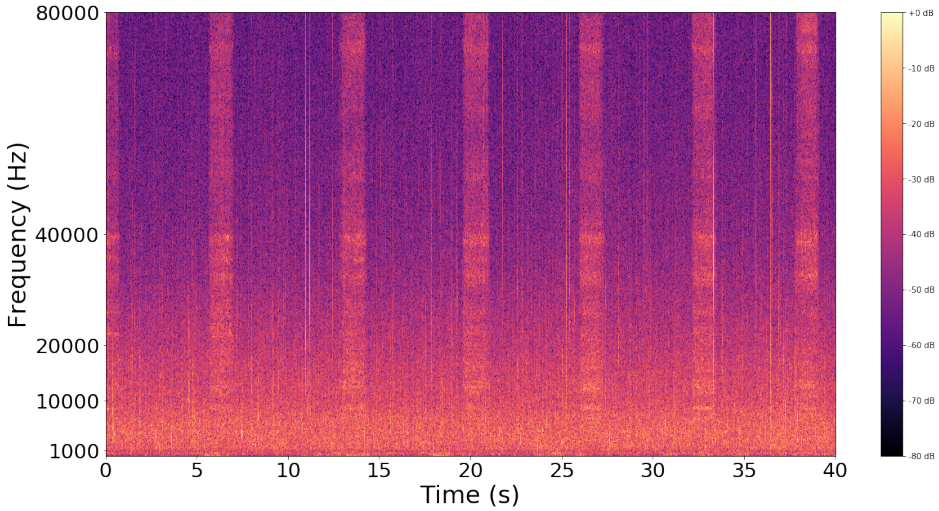

Figure 2. Spectrogram of first 40 seconds of audio from Hydrophone N

From a visual inspection of the spectrogram, a rhythmic broadband acoustic effect is observed, in addition to strong background noise with its spectral content primarily concentrated in the audible frequency range. It is hypothesized that the rhythmic sound is the breathing apparatus, and the background noise can be attributed to water moving past the hydrophones. This is corroborated with what is heard when listening to the provided .wav files.

Part B: Estimation of the Diver's Breathing Rate

One method to estimate the breathing rate is to manually count the breaths from the spectrogram and divide by time period.

For 40 sec shown in Figure 2 this yields: 7 breaths/40sec = 0.175 Hz, falling in the range of 0.16 to 0.33 Hz for the typical respiratory rate for humans, as described in (1).

Another method is to use a frame based approach by computing the powers of segments of the signal over the duration of the recording. Running a peak detector enables identification of frames with high local energy, corresponding to the breaths. Then, the breathing rate can be estimated as the number of peaks found over their time duration.

In order to isolate these breaths, we study the frames in the frequency domain. Figure 3 shows the FFT of a frame with a breath and a frame with no breath. The frames are chosen to be 1 second long, with a 1 second step size.

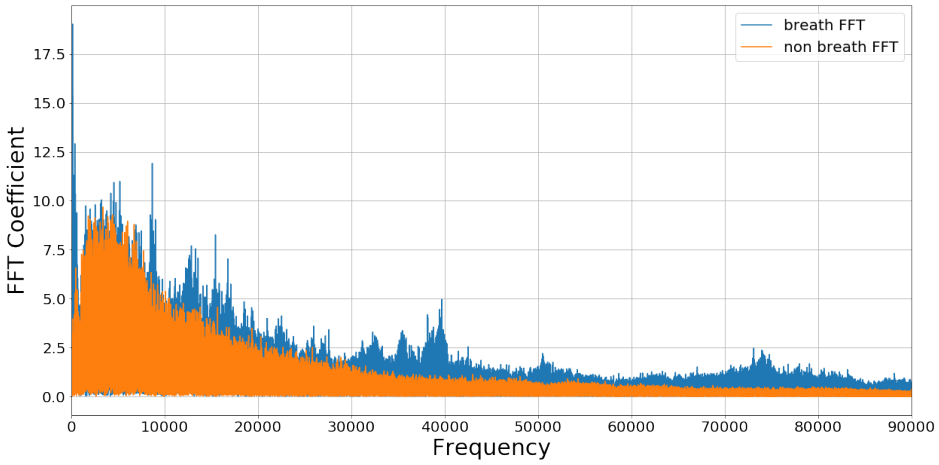

Figure 3. FFT Output of a frame with a breath, and a frame with no breath

Evidently, the breath has more energy in higher frequency bands. The frame power is generated by applying a band-pass filter at the 35-40kHz band, and computing the power of the resulting frame. A peak detector that compares neighboring values is run on the result, and shown in Figure 4. By expressing these peaks on a normalized dB scale, the breaths, especially at earlier times in the recording, can be emphasized.

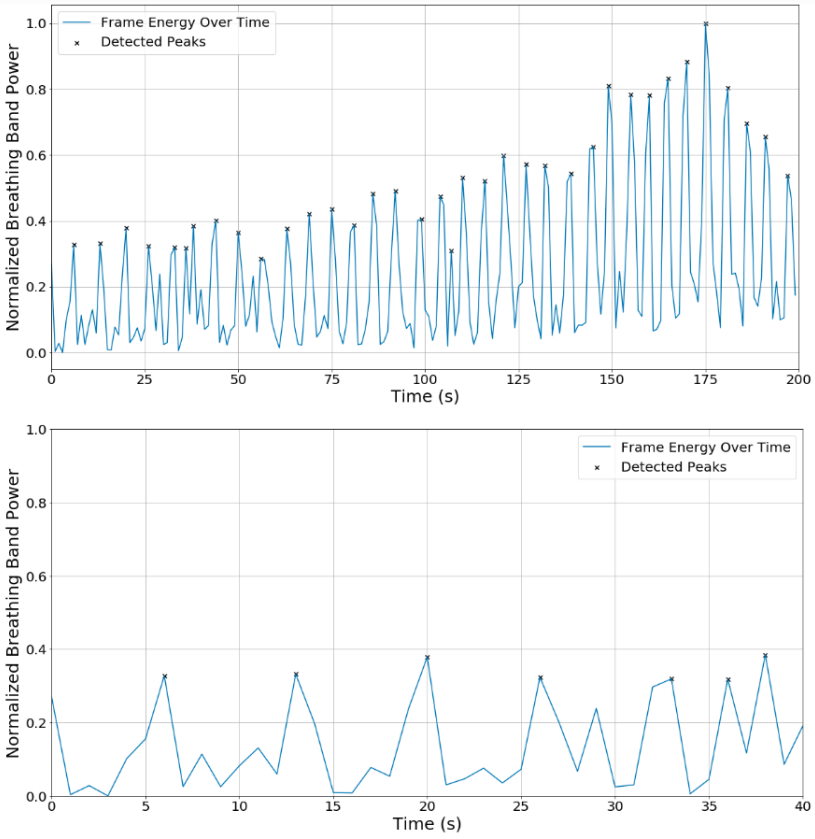

Figure 4. Frame power and peaks over entire recording (above), and for a 40 second audio slice (below)

For the presented 200 seconds, 36 unique peaks are found, resulting in a breathing rate of 0.18 Hz. A breathing rate is also computed for the region with high energy, between 150 and 200 seconds, containing find 9 peaks, yielding a breathing rate of 0.18 Hz.

Given these estimates made across a number of time periods, the breathing rate estimate of the diver is approximately **0.18 Hz**.

TASK 2

Part A: Estimate of the Time of Closest Approach to Hydrophone O

Two approaches are taken to compute the best estimate of time of closest approach to hydrophone O. The first is comparing the frame powers of all three recordings over time. The frames undergo a band-pass filter at the same cutoffs of 35kHz and 40kHz, based on the FFT findings above. Figure 5 shows this progression of frame powers over time. In this example, the powers are left on a linear scale, to emphasize the diver's motion past the array.

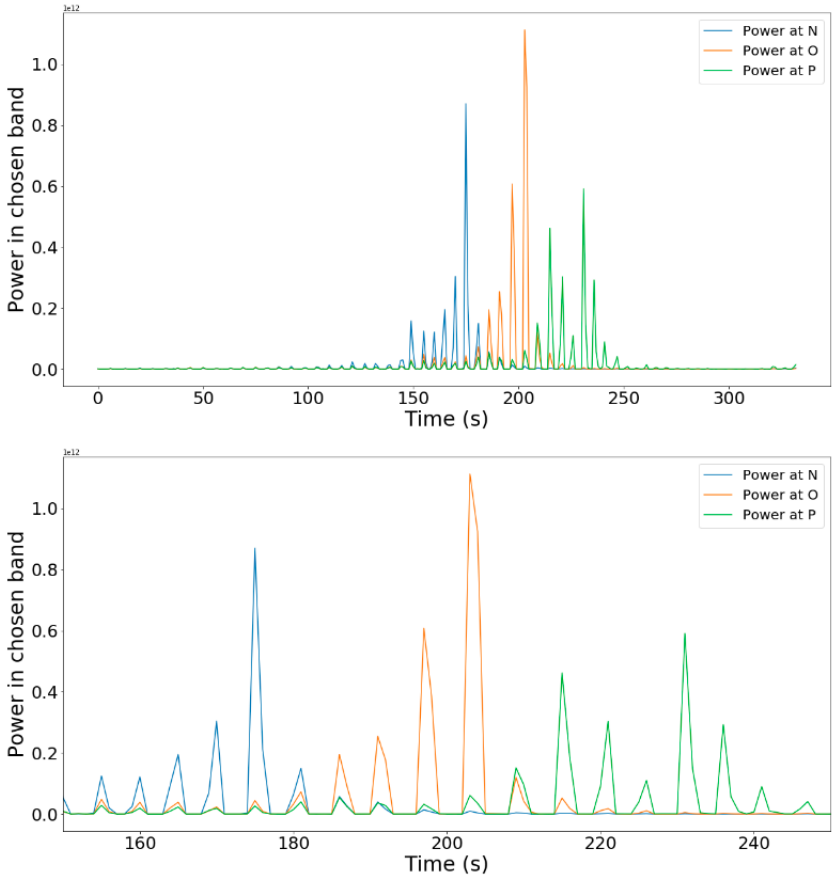

Figure 5. Power of filtered frames over time. Top = full recording length. Bottom = Period with higher peaks

The peak power at microphone O is found at 203 seconds. At P, the peak is 231 seconds. At N, the peak is at 175 seconds. This also suggests that the diver is swimming from a side closer to N, towards O, and then passing P near the end of the journey. This estimation requires us to assume that the breaths are made with a similar intensity by the diver each time, which may not always be the case.

The second, more robust method is by estimating the direction of arrival at each pair of microphones for the duration of the recordings. The horizontal geometry of the hydrophones can be utilized, given the known distance between them, to estimate the direction of arrival of the diver's breaths. This requires a few steps; choosing frames with direct breaths, estimating the timing difference between the frames, and resolving the timing difference to an angle.

Choosing Good Frames

The band-pass filtered frames and their powers as found in Part A are first inspected. It can be seen across the recording that the majority of the breath energy is found in the 150-250s time frame. Hence, it is hypothesized that the diver is far away from the 3 hydrophones until $\sim$150s, and we may use the power seen in the 150-250s range to choose frames with good breaths.

To choose good frames, the computation is modified slightly by computing the energy of the resolved frames in the breathing band by averaging the two signals. This is a valid approach when two measurements of the same quantity are being made at the same time by different receivers, in order to represent their signals as one resolved signal. The direct power present in this resolved signal is used in order to choose frames that may include direct breaths. No band-pass filter is used on the frames in this step, as this may result in changing the temporal structure of the signal which is crucial to maintain in order to properly estimate the timing difference. This results in a slightly different representation of frame power, as shown in Figure 6.

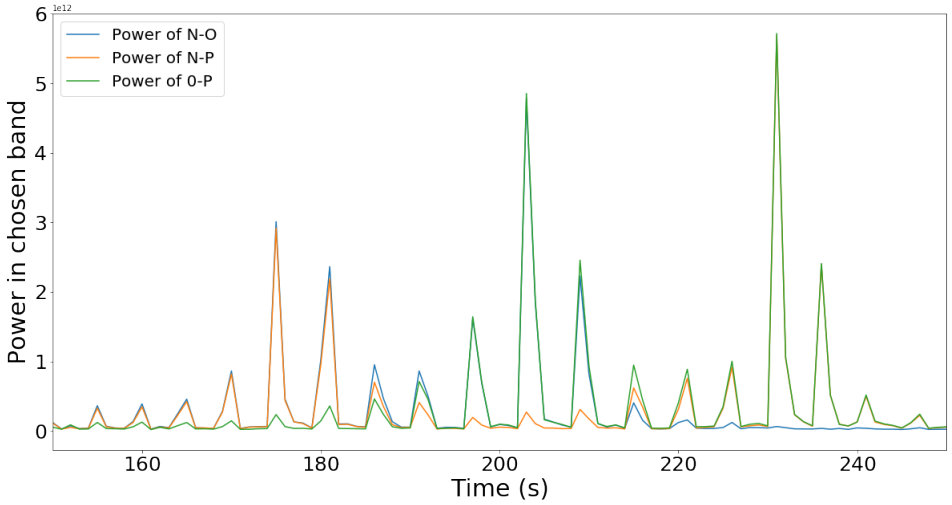

Figure 6. Power of resolved frames over time

The trend of peaks is used to choose a hard threshold for selecting good frames. Good frames are considered to be those with a power greater than 1e9, where all other frames are considered those without direct breaths and are thus ignored.

Timing Difference Estimation

Once the detected audio has been determined to contain direct breath from the diver, the timing difference is estimated between the two frames at each hydrophone.

Since time-domain cross-correlators are computationally expensive and sensitive to reverberation, spectral domain methods are used for this challenge. Binaural timing differences are estimated using the Wiener-Khinchin relation for the cross-power spectrum of two recorded signals x and y (2).

$$G_{xy} = X[f]Y[f]^* \tag{1}$$

This relation is used to estimate cross-correlation output of x and y as per the following generalized formulation, the *argmax* of which indicates the interaural timing difference between the two microphones (3).

$$\hat{R}_{xy} = \int_{-\infty}^{\infty} \psi(f) G_{xy}(f) e^{j2\pi f\tau} df \tag{2}$$

The phase transform (GCC-PHAT) pre-whitens the cross-correlation response using the value of ψ as in Eq (3), providing robustness against reflections in difficult auditory environments. Considering the ambient noise heard in the hydrophone recordings, this technique may be more suitable to the task at hand.

$$\psi_{PHAT}[f] = \frac{1}{|G_{xy}(f)|} \tag{3}$$

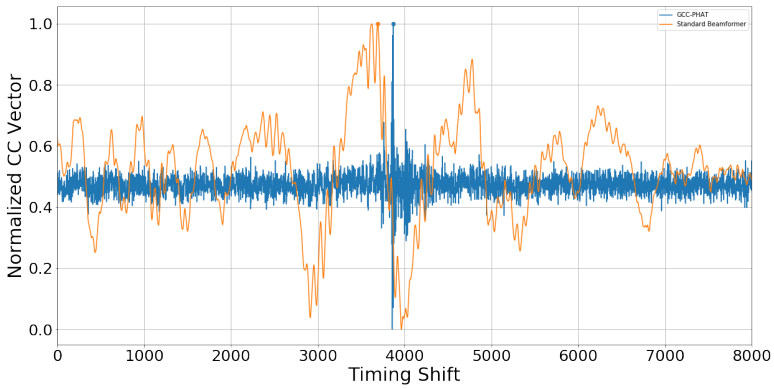

Figure 7. Cross Correlation With Different Estimators

These cross-correlation methods are visualized by their output on a frame of audio from hydrophones N and O. Fig 7 shows the results from a time domain beamforming operation, and the GCC-PHAT.

Evidently, the result from the standard beamformer is noisier and has many local maxima, which can introduce uncertainties when choosing the best value for the time shift between the two signals. The prominence of the singular peak in the GCC-PHAT increases the confidence that it comes from the direct breath of the diver, and so the GCC-PHAT is used in this work as the timing difference estimation method.

Generating the Direction Of Arrival

Once the timing difference has been accurately determined, a geometric model is used to estimate the direction of arrival of the sound source. A simplified description is presented in Fig 8, showing two microphones M1 and M2, separated by a distance D, with two unique path lengths X1 and X2 to a sound source S.

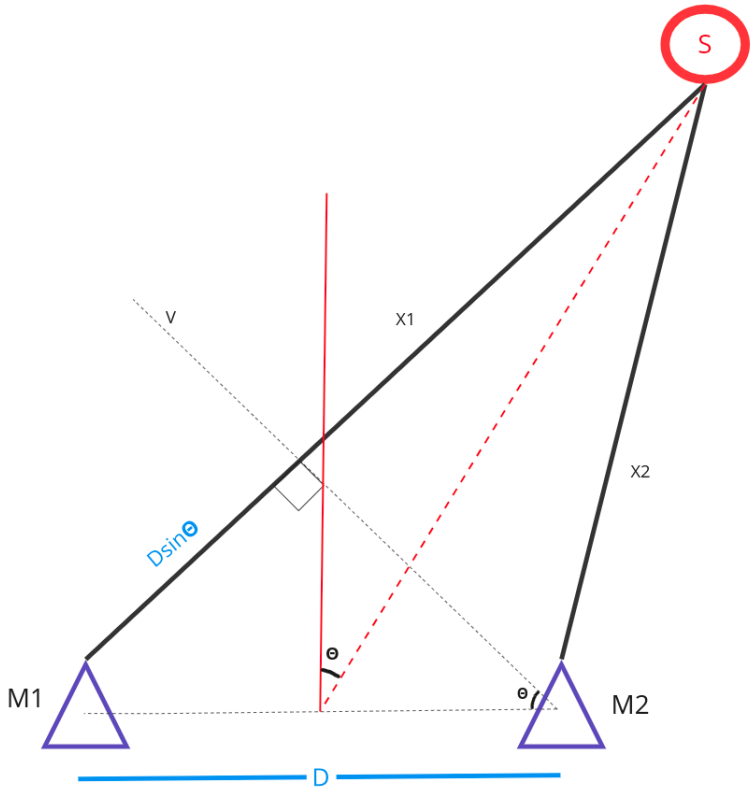

Figure 8. Simple DOA geometry

Given the right triangle made by M1, M2 and the path length X1, with an angle of θ, the opposite side then becomes $Dsin\theta$, given the distance D between the two microphones. This distance represents the extra distance the wavefront must travel to reach M1 once it has reached M2. This distance is directly computed from the timing difference τ, and so the measured quantities are related as in Equation (4), where c denotes the speed of sound.

$$Dsin\theta = c\tau \tag{4}$$

We apply the GCC-PHAT on the selected good frames as described above, to estimate τ in each instance. In order to handle NaN values for frames that do not meet the threshold, a linear interpolation is applied. In order to further suppress the volatility of estimates, a median filter of size 5 is applied to the measurements of timing differences. The median filter also helps remove outliers that may be generatuseded by using a hard threshold that may not be correct for all cases throughout the recording. This transformation generates a smooth trajectory of estimates from the recording that is robust to rapid changes in the acoustic environment.

Since 3 hydrophones are used, it is then possible to generate 3 unique DOA estimates at any given instant. These come from the frame pairs of hydrophones N and O with D = 14m, N and P with D = 28m, and O and P with D = 14m. Using $c = 1528m/s$, with good frame estimates from all three recordings, 3 unique DOA trajectories can be formed. Figure 9 depicts these three trajectories, generated from the smoothed and interpolated timing difference arrays.

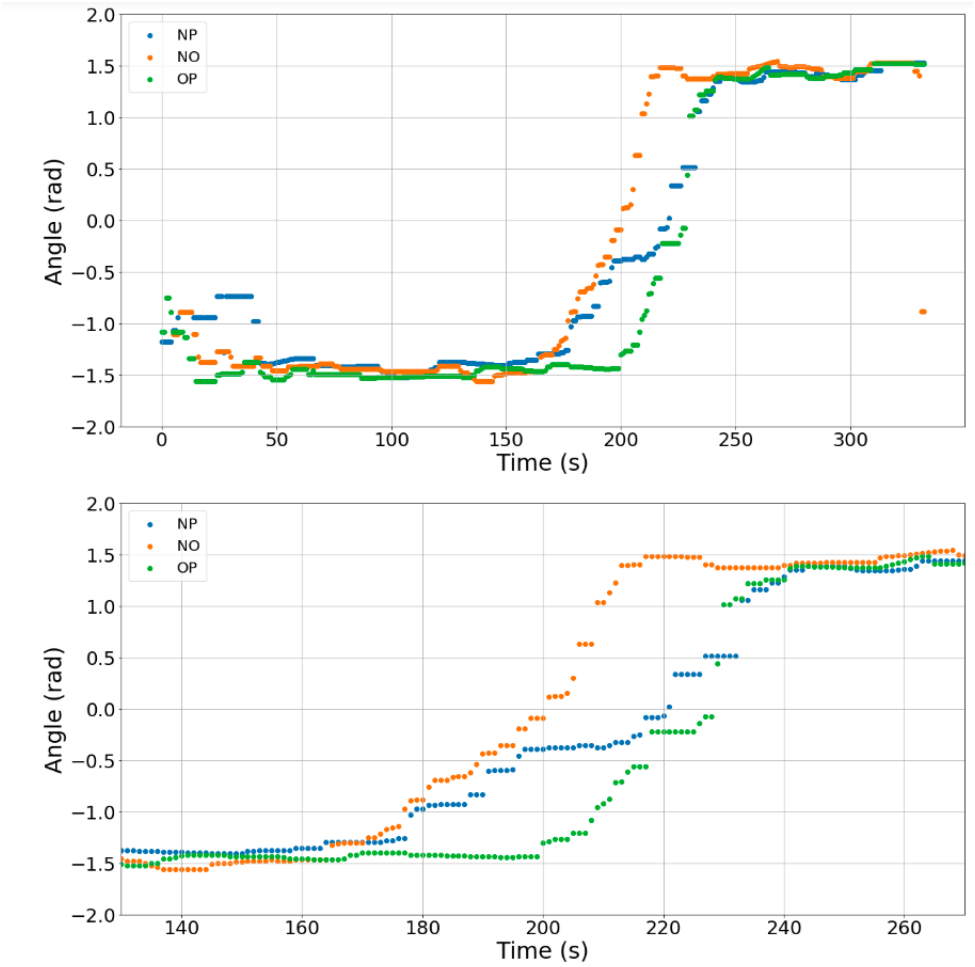

Figure 9. Raw DOA estimates from all 3 hydrophone pairs over entire recording (top) and period of 130-270s (bottom)

From these trajectories, it can be observed that the angles are relatively constant and near to $-\pi/2$ early in the recordings, indicating that the diver is far off to the left side of the array. However, as the diver approaches the array, the angle estimates start to increase and approach approximately $\pi/2$ by the end of the recording, indicating that the diver is passing by the array and ends up far off to the opposite side.

To understand these observations, a simple mathematical model can be developed to estimate the divers trajectory through time. Assuming a constant altitude, y, swim speed, v_{swim}, and starting position, x_0 the relationship between the angle and the horizontal position can be approximately modeled as in Equation (5). As seen in Figure 9, the trajectory curves do resemble an inverse tangent curve, and a model will be developed and compared to these results in a later section, once the estimates of swim speed and altitude have been discussed.

$$\theta(t) = tan^{-1}\left(\frac{v_{swim}t - x_0}{y}\right) \tag{5}$$

To further improve this trajectory and suppress the step-wise nature of the estimates as a result of the median filter, a rolling average with a window size of 5 is applied to the DOA trajectory, resulting in the following estimates for the 130-270s range.

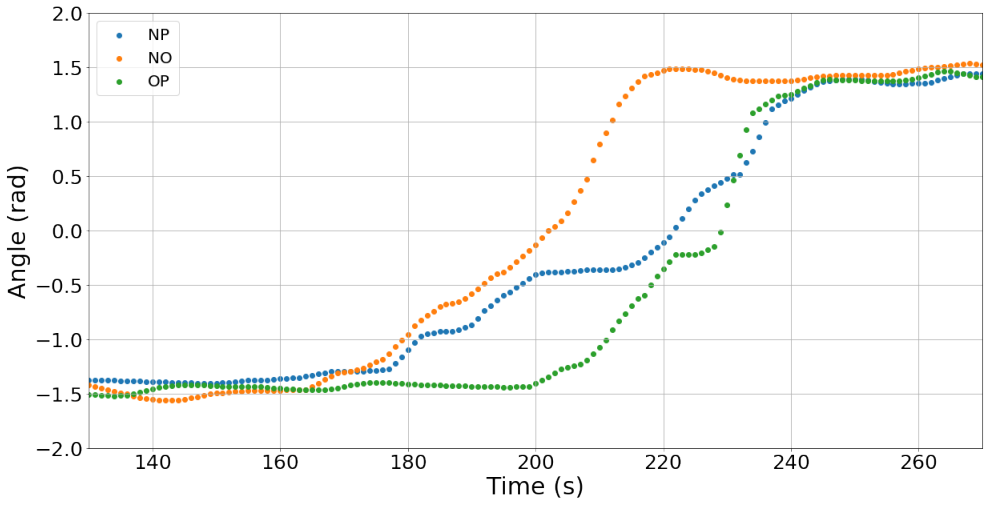

Figure 10. Rolling averaged and median filtered DOA estimates in 130-270s range.

It is seen that the estimates from hydrophone pairs N-O and O-P appear similar in shape, but shifted. This is as expected, given the equal spacing of the hydrophones and the symmetric measurement locations. However, it can be seen that the approximation from hydrophone N-P does not match as closely. It is suspected that this is a result of the increased distance between hydrophones N and P resulting in less resolution and precision in the DOA estimation algorithm. For this reason, hydrophone pairs N-O and O-P are used for the estimates going forward.

Given these DOA estimates and the even spacing of the microphones, we suppose that the time of closest approach to hydrophone O will be when the angle of N-P is close to 0, when N-O is 45 deg or 0.78 rad, and when O-P is -45 deg or -0.78 rad.

For N-P, this occurs at around 222 sec. However, as discussed above, estimates will be made based on pairs N-O and O-P, due to increased resolution.

For N-O, this condition occurs at 210 s. For O-P, it occurs at 214 s. The average of curves N-O and O-P is also taken to find where it approaches 0 rad, as this would be the point at which the diver is at complementary positions with respect to the two hydrophone pairs, and we see that this occurs at 212s. Figure 11 shows the average of these two curves, alongside the curve of N-P, for completeness.

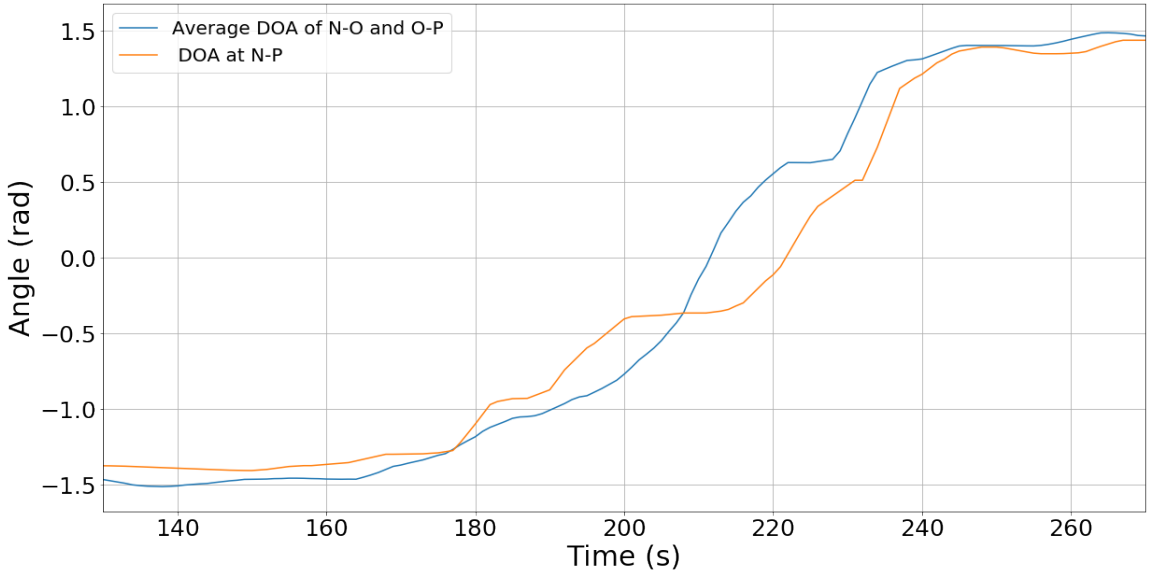

Figure 11. DOA curves of N-P and the average of N-O and O-P

Hence the final estimate is that the diver is closest to hydrophone O at around **212 s** into the recording.

Part B: Estimate of the Altitude at the Closest Point of Approach to Hydrophone O

To estimate the altitude, a geometric model is once again used. The DOA estimates between hydrophone pairs N-O and O-P at equal times are taken, and lines are projected from the center points between the pairs with angles equal to the estimates. The intersection of these projected lines is the estimate for the sound source location on a 2d plane.

If the diver was directly over the center hydrophone, O, the selected angles should be complements. However, due to the sparse nature of the breaths, this is not exactly the case. For this reason, the breath nearest to the time of closest approach is used to estimate the altitude. Figure 12 illustrates the result of this approach.

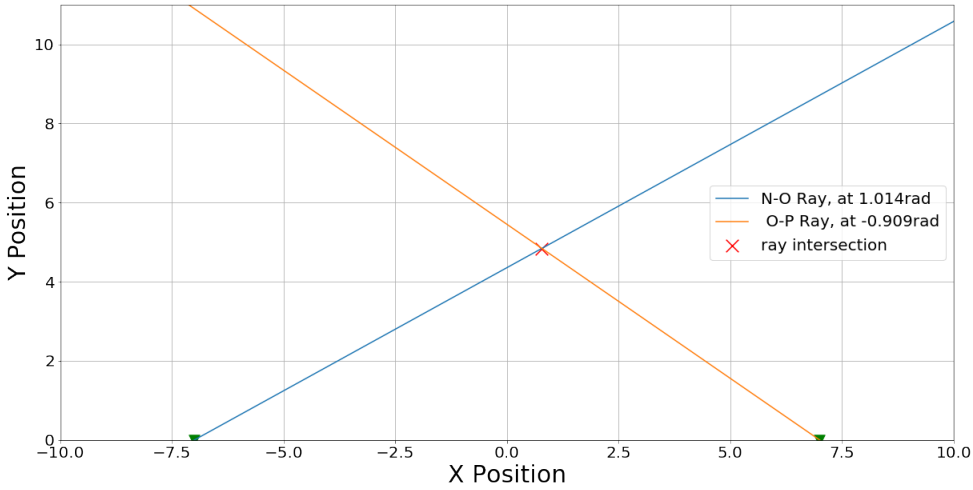

Figure 12. Intersection of rays of N-O and O-P at t=212s

At 212s, the rays intersect at (0.79, 4.85). Hence, the estimate of the altitude of the diver at closest approach is about **4.85m above the hydrophone array**, or **5.85m above the sea floor**. Figure 13 shows the estimated locations of the diver over time from 190s to 228s, using the two rays given by the measurements made at N-O and O-P.

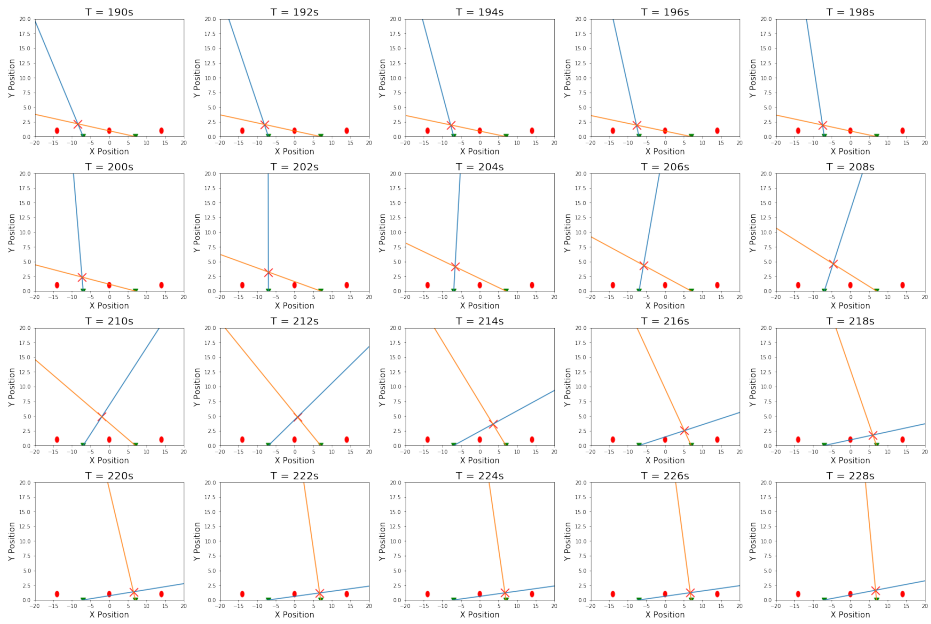

Figure 13. Progression of diver location estimates over time

Although the problem statement seems to mention that the altitude is constant, the results of this method seem to disagree, as seen in Figure 14. The diver seems to be at a range of altitudes from about 2-6 meters from the floor. However, it is unclear if this is the true path the diver takes, or a result of noise influencing the direction estimates.

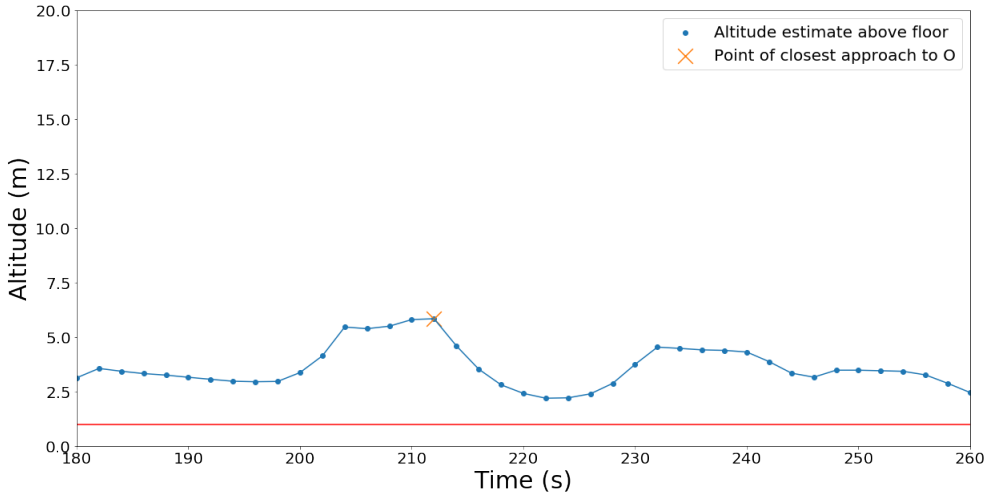

Figure 14. Estimated diver altitude over time. Red line = array height at 1m.

Part C: Estimate of the Diver's Swimming Speed

Similar to the method used for the altitude, the swimming speed can be estimated by using the direction approach. Once again looking at the hydrophone pairs N-O and O-P, a pair of points with nearly identical direction estimates were identified. Given the equidistant spacing between hydrophones, it is known that the distance traveled between these points should be equal to the distance between hydrophones, 14 meters. With this, the time difference between the points with equal direction estimates is taken, and the distance is divided by this. This was done using the direction estimate for hydrophone pair N-O at time 202s and pair O-P and time 229s, with direction estimates of -.003 rad and -.013 rad, respectively. This results in a swimming speed estimate of .518m/s over the selected range of time.

However, it is unclear whether the swim speed is constant, and a similar approach can be taken over multiple sets of points to obtain an estimate of the swimming speed over time. The two curves at N-O and O-P can be treated as the same behaviour from the diver occurring at different points in time. From this, the time it takes the diver to perform the same behaviour at O-P as he did at N-O can be found using a simple cross-correlation. The cross-correlation is applied for the time period of 180-250s, so as to not capture the similar stagnant DOAs from both curves, which may skew the cross-correlation output. It is found that the best timing alignment for the two curves is at 24 seconds. Figure 15 shows the curve N-O shifted by 24 seconds, and the original curve for O-P. Evidently, they are well aligned which suggests that the cross-correlation has worked for this case.

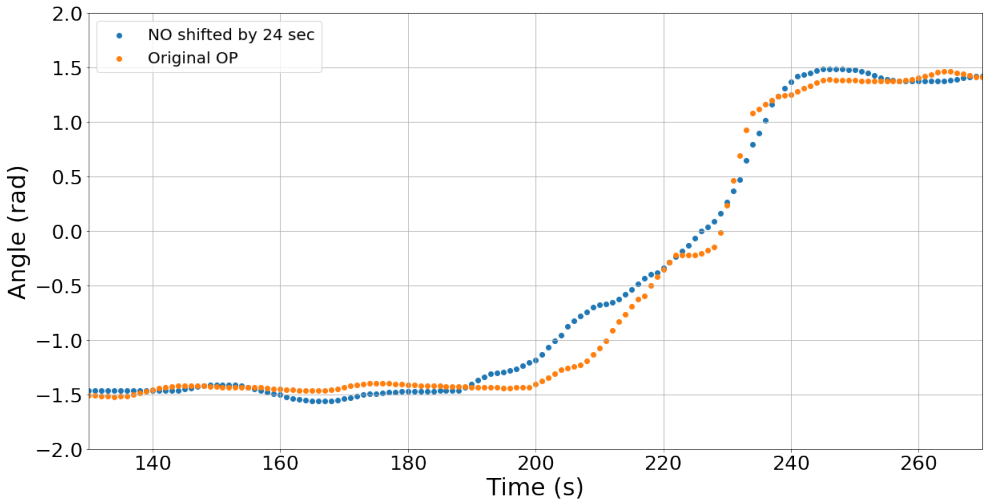

Figure 15. DOA of O-P and 24 sec time-shifted DOA of N-O

Given the timing shift of 24 seconds and a distance between the two pairs of microphones as 14m, the estimate for the speed then becomes 14/24 = **0.583 m/s**. This value is considered a more realistic estimate as it takes into account the behaviour of the DOA estimates across a period of time, rather than a single pair of estimates at one point in time.

MATHEMATICAL VALIDATION

Based on the findings, a mathematical model is proposed to estimate the motion of the diver over time. It is known from basic trigonometric identities that the DOA angle should vary proportionally to the inverse tangent function, assuming a constant altitude and velocity, as described in Equation 5. Inserting the estimate for swimming speed of .583 m/s, the estimated altitude above the hydrophones of 4.85m, and a computed starting position, we can model the angle over time. To estimate this starting position, the zero reference is selected as the location of hydrophone O. We know that the diver is directly over this hydrophone at approximately 212s, so using the assumed constant swim speed estimate results in a starting position of about 123.6m from hydrophone O, in the direction of hydrophone N. This results in the model in Equation 6. This model is plotted alongside the angle estimates of hydrophone pairs N-O and O-P in Figure 16, illustrating that the model falls nicely between the two estimates, which is as expected.

$$\theta(t) = tan^{-1}\left(\frac{.583t - 123.6}{4.85}\right) \tag{6}$$

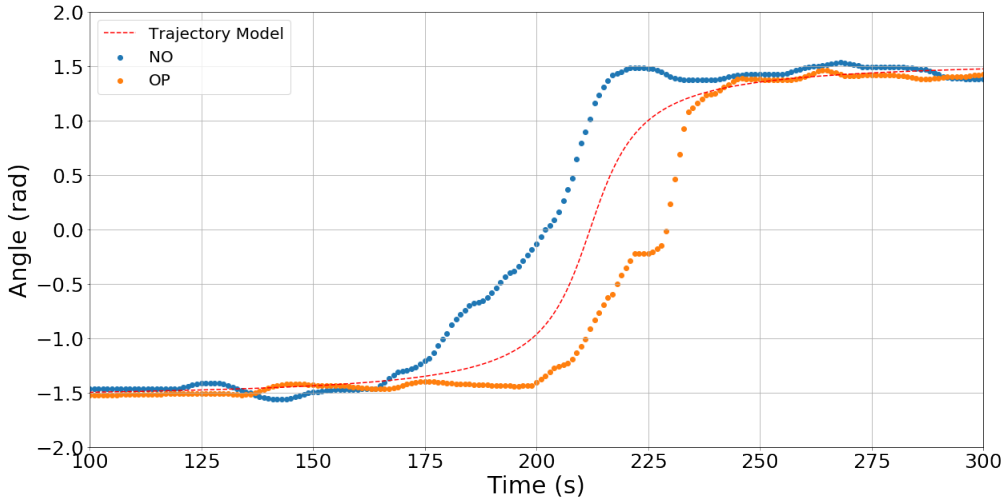

Figure 16. Trajectory Equation and Measured Curves

SUMMARY OF RESULTS

Finally, we report on the estimates for all the tasks in this challenge below.

Task 1

- Task 1a - See Figure 2
- Task 1b - Breathing Rate Estimate: .18 Hz

Task 2

- Task 2a - Time of Closest Approach: t=212s
- Task 2b - Altitude at Closest Approach: 5.85m above the sea floor, 4.85m above the hydrophone array
- Task 2c - Swimming Speed Estimate: .583 m/s

REFERENCES

[1] M. A. Russo, D. M. Santarelli, and D. O'Rourke, "The physiological effects of slow breathing in the healthy human," *Breathe*, vol. 13, no. 4, pp. 298–309, 2017.

[2] E. W. Weisstein, "Wiener-khinchin theorem. From MathWorld–A Wolfram Web Resource." (accessed July. 7, 2022).

[3] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.