

Abstract

Passive sonar is an essential navigational tool for aquatic vessels with which emissions control (EMCON) is of utmost importance, such as submarines and naval surface vessels. Using passive observations to accurately draw conclusions about the surrounding environment has always been a challenge, for many factors can impact the reliability of the data. This paper attempts to address the complexity of aquatic navigation via passive sonar by creating a pipeline of models to autonomously analyze the data. Existing methods of contact classification incorporate various digital signal processing (DSP) methods followed by traditional deep learning models such as general adversarial networks (GANs). This paper uses a basic convolutional neural network (CNN) for the discrimination of signal from noise and the classification of signals. Regarding contact localization, this paper applies 2D image filters to the hydrophonic data to determine the contact relative bearing time series through environmental noise and underwater sound propagation effects. Current techniques for contact distance approximation involve target motion analysis (TMA), requiring assumed information about the contact and often own ship maneuvers. No technique exists to estimate distance given only bearing rate, and this paper attempts to use a discriminator-only LSTM deep learning model to solve this. Finally, using doppler analysis, target heading and speed are approximated. The provided passive sonar analysis pipeline is a starting point towards a fully autonomous navigation system.

Introduction

Sound navigation and ranging (sonar) is the process by which sound waves can be used to analyze the relative placement of objects in an environment. Often used in underwater environments, sonar makes use of the fact that sound waves have a much smaller attenuation in water than electromagnetic waves of all frequencies (including radio waves as used in radar, and infrared light as used in lidar). Thus, sound travels further than other types of waves in similar media, allowing for a better representation of the environment at large. Much like radar and lidar, active sonar relies on the time differential between the emission of a signal and its return to determine the distance to objects in a particular direction. Specific mission sets require stealth, such as operating submarines and unmanned underwater vehicles (UUVs), or researching biologies. The emissive nature of active sonar is thus unviable, and a passive approach is necessary.

Passive sonar relies on the placement of hydrophones in various directions around the hull of a boat, allowing for the capture of sound signature data from surrounding contacts and biologies without the need for sonic emissions. The resulting structure forms a hydrophone array which can use passive beamforming, or the phase delay of the captured signal between each hydrophone to identify the bearing from which the signal is emitted. On naval submarines and surface vessels, the bow array is shaped as a sphere to ensure the collection of data from nearly all angles and bearings. This is opposed to other common hydrophone array structures, such as the linear structure of the towed array which enforces conic symmetry and therefore another layer of complexity requiring TMA. The spherical bow array structure is the structure primarily used for contact classification and localization in this paper due to its simplicity.

The hydrophonic data used in this paper is five dimensional, including the two spatial dimensions of a spherical hydrophone array, the time and frequency domains, and the dependent variable of signal amplitude. The data is preprocessed such that there is one signal amplitude value for each independent variable value combination. As such, the exact sinusoidal signal for each hydrophone is not observed, rather only the max value over a set period of time. To account for long-term contact motion, time series data is only observed in increments of one minute. Thus, many DSP analysis methods are not applicable and signal transformations are given rather than calculated.

Data Simulation

For most of the models described in this paper, real-world data is not readily available in the use cases described above due to the classification of naval vessel sound signatures. As such, a data simulator was developed to recreate the five dimensional hydrophonic data on a minute-by-minute time scale as accurately as possible. The graphic user interface (GUI) PySimple was used to create a usable interface. Figure 1 shows the landing window of the interface with demo input variables applied. Up to 176 different input variables establish the properties of the simulated ship, the surrounding environment, and the observed contacts. The simulation uses the described situation to recreate the theoretical hydrophone observations over time and frequency.

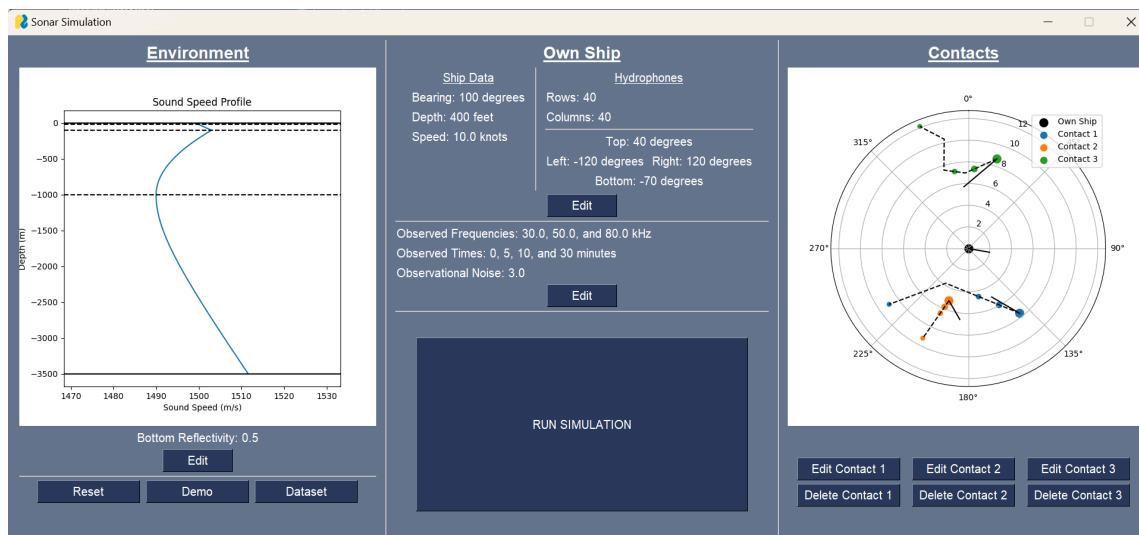


Figure 1: The landing window of the passive sonar simulator

The plot on the left depicts the sound speed profile (SSP) of the environment in which this simulation occurs, demonstrating the speed of sound as a function of water depth for the current ship location. The central column shows the input variables representing the state of the simulated ship. The plot on the right shows the relative location of detected contacts. The solid black lines represent the true course and speed of each simulated ship while the dotted black lines represent the relative course of the detected contacts over the span of thirty minutes. All input variables can be changed using the various edit buttons in each respective section.

The sound speed profile and environmental variables are edited using the window shown in figure 2. On land, sound travels in relatively straight lines with nearly all sound propagation paths consisting of either direct path or reflection (echo). Underwater, sound refracts, resulting in

six possible propagation paths: direct path, surface channel, sound channel, half channel, bottom bounce, and convergence zone. This is due to the varying sound speeds at different depths stemming from variance in temperature, pressure, and salinity. As described by Snell's Law, waves refract when encountering a barrier between two media with different propagation speeds. Accounting for this refraction is essential in predicting sound propagation paths and creating a simulation that accurately shows from which angles contact signals are detected.

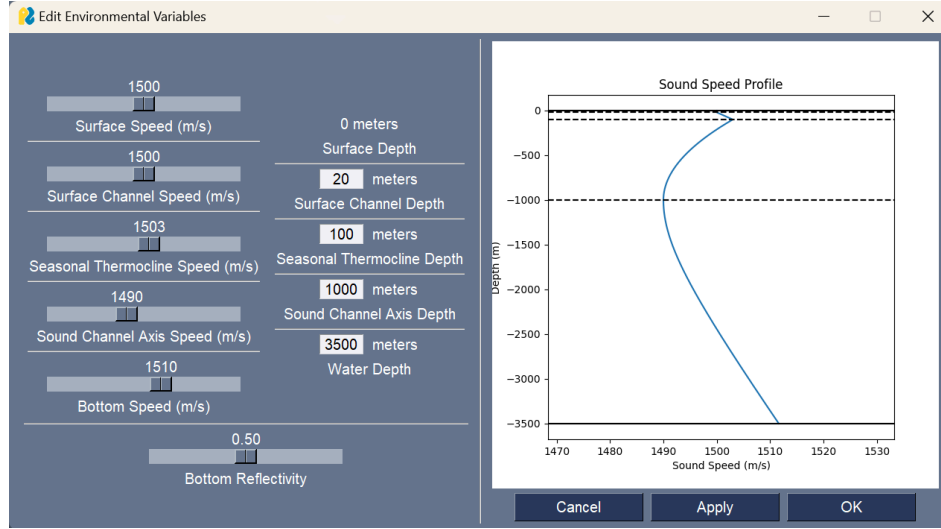


Figure 2: Edit environmental variables window and corresponding SSP

The ocean typically contains set horizontal layers at which sound speed follows a pattern. The surface channel represents a boundary above which sound speed is relatively constant. The seasonal thermocline is the boundary above which the temperature, and therefore sound speed, depends on the season and weather. The sound channel axis typically has the slowest sound speed, and it is the layer above which temperature is the primary influence on speed and below which pressure has the biggest influence. Sound channel propagation occurs when signals are emitted near the sound channel axis and the sound is propagated sinusoidally without reflecting off the surface or bottom. This simulation permits the user to control the speed and depth values of each of these layers as well as at the surface and bottom as shown in figure 2.

Certain assumptions and approximations are made to interpolate the sound speed at each depth in between the horizontal layers. For depths in between the surface and the seasonal thermocline, a simple linear interpolation is used (equations 1 and 2). Depths in between the seasonal thermocline and sound channel axis rely on a quadratic interpolation, enforcing a vertex at the sound channel axis (equation 3). All depths below the sound channel axis are a hyperbolic approximation with the same vertex, ensuring that the speed values are asymptotic (equation 4). The interpolation equations used are shown in equations 1-4 where S, SL, ST, SCA, and B stand for surface, surface layer, seasonal thermocline, sound channel axis, and bottom respectively.

1. $Speed = \frac{S\ speed - SL\ speed}{SL\ depth} * depth + S\ speed$
2. $Speed = \frac{ST\ speed - SC\ speed}{SL\ depth - ST\ depth} * depth + SL\ speed$
3. $Speed = \frac{ST\ speed - SCA\ speed}{(ST\ depth - SCA\ depth)^2} * (depth - SCA\ depth)^2 + SCA\ speed$

$$4. \text{sign}(B \text{ speed} - SCA \text{ speed}) * \sqrt{\left(\frac{(\text{depth} - SCA \text{ depth})^2}{b^2} + 1\right)} * (SCA \text{ speed} - a)^2 + a$$

where $a = SCA \text{ speed} + \frac{1}{2}(SCA \text{ speed} - B \text{ speed})$,

$$b = (SCA \text{ speed} - a) * \frac{SCA \text{ depth} - B \text{ depth}}{a - B \text{ speed}}$$

To estimate sound propagation paths, the sound speed is calculated for each meter of depth. The model is estimated as a series of constant sound speed layers, one meter thick. Depending on the starting angle, the distance the signal travels as it changes depth by one meter is computed using simple trigonometry, representing infinitesimal change dx and dy over time dt. The change in angle is computed using Snell's law and the difference in sound speed at the two media, and the next dx is computed. This process repeats until the signal has reached its target.

Figure 3 shows the interface to change the variables representing the simulated ship, including bearing, depth, and speed, as well as the size and angle of the spherical hydrophone array. Figure 4 shows the interface to adjust the observed frequencies and times to be plotted in narrowband when the simulation is run. The observational noise is the amount of random gaussian white noise applied to each hydrophone as the simulation runs to account for all unpredictable losses.

Figures 3 (left) and 4 (above): Interfaces to adjust the own ship variables and observational variables

Figure 5 shows the interface to edit the variables for each contact. The distance, relative bearing, depth, heading, and speed of the contact are adjustable, and maneuvers known as zig actions can be implemented using the “advanced” button. The contact signals are used to define the sound profile of the contact as shown in the signal frequency spectrum in the bottom left of figure 5. Each signal represents a specific emitter on the contact ship. For example, if a loud generator on board the contact ship runs at 50kHz, the contact signal would have a relatively high volume signal at 50kHz with relatively low noise, for generator signals tend to be cleaner and not fluctuate frequency often. The plot on the bottom right of the figure shows the relative bearing of the contact over time, being recomputed every minute for an hour.

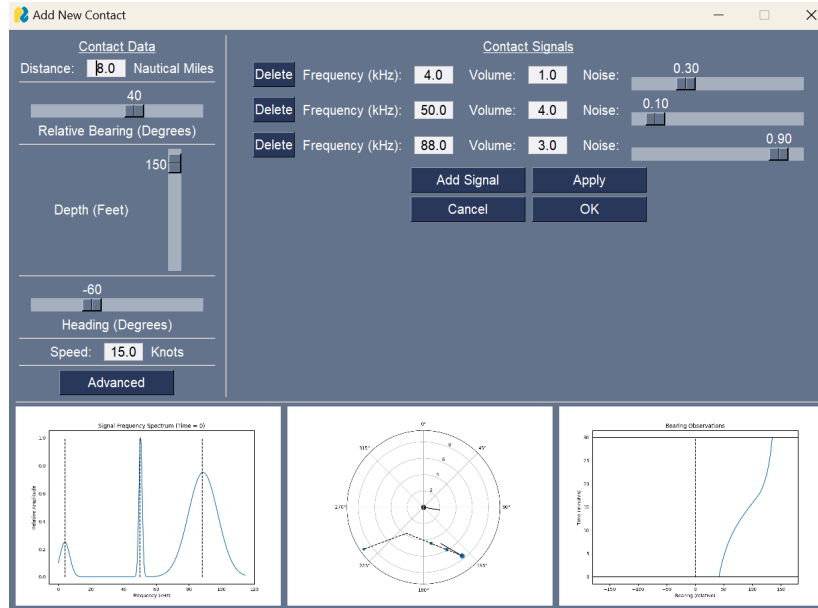


Figure 5: Interface to adjust the variables for a contact

When running the simulation, there are four types of transmission losses that are taken into account, all of which depend on the propagation path and the angle of reception. This simulator assumes that the contact signals are received along the relative bearing of the contact, such that the signal does not refract horizontally during transmission, and that signals are emitted perfectly spherically, so the contact signal extends at all possible angles from the contact. To determine the angle of reception, it is a matter of determining which emission angle causes the contact to perfectly refract to the hydrophone. Then, the angle of reception that corresponds to this emission angle can be recorded. The propagation paths for sample contacts are shown in figure 8.

To determine the exact angle (above or below horizon) from which the signal is detected, a pseudo gradient descent algorithm is performed. All calculations occur on the plane along the relative bearing of the contact and water depth. First the 180 degrees of possible emission angles is linearly subdivided into 16 angle ranges, and all of the angle boundaries of these groups are tested to see if the end of the signal propagation line is above or below the hydrophone. For angle groups in which both boundary angles result in a propagation line that ends above the hydrophone (or vice versa), it is assumed that no angle in that group ends at the hydrophone. For any angle groups that have one boundary angle end above the hydrophone and one below the hydrophone, there is at least one emission angle within that group for which the propagation line ends at the hydrophone. Each of these groups are then tested with gradient descent on the distance between the depth of the hydrophone and the depth of the end of the signal propagation line. The result is an exact emission angle that results in reception at the hydrophone. Finally, the angles of reception is computed using basic trigonometry. Using this method, the angles and bearings of maximum signal reception are known and the total distance of the signal propagation paths are calculated.

The amplitude of the detected signal at each hydrophone is the sum of the amplitudes of the detected signal from each contact at that frequency and time. The signal volume is computed by accounting for all propagation loss mechanisms. First, spreading loss is the result of the

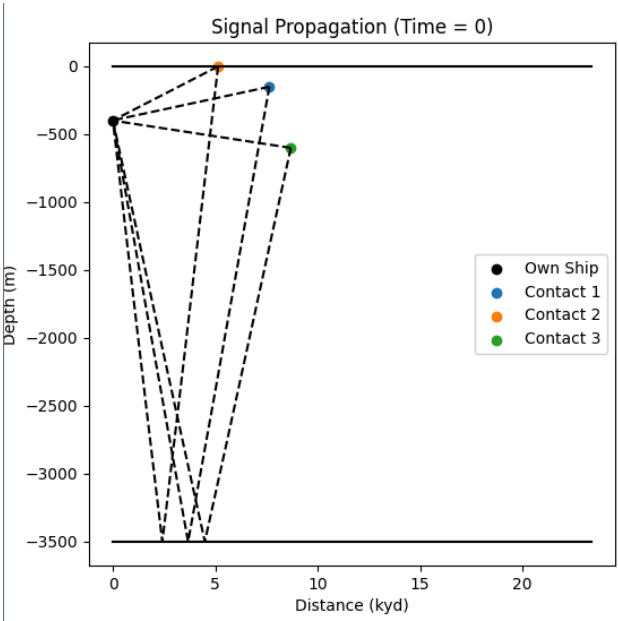
thinning of the signal as it propagates further from the source as a function of distance. At contact distances of less than 1000 meters, it is assumed that the spreading is spherical, and the loss is 20 times the log of the distance in meters. For contact distances greater than 1000 meters, the spreading loss is assumed to be cylindrical, and the loss is 10 times the log of the distance plus 30. Scattering losses are hard to predict, so they are accounted for by adding white noise to the received signals. Scattering and absorption due to reflection, however, is simply computed as 10 times the log of the bottom reflectivity, an input variable that ranges from zero to one. The formula used to compute attenuation losses is a function of sound propagation distance and frequency, as shown in equation 5 below. The total loss for each propagation path is the sum of the different losses, converted into decibels relative to the source. Because the hydrophone array is a sphere, each hydrophone is directed in a different angle and bearing. The angle and bearing of the peak signal reception is known, and the signal amplitude received at each hydrophone can be estimated using trigonometry and the relative angles as shown in equation 6. Once the total loss is known for each hydrophone, it can be scaled as needed to account for observed frequency and time.

$$5. \text{ loss} = \left(\frac{0.0036 * f^2}{f^2 + 3600} + 3.2 * 10^{-7} f^2 \right) * \text{distance}$$

$$6. \text{ loss} = \cos(\text{contact bearing} - \text{hydrophone bearing}) \\ * \cos(\text{contact angle} - \text{hydrophone angle})$$

$$7. f_{\text{observed}} = \left(f * \frac{\text{sound speed} + \text{relative ship speed}}{\text{sound speed}} \right)^2$$

Figures 6-12 are the various outputs of the simulation. Figure 6 shows the exact angle and bearing of each detected signal at each observed time. Figure 7 shows the propagation paths of each contact as a function of radial distance from the hydrophone array in accordance with the SSP and signal propagation mechanisms. Figure 8 shows all of the detected signal frequency spectra at time zero on one plot. The black dotted lines show the theoretical peak frequencies in the spectrum, however several peaks are slightly shifted due to doppler effects and the relative motion between the contact and the observation ship. The relation between the observed frequency and the emitted frequency is shown in equation 7. The noise in the emitted frequency is accounted for with a gaussian curve. Figure 9 also shows the observed frequencies, but on a spectrogram as a function of time. Figure 10 shows the bearing of the observed contacts as a function of time. The zig actions are visible by the sudden discontinuities in the plotted lines. The data from this plot is useful in determining the exact location of contacts, as bearing rate is a key indicator of distance and collision risk. Figure 11 shows the raw broadband observations with color representing intensity of signal for each hydrophone at each time stamp specified in the observational variables. Broadband, as opposed to narrowband, consists of all frequencies and is a representation of the magnitude of the signal. Figure 12 shows the equivalent narrowband data, which is scaled to the presence of each frequency specified in the operational variables. This allows the spectra of each contact to be highlighted for classification.



Figures 6 and 7: Numerical signal reception angles and radial signal propagation plot

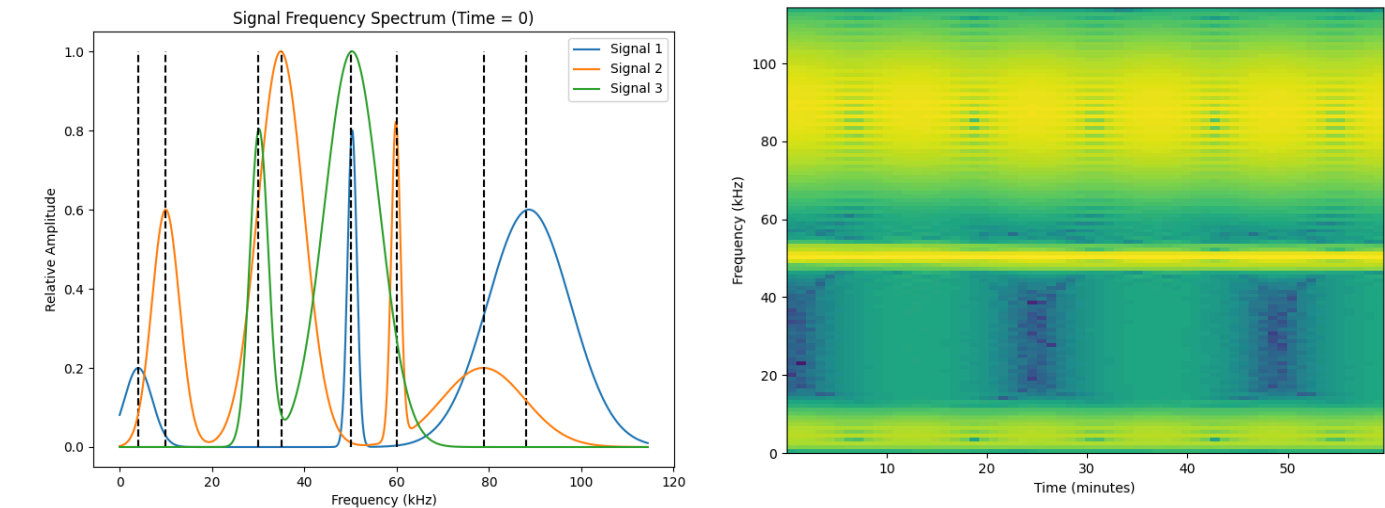


Figure 8 and 9: Signal frequency spectrum and spectrogram

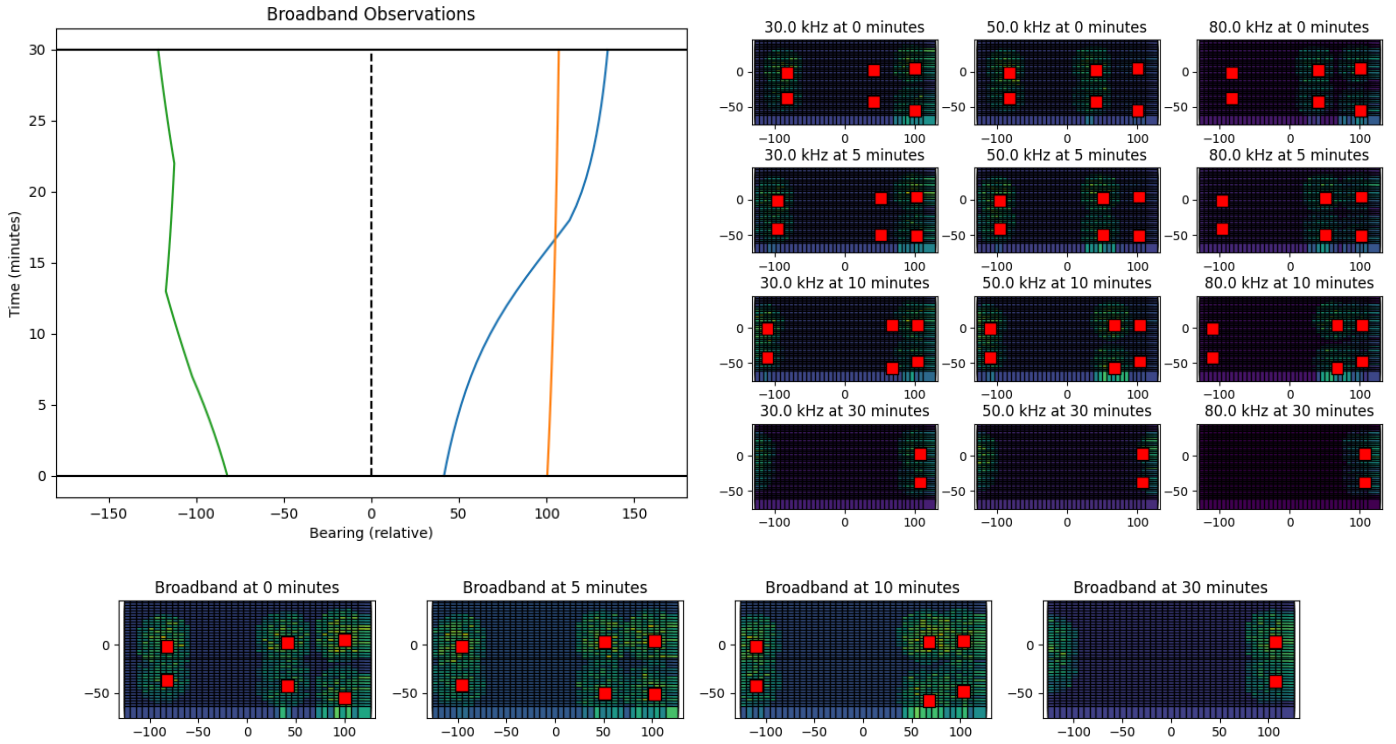


Figure 10 (left), 11 (bottom), and 12 (right): Contact bearings, broadband, and narrowband observations

To facilitate the creation of datasets, a dataset creation feature is implemented so that the requested number of datapoints can be created and saved to a .csv file. Ranges are specified for each variable, values are selected at random, and the simulation is run. The data collected from this simulator is used to train the necessary models for contact classification and localization.

Contact Classification

When operating sonar navigation, particularly in a mission environment such as on a naval vessel, classifying contacts is essential to safety of ship. Differentiating military vessels from cargo vessels from biologics allows for the proper response for the mission. Every ship has a unique sound signature and classifying this is a matter of fourier analysis. Autonomously identifying the frequency spectra that composes a ship's sonic emission can be done through various DSP effects and deep learning concepts.

The first step in the classification process is distinguishing signal from noise, typically done by applying a matched filter to the signal and observing which elements stand out. There is a challenge in building models to perform this classification automatically, and a neural net was trained in an attempt to streamline this process. The provided data comes from a laboratory, where passive observations were made from 0-400 Hz and are labeled as signal or noise. The frequency spectra of each observation was divided into four segments of 100 Hz with each segment consisting of 125 frequency values linearly spaced over the range, and all four segments are used as independent data points. Because the exact frequencies of each range are not recorded, classification needs to occur based on the shape of the frequency spectra rather than the specific frequencies. Ten sample noise spectra and ten sample signal spectra are plotted in figure

13. It is clear in these plots that the signal plots have much sharper spikes at specific frequencies while the noise contains no distinct frequencies.

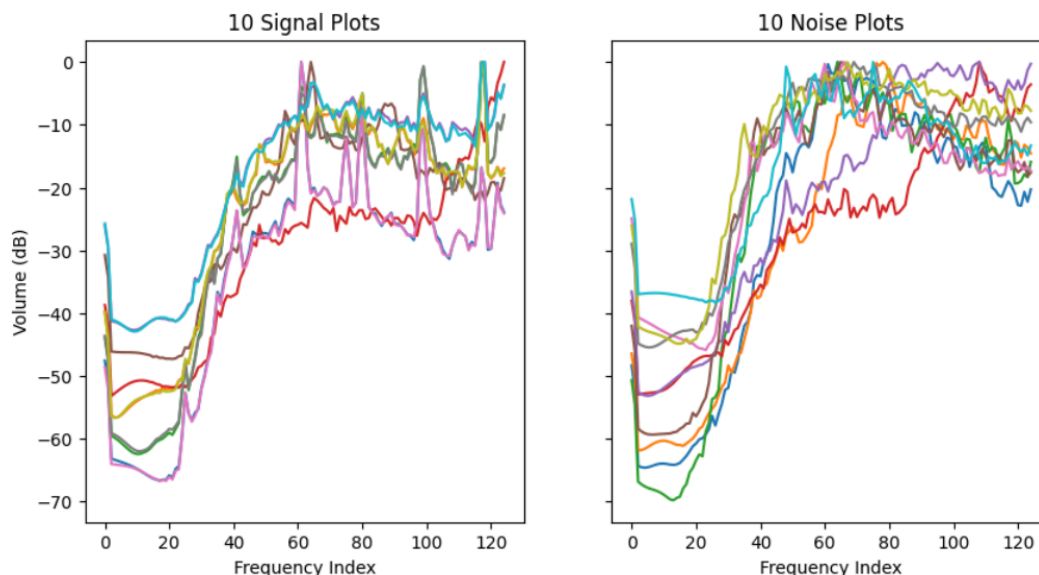


Figure 13: Sample signal and noise frequency spectra

The structure of the basic CNN used is shown in figure 14. The model contains three layers, each with a 1D padded convolution of kernel size 5 and max pool. Rectified linear unit (ReLU) activation functions are used for their ability to efficiently eliminate negative parameter values. Two fully connected layers are used at the end to reduce the size of the output from 240 to 2, representing the two classifications. The data is split into training, validation, and test sets as shown in the table below. Approximately 20% of the data is used for validation and 10% is used for testing. An imbalance in the classifications is apparent, so class weights need to be implemented to prevent overfit.

Dataset	Signal	Noise	Total
Training Set	7361	27717	35078
Validation Set	2158	8365	10523
Testing Set	972	3539	4511
Total	10491	39621	50112

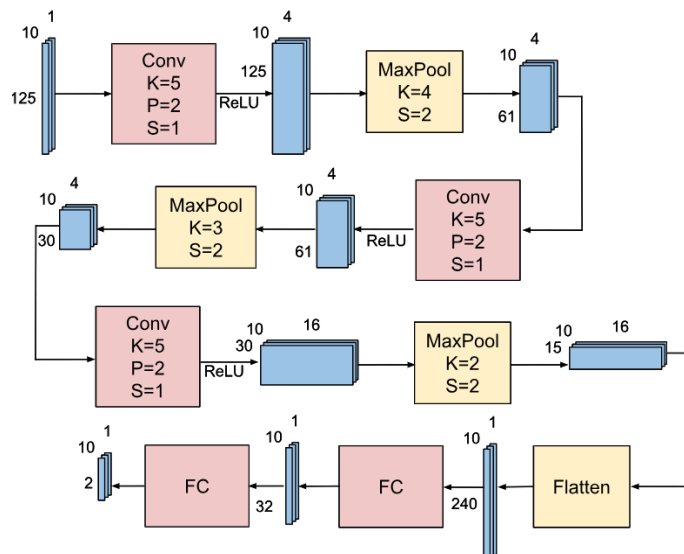


Figure 14 (right): 3-layer CNN used for signal classification

This model was trained for 30 epochs using a batch size of 10, a learning rate of 0.002, and a regularization parameter of 0.0001. Cross entropy loss was used as the loss function. The

results of this training are shown in figure 15. Overall, a maximum validation accuracy of 88.774% was achieved with a minimum loss of 0.2991 after training for 12 epochs. A test accuracy of 85.40% was achieved. These results are better than expected, accounting for the data imbalance and the exclusion of exact frequency values. A more advanced architecture, such as a ResNet may increase the performance, however the simple CNN structure used works because of the clear distinctions in shape between the noise and the signal.

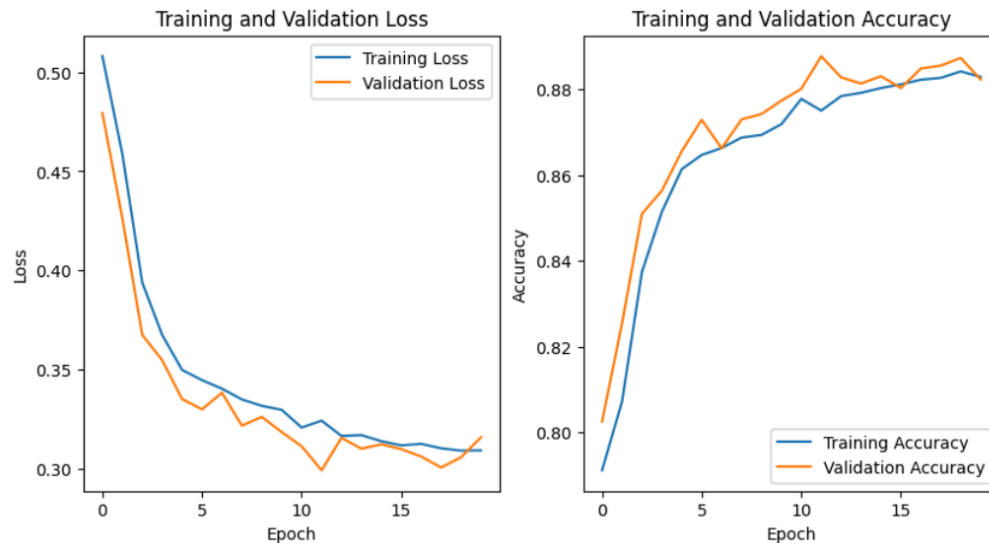


Figure 15: Training and validation loss and accuracy

Because no additional natural data was provided for contact signal classification, a classifier model for different ship signatures was not trained. One option for this classifier is a traditional CNN trained on the frequency spectra, similar to the model trained above. Because ship signatures can rely on a very wide range of frequencies, this type of model can complicate very easily. Additionally, no data augmentation is possible because shifts, flips, etc., all result in an entirely different signature. An alternative involves creating a spectrogram based on the raw data, and applying various computer vision models to the resulting image. Unlike with sound data, image data can be easily augmented and recolored, allowing for the creation of new data points. This model can be studied with additional research.

Contact Localization

Beyond classifying contacts, making proper safety-of-ship maneuvers is impossible unless the exact location of the contacts can be approximated. Because ships navigate on a polar plane, two steps are required for contact localization: determining contact relative bearing and determining contact distance. Finding relative bearing is a much easier task than finding distance, due to the passive nature of the observations. Theoretically, contact bearing can be determined simply by identifying which bearing receives the strongest signal after noise elimination. Due to the relative simplicity of this problem, traditional computer vision algorithms are explored on the hydrophonic data rather than deep learning techniques.

Before bearing can be determined, the noise in the data needs to be accounted for. A traditional 2D window smoothing technique is explored. Hypothetically, replacing each hydrophone signal intensity value with the averages over a window centered at that hydrophone

will account for the random noise, as shown in equation 8. This assumes the noise is random gaussian centered at zero. Figure 16 shows an example of this data smoothing for a particular data point. This particular data point includes three contacts, each with two detected propagation paths, as described in the “Time = 0 minutes” section of figure 6. Each column represents a hydrophone in the array located at the specified bearing and angle, and the height represents the intensity of the received signal. For this particular data point, 30% noise is applied, representing a zero-mean gaussian white noise distribution with a standard deviation of 3.0 dB. The smoothing factor, k , used in this example is 5.

$$8. \text{Intensity}_{I,J} = \frac{1}{k^2} \left(\sum_{i=I-k/2}^{I+k/2} \sum_{j=J-k/2}^{J+k/2} \text{Intensity}_{i,j} \right)$$

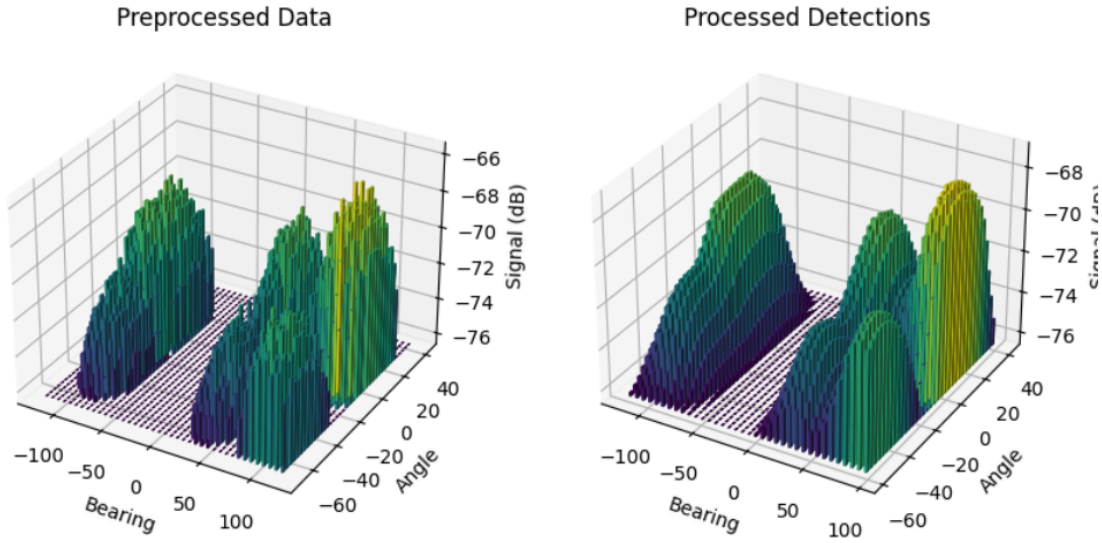


Figure 16: Hydrophone intensity data before and after processing

To determine the exact peak angle and bearing of the detected signals, a gradient ascent technique is applied. Starting with hydrophones linearly spaced in both spatial dimensions, gradient ascent allows for the efficient calculation of which hydrophone represents a local maxima. The hydrophone array used in this study contains 40 columns linearly spaced from -120 to 120 degrees, or 6 degrees between each hydrophone. As such, the hydrophone with the greatest intensity may be up to 3 degrees off from the actual bearing of detection. To account for this, a 2D parabolic interpolation is taken in a 3x3 window around each local maximum. The maximum of the resulting 3D paraboloid is used as the exact bearing of detection.

Figure 17 shows the predicted signal bearings and the ground truth bearings. In general, the predictions are accurate in that there is about one prediction for every ground truth. The differences between the predictions and ground truths are likely due to the window smoothing algorithm. Several of the detection points are close together, so window smoothing causes the peaks to shift towards the intersections of the detection curves. Different windows, such as a circular window, could be used to create a better approximation of the signal detection peak. The array is circular, and these plots are generated using a mercator projection onto a 2D plane. Thus, there is distortion in the shapes and basic window smoothing fails to account for the complexity in the array shape. Additionally, error derives from the interpolation mechanism, for it is

assumed that the local curve in the nine-hydrophone neighborhood surrounding the peak is parabolic. In reality, there are many factors that influence the shape of the detection curve, and more advanced trigonometric models need to be used for a more accurate interpolation.

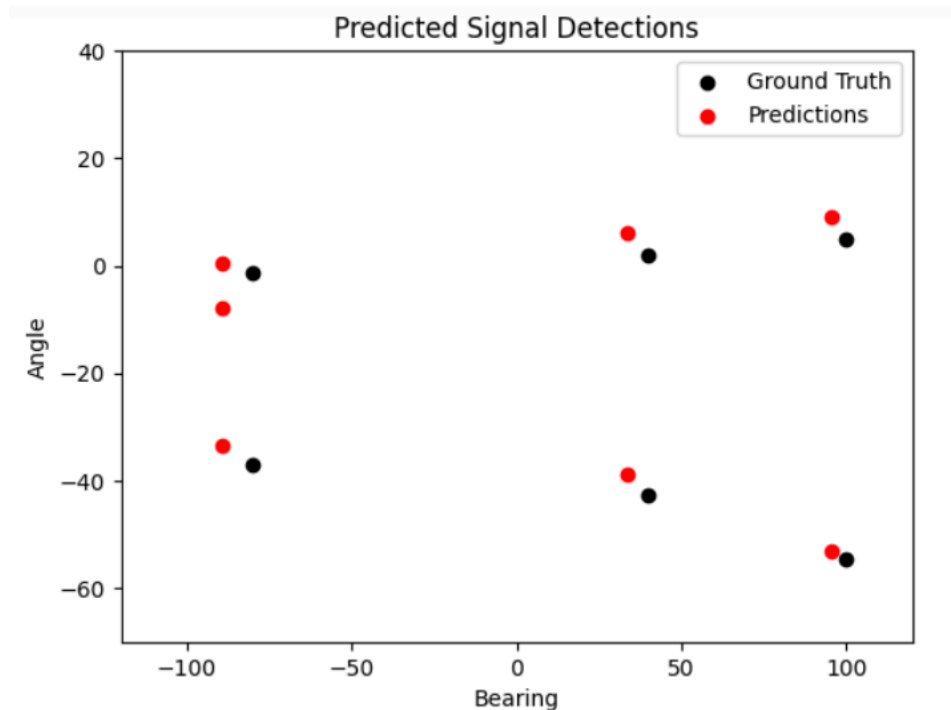


Figure 17: Predicted bearings using window smoothing

Once finding the bearing of each contact at each time value, such as in figure 10, the distances of the contacts need to be approximated. Finding distance using passive observations has proved to be a challenge that remains unsolved today. Currently, a range of possible distances are found using a range of assumptions about the contact, such as speed and direction of travel, and the contact is assumed to be at the closest reasonable distance to ensure safety-of-ship. Certain TMA maneuvers can be performed on a ship to limit this range of distances, typically involving observing bearing change during a turn. However, assumptions still need to be made about the contact. Bearing rate, or the derivative of bearing over time, is a key indicator of contact distance. Ships that are closer generally tend to have greater bearing rates when moving along the same path unless a further ship is moving significantly faster. Ships with a zero or near-zero bearing rate are typically on course for collision. Could a deep learning model be trained to approximate distance over time relying only on the bearing time series data without too many additional assumptions?

Such a model needs to use all previously known bearings and previously predicted distances to make the most accurate prediction possible. This architecture is similar to that of a large language model (LLM) decoder architecture which relies on the previous words in a sentence to predict the next word. Instead of word vectors, the bearing time series data can be the model input, allowing the model to use the bearing rate and acceleration as well as previous distances to predict the next point. A series of 64 LSTM cells are used in this model to provide basic decoder-only memory capabilities.

The data used for this model comes from the simulator. Nearly 200,000 data points were collected with random variable inputs and no zig actions, where each point consists of 30 minutes of relative bearings and the corresponding 30 minutes of distances. Figure 18 shows 10 sample data points with the bearing and distance plots of the same color corresponding to the same contact. An important observation is that distance always increases as time approaches infinity. This is logical, as the distance between two ships will always increase with time in the long term, assuming they are not traveling in perfect parallel. Additionally, observe that if the minimum distance point is within this time frame, the minimum always occurs at the inflection point on the bearing line. This is because the bearing rate (first derivative of bearing line) is always fastest when the contact is the closest to the observational ship. This point is known as the closest point of approach (CPA). Navigators want to maintain a CPA greater than a certain threshold to ensure there is no risk of collision.

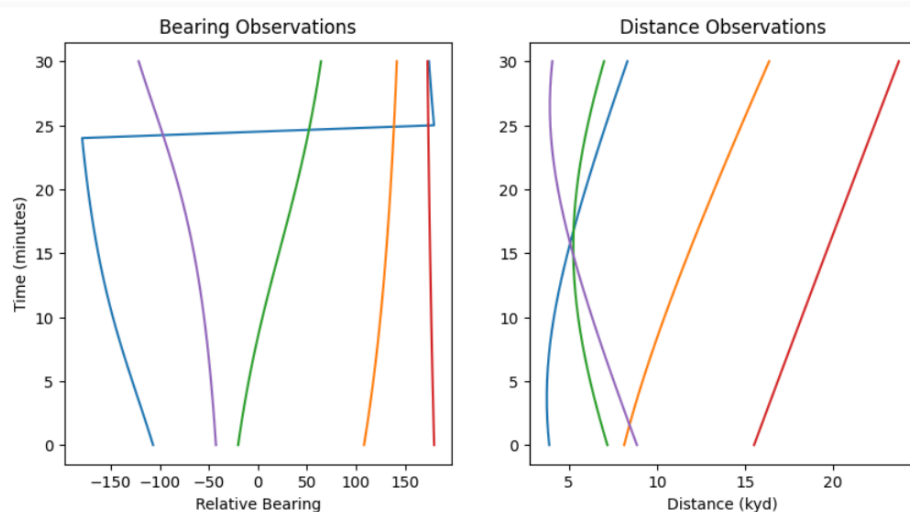


Figure 18: Bearing and distance observations over time

The LSTM model was trained for 30 epochs using 30% of the data for testing and a mean square error (MSE) algorithm for the loss function. With a batch size of 10, learning rate of 0.001, and regularization parameter of 0.0001, the training achieved a minimal loss of 13.2035 at 21 epochs.

Some sample test paths are shown in figure 19. In this plot, five sample bearing time series are passed through the model and both the predicted and ground truth distance plots are shown, with the solid line and dashed lines of the same color representing the ground truth and predictions respectively. The LSTM model works surprisingly well, as it is able to effectively predict the shape of the distance curve over time. The predictions are always increasing as time approaches infinity, and the CPA times are accurate. In many test cases, however, the distance plot is shifted so that the exact distance at a specific timestamp is inaccurate. This is likely because predicting distance depends a lot on the speed of the contact and the direction of movement, both of which were not accounted for in this model.

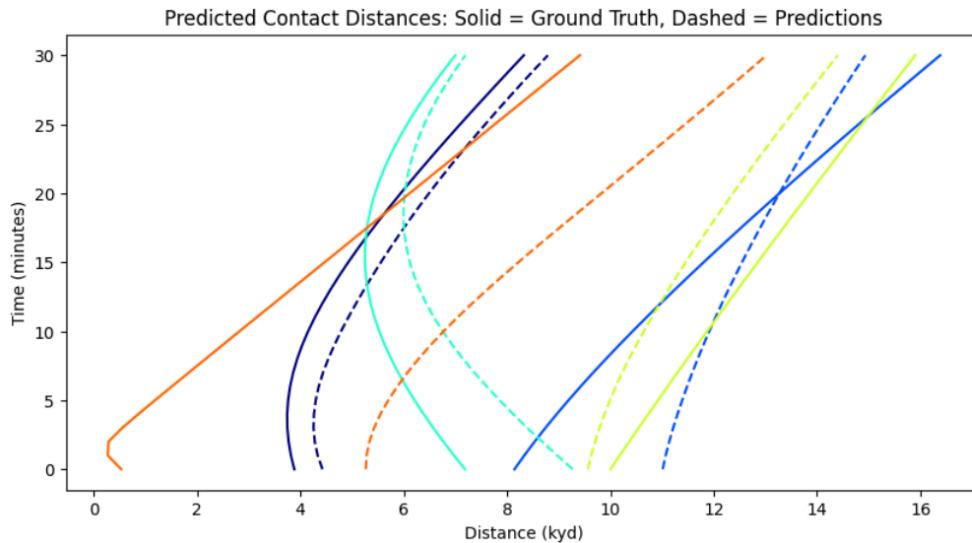


Figure 19: Predicted and ground truth distance curves

With additional tuning and with the implementation of assumptions such as contact maximum speed, this model should be able to approximate contact distance accurately. The benefit of an LSTM model is that it can be easily adjusted to predict given any number of time points. As a result, if a contact has only been visible for a few minutes, a distance prediction can still be made. Regardless, this architecture may be a feasible option for automating contact localization in the future.

Motion Estimation

Understanding relative bearing and distance can allow for the creation of line of relative motion (LRM). The final step in understanding your environment is determining the true course and speed of each contact. This can be accomplished through vector geometry on the LRM by accounting for the observational ship's course and speed. Because of the inaccuracy in the LSTM model, it makes sense to try and compute the true course and speed using a different algorithm and averaging via a kalman filter or other method. Specifically, can true course and speed be determined through knowledge of doppler shifts in signal frequency?

This is tested using the demo simulation with the spectrogram shown in figure 9. A singular contact was used in this simulation for simplicity. The first step in this process is determining peak frequencies at each time stamp from the spectrogram. This is done by finding the local maxima among the frequency values included. Because the frequencies are plotted discretely, and the change in doppler shift over time is likely less than the size of the frequency box, it is necessary to approximate the actual maxima interpolated between the discrete values. This is done using a simple cubic spline and finding which frequency values maximize the function. After finding the local maxima for the first time stamp, a spline can be created around those frequency values in all future time stamps, and the change in frequency can be observed over time. Figure 20 shows the result of this computation on the spectrogram in figure 9. Figure 21 shows a scaled version of the same plot so that the changes in doppler shift over time are more apparent. The contact used in this sample contains three clear frequency peaks around 88, 50, and 4 kHz.

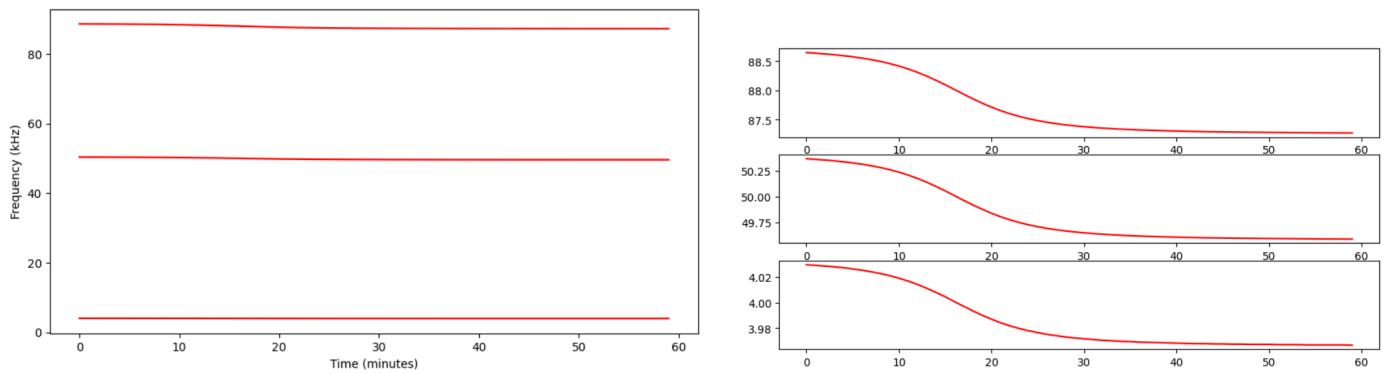


Figure 20 and 21: Peak detected frequencies over time

Clearly, these plots show the change in doppler shifts over time, demonstrating the general relative path of the contact. The inflection point around 16 minutes represents the CPA of the contact. Before this, there is a positive doppler shift, indicating that the contact is moving towards the observational ship. After the inflection point, there is a negative shift, indicating that the contact has passed the CPA and is moving away from the observational ship. It is assumed that the frequency at the inflection point is the true frequency because the contact is not moving closer or further at the CPA. These shifts can be used to determine the relative speed of the contact using equation 7.

Figure 22 shows the result of applying this equation. The relative velocity is plotted for all three peak frequencies and the average is shown by the dotted line. Because higher frequencies are impacted more by the doppler shift, the 88 kHz signal is the most reliable indicator of relative motion. This is shown in the figure, as the 88 kHz signal (signal 2) is the smoothest line. Nonetheless, all three peak frequencies follow relatively the same path, proving that doppler shift can be used to predict relative motion. The CPA is at around 16 minutes as expected, and there is a zig action present at 18 minutes, indicated by the dotted line.

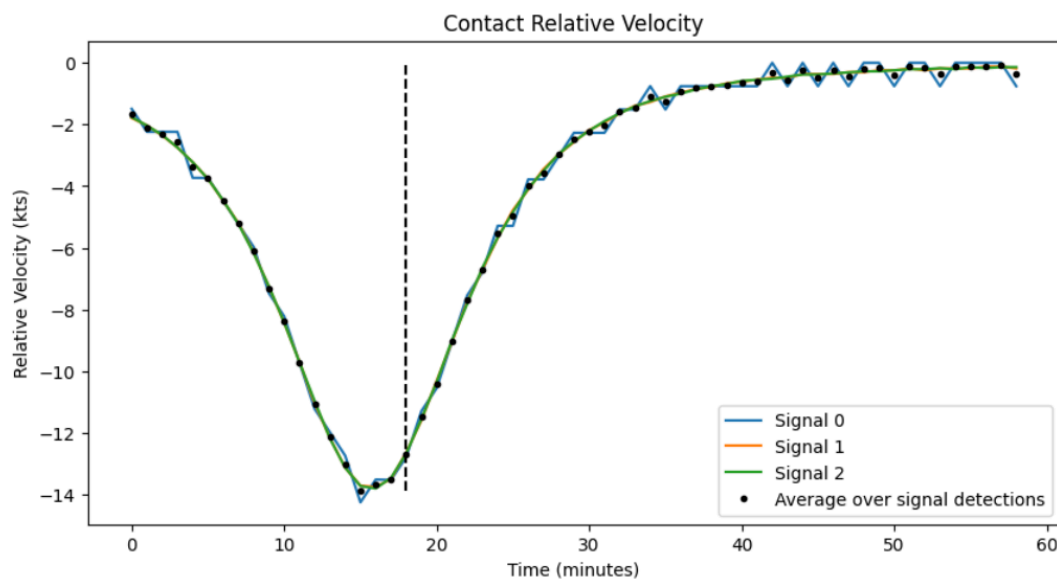


Figure 22: Relative velocity over time calculated from doppler shifts

The contact distance and relative bearing, which can be computed using the methods previously described, are shown in figure 23. The ground truth is this contact travels at 15 kts along a 300 degree bearing before changing course to 235 and increasing speed to 20 kts at 18 minutes. The moment of this zig action, indicated by the black dotted lines, is visible as the bearing and distance curves undergo a sharp change. Figure 24 shows this action on a polar plot.

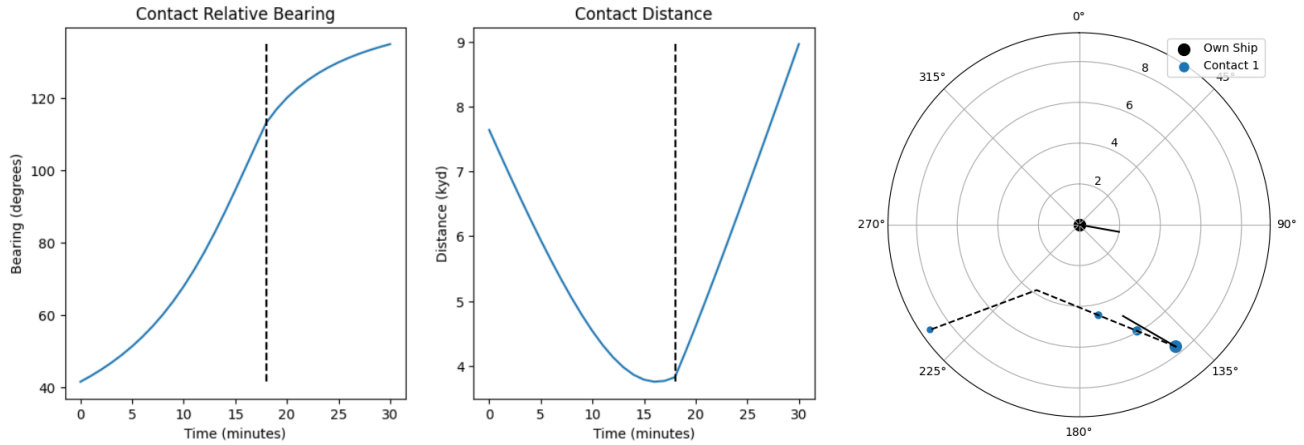


Figure 23 and 24: Ground truth path of contact

Knowing relative bearing, relative speed, and distance, true course and true speed can be computed trigonometrically. Equations 9-13 show the calculations used to approximate true contact course and speed. D is the contact distance values, RB is the relative bearing values, and RV is the relative velocity shown above.

$$9. \text{LRM Heading} = \arctan\left(\frac{D_{i+1} * \sin(RB_{i+1}) - D_i * \sin(RB_i)}{D_{i+1} * \cos(RB_{i+1}) - D_i * \cos(RB_i)}\right)$$

$$10. \text{True Course}_x = \text{Own Ship Speed} * \sin(\text{Own Ship Bearing}) - RV * \sin(\text{LRM Heading})$$

$$11. \text{True Course}_y = \text{Own Ship Speed} * \cos(\text{Own Ship Bearing}) - RV * \cos(\text{LRM Heading})$$

$$12. \text{True Course} = \arctan\left(\frac{\text{True Course}_x}{\text{True Course}_y}\right)$$

$$13. \text{True Speed} = \sqrt{\text{True Course}_x^2 + \text{True Course}_y^2}$$

After performing these calculations for every time stamp, the estimated true course and speed are shown in figure 25. The ground truth as described above is shown by the dashed black lines. While the approximation is not perfect, the general shape is correct. The zig action is correctly found at 18 minutes shown by the sudden jump in the predictions. Generally, the contact true heading is independent of the relative bearing, as relative bearing always changes over time while heading can remain constant. Inaccuracies in the prediction are likely due to the approximations used in finding the relative velocity. Because averages were taken over the three frequency peaks, the relative velocity line is discontinuous. For a more stable prediction, just the relative velocity line found from the largest frequency should have been used.

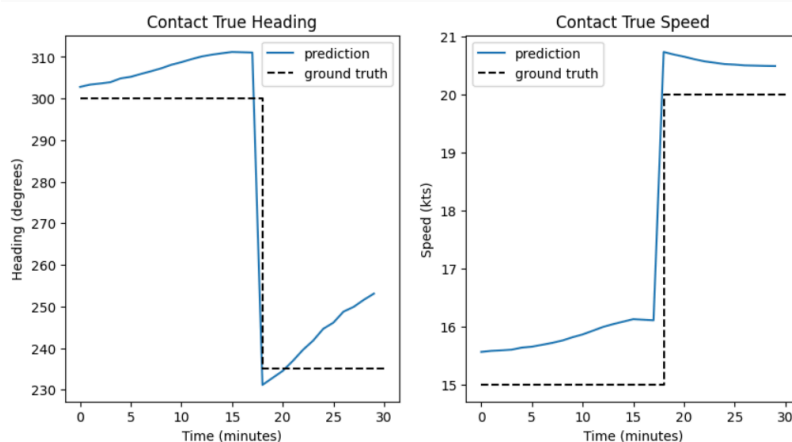


Figure 25: Predicted contact true course and speed

Conclusion

Passive sonar systems are critical for safety-of-ship navigation in mission-critical environments. The clandestine nature of passive systems allows for their use when EMCON is necessary, but creating an understanding of the surrounding environment can be quite a challenge. Through a pipeline of steps proposed in this paper, navigation can be automated via computer vision and deep learning techniques. Contact classification can be accomplished through a simple CNN architecture analysis of spectrogram data, allowing for sonic data to be represented visually. Noise and signal information can also be differentiated using a basic CNN structure. Contact localization provides a greater challenge and can be subdivided into several steps. Determining contact relative bearing is a matter of processing the hydrophonic data and assessing which bearing has the strongest return. Passively determining contact distance is a continuing area of research, and this paper proposes a decoder-only LLM model for distance estimation. Lastly, an analysis of doppler shifts over time can allow for the true bearing and speed of a contact to be approximated.

Due to the limited time span of this research, there are several areas of study that are left for future work. The limited access to data severely inhibits the ability to apply these models to real world scenarios. If given actual discrete time-domain signals from each hydrophone, DSP methods can be applied and the true performance of these models can be tested. For example, after completing a fourier analysis, a spectrogram of a contact's emission can be observed. This allows for a legitimate classifier to be trained on sound signature components such as propeller shape. Due to the spherical assumption of the hydrophone array, no beamforming techniques were employed. Additional work would include applying these models to the linear phased towed array, which relies on a passive beamforming to establish a conical view of the surrounding environment and cover the baffles. In general the use of additional real-world data would permit these models to have a more realistic representation. Overall, the pipeline described in this paper is a starting point for additional research into the automation of passive sonar navigation.