Source localization with temperature variation in a laboratory tank

Natalie Bickmore

A senior thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Bachelor of Science

Tracianne B. Neilsen, Advisor

Department of Physics and Astronomy

Brigham Young University

ABSTRACT

Source localization with temperature variation in a laboratory tank

Natalie Bickmore
Department of Physics and Astronomy, BYU
Bachelor of Science

Environmental variability complicates deep learning applications to ocean acoustics. Transfer learning has the potential to improve the deep learning model's ability to perform in a new environment. This work illustrates the potential of transfer learning for the case of predicting source-receiver range in a laboratory tank with changing water temperature. I trained a one-dimensional four-layer convolutional neural network to predict the source-receiver range. The training dataset consists of spectral levels from linear chirps (50-100 kHz) measured at room temperature. Due to time-dependent ambient noise, data measured on multiple days are combined to form training datasets with 1806 samples. The training process employs early stopping, and the trained models are validated on test datasets also measured in room temperature water. The trained models then process data measured in warmer water, and the error in the predicted ranges increases. The model undergoes transfer learning using a dataset with 51 samples measured at warmer temperatures, and new testing follows. Transfer learning refines the trained model and leads to reduced errors in the predictions. This work provides an example of how transfer learning could be applied to accommodate environmental variation in ocean applications of deep learning.

Keywords: source localization, deep learning, convolutional neural network, transfer learning, underwater acoustics

ACKNOWLEDGMENTS

I would like to thank Brigham Young University for its emphasis on providing opportunities for undergraduate research. I am especially grateful to the generous donors and supporters of the College of Physical, Mathematical, and Computational Sciences, whose financial contributions made this research assistantship possible. I also extend my thanks to the Office of Naval Research for funding the Brigham Young University Underwater Acoustics Lab.

The completion of both this thesis and my undergraduate degree was made possible in large part by Dr. Traci Neilsen. Through consistent meetings, unwavering faith in my abilities, and encouragement to step outside my comfort zone, she has provided invaluable mentorship. I deeply appreciate Dr. Neilsen's availability to her students, both academically and emotionally. She has taught me not only how to be a better research scientist but also a stronger leader and person.

I am also grateful for the contributions and constant support of my fellow BYU lab members. Madilyn Randall, Ben White, Sarianne Winters, Taylor Jackson, and Molly Boseman assisted in data collection and always made the lab an enjoyable place to work. I wish to especially acknowledge Corey Dobbs, whose patient mentoring fostered enthusiasm for research and confidence in my technical skills. The thorough work he produced during his master's program laid the foundation for this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis explores applications of machine learning methods to underwater acoustics. The chapter opens with a discussion on motivation and relevance for machine learning usage in this field. Challenges are addressed and a potential solution using transfer learning is proposed. The chapter also outlines the rest of the thesis, recognizing previous work, and clarifying my contributions.

1.1 Motivation

Source localization is the process of estimating the position of a sound source relative to the recording sensor [1]. Passive source localization (passive sonar) refers to the detection of sound sources without emitting any signals, relying solely on incoming sound waves. Active source localization (active sonar) involves emitting a signal and then analyzing the reflected waves to determine the source location. Source localization has both defense-related and environment-related applications. Implications on the defense side would include the discovery of mines, the location of submarines in the deep ocean, and the detection of intruders near the coastline [2]. Environment-related impacts would include monitoring fish stocks and observing marine mammals or other species (especially in response to climate change). Figure 1.1 from Sinay [3], a maritime technology

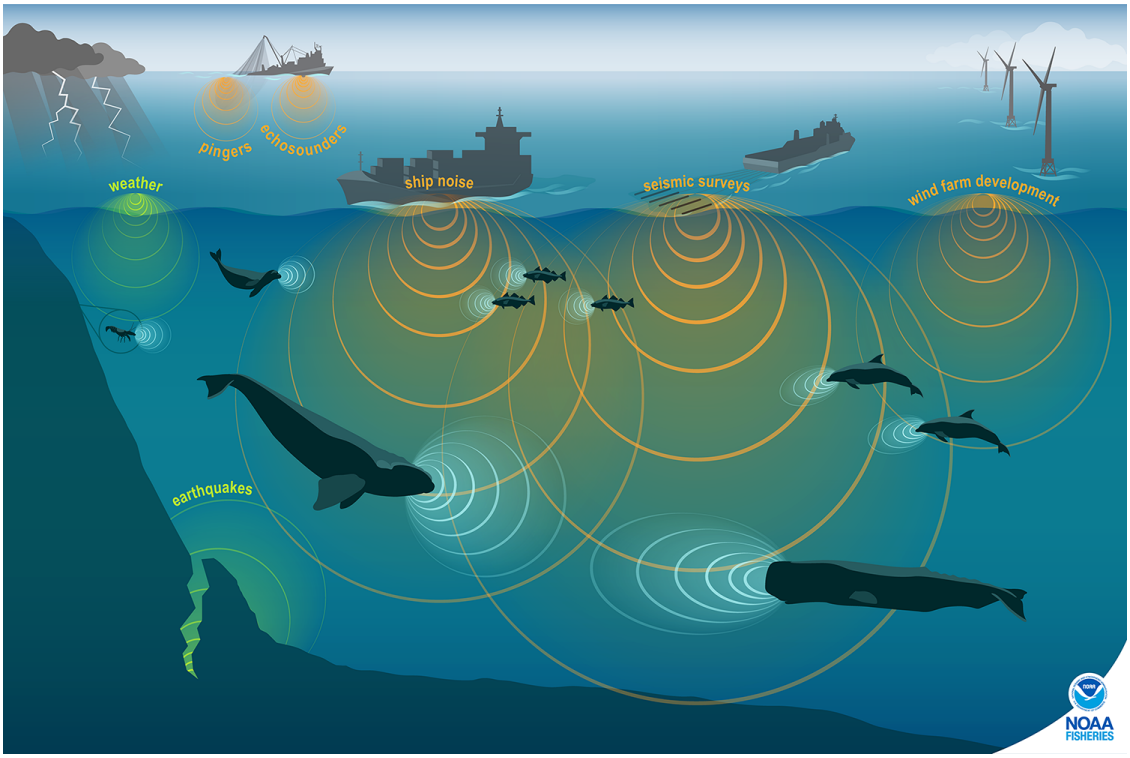

Figure 1.1 Illustration of active acoustic sources in the ocean, including sound emissions from marine animals, weather, and a surface vessel. This image conceptually represents active source localization, where emitted signals are used to analyze and detect the location of underwater sources. Source: National Oceanic and Atmospheric Administration (NOAA) [4]

company that provides data solutions for environmental monitoring, depicts a scene of source localization in the ocean. The figure shows the active source localization process with a surface vessel, a submarine, a dolphin, a pod of fish, and a whale.

For several decades, the leading technique for source localization in underwater acoustics has been signal processing [5]. However, as noted in a review on advancements in machine learning for underwater acoustics, "machine learning has been applied to passive sound source localization since the early 1990s" [6]. Machine learning has been applied to underwater acoustics in various subdomains, ranging from source target recognition to underwater communication [7]. One prominent application of machine learning is source localization.

An approach for passive source localization in underwater acoustics is deep learning, a subset of machine learning that employs artificial neural networks to recognize patterns. One type of neural network that is especially good at detecting patterns is a convolutional neural network (CNN). Supervised training of these networks relies on labeled data: data where the ground truth is known for the property that is being learned by the network. Deep learning-based methods hold significant potential for sonar technology, enabling efficient source tracking and localization. In addition, it has applications in geoacoustic inversion, a technique used to characterize the seabed based on acoustic data.

Examples of deep learning usage in underwater acoustics for source localization have grown over the past decade. Two such examples include Huang *et al*. [8] and Wang *et al*. [9]. Huang *et al*. studied source localization in a shallow water environment using a 30 hydrophone vertical line array with both simulation data and real data. Wang *et al*. also used a vertical line array but with 21 hydrophones. The results for Wang *et al*. were collected for shallow, deep, broadband, and narrowband cases. The results of these two studies are compared with the results of this thesis in Chapter 4.

Two major challenges impede the application of deep learning to oceanic source localization: environmental variability and scarcity of labeled data. There are large databases of unlabeled underwater acoustic data, but source localization through deep learning requires labeled underwater acoustic datasets. The lack of large labeled datasets presents a significant obstacle to creating a robust deep learning model because it requires extensive labeled data to train supervised learning models. Large labeled underwater acoustic datasets are much more difficult to obtain for each scenario.

Even with labeled data, training a reliable model is very difficult due to the fact that ocean environments are characterized by significant noise, signal distortions, and complex propagation phenomena [6]. These complexities include variations in water depth, ambient noise caused by

surface and wind activity, temperature fluctuations, and differences in seafloor material. This variability complicates the development of robust machine learning models capable of generalizing across diverse oceanic conditions. Addressing these issues requires the development of a method to adapt deep learning models to variable environments.

1.2 Transfer Learning Approach

One approach to address the challenges of ocean variability and limited labeled data in underwater source localization is transfer learning. Transfer learning is a technique that enhances a deep learning model by leveraging knowledge from a model trained on a large dataset [10]. This approach allows accurate results to be achieved even when only a small dataset is available, making it particularly advantageous in situations where labeled data is scarce.

The key advantage of transfer learning lies in its ability to modify an existing model to suit a new task rather than building a new model from scratch. By modifying an existing model, it reduces the need for extensive data collection and training. This technique works by sharing certain deep neural network parameters from the original model while relearning others specific to the new context [6]. This adaptability makes transfer learning a powerful tool for overcoming the limitations of data availability and variable environmental conditions in underwater acoustics.

Hu *et al*. [11] tested a transfer learning approach for dolphin sound recognition. They found that transfer learning performed more than 15% higher in average accuracy than the usual machine learning methods. The dolphin sound recognition model was a classification model (discrete categories), while this thesis uses a regression model (continuous values). A different study, Wang *et al*., tested transfer learning for underwater acoustic source ranging. The study initially used synthetic datasets to train the model and refined the model with real deep-sea data. Unlike this thesis in which the source-receiver range is found using regression, Wang *et al.* discretized the

source ranges into bins, creating a classification problem. Wang *et al.* used a probabilistic Gaussian distribution model which makes comparison difficult with this thesis. [12].

1.3 Thesis Outline

In this thesis, I share the process and results of applying transfer learning to predict source-receiver range in a laboratory tank when temperature fluctuations are introduced. The primary objective is to evaluate the effectiveness of transfer learning as a method for source localization, specifically in accounting for environmental variability such as temperature changes. In Chapter 2, I provide information on the measurements including frequency range, temperature, and the varying ambient noise. I also describe the neural network architecture, focusing on validation methods and early stopping. In Chapter 3, I discuss the results of comparing prediction errors between deep learning models and transfer learning models. This comparison demonstrates that transfer learning effectively reduces prediction errors in the estimation of source-receiver range under temperature fluctuations in a controlled laboratory setting. In addition, Chapter 4 concludes the thesis and explores the applications of these findings and outlines potential directions for future research in this field.

1.4 Previous Work

The foundational work for this project was conducted by Corey Dobbs during his undergraduate and graduate studies at Brigham Young University (BYU) [13]. He developed the original code used for training, testing, and performing transfer learning on the models. Corey's contributions also included optimizing the neural network architecture, such as determining the ideal number of samples for the model. He also conducted training and testing on multiple models to refine their performance.

My role in this deep learning project involved building upon the existing framework. I modified

the CNN architecture to improve the implementation of early stopping, ensuring better prevention of overfitting. Additionally, I designed a more efficient data processing method by applying band-pass filtering based on the frequency range. Beyond code optimization, I participated with other lab members in collecting new data from the laboratory tank. I used these updated datasets to train and test the models.

Chapter 2

Methods

This chapter outlines the application of transfer learning to data collected in a laboratory tank. The first section details the experimental setup, including the features of the laboratory tank, the measurement chain, and the relevant measurement variables. The following section discusses the architecture of the CNNs used for training, testing, and transfer learning. It provides an overview of key model parameters, such as the number of layers, early stopping techniques, and the number of epochs.

2.1 Experimental Setup

Measurements were taken in the BYU underwater acoustics laboratory to train and test multiple deep learning models. The tank is used as a surrogate for the ocean. It is a controlled environment that I can use to test specific variables. Standard measurements were taken for training models and variation measurements were taken for testing and applying transfer learning to the trained models.

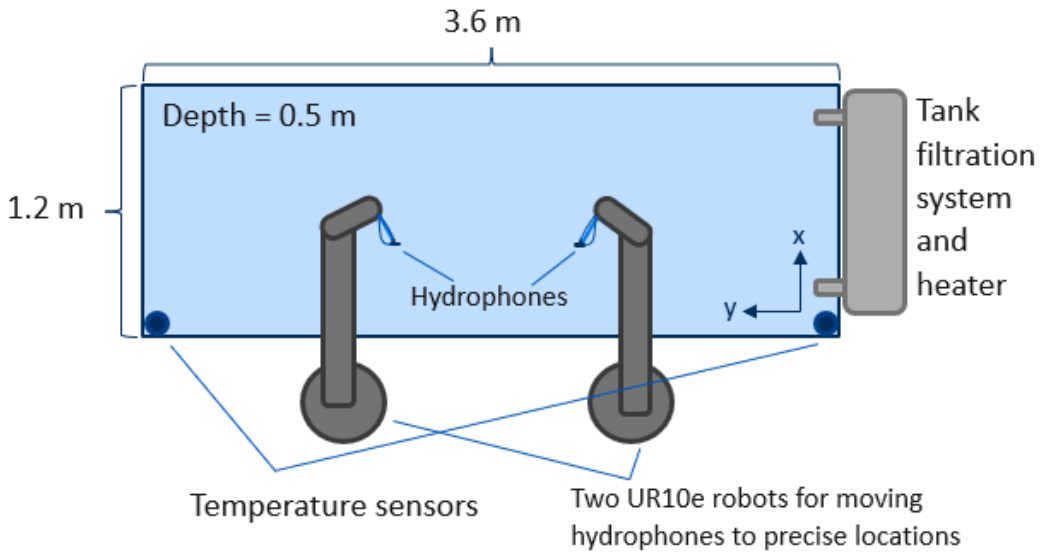

Figure 2.1 Laboratory tank schematic from top view. BYU Laboratory acrylic tank with two UR10e robots, tank filtration system and heater, and labeled dimensions in x, y, and z. The left robot is the source and the right robot is the receiver. The receiver robot is on a track that extends its reach in the y direction. This tank is used as a controlled environment to take data to use for deep learning.

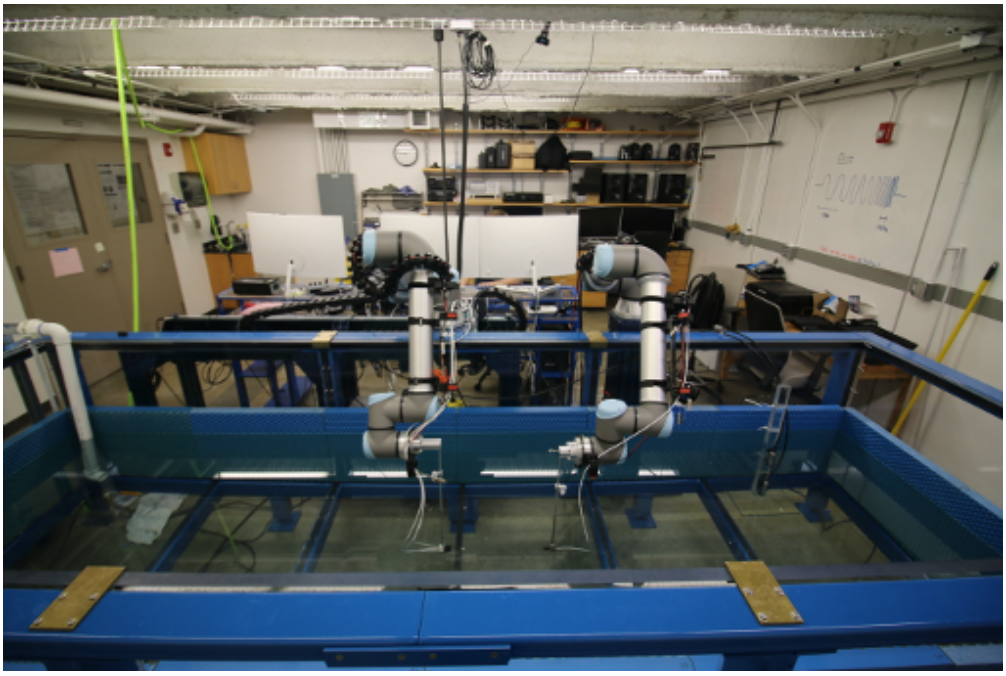

Figure 2.2 Laboratory tank setup. BYU Laboratory tank with two UR10e robots and hydrophones in the water at the end of robot attachments. Blue panels are anechoic panels to reduce reflections. This is the side view of the tank.

2.1.1 Measurement Configuration

The laboratory features an acrylic tank with dimensions of 1.2 meters (x-direction), 3.6 meters (y-direction), and 0.9 meters (z-direction, that is, depth). The depth of the water in the tank is maintained at 0.5 meters, unless intentionally varied. Figure 2.1 represents a schematic of the laboratory tank that includes the robot positioning system and the position of the axes. Acrylic was chosen for the tank because its acoustic impedance is closer to that of water compared to other materials. The tank walls are lined with polyurethane anechoic panels (Aptile SF5048) by Precision Acoustics, designed to reduce sound reflections. Figure 2.2 is an image of the laboratory tank showing both the robots and the anechoic paneling.

The tank is equipped with both a filtration and a heating system. The filtration system utilizes a particle filter and UV treatment to maintain water cleanliness. The heating system is capable

of adjusting the water temperature between 19 and 38°C. In addition, the tank undergoes regular chemical treatments to stabilize the levels of pH and alkalinity. These chemical treatments prevent algae from growing in filtration pipes and tank water, ensuring consistent acoustic measurements.

Two UR10e Universal Robotics robots are positioned just outside the tank, extending over the water. Both robots are capable of moving along six axes, allowing the attached hydrophones to be positioned within a range of 1.3 meters. The receiver robot is mounted on a Vention track, adding a seventh axis that extends its movement by 1.4 meters in the y-direction. The robots are highly precise, with a repeatability of 0.01 mm. In order to make this laboratory setup comparable to a real ocean environment, the wavelength size needs to be scaled down to match the smaller dimensions of the tank. A smaller wavelength is achieved by using signals with higher frequencies. Further details on the laboratory setup can be found in Cameron Vongsawad's master's research at BYU [14].

2.1.2 Data Acquisition

A key aspect of the experimental setup is the measurement chain, which describes the process by which a signal is generated by the computer, transmitted as a physical signal through the laboratory tank, and received back as an electrical signal. Figure 2.3 depicts the measurement chain process used in the BYU underwater acoustics lab. The measurement chain relies on custom LabView software called Easy Spectrum Acoustics Underwater (ESAU), developed by Adam Kingsley for use in the laboratory. The signal generated by ESAU is sent to the Spectrum M2p.65xx-x4 DAQ system. The Spectrum DAQ card then converts the digital signal to analog. This analog signal is amplified by a TEGAM 2350 amplifier, which increases the amplitude of the signal by a factor of fifty. After passing through an impedance matching transformer, the signal is transmitted from the source B&K8103 (Brüel and Kjær) hydrophone to the receiver B&K8103 hydrophone.

For these measurements, the source is produced by a single B&K 8103 attached to the robot to the left of Figure 2.1. The sound is received by a small horizontal array of three B&K 8103

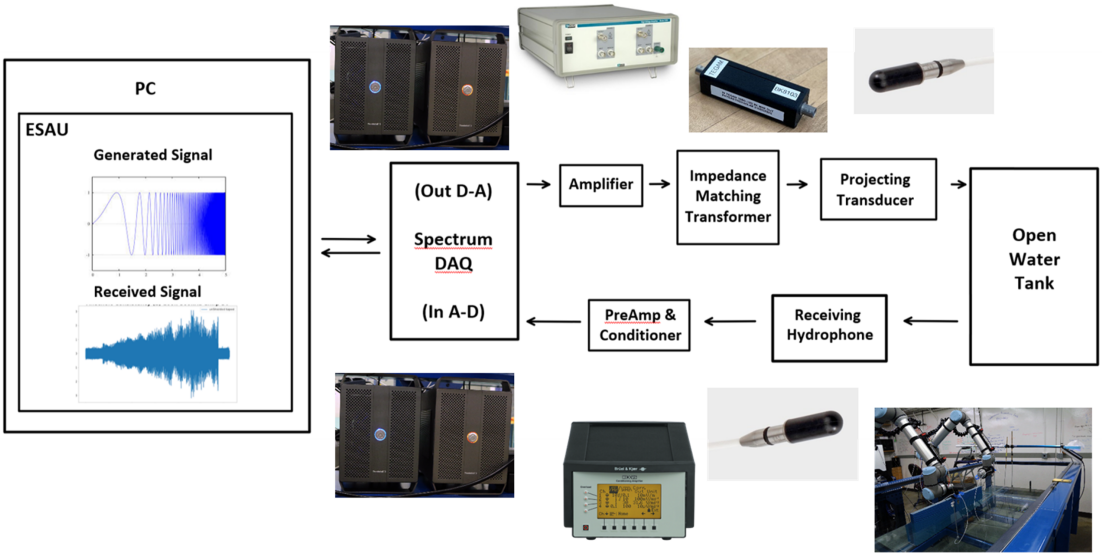

Figure 2.3 Measurement chain. Measurement chain for BYU laboratory experimental measurements using custom LabView software, ESAU (Easy Spectrum Acoustics Underwater). The process begins on the left and follows along the top row. The computer generated signal is amplified and sent to the source B&K8103 in the tank which is then received by another B&K8103. This received signal is then converted with the preamplifier and signal conditioner and read in by the computer.

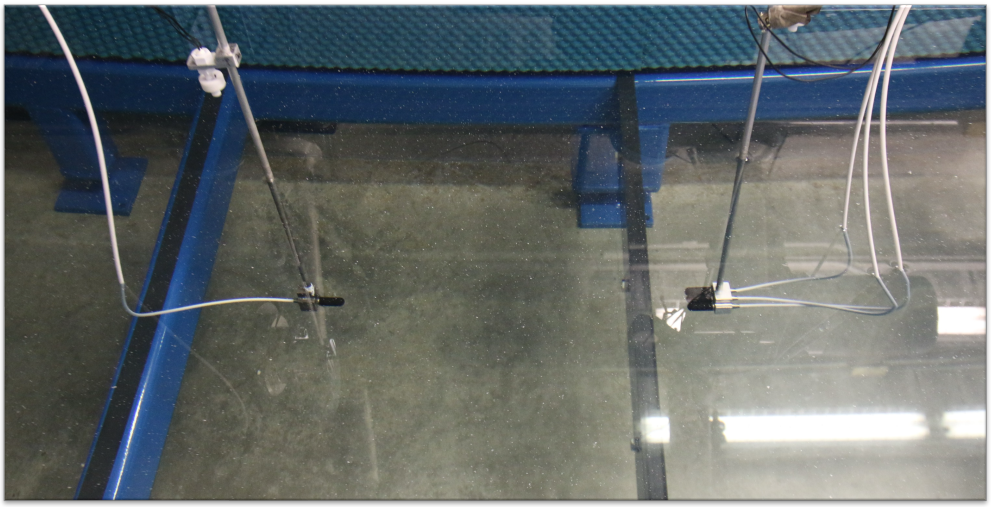

Figure 2.4 Hydrophone attachments in the tank. The left attachment is the source and the right attachment is the receiver. When using three channels, all three receiving hydrophones, spaced 1 cm apart, are used. When using only one channel, only the middle receiving hydrophone is used.

hydrophones; the small array is oriented along the x direction in Figure 2.1. The receiver robot holds three receiving hydrophones, each 1 cm apart. Figure 2.4 shows an image of the source and receiver hydrophone attachments. The source hydrophone is on the left with only one channel (one hydrophone). The receiver is on the right with three channels (three hydrophones). When using one receiver channel, only the middle hydrophone is used. The received signals are sent to the B&K NEXUS Conditioning Amplifier, where the charge sensitivity (specific to the hydrophone model) and the transducer sensitivity (10 mV/Pa) are manually set. The final step in the measurement chain is when the signal is sent through the input Spectrum DAQ card and converted back from analog to digital. ESAU records the data as bin files due to this conversion.

The frequency band of interest for this work is 50 to 100 kHz. As mentioned earlier, higher frequencies are used to compensate for the smaller physical dimensions of the tank to create smaller wavelengths. This ensures that the acoustic behavior, such as reflection, diffraction, and attenuation, scales appropriately, allowing the laboratory setup to mimic the spatial and temporal characteristics

of sound propagation in the real ocean. This frequency range is much higher than the audible hearing range, 20 to 20,000 Hz. ESAU (Easy Spectrum Acoustics Underwater) was used to generate a 1-second-long linear chirp across this frequency band. The signal was transmitted from the source B&K8103 in the tank, with 20 ms of padding (zeros) at each end. The signal was 1 second long, so the entire recorded signal was 1.04 seconds with leading and trailing zeros. The sampling frequency was 1 MHz and the input amplitude was 3600 mV. The source transducer was at [x=0.6, y=2.2, z=0.25], and the center receiving transducer was at [x=0.6, y=2.05-0.55 (with 301 or 51 points), z=0.25], unless otherwise varied.

Measurements were taken on different days to account for ambient noise in the tank. Figure 2.5 illustrates the amplitude of ambient noise in the tank over time on three different days to highlight differences. The plot on the left is the waveform that plots the signal amplitude as a function of time. The waveform plot only shows the first 20 ms of padding before the signal was sent. The plot on the right shows the power spectral density as a function of frequency. This plot shows the strength of the signal at each frequency. Each color represents a different day. Data were taken on different days so that the model could be trained with data from multiple days at once. Combining data across different days creates a level of inherent variability in the data, allowing the model to learn to account for variability during training.

The measurement data were processed using signal processing techniques, including the comparison of narrowband spectra for each source-receiver range. A fast Fourier transform was applied to the signal. The inputs for the deep learning algorithms were the band-passed power spectral levels from 50-100 kHz, which were the frequencies in the source signal.

2.1.3 Standard Measurements

Measurements were collected and classified into predetermined datasets as training or testing. Figure 2.6 shows the waveforms and spectral levels of the source (a and c) and the received (b

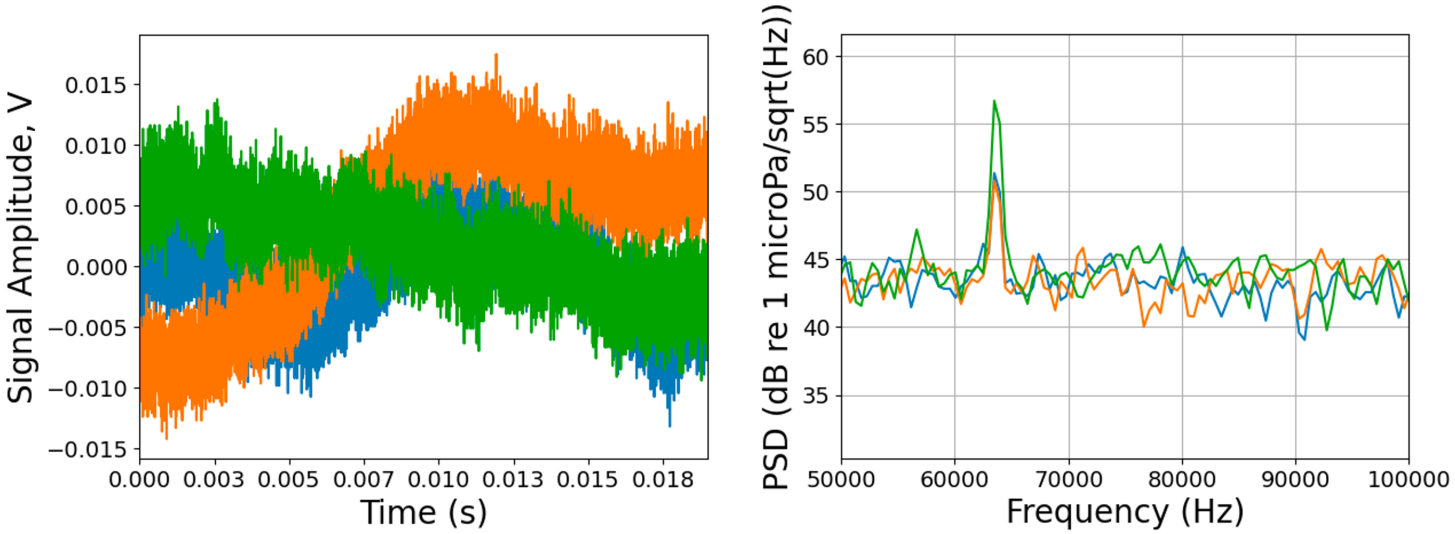

Figure 2.5 Ambient noise in laboratory tank on three different days. Each color represents a different day. On the left is the waveform, a function of time. The right is the time-averaged power spectral density levels, a function of frequency. Power spectral density levels are a measure of the strength of the signal at each frequency. The variation across different days helps the model learn to account for variability. The model can better account for variability because training data includes data from different days with varying ambient noise levels.

and d) signals. In plot a, the wave is so condensed that it looks more like a box than a traditional wave. The source and receiver hydrophones are 0.155 m apart in the y direction, which is the closest source-receiver range. Figure 2.7 shows the time-averaged power spectral density levels of the received signals at two different source-receiver ranges. The plot on the left with higher PSD values is at a source-receiver range of 0.155 m (the same source-receiver range as Figure 2.6). The right plot is at a source-receiver range of 1.65 m, meaning the source and receiver hydrophones are farther

apart, resulting in lower PSD values. This figure shows how the received signal changes as the source-receiver range changes.

The training datasets consist of 301 points, each representing a distinct source-receiver range. The water depth was maintained at 0.5 meters in the z-direction, and the average temperature of the tank was approximately 20 °C, corresponding to room temperature. For each of the specified source-receiver ranges (spaced 5 mm apart), the recorded sounds are saved. To improve the accuracy of the model, some measurements included a ±0.02 m offset in the receiver's y-position.

Models were tested using both one and three channel data. The number of channels refers to the number of receiving hydrophones being used in training. The receiver robot has an attachment with three spots, each 1 cm apart. When training a model with one channel, only data from the center hydrophone were used. A three channel setup is most often used for source localization because it provides opportunity for other signal processing techniques, such as beam forming, to determine the location. In this project, I compared the use of one receiving hydrophone with three receiving hydrophones to see if there was a significant difference. The spectral levels for these measurements, labeled with their respective source-receiver ranges, were used to train a deep learning model.

Table 2.1 Table describing number of variation measurements. The variation measurements used to test the robustness of the deep learning model and potentially apply transfer learning are receiver x position, receiver z position, water height, and temperature. The control values are the values used in the standard measurements. The offset values are the variations in measurements from the control value. Transfer learning was only performed with temperature variation measurements.

Variable	**Control**	**Offset**
Receiver x	0.6 m	± 1 mm, ± 2 mm, ± 5 mm, ± 1 cm, ± 2 cm
Receiver z	0.25 m	± 1 mm, ± 2 mm, ± 5 mm, ± 1 cm, ± 2 cm
Water Height	0.5 m	+ 2 cm, + 4 cm, + 5 cm
Temperature	20 C	+ 1 C, + 2 C, + 4 C

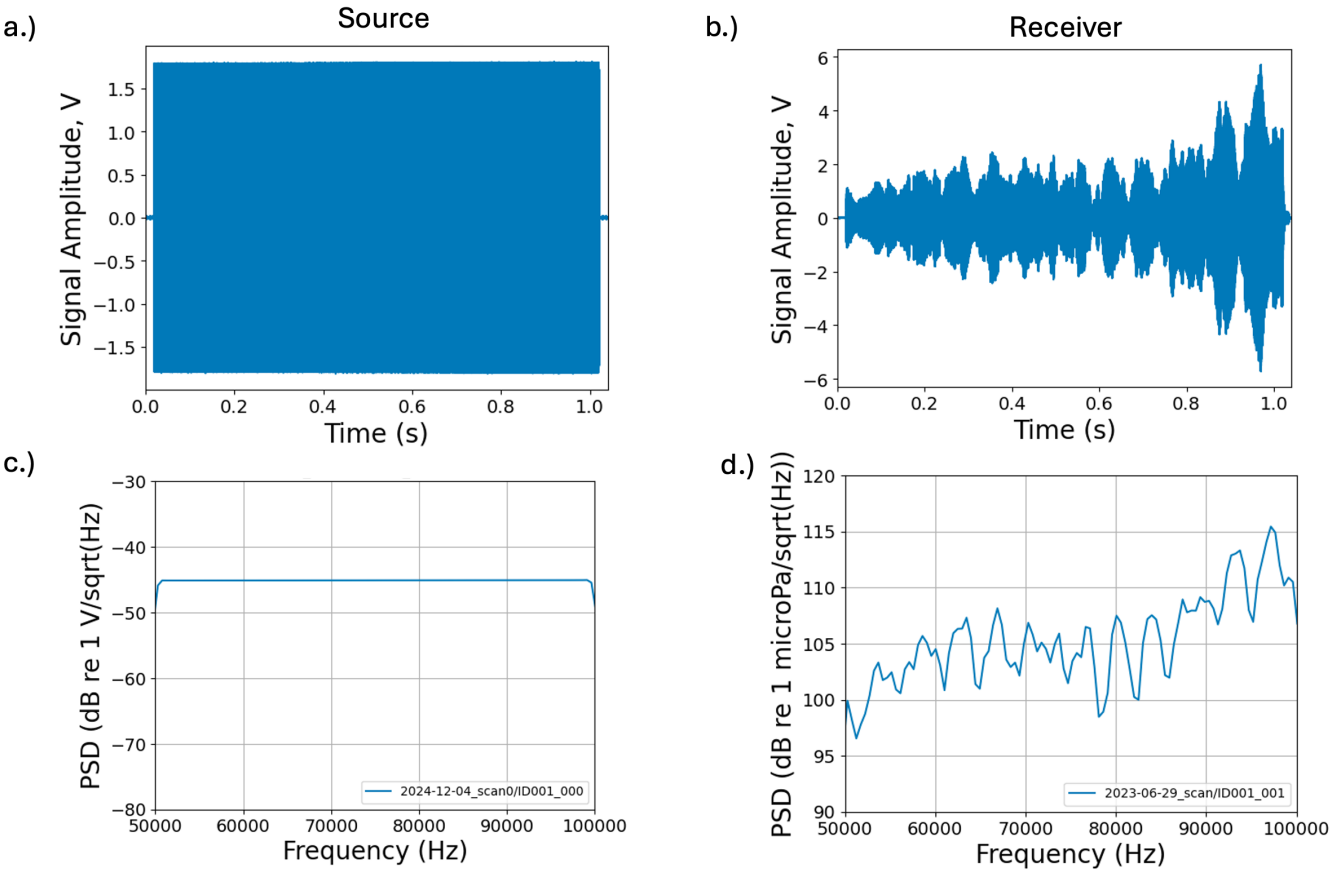

Figure 2.6 Source versus receiver. Plots a. and c. are the monitor signal, the source signal sent to the Tegam power amplifier but multiplied by 12 due instead of 50. Plots b. and d. are the received signal for one hydrophone. The source and receiver hydrophones are 0.155 m apart in the y direction. Plots c. and d. are the time-averaged power spectral density levels. Plots a. and b. are the waveforms. In plot a, the wave is so condensed that it looks more like a box than a traditional wave. Plot c. shows a linear chirp because the horizontal line means that each frequency has equal signal strength. The received signal is a modified version of the source signal showing the affects of the sound traveling through the water in the laboratory tank.

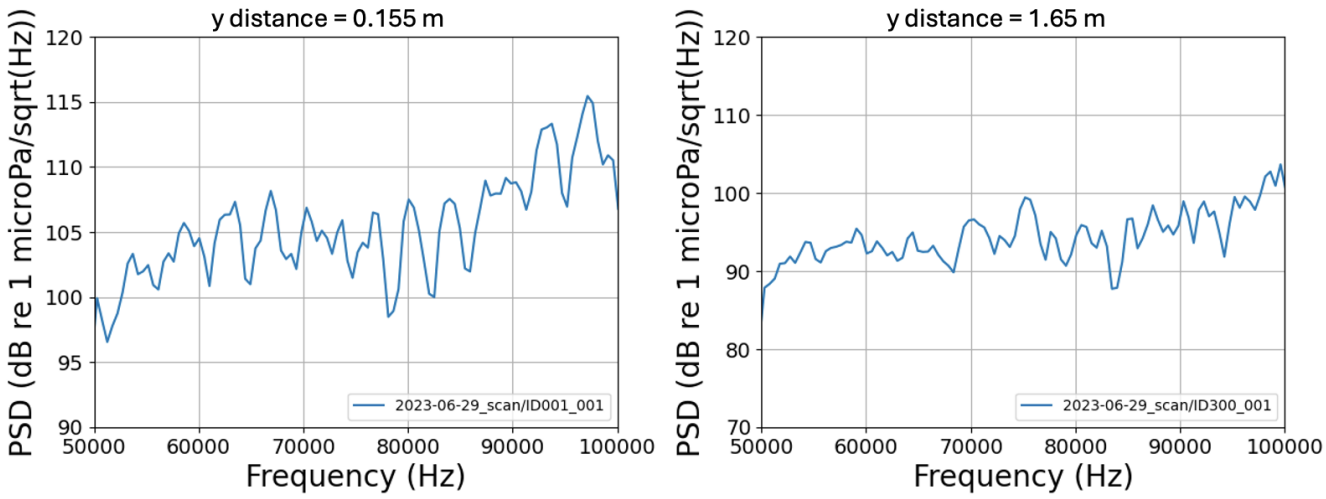

Figure 2.7 Two power spectral density (PSD) plots at varying source-receiver ranges. The left plot is at a source-receiver range of 0.155 m in the y direction. The right plot is at a source-receiver range of 1.65 m in the y direction. The closer source-receiver range, the left plot, has higher PSD values than the right plot because the two hydrophones are closer to each other. This change in PSD shows how the received signal changes as the source-receiver range changes.

2.1.4 Variation Measurements

To design more accurate source-localization deep learning models that account for ocean variability, intentional variations are required. These measurements serve to test deep learning models and facilitate the application of transfer learning. The datasets used for testing and transfer learning consist of 51 data points. Four variables were intentionally varied in this project: the x-position, z-position, water depth, and temperature. The x-position of the receiver was adjusted from standard measurements by increments of 1 mm, 2 mm, 5 mm, 1 cm, and 2 cm. Similarly, the z-position of the receiver was varied by the same five increments. Measurements were also taken with the water depth varied by 0.05 m, 0.04 m, and 0.02 m. Furthermore, measurements were performed when the water was heated with the heater to 1, 2, and 4 degrees above the 20 °C water that corresponds to room temperature. Table 2.1 presents a concise representation of the variation data for all four variables. The table outlines the control and offset values for each variable. The control values are the values used for standard measurements and the offset values create the data used for variation measurements.

2.2 Machine Learning

After the measurements are taken, the data is transformed using the FFT to obtain spectral levels and then undergoes processing through bandpass filtering. The results are then used in the deep learning algorithm as the input. Several parameters have been adjusted to optimize the performance of the model.

2.2.1 Network Architecture

The deep learning algorithm used in this thesis employs a one dimensional four-layer CNN, designed as a regressive model rather than a classification model [15]. A regressive model predicts continuous

numerical values, instead of categorizing results into specific classes as with a classification model. Figure 2.8 from Corey Dobb's thesis [13] displays the sizes of both the convolution and fully connected layers. As shown in the far right of the figure, MSE is the loss function for this network because I am using a regression model. The initial learning rate is set to $1e{-}3$, with adjustments made during training using a cosine scheduler. The learning rate is the value that controls the step size in adjusting the weights during training. The scheduler allows the learning rate to be dynamically adjusted.

The batch size determines how many training samples are processed at once, and for this

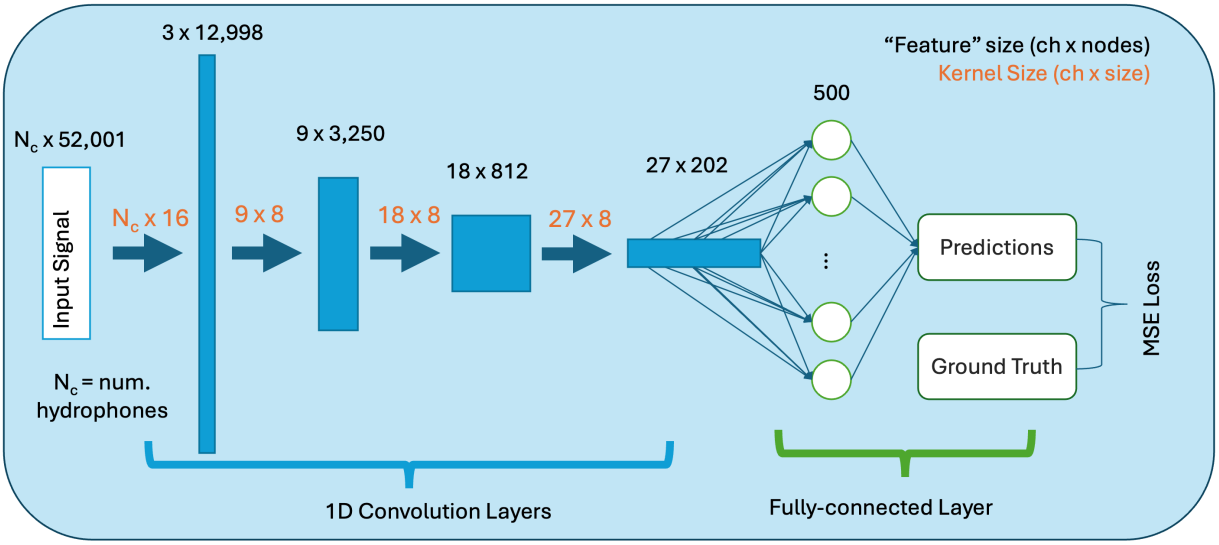

Figure 2.8 Network architecture. A schematic of the 1D network architecture with four convolutional layers and two fully connected layers. The kernel is a matrix that slides over the input data and calculates weighted sums. The kernel and feature sizes of each layer are expressed. The kernel size is represented by the number of filters x filter length. The feature size is represented by the number of channels x feature length. The loss is determined at the end by comparing the predicted source-receiver range to the known source-receiver range for the data.

network, it is set to 32. Although batch normalization is included, max pooling is not incorporated. Batch normalization makes training faster and more stable by adjusting the outputs of each layer. Max pooling reduces the size of the spatial dimensions of the CNN output while preserving the key information. A 5-fold cross-validation is used, with an 80/20 training-testing split. An 80/20 training-testing split means that 80% of the data is used for training, while the remaining 20% is reserved solely for validation. This 20% of the data has not been seen by the model but is randomly selected from the data used for the training. Each of the five folds uses different 20% of the training data as the data for the validation testing. The Adam optimizer, with an epsilon value of 0.5, is also used. The primary function of the optimizer is to conduct the backpropagation step when the weights are updated.

2.2.2 Epochs Versus Early Stopping

Neural networks are trained for multiple epochs. In one epoch, all training data are drawn in batches and passed through the network. There are different ways to decide how many epochs are completed. One method is to specify the number of epochs to be completed, while another is to employ early stopping. Early stopping is a technique used to improve model generalization and reduce overfitting. The loss examined during early stopping is the root mean squared error (RMSE). RMSE is given by:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2.1}$$

where N is the number of samples, y_i is the true value of the ith data point, and $\hat{y}_i$ is the predicted value for the same data point. Early stopping occurs when the validation RMSE decreased by less than a specified threshold over a specified number of epochs, referred to as the stopping patience. For this work, the threshold is 0.001 and the stopping patience is 5 epochs. The model will then save the version that corresponds to the lowest RMSE observed during training, before the five epochs without improvement. The version that initiates early stopping corresponds to the point

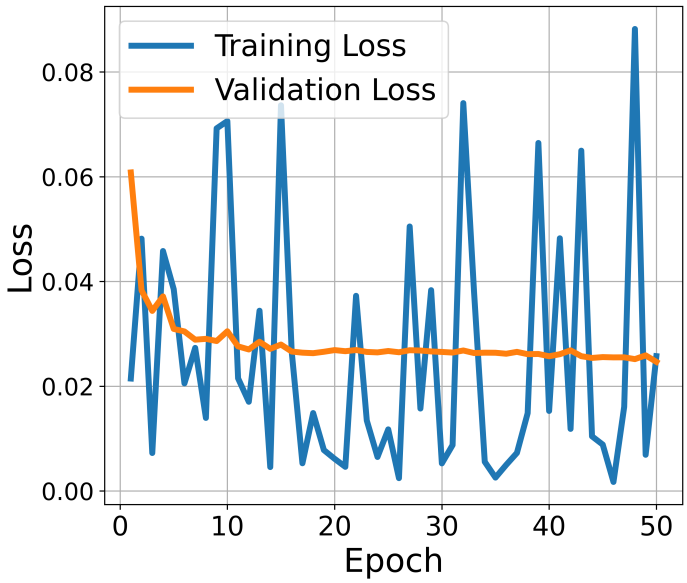

Figure 2.9 Loss curve for base model. The loss is the difference between the prediction from the algorithm and the true value. The number of epochs corresponds to the amount of times the entire training dataset is used to train the model, one complete run. The training loss is the loss acquired after each epoch for the training data while the validation loss is the loss acquired after each epoch for the validation data. The validation loss levels out to a horizontal line at about epoch number 15. The y axis scale is very small. The loss curve shows that optimal number of epochs is about 15.

where the loss curve is at its minimum.

Early stopping is employed to avoid overfitting and promote better generalization. Overfitting occurs when a model learns the specific details of the training data too well, failing to generalize to new, unseen situations. Overfitting results in good validation performance but poor testing performance on new data. Allowing the model to run for too many epochs can exacerbate overfitting. For this project, the maximum number of epochs is set to 15. Figure 2.9 shows the loss curve of a model trained with a maximum of 50 epochs. The training loss is the loss acquired after each epoch for the training data, while the validation loss is the loss acquired after each epoch for the validation data. The validation loss line flattens out starting at about 15 epochs, which is why that value was chosen.

2.2.3 Training Dataset

The machine learning parameters and the corresponding decisions are described in detail in Corey Dobbs' thesis [13]. In his work, the ideal number of samples for the training dataset is discussed. The challenge is to find a balance between having too many samples and not enough. Too few samples result in a failure to capture the variation in normal room temperate data and the time-varying ambient noise, while too many samples can lead to excessive repetition and extended training times. His study concluded that training one model should use data collected over six days. Each model is trained on data for multiple days due to the varying ambient noise in the tank shown in Figure 2.5. Training on data from different days provides inherent variability in the training data, which increases the generalizability of the model for variation data. With 301 samples in each of those days, the training dataset, one training model incorporates 1,806 different samples.

2.2.4 Evaluating Performance

The label learned with these models is the source-receiver range. The metrics include both the root mean squared error (RMSE) and the mean absolute percent error (MAPE). The preferred method used to evaluate the performance of a model is MAPE. MAPE is emphasized over RMSE because it accounts for the relative distance to the object. Because MAPE accounts for relative locations, it allows us to make a better connection to what would happen if source localization were applied in an ocean scenario. For example, if the receiver is 10 km away from the source, an error in the range of 500 m is relatively small compared to the total distance, a 5% MAPE. However, a 500 m error would be considered a large error if the source-receiver range was only 1 km, resulting in a 5% MAPE. The equation for MAPE is

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{2.2}$$

where n is the number of samples, y_i is the true value of the ith data point, and $\hat{y}_i$ is the predicted value of the ith data point. In naval applications, an error of 10% in the range is considered acceptable. In reverberant spaces, such as a laboratory tank, larger errors are expected.

2.2.5 Transfer Learning

Transfer learning is applied to previously trained deep learning models with the aim of achieving better generalizability on new test data samples than is obtained using the base model. Some hyperparameters remain the same as in training, while others are modified. The number of runs, learning rate, scheduler, and batch size are unchanged. Both models incorporate batch normalization but exclude max pooling. For transfer learning, the epsilon value for the Adam optimizer is set to 0.01. Early stopping is disabled in transfer learning. During transfer learning, only the parameters (or weights) of the very last layer of the original model are updated based on the 51-point transfer dataset. Modification of the base model enables transfer learning to adapt the model to account for the introduced variations.

Chapter 3

Results

This chapter presents the results of testing on the base deep learning model, as well as the outcomes from applying transfer learning to the model for predicting source-receiver ranges. The comparison of the MAPE between these two approaches is made for temperature variations.

3.1 Validation Testing Results

The training data are divided into training and validation sets with an 80/20 split. The network does not use the validation data for training, but rather for testing. The results of the validation testing are evaluated as the model is trained as a way to monitor the progress of the training. Consequently, the validation testing data share similar environmental factors, such as ambient noise, temperature, and position, since they are drawn from the same measurements taken on the same day. The five-fold cross validation used in training results in five trained models. The average validation MAPE is the average validation MAPE across those five trained models, whereas the best MAPE is the lowest MAPE of those five.

Figure 3.1 shows the predicted labels versus the actual labels of the validation and generalized testing of the base model for one of the five folds. The labels are the source-receiver ranges. The

predicted labels are the source-receiver ranges estimated by the deep learning model. The actual labels are the true source-receiver ranges of the data. The plot on the left showing the validation testing results contains 361 samples (1806 divided by 5). The red dashed line is the $y = x$ line, where the predicted and actual labels are the same. A result on the red dashed line means that the model is correctly producing the source-receiver range. As mentioned above, the US Navy classifies a MAPE of 10% as good, although larger values are expected in reverberant laboratory tank environments.

The average MAPE results for validation testing for three channels, with the maximum number of epochs being set to 15 (which was found to be ideal) and the updated early stopping criteria, ranged from 11.66% to 14.43%. The best MAPE results for these models ranged from 10.07% to 12.03%. These average MAPE values represent an improvement over earlier validation results obtained with a previous version of early stopping from Corey Dobbs thesis [13], which produced values between 14.06% and 15.57%. When the maximum number of epochs was increased to 1000, the validation testing results improved further, with MAPE values ranging from 4.29% to 4.82%. However, generalized testing did not show a decrease in MAPE, suggesting that the model may have suffered from overfitting.

3.2 Generalized Testing Results

Generalized testing results are different from validation testing results. Generalized testing results come from testing a trained model on a new dataset that the model has never seen before. These datasets for our experiments contain 51 samples, as opposed to the 1806 samples used for training and the 361 samples used for validation testing. The generalized testing datasets were taken on different days than the training datasets. Data taken on different days always exhibited slight variability due to ambient noise in the tank, which varied from day to day. As mentioned above,

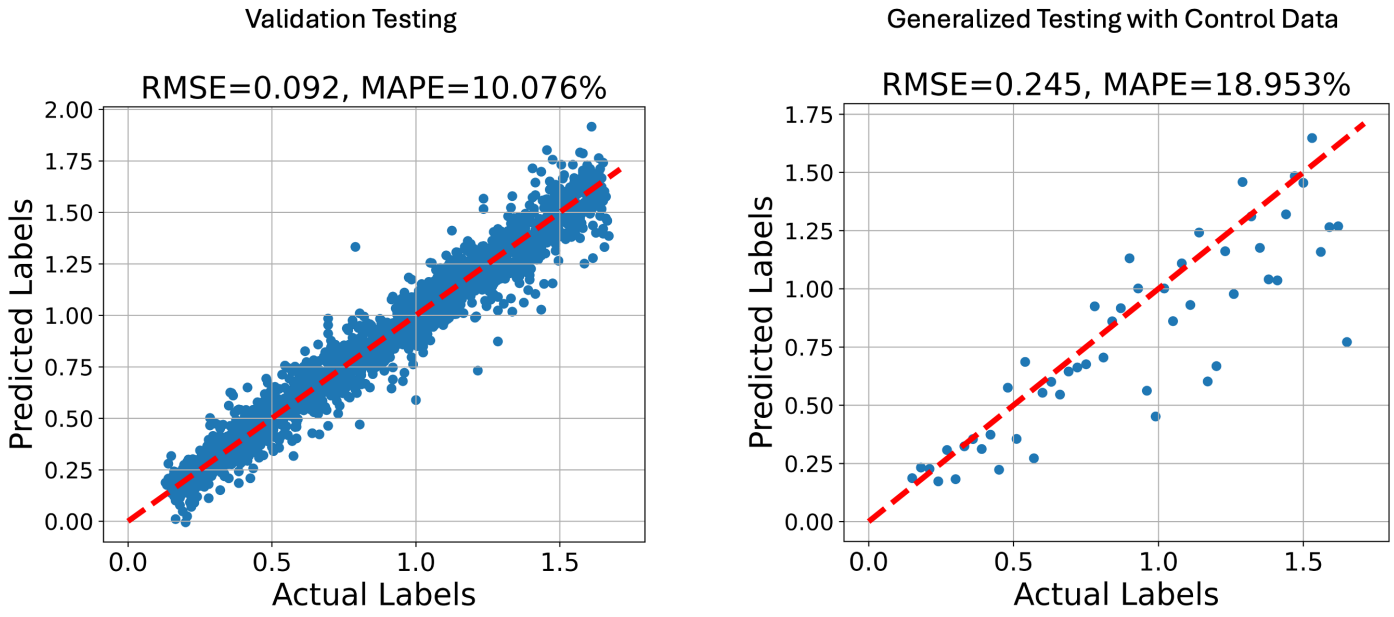

Figure 3.1 Predicted Versus Actual for Base Model. The left plot shows predicted labels versus actual labels for the validation testing results on the base model with 1806 samples. The labels are the source-receiver ranges. The predicted labels are the source-receiver ranges estimated by the deep learning model. The actual labels are the true source-receiver ranges of the data. The right plot shows the results with the same model for generalized testing on a control data sample. The validation testing has 361 points (1806 divided by 5 due to five-fold cross-validation) while the generalized testing only has 51 points. The validation testing has lower MAPE and RMSE values than the generalization plot. These are the results for one out of the five folds of training for standard measurements.

Figure 3.1 shows the predicted labels versus the actual labels of the validation and generalized testing of the base model with the red line being where the model is correct. The right plot showing generalized testing results contains 51 samples. These plots are from one of the five folds of training for the base model. The generalized testing data used for this example plot were standard data taken on a different day than the trained model. Because the data used for generalized testing

Table 3.1 Base model 1 testing MAPE and RMSE. The first two rows show validation MAPE and RMSE results on base model 1 with 1806 samples that include the generalized testing day. The next two rows show the results for generalized testing on room temperature data. Models trained on one channel (one receiving hydrophone) and three channels (three receiving hydrophones) of data are shown. There is never more than a 3% difference in MAPE and a 0.01 difference in RMSE between 1 and 3 channels. The generalized testing values are about 5% less than the generalized testing on base model 2. The average values come from the results of all five folds, the best values are the lowest values from one of the five folds.

Model (Base Model 1)	Average MAPE	Best MAPE	Average RMSE	Best RMSE
1 Ch Val	13.89	13.09	0.12	0.11
3 Ch Val	12.53	11.38	0.11	0.1001
1 Ch Room Temp Test	15.94	10.84	0.13	0.103
3 Ch Room Temp Test	15.2	13.56	0.13	0.107

were standard data, the only variation in the data is ambient noise and the number of samples. The generalized testing MAPE value was higher than the validation testing MAPE value. Higher generalized testing MAPE is what I would expect because the validation testing data is part of the training dataset which means the model has already seen data from the same day with the same ambient noise levels.

Generalized testing on room temperature data was performed on two different models. The first model included data from the same day that the 51 point generalized data was taken, this will be referred to as the base model 1. Because data from the same day was seen during training, the model had already accounted for the specific ambient noise in the tank on that day. Base model 1 did not include data with the 0.02m y offset and included data from days 1, 2, 3, 4, and 5. Table 3.1 compares the validation and generalized testing values for models trained, without variation, on one channel and three channels for base model 1. The generalized testing average MAPE results were around 15%. The second model, base model 2, did not include data from the same day as

Table 3.2 Base model 2 testing MAPE and RMSE. The first two rows show validation MAPE and RMSE results on base model 2 with 1806 samples that include the generalized testing day. The next two rows show the results for generalized testing on room temperature data. Models trained on one channel (one receiving hydrophone) and three channels (three receiving hydrophones) of data are shown. There is never more than a 3% difference in MAPE and a 0.01 difference in RMSE between 1 and 3 channels. The generalized testing values are about 5% greater than the generalized testing on base model 1. The average values come from the results of all five folds, the best values are the lowest values from one of the five folds.

Model (Base Model 2)	Average MAPE	Best MAPE	Average RMSE	Best RMSE
1 Ch Val	13.38	12.09	0.13	0.11
3 Ch Val	11.73	10.07	0.1007	0.09
1 Ch Room Temp Test	20.63	16.87	0.202	0.18
3 Ch Room Temp Test	17.91	13.08	0.19	0.13

the generalization data. Base model 2 included data from days 2, 5, 6, and 7 while the generalized testing dataset is from day 1. Because the generalized testing dataset is from a different day than all the training data, Table 3.2 shows the average MAPE results closer to 19%. Base model 2 generalized testing results were about 4% higher than base model 1 generalized testing results. Base model 2 is a more realistic model because it includes the training data with a 0.02m offset in the receiver y position and variable ambient noise. The results from base model 2 are more representative of a source localization model that accounts for real ocean variability.

For both base model 1 and base model 2, shown in Table 3.1 and Table 3.2, there is never more than a 3% variation in MAPE values. This 3% MAPE variation illustrates the minimal difference in results using one channel versus three channels. Three channel data are often used in source localization because it provides opportunity for beam forming in signal processing. One channel measurements require less equipment because only one receiver hydrophone is needed. Another surprising result from the comparison between one and three hydrophones is that the fold

Table 3.3 Generalized variation results. The left column represents the amount of offset from the control value in the measurements. The reported values for the three variables are the differences between generalized testing on the variation data with that specific offset and generalized testing on standard data. Testing was done on base model 2 with three channels. There is not a significant shift in average MAPE values between both types of generalized testing, hence why transfer learning was not applied with these variables. Temperature had a more significant difference in MAPE values and therefore led to transfer learning.

Offset	**Receiver x**	**Receiver z**	**Water height**
1 mm	3.36	4.27	—-
2 mm	2.98	4.41	—-
5 mm	7.39	2.23	—-
1 cm	3.65	5.03	—-
2 cm	1.05	1.31	2.12
4 cm	—-	—-	-1.1
5 cm	—-	—-	0.42

that produced the best validation value did not always produce the best generalization value. The discrepancy suggests that validation performance alone may not reliably indicate generalization ability, highlighting the need for robust evaluation across multiple data splits.

As mentioned earlier, there were four categories of variation testing data: change in receiver x position, change in receiver z position, change in water height, and change in water temperature. Table 3.3 shows the generalized testing results for x receiver position, z receiver position, and water height. The values recorded in the table are the differences between generalized testing on the variation data and generalized testing on standard data that has the same configuration as the original training data. Base model 2 with three channels was used for Table 3.3 so the testing values were evaluated against a control generalized testing value of 17.91%. The change in x position led to an increase 1.09% to 7.39% in MAPE from generalized testing at $x = 0.6$ m. The z position

testing follows a similar pattern. There was a 1.31% to 5.03% increase in MAPE for changing the z position from z = 0.25 m. For shifts in water height (WH), the generalized testing MAPEs from three different heights were off by 0.42% to 2.12% from generalized testing at WH = 0.5 m. The difference in generalized testing MAPE values for the temperature measurements on one channel and three channels spanned from 5.27% to 12.17%. The average MAPE increase for three channels was 3.69% for x, 3.45% for z, 1.21% for WH, and 9.09% for temperature. Transfer learning was not applied to the first three variables because the average MAPE differences were not as large as the average MAPE difference for temperature variation.

3.3 Transfer Learning Results

The goal of transfer learning is to provide a deep learning technique for source localization that can account for ocean variability better than generalized testing while still using a smaller set of labeled measurements. Transfer learning uses a smaller set of variation data to modify the last layer of weights on an existing model. The existing model to which transfer learning was applied was base model 2. Transfer learning was performed on the change in temperature values for 22 °C and 24 °C. Transfer learning was performed on temperature measurements because they performed the most poorly out of all four variables in generalized testing. Since transfer learning modifies an existing model, it provides both validation and generalized testing results because it creates a new model. The generalized testing results come from testing the new transfer learning model on another set of data measured at the higher temperature. The generalized testing data are temperature data that were not used to train the model. Figure 3.2 and Figure 3.3 show an improvement in generalized testing results with transfer learning by comparing predicted and actual labels for 22 °C and 24 °C respectively. A 15% improvement is demonstrated for one fold at 22 °C with three channels, and a 45% improvement is presented for one fold at 24 °C with one channel. For 22 °C, the post transfer

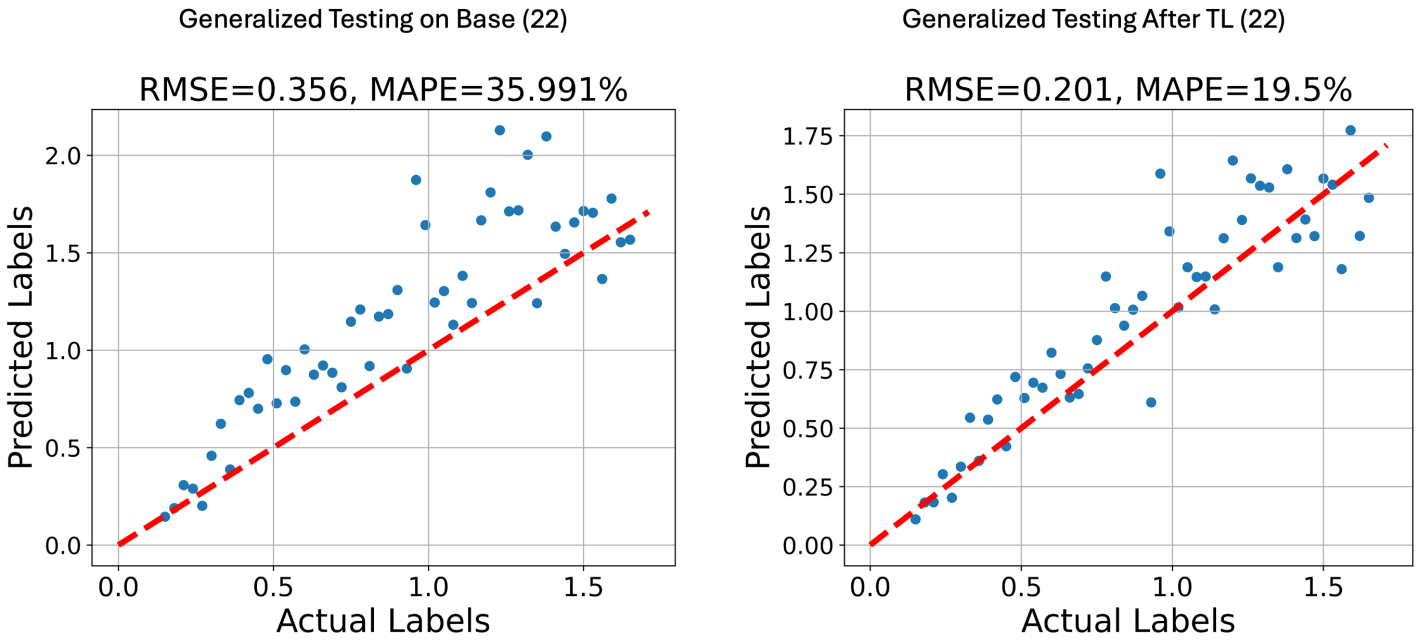

Figure 3.2 22 degrees generalized testing results. The left plot shows generalization results for one fold of base model 2 with three channels before transfer learning. The right plot shows the generalization results after transfer learning with a 22 degrees Celsius dataset. After transfer learning, the MAPE and RMSE values are significantly lower and the data points are closer to the red line. For this specific fold, there is about a 15% improvement in the MAPE after transfer learning.

learning generalization results were within 1.59% of the 17.91% control generalization results on base model 2 without temperature variation. For 24 °C, the post transfer learning generalization results were within 2.56% of the 20.63% control generalization results on base model 2 without temperature variation.

Table 3.4 presents the results at 22 °C for both one channel and three channels. The first two columns show the generalized testing results of temperature data on base model 2. The next two columns show the validation testing results of the transfer learning model. The transfer learning

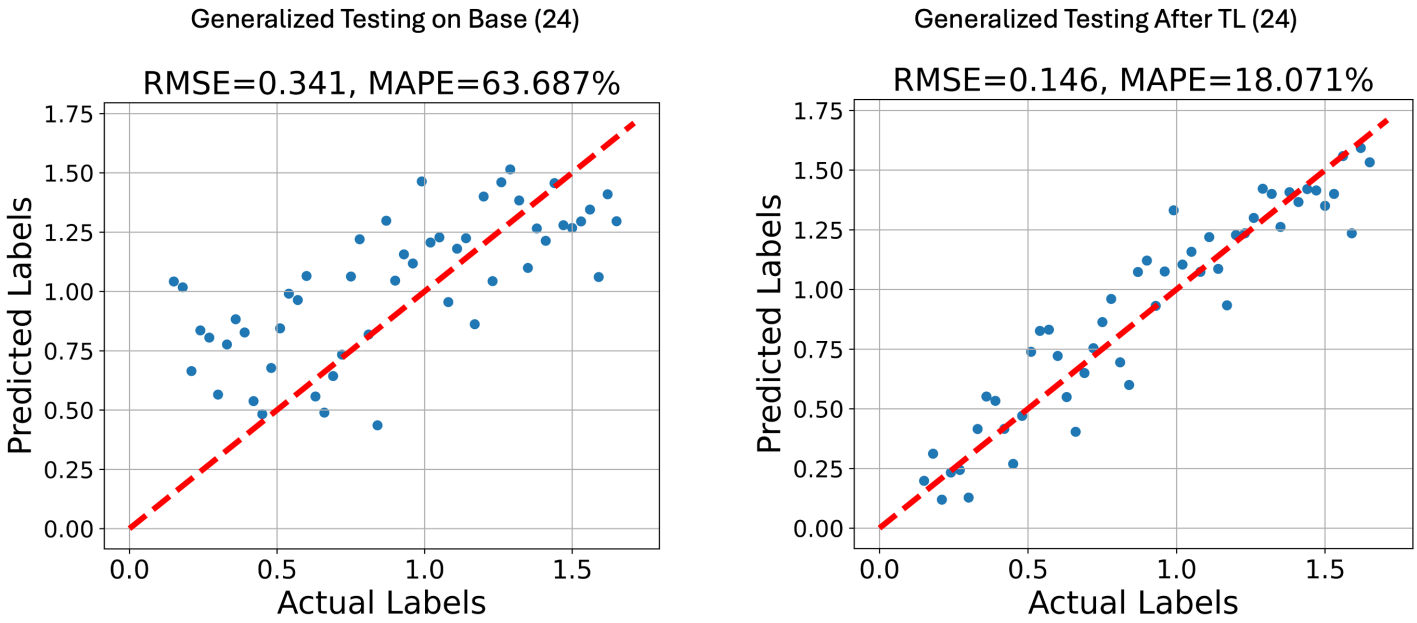

Figure 3.3 24 degrees generalized testing results. The left plot shows generalization results for one fold of base model 2 with one channel before transfer learning. The right plot shows the generalization results after transfer learning with a 24 degrees Celsius dataset. After transfer learning, the MAPE and RMSE values are significantly lower and the data points are closer to the red line. There is a larger improvement in MAPE values with transfer learning at 24 degrees Celsius than 22. For this fold, there is about a 45% improvement in the MAPE after transfer learning.

model is a result of applying transfer learning with 22 °C data on base model 2. The last two columns represent the results of generalized testing using new 22 °C on the 22 °C transfer learning model. Table 3.5 presents the same data but at 24 °C. The 24 °C transfer learning model is a result of applying transfer learning with 24 °C data on base model 2. The last two columns represent the results of generalized testing using new 24 °C on the 24 °C transfer learning model. In both tables, the generalized testing MAPE results after transfer learning are significantly, between about

Table 3.4 22 degree transfer learning results. MAPE validation and generalization results at 22 degrees Celsius for model trained with one channel versus three channels using base model 2. The transfer learning model is a result of applying transfer learning with 22 °C data on base model 2. One and three channel results were within at least 4% of each other. Average generalized testing MAPE results decreased by about 10% for one and three channels after transfer learning showing improvement.

Channels	Ave. Gen MAPE	Best Gen MAPE	Ave. Val MAPE After TL	Best Val MAPE After TL	Ave. Gen MAPE After TL	Best Gen MAPE After TL
1	26.56 ± 6.21	20.31	14.56 ± 2.2	12.5	19.88 ± 2.54	16.09
3	25.98 ± 6.15	18.98	12.61 ± 2.85	8.76	17.44 ± 2.34	14.55

Table 3.5 24 degree transfer learning results. MAPE validation and generalization results at 24 degrees Celsius for model trained with one channel versus three channels using base model 2. The transfer learning model is a result of applying transfer learning with 24 °C data on base model 2. One and three channel results were within at least 13% of each other. Average generalized testing MAPE results decreased by about 15% after transfer learning for three channels. Results decreased by about 30% after transfer learning for one channel.

Channels	Ave. Gen MAPE	Best Gen MAPE	Ave. Val MAPE After TL	Best Val MAPE After TL	Ave. Gen MAPE After TL	Best Gen MAPE After TL
1	43.84 ± 13.69	33.19	14.65 ± 2.37	13.05	16.92 ± 1.26	15.02
3	30.09 ± 7.97	20.61	12.81 ± 4.2	7.58	15.47 ± 3.89	9.2

10% and 30%, better than the generalized testing MAPE results before transfer learning for the models with one and three channels. The average generalized testing MAPE values have improved for the four rows after transfer learning. These values are not more than 4% off from the standard generalized testing values for three channels and one channel shown in Table 3.2, 17.91% and 20.63%.

Chapter 4

Conclusion

The purpose of this thesis was to evaluate the effectiveness of transfer learning as a deep learning method for handling ocean variability in source localization. The objective was to determine whether transfer learning could produce a more robust model than previously trained deep learning models. Transfer learning is particularly advantageous in real-world applications because it requires only a small amount of newly labeled data. The results corroborate the effectiveness of transfer learning, showing a 10 to 30% decrease in MAPE values for generalized testing after applying transfer learning.

4.1 Overview

In this thesis, machine learning models were trained on measured data collected in an acrylic tank. Temperature variations were introduced and smaller datasets were used to train transfer learning models. This study detailed the equipment, laboratory setup, and signal processing techniques used. In addition, the network architecture parameters and the performance evaluation methods were discussed.

To assess the effectiveness of transfer learning, results from the base model are compared with

results after applying transfer learning to the base model. The results indicate that the transfer learning model consistently outperformed the base model, demonstrating lower MAPE values when tested on novel data.

4.2 Discussion

The data demonstrates that transfer learning methods outperformed conventional deep learning methods for source localization in a laboratory tank with varying temperatures. These findings agree with previous research on dolphin sound recognition [11], where transfer learning also yielded superior results compared to existing models. However, while the dolphin sound recognition study employed a classification model (recording average accuracy), this thesis used a regression model (recording MAPE), making direct comparisons challenging.

Putting the MAPE values of this thesis into perspective involves comparing the results with the Wang *et al*. [9] and Huang *et al*. [8] studies mentioned in Chapter 1. Both of these studies did not apply transfer learning with variation in their data; they simply trained and tested a model for source localization using deep learning. A wavelength to water depth ratio is used to better compare the results between studies by relating scenarios. Table 4.1 details the comparisons of the three acoustic environments: this thesis, Wang *et al*., and Huang *et al*. Because both papers used a vertical line array of hydrophones, the water depth is the center water depth of the vertical line array. The table shows the setup in Wang *et al*. in terms of minimum and maximum source receiver ranges is more similar to this thesis.

The model developed in Huang *et al*. was used to predict significantly larger ranges in terms of wavelength ratio compared to this thesis. Their validation testing MAPE results were about 1% while their generalized testing MAPE results were about 5%. This model used a 30 hydrophone vertical line array, while the model in this thesis only uses 1 or 3 hydrophones, so poorer performance

for this thesis would be expected. In the data for this project without variation, validation testing MAPE results were about 13% and generalized testing MAPE results were about 19% (average MAPE values between 1 and 3 hydrophones).

The testing values and ranges relative to the wavelength ratio presented in this thesis are more closely aligned with the results in Wang *et al*. Wang *et al*. used a 21 hydrophone vertical line array. However, similar to this thesis, Wang *et al*. used spectral values as the model input. Their generalized testing MAPE values ranged between about 10% and 14% which is much closer to the 19% recorded in this thesis. This comparison shows that this thesis got close MAPE values to a similarly scaled problem using only 1 and 3 hydrophones at one depth rather than 21 hydrophones across depth.

Table 4.1 Comparison of frequency bands, water depths, and source-receiver ranges (in terms of wavelength) between this work and two reference papers. Putting values in terms of wavelength allows for more accurate comparison between situations. The setup in Wang *et al*. is more similar to this thesis for minimum and maximum source receiver ranges than Huang *et al*.

Parameter	This Thesis	Wang *et al*.	Huang *et al*.
Frequency Band (Hz)	50k-100k	100-300	50-1000
Wavelengths (m)	0.03-0.015	15-5	30-1.5
Water Depth (in λ)	16-32	6-20	1-19
Min Source-Receiver Range (in λ)	5-10	66-200	333-6666
Max Source-Receiver Range (in λ)	55-110	333-1000	95-19000

In this study, transfer learning models achieved at least a 10% improvement over their base model results for generalized testing performance on unseen data. Table 4.2 summarizes the transfer learning results for three channels while Table 4.3 summarizes the transfer learning results for one channel. These two tables demonstrate that transfer learning improved performance by lowering

Table 4.2 3 channel transfer learning results. The base model is set at 20 degrees Celsius. The decreased MAPE values between generalized base testing and generalized testing after transfer learning show an improvement in model performance after transfer learning by about 10% and 15%. The resulting MAPE values after transfer learning are similar, within 2%, to the generalized base testing MAPE value on the 20 degree data.

Water Temperature (C)	Base Val MAPE	Base Gen MAPE	After TL Gen MAPE
20	11.73 ± 1.55	17.91 ± 2.81	—-
22	—-	25.98 ± 6.15	17.44 ± 2.34
24	—-	30.09 ± 7.97	15.47 ± 3.89

Table 4.3 1 channel transfer learning results. The base model is set at 20 degrees Celsius. The decreased MAPE values between generalized base testing and generalized testing after transfer learning show an improvement in model performance after transfer learning by about 7% and 30%. The resulting MAPE values after transfer learning are similar, within 4%, to the generalized base testing MAPE value on the 20 degree data. These results are similar to transfer learning results on 3 channel data. The generalization testing results after transfer learning are within 4% of each other.

Water Temperature (C)	Base Val MAPE	Base Gen MAPE	After TL Gen MAPE
20	13.38 ± 1.92	20.63 ± 3.31	—-
22	—-	26.56 ± 6.21	19.88 ± 2.54
24	—-	43.84 ± 13.69	16.92 ± 1.26

MAPE values. The generalized testing MAPE values after transfer learning (rows 2 and 3 of column 4) were within 2% of the base generalization testing for 3 channels and within 4% for 1 channel (row one of column 3). The generalized testing MAPE values after transfer learning were within about 4% of each other for 1 and 3 channels between Table 4.2 and Table 4.3. Furthermore, modifying Corey Dobbs' model with early stopping and bandpass filtering led to a reduction of 2 to 3% in MAPE values for validation testing on the original models [13].

Only the temperature varied position data was used to train a transfer learning model. For the x and z position variation data, along with water height data, the generalized testing results for the original model differed by no more than 7% from base model 2, indicating that transfer learning was unnecessary for these variables. Given this minimal change in MAPE, it was determined that the original model was robust enough to handle these variations.

The fact that transfer learning improved generalized testing performance has important implications for real-world applications. In practice, one would first train a base model on a large labeled ocean dataset. A smaller synthetic dataset, designed to represent the specific expected ocean variability, would then be generated by a sound propagation model. Transfer learning would be applied to the base model using this specialized dataset. When the month and place of an area of operations are determined, the model after transfer learning should be prepared to yield better predictions in the area of interest. The findings of this thesis suggest that this transfer learning approach could improve the predictions of the source-receiver range under real-world ocean conditions.

Despite its promise, this transfer learning method has several limitations due to the inherent complexity of ocean environments. These limitations include spatial and temporal variations in ambient noise, water depth (bathymetry), seafloor composition, and water column sound speed. Such factors create a highly dynamic environment that presents significant challenges for deep learning-based source localization. Another limitation is that the MAPE values in this study have not fallen below approximately 10%, the value the Navy deemed acceptable. This minimum MAPE threshold is likely due to the remaining reverberation in the tank that is not absorbed by the anechoic panels.

4.3 Future Work

The next step is to transform this thesis into a published paper. Following that, future research should explore the application of transfer learning to additional sources of variability. One potential extension involves incorporating seafloor material variations. Testing seafloor variations would require the instillation of different materials on the bottom of the acrylic laboratory tank, such as plastic, aluminum, sand, or a sloped surface. Furthermore, the current hyperparameters of the CNN architecture could be further optimized to improve the robustness of the model. Continued development in the field of source localization in underwater acoustics could lead to advances in defense and environmental sonar applications, enabling deep learning models to adapt more effectively to ocean variability.

Bibliography

[1] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, "A survey of sound source localization with deep learning methods," The Journal of the Acoustical Society of America **152,** 107–151 (2022).

[2] B. G. Ferguson, "Defense Applications of Acoustic Signal Processing Acoustic signal processing for enhanced situational awareness during military operations on land and under the sea,", 2019.

[3] Sinay, "International Standards for Noise Monitoring: Shipowners and Operators' Responsibility,", 2023, accessed: 2025-04-16.

[4] National Oceanic and Atmospheric Administration (NOAA) Fisheries, "Ocean Noise,", https://www.fisheries.noaa.gov/national/science-data/ocean-noise, 2025, last updated by Office of Communications, NOAA Fisheries. Accessed 16 Jun 2025.

[5] A. V. V. D. Berg and R. J. A. Buwalda, "Source localization in underwater sound fields," Technical report (1994) .

[6] H. Niu, X. Li, Y. Zhang, and J. Xu, "Advances and applications of machine learning in underwater acoustics," Intelligent Marine Technology and Systems 1 (2023).

[7] Z.-H. Michalopoulou, P. Gerstoft, B. Kostek, and M. A. Roch, "Introduction to the special issue on machine learning in acoustics," The Journal of the Acoustical Society of America **150,** 3204–3210 (2021).

[8] Z. Huang, J. Xu, Z. Gong, H. Wang, and Y. Yan, "Source localization using deep neural networks in a shallow water environment," The Journal of the Acoustical Society of America **143,** 2922–2932 (2018).

[9] Y. Wang and H. Peng, "Underwater acoustic source localization using generalized regression neural network," The Journal of the Acoustical Society of America **143,** 2321–2331 (2018).

[10] F.-X. Ge, Y. Bai, M. Li, G. Zhu, and J. Yin, "Label distribution-guided transfer learning for underwater source localization," The Journal of the Acoustical Society of America **151,** 4140–4149 (2022).

[11] X. Hu, Z. Xu, X. Wu, Y. Yue, and W. Su, "Trans-CNN: A Transfer Learning Based Approach for Dolphin Sound Recognition," In *Proceedings of 2023 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC 2023*, (Institute of Electrical and Electronics Engineers Inc., 2023).

[12] "Deep transfer learning for source ranging: Deep-sea experiment results," The Journal of the Acoustical Society of America **146,** EL317–EL322 (2019).

[13] C. E. Dobbs and C. Emerson, "Deep Learning for Source Localization in a Laboratory Tank," BYU Scholars Archive (2024).

[14] C. T. Vongsawad and C. Taylor, "Development and Characterization of an Underwater Acoustics Laboratory Via in situ Impedance Boundary Measurements," BYU Scholars Archive (2021).

[15] "Source localization in the deep ocean using a convolutional neural network," The Journal of the Acoustical Society of America **147,** EL314–EL319 (2020).

Index