Passive Sonar Localization of a Moving Contact with Interference

A Thesis

Submitted to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in the

Department of Electrical Engineering

University of Saskatchewan

Saskatoon, Saskatchewan

by

Keith D. Jeffrey

Spring, 1995

Copyright

UNIVERSITY OF SASKATCHEWAN

Passive Sonar Localization of a Moving Contact with Interference

Candidate: Keith D. Jeffrey

Supervisor: Dr. J. E. Salt

M.Sc. Thesis Submitted to the
College of Graduate Studies and Research
Spring, 1995

Abstract

Cross correlation of received acoustic signals is a common method for estimating the location of an underwater contact in passive sonar systems. Movement by the contact during the observation period and the presence of an interfering coherent sound source degrade the cross correlation and impair the accuracy of the location estimate.

It is necessary to compensate for contact motion in order to avoid degradation of the correlation. One approach is to compress or expand the time axis of one of the received signals prior to correlating, with a scaling factor that compensates for that motion. This approach is computationally expensive.

The Select-Correlate-Sum method, proposed in this work, uses an alternate approach to compensating for contact motion that is twice as efficient. This approach uses a scaling factor to select appropriate short-duration extraction pairs from signals received from two sensors. These uncompensated extraction pairs are correlated and then averaged to obtain an adequately long total correlation time.

A scaling factor that properly compensates for contact motion also decorrelates a coherent interfering signal, causing it to affect the correlation like noise. In this way the compensation simultaneously avoids the correlation degradation caused by contact motion and the interference.

Expressions are derived for the time delay and for the scaling factor which maximize the magnitude of the correlation peak. These expressions are verified using computer simulations.

The performance of the Select-Correlate-Sum method is compared to the time-axis scaling method. The Select-Correlate-Sum method yields a somewhat higher variance in location estimate for a given observation period. However, to achieve a

given level of location estimate variance, the Select-Correlate-Sum method requires about half the number of computations.

Acknowledgements

I would like to thank Dr. Salt for his supervision of this work. His assistance and sound advice, during both the research and the preparation of the thesis, are greatly appreciated.

I wish to acknowledge the support of the Department of Computing Services, especially Mr. Dean Jones and Mr. Ed Pokraka, who encouraged me in this undertaking and who permitted me sufficient flexibility in my other responsibilities.

I would like to express my gratitude to Professor J.W.R. Griffiths and Dr. T.A. Rafik for their assistance and hospitality during a four month visit to the Sonar Laboratory of Loughborough University of Technology, U.K.

Mostly I wish to thank my wife, Noreen, whose encouragement, support and patience made the completion of this thesis possible.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Overview of Passive Sonar Localization

The work documented in this thesis is an extension to a large body of research on the use of passive sonar in the detecting and locating of underwater sound sources. The common application of techniques in this field is to accurately determine the location of submarine warships. Active sonar, in which sound impulses called pings or chirps are sent out through the water to be returned as reflections from the submarine, is familiar to many people. Active detection has the inherent disadvantage of informing the submarine of its detection. Passive sonar, in contrast, does not transmit any sound impulses but instead listens only to the sounds generated by the submarine and its movement through the water. Passive detection is accomplished secretly. Passive sonar has the additional militarily strategic advantage of not disclosing the existence or location of the equipment searching for the submarine.

A moving submarine emits a variety of sounds from its engines, its turning propellers and from the motion of water flowing over the surface of the hull. These sound emissions propagate through the water and can be received by hydrophone sensors. The sounds arriving at multiple hydrophones are compared for coherence in order to detect a submarine, to determine a compass bearing to the submarine and ultimately to estimate the actual location of the submarine. Hydrophone arrays range from two sensors to long linear arrays of hundreds of sensors and to three dimensional clusters of sensors.

A long linear array of hydrophones provides much more signal gain and can yield a high degree of accuracy in determining the compass bearing to the submarine. Several compass bearings obtained from different parts of a long array can be used with triangulation to estimate the actual location of the submarine and not just its

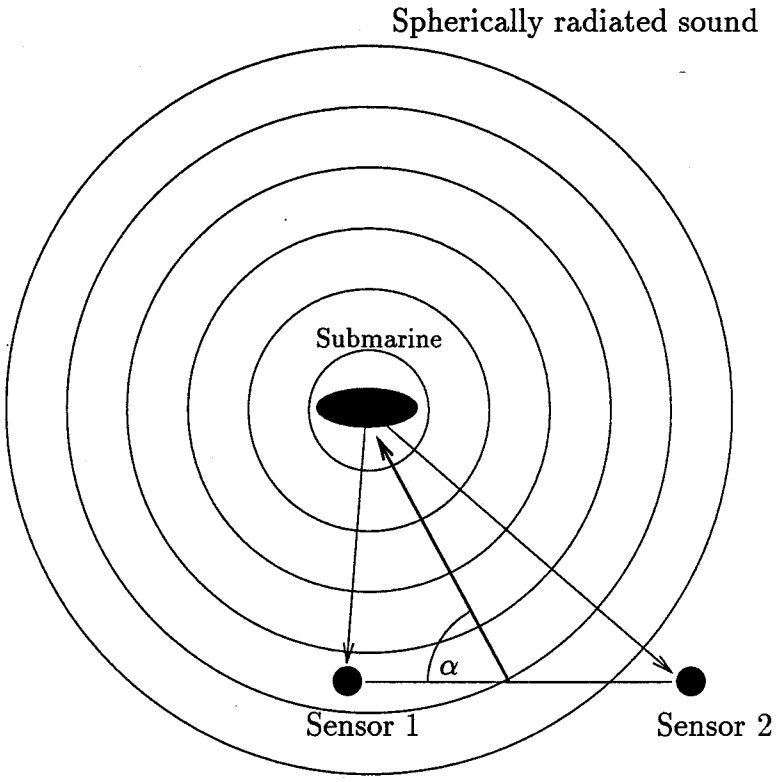

Figure 1.1 Plan view of an ocean region with a spherically radiating sound source and two hydrophone sensors.

bearing. Linear arrays can be laid on the bottom of the ocean, towed behind a moving vessel or attached to the external sides of the vessel hull. A horizontal linear array can directly determine a bearing, but it cannot directly determine the depth of the submarine. A vertical linear array can provide depth information, but not compass bearing. Three dimensional arrays of hydrophones can determine both bearing and depth.

What holds the clue to the location of the submarine is the timing of the arrival at the hydrophone sensors of sounds from the submarine. Consider the scenario where two hydrophone sensors have been deployed near a submarine, as shown in Figure 1.1. The figure shows a plan view of a region of open ocean at a particular instant in time where a submarine is moving past two hydrophone sensors which are part of a system

to determine the location of the submarine (localization system). Sounds emitted by the moving submarine are radiated in a spherical pattern at a speed characterized by the temperature and salinity of the water, but assumed to be constant in this region of ocean. Given the illustrated positions, the sound arrives at Sensor 1 first. At some later time, proportional to the extra distance that the sound must travel, the same sound arrives at Sensor 2. In examining the sounds captured by the two hydrophones, Sensor 2 can be seen to receive a delayed version of the sound received by Sensor 1. The amount of this delay, known as the time difference of arrival (TDOA) is the key information to be determined in passive sonar localization systems.

Early hydrophones employed air-filled tubes which channelled the sounds received by the sensors back to a human operator. The operator could lengthen the tubes, thereby introducing delay into one or other of the hydrophone signals. The amount of delay that had to be introduced in order for the sounds from the two channels to seem simultaneous to the human operator, corresponded to the TDOA.

Now hydrophones are typically underwater microphones which convert the underwater sounds into electric signals which are sampled, digitized and subsequently processed using digital computers. The TDOA can be computed by cross correlating the sample sequences from the two sensors. The correlation compares the sample sequence from one sensor with progressively shifted versions of the sample sequence from the other sensor. The correlation will produce a peak (the comparison will be best) at a shift which corresponds to the TDOA of the emitted sounds at the two sensors.

Sound radiated in a spherical pattern from a submarine might also reach the sensors after having been reflected at the surface or ocean floor boundaries. Reflections from the ocean floor may or may not be significant depending on the depth of the ocean and the makeup of the ocean floor. Bottom reflections are not significant in deep water or over a soft, muddy bottom. Surface reflections, though, are almost always significant and frequently used to assist with localization.

Figure 1.2 shows a sectional view of a region of ocean with the two submerged

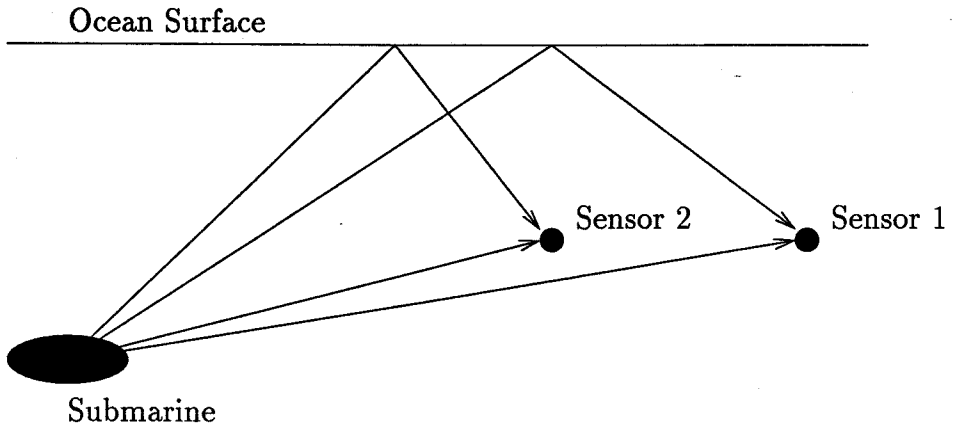

Figure 1.2 Sectional view of an ocean region showing direct and surface reflected sound paths to hydrophone sensors.

hydrophone sensors and a submarine. It is assumed that the ocean is very deep in this region and that no significant sound is returned from the ocean floor. In addition to the direct paths from the submarine to the sensors, sound arrives at the sensors on paths which are reflected at the water-air surface boundary. When reflected sound is present in the received signals, the cross correlation will produce additional peaks corresponding to other TDOAs attributable to the depth of the submarine relative to the sensors. In the general case, the cross correlation of sequences from the two-sensor array will show four peaks corresponding to the four TDOA values possible between the two sensors and two sound paths (direct 1 and direct 2, surface 1 and surface 2, direct 1 and surface 2, surface 1 and direct 2). Autocorrelation of each sensor's signal can also be used to get additional time differences (direct 1 and surface 1, direct 2 and surface 2). These TDOA values provide almost enough information to estimate the location (position and depth) of the submarine using just two submerged sensors. Additional information is required to determine which of two possible locations is the actual location of the submarine.

The two sensor array is important despite its lower sensitivity relative to larger arrays. Linear arrays attached to a ship or towed have the strategic disadvantage that they are constrained to the same position in the ocean as the ship. To remove that

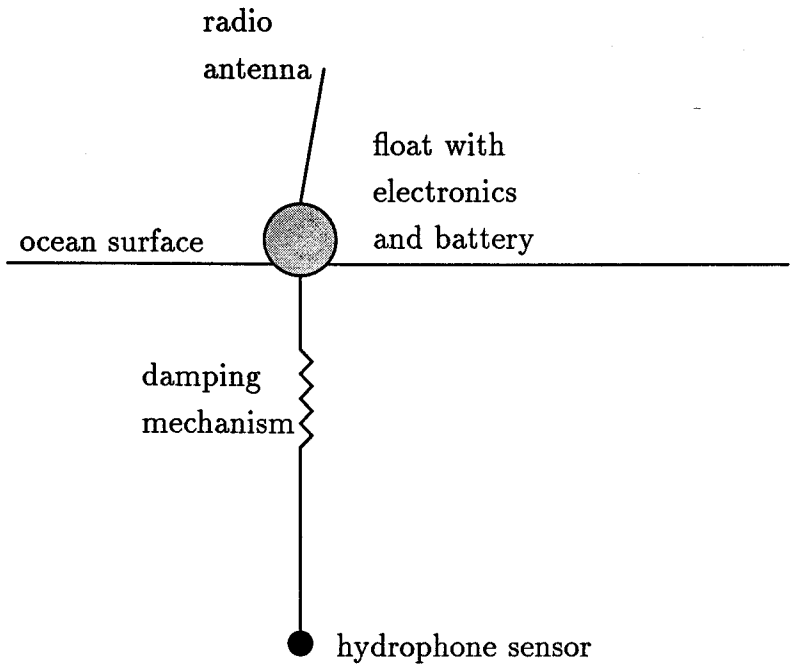

Figure 1.3 Diagram of a sonobuoy

constraint, sonobuoys are deployed. A sonobuoy is an apparatus that, upon deployment, floats on the surface of the water, dangles a hydrophone at a predetermined depth below the surface, raises an antenna and transmits a radio signal copy of the sounds that the hydrophone picks up. A sonobuoy is depicted in Figure 1.3.

Sonobuoys can be deployed by dropping them over the stern of a ship or dropping them from an airplane or helicopter. They can, therefore, be deployed where they are needed. Sonobuoys are not recoverable. After a short period of time the battery dies, a fuse burns to destroy the float and the apparatus sinks to the ocean floor. Clearly, if localization of a submarine can be accomplished with just two sonobuoys, then the cost will be much less than using larger numbers of them.

There are many factors which can complicate localization and reduce the accuracy of the estimate. The oceans are full of acoustic noise generated by surface waves and aquatic life forms. This noise comes from an almost infinite number of separate sources and is essentially incoherent when received by the hydrophones—the noise from these sources received by one sensor is not a time shifted version of the noise

received at other sensors. This noise will introduce randomness in the location of the correlation peak, and therefore into the estimation of the TDOA. The path of sound in the real ocean is not always a straight line. The salinity of the water varies with the depth and the change in density results in some refraction of the sound waves. The temperature of the water also varies and causes differences in the speeds that sound travels through the water. These factors introduce errors into the estimation of the actual location of the submarine. Uncertainty in knowing the exact location of the sensors in the ocean also introduces variance in the estimate. Finally, the motion of the submarine itself and the existence of other ships cause significant problems in using many of the localization algorithms based on cross correlation.

1.2 Problems Due to Interference and Source Motion

A strong second sound emitting ship in proximity to the sensors can overwhelm the sound from the submarine and make accurate localization difficult. A second source will have its own TDOA and peak in the cross correlation of the signals from the sensors. If the interfering source is strong, then the correlation will be confused. The signals from the two sources affect each other in the correlation, the correlation peaks will be skewed and their location on the axis of the correlation variable will not be truly representative of the respective TDOAs. General correlation to determine TDOA does not work well in the presence of strong interference.

Movement of the submarine during the period of observation also causes some problems in estimating the TDOA. Movement, in general, will result in a TDOA that varies with time. While it can often be assumed that the motion of the submarine is not significant during a short observation period, if the observation time must be long then that assumption is not valid. Movement during the observation has the effect of smearing the correlation peak. The top of the peak will be broadened, making it more difficult to pinpoint the maximum and making the maximum much more susceptible to dislocation by noise.

Interference and motion are, however, two frequently encountered conditions in

the real problem of localizing a submarine during hostilities. A submarine will approach a surface ship to bring it within range of torpedoes. The surface ship is a very strong emitter of sound and normally both it and the submarine will be in motion. Alternatively surface ships try to detect, localize and track submarines to remain aware of their location or to attempt to destroy them. An algorithm that can determine a submarine's location under these two conditions is, therefore, of interest.

The submarine will try to remain undetected by the surface ship. One way that it can do this is to trail behind the ship. The very loud sound generated by the surface ship's propellers creates a blind spot directly aft of the ship in which hydrophones attached to the sides of the ship's hull cannot detect a submarine. Ships will occasionally want to check that blind spot to determine whether they are being closely followed by a submarine. One technique is to turn the ship sharply so that the hull sonar can search the path the ship had just been on. The submarine will be able to detect this ship's maneuver, will know that it is being sought and can take measures to remain undetected.

Another technique to detect a following submarine would be to drop sonobuoys over the stern of the ship and continue along the same course. The submarine will not be able to detect the deployment of the sonobuoys and it will be observable in the vicinity of the sonobuoys. Since the sonobuoys are deployed directly aft of the ship, the geometry has the ship and two sonobuoys collinear. In this case, the sound from the ship arriving at the two deployed sonobuoys has a constant TDOA despite the ship's continuing motion. The submarine will be located either on the far side of both sonobuoys from the ship or caught between the two sonobuoys. In the former case, if the submarine is following directly behind the ship, its TDOA at the two sensors is constant. However, this constant TDOA is quite different from that of the ship, so a straightforward correlation will yield clear and separate peaks. For the case where the submarine is between the sensors, the motion of the submarine results in a fast-changing TDOA for the submarine.

The scenario that catches the submarine between two sensors, which is the more

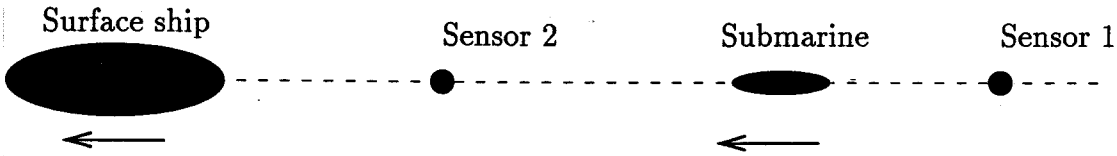

Figure 1.4 Scenario to study the case of a moving submarine with a strong interfering surface ship.

difficult from the signal processing point of view, will be used to demonstrate the proposed algorithm in this thesis. It is a plausible scenario and it has several advantages. The ship's TDOA is constant, the submarine's TDOA is fast-changing and, finally, the motion of the submarine between the sensors produces a linearly varying TDOA. These conditions make it a good scenario to use to demonstrate the algorithm.

1.3 Exploiting Source Motion to Suppress Interference

Both interference and motion have been identified as problems for general correlation localization algorithms. Interference and motion have also been identified as likely conditions to occur in submarine warfare. The proposed algorithm exploits the motion as a way to accurately localize a submarine in the presence of strong interference. The scenario is as illustrated in Figure 1.4, a plan view of a region of open ocean. A surface ship under way, suspecting that it may be closely followed by a submarine, has recently dropped two sonobuoys over its stern. The submarine, unaware of the sonobuoy's deployment continues to closely follow the ship at a constant speed. The ship examines the signals from the sonobuoys during the period that the submarine is between the two sensors.

Although the ship is in motion, due to its location being collinear with the two sonobuoys, it has a constant TDOA which is the maximum possible for any particular separation distance of the two sonobuoys.

The submarine's position is changing rapidly. It is moving towards Sensor 2 and away from Sensor 1. The TDOA of sound from the submarine is changing

linearly. The ship wants to determine the location of the submarine at the end of the observation period, based on the TDOA at that instant in time.

The signal received by Sensor 2 will be a time-compressed version of the sound emitted by the submarine and, similarly, the signal received by Sensor 1 will be a time-expanded version. If the submarine is moving towards a sensor, the time sound takes to travel from the submarine to the sensor becomes progressively shorter. This causes each successive bit of the signal to arrive at the sensor a little earlier than it would have had the submarine not been moving. This can be thought of as compressing the time scale of the received signal relative to the time scale of the sound actually emitted. This is just another way of looking at the Doppler effect. Instead of considering the signals to be shifted in frequency, the signal time scales are considered to be compressed or expanded.

Since the signal from the submarine received at Sensor 2 is compressed and that received at Sensor 1 is expanded, the two are no longer just shifted replicas of each other and a straightforward correlation will not yield a clear peak for an accurate estimation of the TDOA for the submarine's emitted sound. Since the TDOA for the ship is not changing, a straightforward correlation would yield an estimate for its TDOA.

The proposed algorithm will purposefully expand the signal received at Sensor 2 (which is an compressed version of the emitted sound) until it is a good replica of the signal received at Sensor 1 (which is an expanded version of the emitted sound). Since the true time scale of the emitted sound is unknown, the signal received at Sensor 1 is chosen as a reference. The amount of expanding required, to make the Sensor 2 signal a replica of the Sensor 1 signal, corresponds to the rate of change in the TDOA of the submarine. That rate of change is itself proportional to the speed of the submarine. Once the Sensor 2 signal is sufficiently expanded, it is a replica of the Sensor 1 signal and the two can be correlated. In this way compensation is made for the motion of the submarine.

Before the Sensor 2 signal had been expanded, correlation of the two received

signals yielded a good estimate of the TDOA for the surface ship. After expanding, good replicas of the ship sound no longer exist in the two received signals and the correlation does not produce a pronounced peak at the TDOA for the ship. Without the interference of lobes in the correlation from the ship sound, the determination of an accurate TDOA for the submarine is unhindered. The algorithm yields an accurate TDOA for a moving submarine in the presence of strong interference from the surface ship.

The location and speed of the submarine will not usually be known, so the amount of expanding required will not be known. The algorithm will require a search for the motion-compensating scaling factor that results in the best correlation of signals originating from the submarine. The result of that search will be the scaling factor that best compensates for the changing TDOA, and the location of the correlation peak at that compensating factor will yield the TDOA for the submarine at the end of the observation period (the clue to the submarine's location).

1.4 Focus On Computational Efficiency

Expanding the time scale for each sample in the digitized sample sequence for every guess at the expanding factor is a very computationally expensive operation. There is some urgency in localizing the submarine not only because of the danger of attack, but also because the submarine is moving rapidly and delay in determining its location after observation may reduce the usability of the estimate since the submarine may have changed speed or course in the interim.

The proposed algorithm seeks to be much more computationally austere relative to full time scale expansion. The price of the computational savings will be some loss of accuracy or increased variance in the estimate of TDOA.

The algorithm proposes to correlate only short-time extractions from the sample sequences, where the length of the extractions is short enough so that time scale expansion is not required. The short extractions, one from the Sensor 1 sample

sequence and the other from Sensor 2, will be approximate replicas of one another. To provide an adequately long correlation time, many of these extraction pairs will be correlated and averaged. Successive guesses at the scaling factor direct the selection of the data points to make up the extraction pairs. The avoidance of scaling to expand a sample sequence will result in much greater computational efficiency.

1.5 Simplifications Used in this Investigation

In order to adequately present and verify the algorithm, a number of assumptions and simplifications are made. The sounds received at the two sensors from the submarine and the surface ship are synthesized as random low frequency sample sequences. This is a common practice in the development of passive sonar algorithms. This permits the algorithm to be tested on a large number of different data sequences. Localization based on two sensor arrays has been adequately dealt with in [1]. The geometry is confined to the scenario described above. All of the essential elements of the algorithm can be investigated with this scenario. That scenario presents only a linearly varying TDOA. Only direct sound paths from the submarine and ship to each of the sensors are considered. The reflected or refracted paths are neglected. While they are generally present, a discussion of surface reflections would needlessly complicate the presentation of the algorithm. Finally the uncorrelated ocean noise will be omitted from the development. The noise would have an effect on the variance in the estimation of the TDOA, but its presence does not affect the fundamentals of the algorithm

After describing the algorithm in detail (Chapter 2) and developing the mathematical expressions for the optimum scaling factor and TDOA (Chapter 3), several of the key decisions and problems in the algorithm will be investigated more thoroughly (Chapter 4): Selection of the length of the extractions; selection of the observation period length; alignment of the multiple correlations prior to summing; number of extraction pairs to sum. In Chapter 5 the theory will be validated using computer simulations. Chapter 6 will discuss some properties of the algorithm and compare

the algorithm and full time scaling, in terms of accuracy and computational cost. Finally, Chapter 6 provides some concluding remarks and a discussion of areas for future investigation.

2. Approach to the Problem

2.1 Statement of the Problem

The objective of this work is to develop an algorithm for efficiently determining the location of a moving underwater source in the presence of a strong interfering sound source.

To keep the discussion clearer, the scenario presented in the introduction will be used throughout the development. In that scenario, a surface ship (the interferer) is attempting to locate a following submarine (the contact). The ship has recently deployed two sonobuoys over the stern, and the contact is assumed to be in motion between the two sonobuoys. This scenario is pictured in Figure 2.1. The hydrophone sensors and the contact are all assumed to be at the same ocean depth of 200 meters. The sensors are separated by 500 meters. The ocean medium is considered sufficiently homogeneous so that sound travels in straight paths at a constant speed over short distances.

The characteristics of the sounds emitted by the contact and the interferer will, in general, be only partially known. These characteristics depend on the type of vessel, its speed and, in the case of a submarine, its depth. They are commonly modelled as zero mean, Gaussian, stationary stochastic processes [2, 3], with large tonal components at a frequency related to the propeller turning speed. The acoustic signals emitted from the contact and the interferer are independent. Only the outputs from the sonobuoys are available for processing, and these are sequences of discrete-time, digital samples of the sound detected by the sensors. The sampling at the sensors is synchronized. For the purposes of this study, the sampling rate is 8000 samples per second, the tonal components caused by the spinning propellers are assumed to have been filtered out [4, 5, 2, 6], and the broadband component is assumed to be filtered

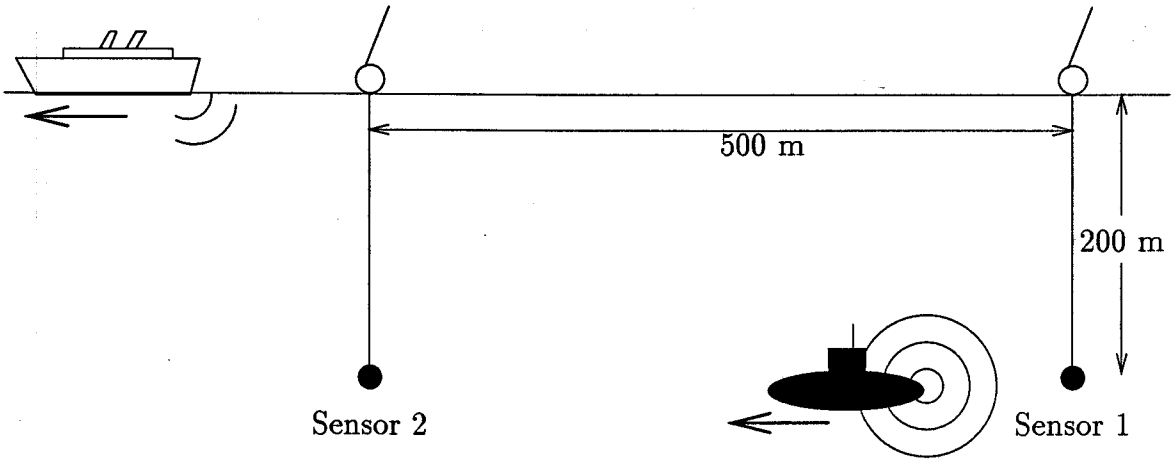

Figure 2.1 Sectional view of scenario geometry

above 400 Hz.

2.2 General Correlation Technique

The Select-Correlate-Sum algorithm, being developed in this paper, is one of several techniques that uses correlation of received acoustic signals to determine the time difference of arrival at the two sensors of sound from a contact. The role of correlation in these techniques will be discussed in this section.

For the moment, consider the following simplifications to the scenario: There is no interferer present and the contact is not moving (but still radiating acoustic signals). Then the sound received at the two sensors are time delayed versions of the sound emitted by the contact. The parameters of the sound are as described in Section 2.1. A typical cross correlation of sample sequences from the two sensors, $R_{12}(\tau)$, is plotted in Figure 2.2. The shape of the correlogram is very similar to that from an autocorrelation of a broadband signal, which is to be expected since the two sequences are largely time-delayed versions of each other. The maximum occurs at the peak of a main lobe whose width from zero-crossing to zero-crossing is proportional to $1/B$, where B is the bandwidth of the signals (400 Hz in this example). The magnitude of

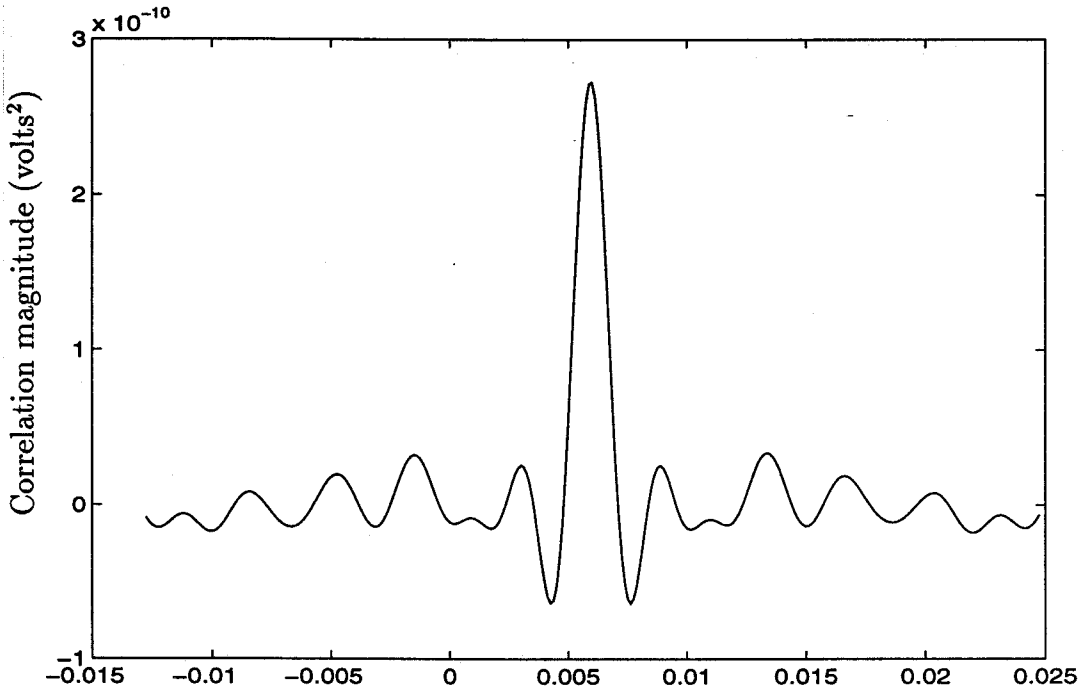

Figure 2.2 Cross correlation of signals received from a still source without interference.

the maximum of the correlation is equal to the square root of the product of powers in the signals received at the two sensors. Numerous additional lobes, called side lobes, have a smaller magnitude.

Whereas the location of the peak in an autocorrelation is always at the origin, i.e., corresponding to a shift $\tau = 0$, the location of the peak in the cross correlation will, in general, occur at a non-zero shift. The location of the peak corresponds to the time delay between replicas of the sound—the time difference of arrival (TDOA). The TDOA can be used in localization systems to determine an estimate of the compass bearing to the contact and an estimate of the location of the contact.

The correlation function, $R_{12}(\tau)$, is not completely known, but can be estimated with a time-average correlation, $\hat{R}_{12}(\tau)$. The assumption is made that the signal is correlation-ergodic—that the autocorrelation time-average estimate, $\hat{R}_{cc}(\tau)$, will approach the autocorrelation function, $R_{cc}(\tau)$, in the limit as T approaches infinity. Because the correlation time, T, is not infinite, the location of the main lobe peak

is just an estimate of the TDOA. The signals are random in nature and the shape of the correlogram and the peak location will vary with the actual data contained in the sequences. A longer correlation time will improve the time average estimate of the correlation and the estimate of the TDOA.

Locating the peak of the main correlogram lobe is more precise if the lobe is narrow and sharp. A wider signal bandwidth would result in a narrower lobe, but high sound frequencies are rapidly attenuated in the ocean and are not usually significantly present in the signals as they are received.

In later sections, the effects of noise, contact motion and interference on the correlation and the estimate of TDOA are explained. First, however, the next section is dedicated to a description of the units used to express quantities related to sound information used in sonar.

2.3 Units for Sound Levels

For acoustic analysis, underwater sound sources are considered as point sources which radiate a pressure wave equally in all directions. The pressure wave propagates through the ocean medium as the back and forth motion of the molecules. The wave is detectable as pressure changes by a hydrophone sensor. As the wave propagates, there will be a flow of energy past the hydrophone which can be expressed as energy per second across a unit area, and which is defined to be the intensity of the wave. The equation for acoustic intensity is given by

$$I = \frac{\langle p^2 \rangle}{\rho c} \tag{2.1}$$

where $\langle \cdot \rangle$ indicates a time average, p is pressure in newtons/m^2, ρ is the density of water in kilograms/m^3, c is the speed of sound in water as m/s and I is the intensity (power per unit area) in watts/m^2.

The intensity at a range of r meters from a spherically radiating compact source

of total power P watts is

$$I = \frac{P}{4\pi r^2}. \tag{2.2}$$

To handle the wide range of intensities, acoustic analysis uses a decibel notation, expressing intensities relative to an intensity reference standard. That standard is for an acoustic pressure of one micropascal. For $\rho = 1$ gram/cm^3 and $c = 1500$ m/s, that acoustic pressure results in an intensity of 0.667 x 10^{-18} watts/m^2.

In underwater acoustics it is customary to refer to the strength of a compact sound source in terms of the intensity at a distance of one meter from the hypocenter and expressed relative to the reference intensity. Converted to decibels, this relative source intensity is called the source level (SL) and is denoted with SL//μPa@1m. The source level of a P watt compact source is defined as

$$\text{SL//}\mu\text{Pa@1m} = 10\log\left[\frac{I_{1m}}{I_{ref}}\right] \tag{2.3}$$

$$= 170.77\text{dB} + 10\log\frac{P}{1 \text{ watt}} \tag{2.4}$$

where P is the power of the source in watts.

A broadband acoustic signal radiates power over the frequency band without having any power concentrated at a single frequency. The power distribution is usually conveyed as a plot of the intensity of radiated power in a 1 Hz bandwidth versus the center frequency of the 1 Hz band. This spectral density is denoted SL//μPa/1Hz@1m, and has units of 1/Hz.

2.4 Effect of Noise on Correlation

Although uncorrelated noise sources (such as noise in the hydrophone electronics, noise related to the flow of water past the hydrophone sensor, suspension and cable, and ambient ocean noise) are not given much consideration in this thesis, an examination of the effect of noise on the correlation of the signals from the contact

is required. The Select-Correlate-Sum processing method being developed alters the effect of the interferer to an effect equivalent to that of uncorrelated noise of a similar total power.

Noise corrupts the sample sequences and degrades the similarity of the delayed versions of the sound emitted by the contact. This in turn affects the degree of correlation and the shape of the correlogram. Most importantly, noise introduces additional variance into locating the peak of correlogram main lobe and therefore into estimating the TDOA. The sharpness of the peak, i.e., the narrowness of the main lobe, affects the degree to which noise can affect the location of the peak. A main lobe with a larger magnitude second derivative in the vicinity of the peak is less affected by noise.

To demonstrate the effect of noise on the estimate of the TDOA, various levels of uncorrelated noise were added to sample sequences received from a non-moving contact. The noise sources were considered as point sources located an equivalent distance from the sensor as the contact, however each noise source affected only one sensor. A plot of the variance of the TDOA estimate versus the source to noise Source Level ratio is shown in Figure 2.3. A straightforward correlation was used over a 1 second period. This figure clearly shows the decrease in TDOA variance with increasing Signal to Noise Ratio (SNR).

2.5 Effect of Contact Motion on Correlation

Movement of the contact during the correlation time has the effect of flattening the peak of the correlogram main lobe. This exacerbates the effect of noise and interference.

Consider that the contact is moving during the observation period along a path away from Sensor 1 towards Sensor 2. The TDOA is changing throughout the correlation time. This has the effect of smearing the correlogram, broadening the width of the lobes or flattening the peak. The effect of increasing contact speed on the

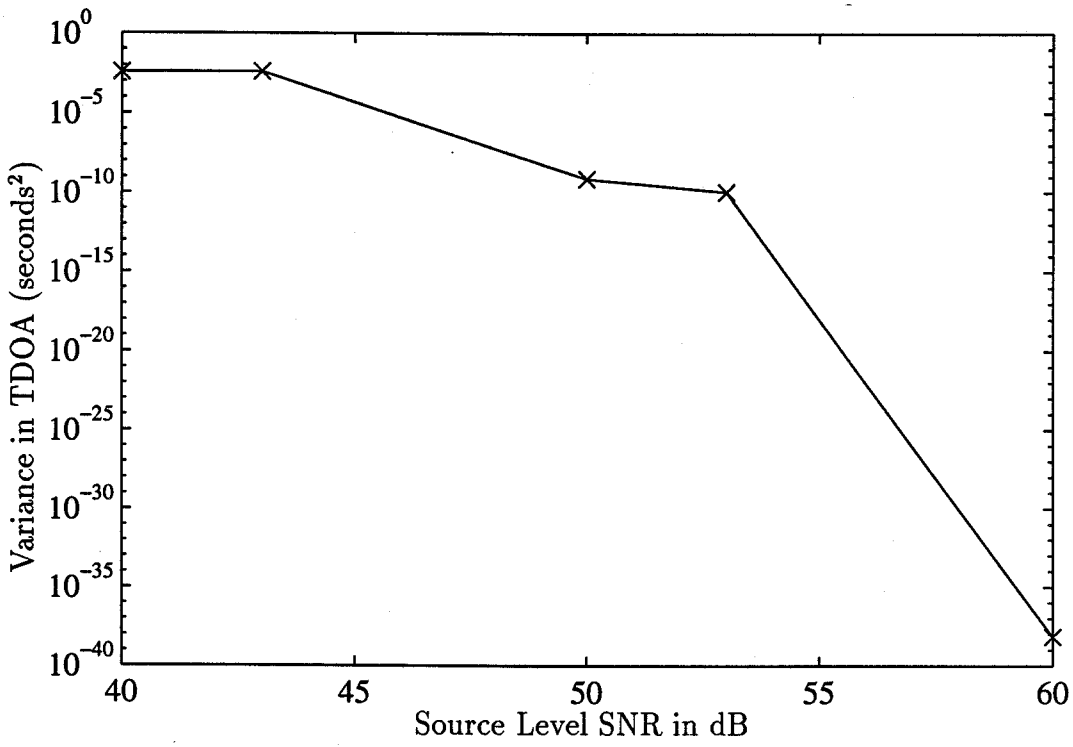

Figure 2.3 Plot of TDOA variance vs source:noise Source Level ratio in dB.

shape of the correlogram is shown in Figure 2.4. The figure shows correlograms for a contact moving at 0, 5 and 10 m/s. The three correlograms are arbitrarily aligned at a correlation variable shift $\tau = 0$ for comparison. The correlation time for each case is $100/B$, where B is the bandwidth. Compared to the correlogram for the stationary contact, the correlogram of the contact moving at 5 m/s shows slight smearing, while a large amount of smearing is seen in the correlogram for the contact moving at 10 m/s.

The consequence of smearing is that it becomes more difficult to precisely estimate the TDOA because the main lobe is broader and the peak is more susceptible to being displaced by noise. The variance in the estimate of the TDOA increases due to contact motion. The variance of the TDOA estimate for a selection of contact speeds is plotted in Figure 2.5, for a SNR of 100 dB.

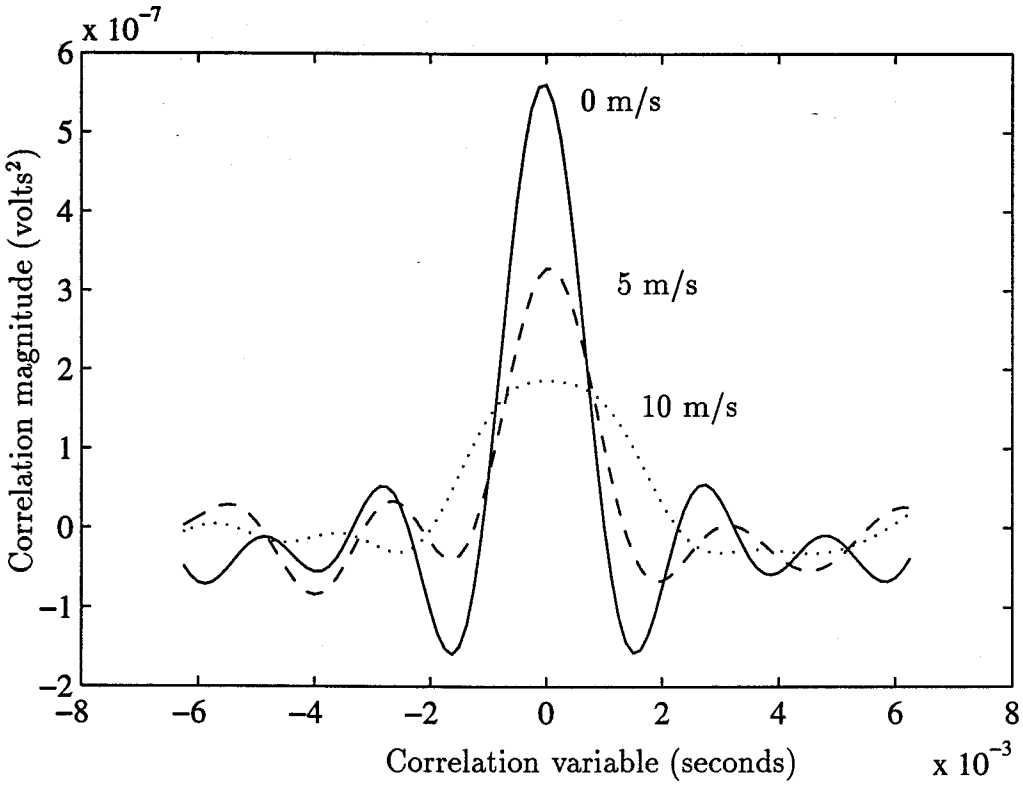

Figure 2.4 Cross correlations for contact speeds of 0, 5 and 10 m/s. All correlograms based on observation period of $100/B$.

For a still contact, increasing the length of the correlation time is effective in reducing the variance in TDOA estimate caused by noise. However, for a moving contact, lengthening the correlation time also increases the total amount of change in TDOA. This exacerbates smearing of the correlogram and increases variance in the TDOA estimate.

2.6 Effect of Interference on Correlation

A strong interference will adversely affect the accuracy of the TDOA estimate. Sound from both the contact and the interferer are represented in the sequences being correlated. Since the two sources are independent, the resulting correlogram is like the sum of two correlograms obtained by separately correlating the contact and interferer components. A local peak in the contact correlogram could be increased, decreased or slightly shifted when combined with the correlogram from the interferer. If the

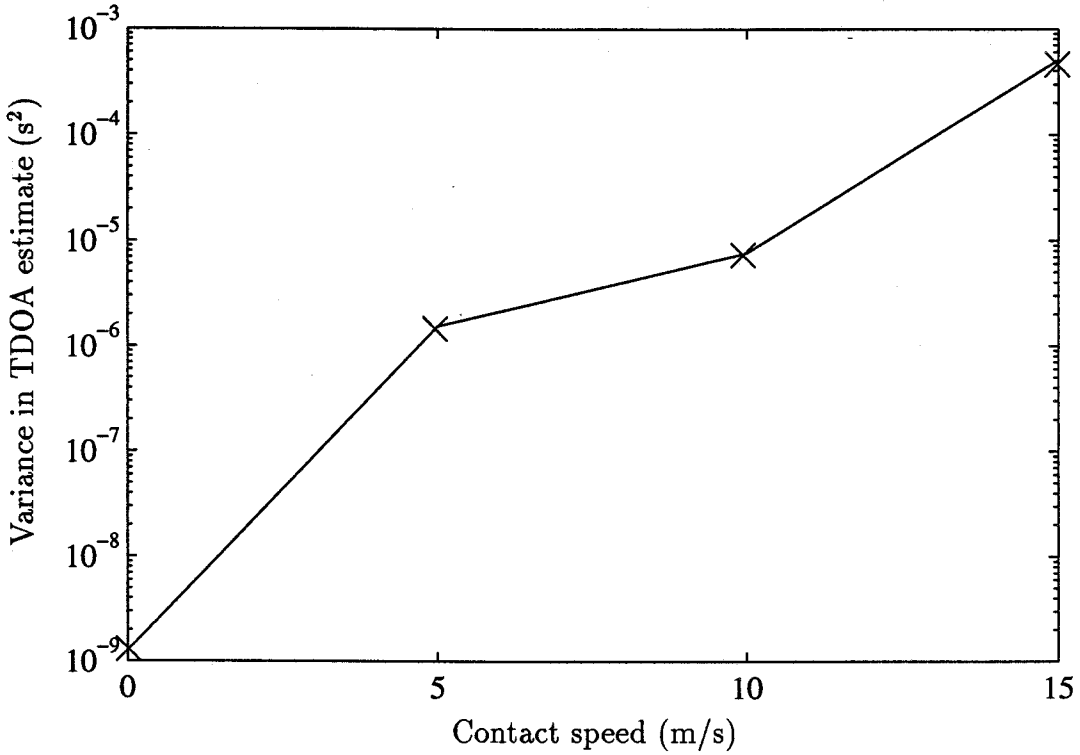

Figure 2.5 Variance in TDOA estimates vs contact speed using a SNR of 100 dB. All correlations based on observation period of $100/B$.

interferer is strong, then many of the side lobes in the interferer correlogram could be large enough to affect the location of the main lobe of the contact correlogram. A simulation of the effect of a strong interferer on the correlogram of the contact is illustrated in Figure 2.6.

The figure shows how, in the correlation of signal comprising sound from both contact and large interferer, the location of the peak of the main lobe for the contact is displaced by a strong side lobe from the interferer. The large main lobe of the interferer correlogram is not found in this range of the correlation variable.

The degree of the effect of the interferer is determined by the strength of the interferer and by the sharpness of the contact correlogram main lobe.

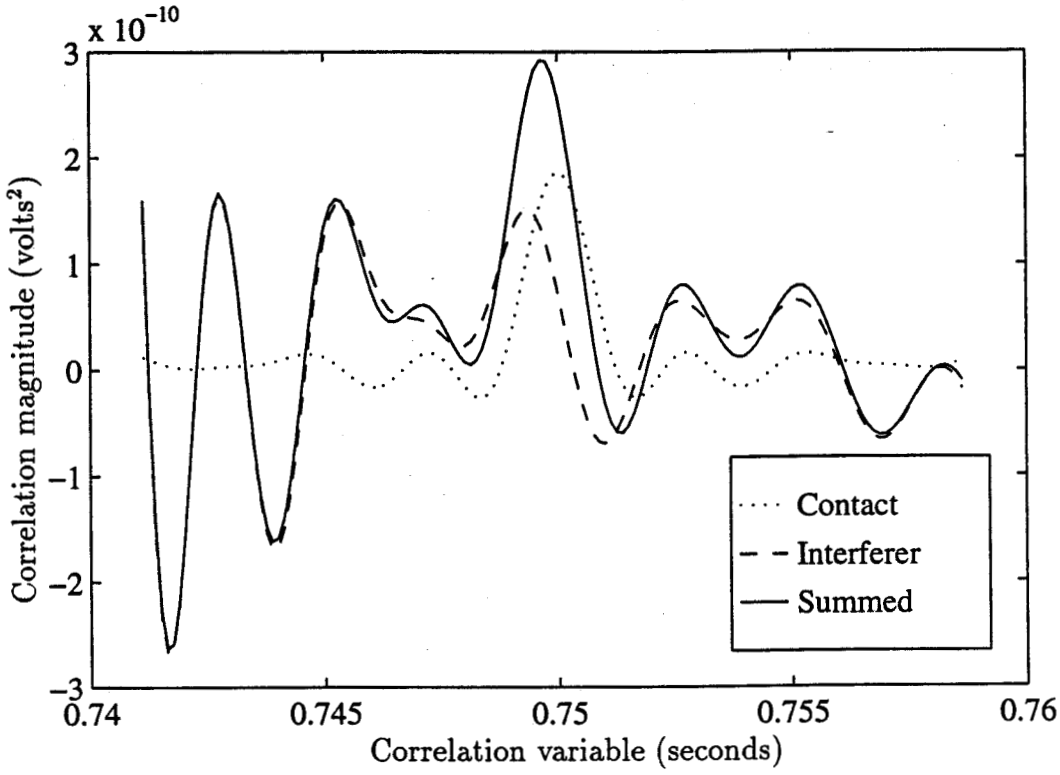

Figure 2.6 Example correlograms for contact, very large interferer and the resulting combined correlogram.

2.7 Effect of Time Scaling During Signal Processing

The negative effects of contact motion and interference on the estimation of TDOA can be reduced by time scaling one of the received sample sequences to compensate for the contact motion. In the scenario, the contact signal received by Sensor 1, $s_{c1}(t)$, is like a version of the emitted sound where the time scale has been stretched. This is because the contact is moving away from Sensor 1 and sound emitted later takes incrementally longer to reach Sensor 1. Similarly, sound from the contact received at Sensor 2, $s_{c2}(t)$, has a compressed time scale relative to the emitted sound. This is shown, in exaggerated fashion, in Figure 2.7. The signals $s_{c1}(t)$ and $s_{c2}(t)$ are not good replicas of one another and lobes in the correlogram of their cross correlation would be smeared, as discussed earlier.

The time scale of $s_{c2}(t)$ can be stretched until it becomes a good replica of $s_{c1}(t)$. The choice of $s_{c1}(t)$ as a reference signal is arbitrary. The amount of scaling necessary

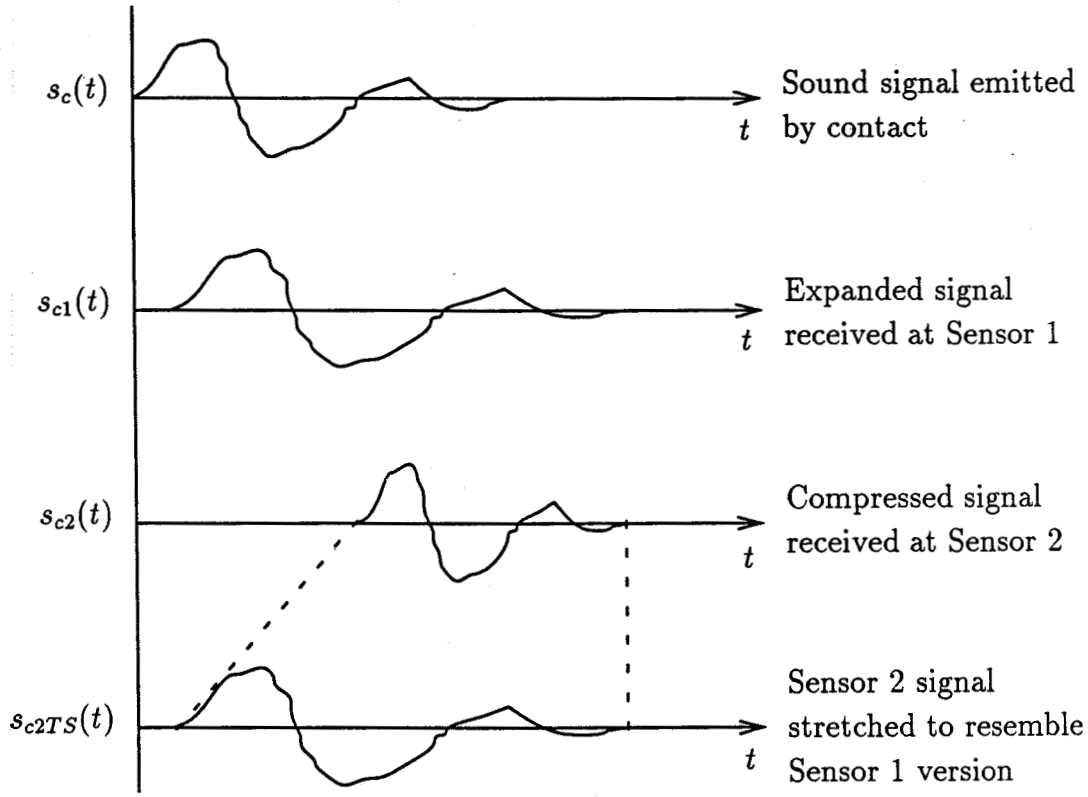

Figure 2.7 Sketch of time signals emitted by contact, received by Sensor 1, received by Sensor 2, and Sensor 2 signal stretched prior to correlating with the signal received by Sensor 1.

to make $s_{c2}(t)$ like $s_{c1}(t)$ is proportional to the rate of change of the TDOA which is proportional to the speed of the contact. It is not important that neither $s_{c1}(t)$ nor $s_{c2}(t)$ are good replicas of the emitted sound. Once $s_{c2}(t)$ has been stretched sufficiently, say to $s_{c2TS}(t)$, it is correlated with $s_{c1}(t)$ to yield a correlogram in which the peak for the contact is not smeared. Without smearing, the peak remains sharp and less susceptible to being dislocated by noise.

Scaling of the time axis of the signal from Sensor 2 affects the interferer component as well as the contact component. Because of its position relative to the two sensors, the interferer's motion does not introduce a change in the TDOA. The TDOA for the interferer is the maximum possible given this spacing between the sensors and will remain maximum as long as the interferer remains collinear with the sensors.

Since the TDOA is not changing, a cross correlation of interferer component, $s_{i1}(t)$ and $s_{i2}(t)$, produces an unsmeared correlogram peak, that is if there is no time axis scaling.

However, the interferer component, $s_{i2}(t)$, is scaled along with the contact component, $s_{c2}(t)$. The scaled interferer component, $s_{i2TS}(t)$, is no longer a good replica of $s_{i1}(t)$ and correlation of these two components produces a correlogram with flattened main and side lobes. The smeared lobes can be considered to have been averaged, or low-pass filtered, with the result that the peaks are less pronounced and more gradual. The location of the contact correlogram's main lobe is affected less by the smeared interferer side lobes, so the contact TDOA can be more accurately determined.

Certain scaling factors diminish the effect of the interferer and enhance the TDOA estimation for the contact. Other scaling factors (for example, no scaling in this scenario) diminish the correlogram of the contact and peaks the correlogram for the interferer.

2.8 Search for Scaling Factor and TDOA

Since the speed of the contact is unknown, the scaling factor required to compensate for contact motion must be estimated. Successive scaling factors are hypothesized and tested. Better guesses at the scaling factor make the scaled signal, $s_{c2TS}(t)$, a better replica of $s_{c1}(t)$ and result in a more pronounced contact main lobe in the correlogram. Guesses that are far removed from the scaling factor that compensates for interferer motion will result in a significant filtering effect on the lobes contributed by the interferer. Each scaling factor guess results in a separate correlogram. The set of correlograms from all guesses, placed side by side forms a surface. The maximums of the surface indicate the best estimates for TDOA and rate change of TDOA for the contact and the interferer. This is commonly referred to as an ambiguity surface, since the two peaks must be differentiated using additional information. Knowledge of the scenario will dictate which maximum indicates the parameters for the contact or interferer. For example the speed of the surface ship (the interferer) will be known.

Also, due to its location collinear with the two sensors the rate of change of TDOA for the interferer will be zero. Additionally, it can be assumed that the following contact has a speed close to that of the interferer, from which an approximate rate change of TDOA can be derived.

2.9 The Select-Correlate-Sum Algorithm

This section introduces the fundamentals of the Select-Correlate-Sum algorithm.

Scaling the time axis of the sample sequence from Sensor 2 is a computationally expensive process which must be repeated for each unique guess at the scaling factor. Scaling of a discrete sample sequence can be done by calculating new arrival times for each sample using a function of time and the scaling factor. The resulting sequence is then interpolated at uniform intervals.

Time axis scaling can be avoided by correlating unscaled extractions that are sufficiently short so that smearing is not a significant problem. Over a very short interval, $s_{c1}(t)$ and $s_{c2}(t)$ will be almost delayed replicas of one another. The scaling factor is used to calculate the location in the Sensor 2 sequence of the extraction that is proposed as the replica of a short extraction from the Sensor 1 sequence. Each scaling factor proposes a slightly different extraction from the Sensor 2 sequence. The extractions are cross correlated for each scaling factor and an ambiguity surface is prepared as before.

A short extraction, however, provides an insufficiently long time-averaged cross correlation to give a reliable estimate of the desired TDOA. So numerous different extractions pairs are respectively correlated throughout the observation period. This is shown in Figure 2.8.

In Figure 2.8, $s_{c1}(t)$ and $s_{c2}(t)$ are the expanded and compressed versions of the emitted contact sound as received at Sensor 1 and Sensor 2, respectively, in the absence of interference. Taken as a whole, they are clearly not good replicas of one another. Three extractions from $s_{c1}(t)$ are indicated as the parts of $s_{c1}(t)$ visible

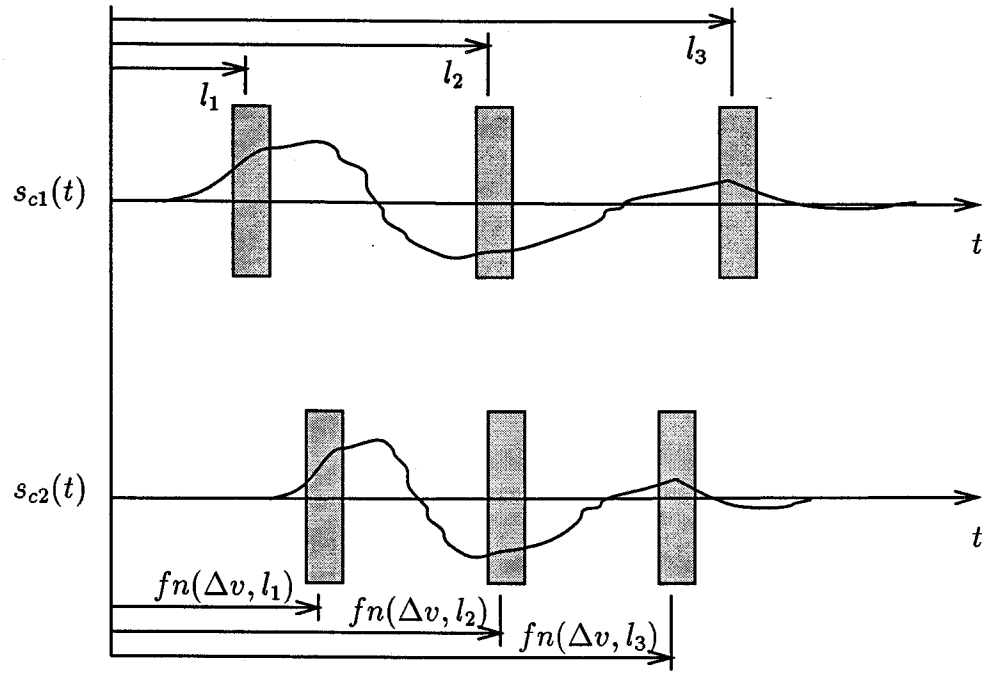

Figure 2.8 Representations of $s_{c1}(t)$ and $s_{c2}(t)$, expanded and compressed versions of the emitted contact sound. Short extractions from each signal, indicated as the parts of the signals within the shaded regions, are correlated as matching pairs.

within the shaded regions, centered at times l_1, l_2, and l_3 from an arbitrary time origin. Extractions from $s_{c2}(t)$ are selected using a function of the currently proposed scaling factor and the times of the matching extractions from $s_{c1}(t)$, l_1, l_2, and l_3. These extractions are indicated as the parts of $s_{c2}(t)$ within the shaded regions. It can be seen that the first extractions from $s_{c1}(t)$ and $s_{c2}(t)$ are nearly time-shifted replicas of one another. The same can be said of the other extraction pairs. The short correlation of each extraction pair contributes to a correlogram. In total, the numerous extractions amount to a sufficiently long correlation time.

The computational savings are the result of not having to resample the Sensor 2 sample sequence to stretch its time scale. Additional efficiency could be attained by only computing points on the aggregate ambiguity surface necessary to identify the maximum, for example through the use of a "hill-climber" algorithm.

2.10 Related Studies

There have been several studies which propose methods to compensate for contact motion.

Scarbrough [3] described a method using a frequency domain approach. Instead of correlating time series, the method takes the product of the Fourier transform of segments from each sensor. Compensating for contact motion is accomplished by applying an appropriate phase rotation to the cross power spectrum estimate for each segment. The estimates from all segments are then averaged using appropriate weighting functions before taking the inverse Fourier transform.

Betz [2] describes a method that correlates the time series within each segment once, and subsequently shifts those correlations prior to summing as a way of compensating for contact motion. Kuhn et al. [5] made a comparison of these time domain and frequency domain approaches. That paper points out that the time domain method lends itself better to cases with a large time delay. The frequency domain method can readily handle cases with time delay of just one-half of the segment length.

None of these methods address the issue of a strong coherent interferer, which is a key factor in this thesis. They do however describe large computational savings compared to scaling every data point prior to a lengthy correlation.

Izzo et al. [7] did address interference but using a method that required the interference to by cyclostationary.

2.11 Summary

This chapter has presented a context and an overview of the Select-Correlate-Sum algorithm to determine an estimate of the TDOA for the contact in the scenario discussed. The Select-Correlate-Sum (SCS) algorithm selects short extractions in such a way as to compensate for contact motion, correlates the unscaled extractions and then sums the results of the correlations. Correlating short extractions minimizes

the effect of smearing caused by contact motion during the correlation time. Using a large number of extractions provides an adequately long aggregate correlation time to reduce the variance of the estimate caused by noise. The effect of the strong interferer is reduced by exploiting the motion of the contact and interferer which results in different rates of change of TDOA. The motion must be observed over a sufficiently long period of time so that adequate averaging of the interferer side lobes occurs. The issues related to choosing the length of the extraction, the number and selection of extractions and the length of the observation period will be discussed in Chapter 4. The computational savings resulting from not scaling a sample sequence was introduced.

This description of the SCS algorithm will be developed mathematically in Chapter 3.

3. Mathematical Development of the Select-Correlate-Sum Algorithm

The previous chapters have provided a description of the problem of efficiently determining the time difference of arrival of a moving contact in the presence of a strong interferer. As well, the previous chapters have outlined the workings of the Select-Correlate-Sum algorithm that is being proposed in this paper. This chapter will develop the mathematical basis of that algorithm.

A mathematical expression will be developed for the ensemble average cross correlation function by taking the expected value of the time-average cross correlation of the signals received at the two sensors. Expressions will be derived for the value of the scaling factor that produces the largest peak in the ensemble cross correlation function and for the shift at which that peak occurs. These values are estimates, respectively, of the time scaling caused by contact motion and the time difference of arrival (TDOA) of sound from the contact. The TDOA estimate is of primary importance, although there is useful contact velocity information in the scaling factor.

Finally in this chapter, the relationship between the mathematical basis and the implementation of the Select-Correlate-Sum algorithm is discussed. The details of the implementation are deferred until Chapters 4 and 5.

3.1 Scaling a Continuous Time Signal

The signal from the hypothetical contact is broadband, limited to 400 Hz, zero-mean, Gaussian and assumed essentially stationary over the observation periods that are used. The signal from the contact is denoted $s_c(t)$. The signal has units of μPa; the sound level emitted by the contact is expressed as SL dB//μPa/1Hz@1m. A

single coefficient is used to indicate all of the channel factors influencing the signal as it propagates through the water and is received by the hydrophone sensor. The coefficient includes spreading and absorption losses and hydrophone conversion. The coefficient $a_{c1}(t)$ denotes the losses *en route* to Sensor 1. The units of the coefficient are such that the product, $a_{c1}(t)s_c(t)$ has units of volts. For the channel from the contact to Sensor 2, the coefficient is $a_{c2}(t)$.

Similarly, $s_i(t)$ indicates the signal from the interferer. In this study, it has the same signal characteristics as the signal from the contact. The channel coefficients for the interferer are denoted by $a_{i1}(t)$ and $a_{i2}(t)$. The contact and interferer signals are independent. Ambient ocean noise picked up by Sensor 1 and Sensor 2 are denoted, respectively, as $s_{n1}(t)$ and $s_{n2}(t)$ and have units of SL dB//μPa/1Hz. The coefficients $b_{n1}(t)$ and $b_{n2}(t)$ cover the hydrophone conversion and have units such that the product $b_{n1}(t)s_{n1}(t)$ has units of volts. The noise signals are uncorrelated with themselves and with each of the contact and interferer signals.

The signals available at the outputs of hydrophones 1 and 2, denoted $s_1(t)$ and $s_2(t)$ respectively, can be expressed by:

$$s_1(t) = a_{c1}(t)s_c(t - d_{c1}(t)) + a_{i1}(t)s_i(t - d_{i1}(t)) + b_{n1}(t)s_{n1}(t) \tag{3.1}$$

and

$$s_2(t) = a_{c2}(t)s_c(t - d_{c2}(t)) + a_{i2}(t)s_i(t - d_{i2}(t)) + b_{n2}(t)s_{n2}(t), \tag{3.2}$$

where $d_{c1}(t)$, $d_{c2}(t)$, $d_{i1}(t)$ and $d_{i2}(t)$ are the time delays that the signals arriving at instant t have already experienced in travelling from the contact to Sensors 1 and 2 and from the interferer to Sensors 1 and 2 respectively.

The autocorrelation functions for the contact and interferer signals are given by:

$$R_{cc}(\tau) = \mathrm{E}[s_c(t)s_c(t - \tau)]$$

and

$$R_{ii}(\tau) = \mathrm{E}[s_i(t)s_i(t-\tau)].$$

Consider the time delay functions used in (3.1) and (3.2). These time delay functions can be approximated with the linear terms of a Taylor series expansion. The expansion will be taken about time $t = T/2$, the midpoint of the observation period of length T. Using the first two terms of the expansion yields the following:

$$d_{c1}(t) \approx v_{c1} \cdot (t - T/2) + D_{c1} \tag{3.3}$$

$$d_{c2}(t) \approx v_{c2} \cdot (t - T/2) + D_{c2} \tag{3.4}$$

$$d_{i1}(t) \approx v_{i1} \cdot (t - T/2) + D_{i1} \quad \text{and} \tag{3.5}$$

$$d_{i2}(t) \approx v_{i2} \cdot (t - T/2) + D_{i2} \tag{3.6}$$

where

$$v_{c1} = \left.\frac{d\, d_{c1}(t)}{dt}\right|_{t=T/2} \tag{3.7}$$

$$v_{c2} = \left.\frac{d\, d_{c2}(t)}{dt}\right|_{t=T/2} \tag{3.8}$$

$$v_{i1} = \left.\frac{d\, d_{i1}(t)}{dt}\right|_{t=T/2} \quad \text{and} \tag{3.9}$$

$$v_{i2} = \left.\frac{d\, d_{i2}(t)}{dt}\right|_{t=T/2} \tag{3.10}$$

and

$$D_{c1} = d_{c1}(T/2) \tag{3.11}$$

$$D_{c2} = d_{c2}(T/2) \tag{3.12}$$

$$D_{i1} = d_{i1}(T/2) \quad \text{and} \tag{3.13}$$

$$D_{i2} = d_{i2}(T/2). \tag{3.14}$$

Thus v_{c1}, for example, represents a constant rate of change in the delay of the signal

travelling from the contact to Sensor 1. The actual delay that the signal received at time $t = T/2$ experienced in reaching Sensor 1 is D_{c1}.

Before correlating, $s_2(t)$ must be scaled to compensate for contact motion during the observation period. The scaling of the time variable of $s_2(t)$ to give $s_{2ts}(t)$ is done using the constant Δv. This is a single linear scaling factor that is proposed to compensate for the rates of change of delay, v_{c1} and v_{c2}, from the contact to both sensors. The scaling is then given by

$$\begin{aligned} s_{2ts}(t) &= s_2(t \cdot (1 + \Delta v)) \\ &= a_{c2}(t(1 + \Delta v))s_c(t(1 + \Delta v) - d_{c2}(t(1 + \Delta v))) + \\ &\quad a_{i2}(t(1 + \Delta v))s_i(t(1 + \Delta v) - d_{i2}(t(1 + \Delta v))) + b_{n2}(t)s_{n2}(t). \quad (3.15) \end{aligned}$$

The assumption will be made that the loss functions $a_{sj}(t)$ and $b_{sj}(t)$, where $s = \{c, i\}$ and $j = \{1, 2\}$, vary only slowly so that

$$\begin{aligned} a_{sj}(t(1 + \Delta v)) &\approx a_{sj}(t) \text{ and} \\ b_{sj}(t(1 + \Delta v)) &\approx b_{sj}(t) \text{ for } 0 \leq t \leq T. \end{aligned}$$

This assumption will be valid if ΔvT is small and if the contact and interferer are not very near either sensor. The noise signal loss coefficients, $b_{sj}(t)$ are not related to contact or interferer motion and can readily be assumed to be constant over the observation period.

Substituting into (3.15) the Taylor expansions from (3.4) and (3.6), and using the above assumption yields

$$\begin{aligned} s_{2ts}(t) &\approx a_{c2}(t)s_c(t(1 + \Delta v) - v_{c2} \cdot t \cdot (1 + \Delta v) - D_{c2} + v_{c2}T/2) \\ &\quad + a_{i2}(t)s_i(t(1 + \Delta v) - v_{i2} \cdot t \cdot (1 + \Delta v) - D_{i2} + v_{i2}T/2) + b_{n2}(t)s_{n2}(t) \\ &= a_{c2}(t)s_c(t(1 + \Delta v)(1 - v_{c2}) - D_{c2} + v_{c2}T/2) \\ &\quad + a_{i2}(t)s_{i2}(t((1 + \Delta v)(1 - v_{i2}) - D_{i2} + v_{i2}T/2) + b_{n2}(t)s_{n2}(t). \end{aligned}$$

The algorithm now calls for the cross correlation of $s_1(t)$ and $s_{2ts}(t)$. Only a time-average cross correlation is possible since only a single sample function is available. The signals are treated as correlation-ergodic. The cross correlation over the observation period T is given by

$$\begin{aligned}
\hat{R}_{12}(\tau) &= \frac{1}{T}\int_0^T s_1(t)s_{2ts}(t-\tau)dt \\
&= \frac{1}{T}\int_0^T s_1(t)s_2((t-\tau)(1+\Delta v))dt \\
&= \frac{1}{T}\int_0^T [a_{c1}(t)s_c(t(1-v_{c1}) - D_{c1} + v_{c1}T/2) \\
&\qquad + a_{i1}(t)s_i(t(1-v_{i1}) - D_{i1} + v_{i1}T/2) + b_{n1}(t)s_{n1}(t)] \times \\
&\qquad [a_{c2}(t-\tau)s_c((t-\tau)(1+\Delta v)(1-v_{c2}) - D_{c2} + v_{c2}T/2) \\
&\qquad + a_{i2}(t-\tau)s_i((t-\tau)(1+\Delta v)(1-v_{i2}) - D_{i2} + v_{i2}T/2) \\
&\qquad + b_{n2}(t)s_{n2}(t)]\,dt.
\end{aligned}$$

Since the averaging time, T, is finite, $\hat{R}_{12}(\tau)$ is a random variable. The expected value of the time-average cross correlation is

$$\begin{aligned}
&\mathrm{E}[\hat{R}_{12}(\tau)] \\
&= \frac{1}{T}\int_0^T a_{c1}(t)a_{c2}(t)\mathrm{E}[s_c(t(1-v_{c1}) - D_{c1} + v_{c1}T/2)s_c((t-\tau)(1+\Delta v)(1-v_{c2}) \\
&\qquad - D_{c2} + v_{c2}T/2)]dt \\
&\quad + \frac{1}{T}\int_0^T a_{i1}(t)a_{i2}(t)\mathrm{E}[s_i(t(1-v_{i1}) - D_{i1} + v_{i1}T/2)s_i((t-\tau)(1+\Delta v)(1-v_{i2}) \\
&\qquad - D_{i2} + v_{i2}T/2)]dt \\
&= \frac{1}{T}\int_0^T a_{c1}(t)a_{c2}(t)R_{cc}(t(1-v_{c1}) - D_{c1} + v_{c1}T/2 - (t-\tau)(1+\Delta v)(1-v_{c2}) \\
&\qquad + D_{c2} - v_{c2}T/2)dt \\
&\quad + \frac{1}{T}\int_0^T a_{i1}(t)a_{i2}(t)R_{ii}(t(1-v_{i1}) - D_{i1} + v_{i1}T/2 - (t-\tau)(1+\Delta v)(1-v_{i2}) \\
&\qquad + D_{i2} - v_{i2}T/2)dt \\
&= \frac{1}{T}\int_0^T a_{c1}(t)a_{c2}(t)R_{cc}(t(v_{c2}-v_{c1} - \Delta v(1-v_{c2})) + \tau(1+\Delta v)(1-v_{c2}) \\
&\qquad + D_{c2} - D_{c1} + (v_{c1}-v_{c2})T/2)dt
\end{aligned}$$

$$+\frac{1}{T}\int_0^T a_{i1}(t)a_{i2}(t)R_{ii}(t(v_{i2}-v_{i1}-\Delta v(1-v_{i2}))+\tau(1+\Delta v)(1-v_{i2}) + D_{i2}-D_{i1}+(v_{i1}-v_{i2})T/2)dt. \tag{3.16}$$

This expression describes the expected result of cross correlating finite length signals picked up by the sensors. It can be seen to consist of the sum of versions of the autocorrelation of the contact signal and of the interferer signal. Since the signals are all independent, any cross correlation of them is zero and so cross correlation terms drop out.

The premise is that a maximum of this expression will occur for a scaling factor, Δv, that best compensates for the contact motion during the time-average period and for a τ correlation shift that corresponds to the TDOA of the contact at time T, the end of the observation period. This premise will be tested by examining the arguments of the autocorrelations $R_{cc}(\cdot)$ and $R_{ii}(\cdot)$.

3.1.1 Ignoring Interference

Consider just the integral of (3.16) containing $R_{cc}(\cdot)$. Since $R_{cc}(0) \geq R_{cc}(\tau)$ for all τ, that term will clearly be maximum if the argument of $R_{cc}(\cdot)$ is zero for all t in the integration interval. The argument is zero for all t if and only if

$$v_{c2}-v_{c1}-\Delta v(1-v_{c2})=0$$

and

$$\tau_c(1+\Delta v)(1-v_{c2})+D_{c2}-D_{c1}+(v_{c1}-v_{c2})T/2=0,$$

where τ_c is the value of the correlation variable τ which forces argument of $R_{cc}(\cdot)$ to zero. It is clear from these two equations that, in the absence of interference, $\mathrm{E}[\hat{R}_{12}(\tau)]$ is maximum if and only if

$$\Delta v = \frac{v_{c2}-v_{c1}}{1-v_{c2}} \tag{3.17}$$

and

$$\tau_c = \frac{D_{c1} - D_{c2} + (v_{c2} - v_{c1})T/2}{\left[1 + \frac{v_{c2}-v_{c1}}{1-v_{c2}}\right](1 - v_{c2})} = \frac{D_{c1} - D_{c2} + (v_{c2} - v_{c1})T/2}{1 - v_{c1}} \tag{3.18}$$

According to (3.18), in the absence of interference, the global maximum of the cross correlation occurring at a correlation shift, τ_c, is expected to indicate a biased estimate of the TDOA for time T.

3.1.2 Considering Interference

A similar argument can be made for the $R_{ii}(\cdot)$ term in (3.16) in the absence of contact signal. A different global maximum, at a different Δv, will indicate an estimate of the TDOA at time T for the interferer when

$$\Delta v = \frac{v_{i2} - v_{i1}}{1 - v_{i2}}$$

and

$$\tau_i = \frac{D_{i1} - D_{i2} + (v_{i2} - v_{i1})T/2}{\left[1 + \frac{v_{i2}-v_{i1}}{1-v_{i2}}\right](1 - v_{i2})} = \frac{D_{i1} - D_{i2} + (v_{i2} - v_{i1})T/2}{1 - v_{i1}}.$$

To differentiate, the scaling factor which maximizes the correlation for the contact is denoted Δv_c and the factor which maximizes the correlation for the interferer, Δv_i.

If both contact and interferer signals are present then the expected value of the cross correlation (3.16) contains contributions from both integrals. However, if Δv_c is significantly different from Δv_i then both the integrals will not produce a large correlation peak for any one value of Δv. That is, a proposed Δv close to Δv_c will produce a large correlation peak for the contact. That same Δv will not produce a strong correlation for the interferer.

If the integration time is such that

$$|\Delta v_c - \Delta v_i| \cdot T \gg \frac{1}{\text{bandwidth}} \tag{3.19}$$

then, for a Δv near Δv_c, the interferer will affect the peak of the contact correlation as if it were uncorrelated noise with the same total power. Then the interferer will have been largely suppressed. This will be explained in Chapter 4 in the discussion of smearing of a correlation.

The converse situation is also true. A scaling factor near the optimum for the interferer will suppress the contact. If Δv_c and Δv_i are too similar, then it may not be possible to resolve the TDOA for the contact in the presence of strong interference within a reasonable time.

Each proposed Δv produces a correlogram. The set of correlograms from all attempted Δv values produces a surface, with the set of Δv values forming one axis and the correlation variable τ forming the other. The height of the surface is the value of $\mathrm{E}[\hat{R}_{12}(\tau)]$. The surface will, in general, show two distinct peaks, and is frequently called an ambiguity surface. Resolving the ambiguity of the two peaks requires *a priori* knowledge in order to identify one peak as corresponding to the contact and the other to the interferer.

The development so far has been for the case where every point of $s_2(t)$ is scaled along the time axis. While it will result in high quality correlations, this has the disadvantage that it is computationally intensive to implement. The next section will develop more computationally efficient expressions for Δv_c and TDOA estimate for the contact.

3.2 Scaling the Midpoint of Signal Segments

This section will develop similar mathematical expressions to those developed in the last section, but without scaling every data point in the Sensor 2 signal. Instead, unscaled segments of the two signals will be used in short time-average correlations. The integration time is broken up into intervals that are sufficiently short that each segment from $s_2(t)$ is essentially a coherent replica of the corresponding segment from $s_1(t)$. The benefit of this approach is a reduction in the number of computations

required for scaling.

The expressions for the signals received from the two sensors $s_1(t)$ and $s_2(t)$, for the autocorrelations $R_{cc}(\tau)$ and $R_{ii}(\tau)$, remain the same as described in Section 3.1. Since the ambient noise terms and the signal loss coefficients were shown not to contribute to the expressions for TDOA or Δv, they are omitted from the start. Additive zero-mean independent noise sources do not affect the mean cross correlation output [8].

$$s_1(t) = s_c(t - d_{c1}(t)) + s_i(t - d_{i1}(t)) \tag{3.20}$$

$$s_2(t) = s_c(t - d_{c2}(t)) + s_i(t - d_{i2}(t)) \tag{3.21}$$

$$R_{cc}(\tau) = \mathrm{E}[s_c(t)s_c(t-\tau)]$$

$$R_{ii}(\tau) = \mathrm{E}[s_i(t)s_i(t-\tau)].$$

Assuming $s_1(t)$ is partitioned into N time segments with the partition $\mathcal{P} = \{0, t_1, t_2, t_3, \ldots, T\}$, the time axis scaled Sensor 2 signal is given approximately by

$$s_{2ts}(t) \approx s_2(t + \gamma_k \Delta v), \quad \text{for } t_k \leq t \leq t_{k+1}, \quad k = 0 \ldots N-1,$$

where

$$\gamma_k = \frac{t_k + t_{k+1}}{2}.$$

In this way scaling is done for the center point of each segment, but the data points within each segment are not scaled.

The delay functions in (3.20) and (3.21) are approximated with the linear terms of a Taylor Series expansion as before. This time, however, it must be considered that there will be a number of short-time correlations, and slightly different delay function expansions for each. The expansion is taken about γ_k, the midpoint of each segment of the partition, $\mathcal{P}$, as follows:

$$
\begin{aligned}
d_{c1}(t) &\approx v_{c1}^k \cdot (t-\gamma_k) + D_{c1}^k \\
d_{c2}(t) &\approx v_{c2}^k \cdot (t-\gamma_k) + D_{c2}^k \\
d_{i1}(t) &\approx v_{i1}^k \cdot (t-\gamma_k) + D_{i1}^k \quad \text{and} \\
d_{i2}(t) &\approx v_{i2}^k \cdot (t-\gamma_k) + D_{i2}^k
\end{aligned}
$$

where

$$
\begin{aligned}
v_{c1}^k &= \left.\frac{d\, d_{c1}(t)}{dt}\right|_{t=\gamma_k} \\
v_{c2}^k &= \left.\frac{d\, d_{c2}(t)}{dt}\right|_{t=\gamma_k} \\
v_{i1}^k &= \left.\frac{d\, d_{i1}(t)}{dt}\right|_{t=\gamma_k} \quad \text{and} \\
v_{i2}^k &= \left.\frac{d\, d_{i2}(t)}{dt}\right|_{t=\gamma_k}
\end{aligned}
$$

and

$$
\begin{aligned}
D_{c1}^k &= d_{c1}(\gamma_k) \\
D_{c2}^k &= d_{c2}(\gamma_k) \\
D_{i1}^k &= d_{i1}(\gamma_k) \quad \text{and} \\
D_{i2}^k &= d_{i2}(\gamma_k),
\end{aligned}
$$

where the superscript k refers to the k^{th} segment of $\mathcal{P}$.

Momentarily ignoring the interferer, the time-average cross correlation of $s_1(t)$ and $s_{2ts}(t)$ is given by

$$
\begin{aligned}
&\hat{R}_{12}(\tau) \\
&= \frac{1}{T} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} s_1(t) s_{2ts}(t-\tau) dt
\end{aligned}
$$

$$\begin{aligned}
&\approx \frac{1}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} s_1(t)s_2(t-\tau+\gamma_k\Delta v)dt \qquad (3.22)\\
&= \frac{1}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} s_c(t-d_{c1}(t))s_c(t-\tau+\gamma_k\Delta v-d_{c2}(t-\tau+\gamma_k\Delta v))dt\\
&= \frac{1}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} s_c(t(1-v_{c1}^k)-D_{c1}^k+v_{c1}^k\gamma_k)\\
&\qquad \times s_c((1-v_{c2}^k)(t-\tau+\gamma_k\Delta v)+v_{c2}^k\gamma_k-D_{c2}^k)dt
\end{aligned}$$

where $k = 0...N-1$, $t_0 = 0$ and $t_N = T$. Thus (3.22) is seen to be the sum of a number of short time-average correlations.

Still ignoring interference, the expected value of this time average cross correlation function is

$$\begin{aligned}
&\mathrm{E}[\hat{R}_{12}(\tau)]\\
&\approx \frac{1}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} \mathrm{E}[s_c(t(1-v_{c1}^k)-D_{c1}^k+v_{c1}^k\gamma_k)\\
&\qquad s_c((1-v_{c2}^k)(t-\tau+\gamma_k\Delta v)+v_{c2}^k\gamma_k-D_{c2}^k)]dt\\
&= \frac{1}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} R_{cc}(t(1-v_{c1}^k)-D_{c1}^k+v_{c1}^k\gamma_k\\
&\qquad -[(1-v_{c2}^k)(t-\tau+\gamma_k\Delta v)+v_{c2}^k\gamma_k-D_{c2}^k])dt\\
&= \frac{1}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} R_{cc}(t(v_{c2}^k-v_{c1}^k)-\gamma_k\Delta v(1-v_{c2}^k)+\tau(1-v_{c2}^k)\\
&\qquad +(v_{c1}^k-v_{c2}^k)\gamma_k-D_{c1}^k+D_{c2}^k)dt. \qquad (3.23)
\end{aligned}$$

Since it is assumed that the contact maintains its velocity throughout the observation period, $v_{c1}^k = v_{c1}$ for all k. Similarly, $v_{c2}^k = v_{c2}$, $v_{i1}^k = v_{i1}$ and $v_{i2}^k = v_{i2}$ for all k.

Again, (3.23) is maximum when the argument of $R_{cc}(\cdot)$ is zero for all t in each integration integral $[t_k, t_{k+1}]$. This is the case if and only if

$$t(v_{c2}-v_{c1})-\gamma_k\Delta v(1-v_{c2}) = 0$$

and

$$\tau(1-v_{c2})+(v_{c1}-v_{c2})\gamma_k-D_{c1}^k+D_{c2}^k = 0.$$

From these two conditions, we see that, in the absence of interference, $\mathrm{E}[\hat{R}_{12}(\tau)]$ is maximum when

$$\Delta v = \frac{(v_{c2} - v_{c1})t}{(1 - v_{c2})\gamma_k} \tag{3.24}$$

and

$$\tau_c^k = \frac{D_{c1}^k - D_{c2}^k + (v_{c2} - v_{c1})\gamma_k}{1 - v_{c2}}, \tag{3.25}$$

where τ_c^k is the value of the correlation variable τ which maximizes $R_{cc}(\cdot)$ in the interval $[t_k, t_{k=1}]$.

It can readily be seen from (3.24) that any Δv is completely accurate only for the instant when $t = \gamma_k$, and that for other values of t in the integration interval the scaling will be in error. If, however, the integration interval is sufficiently small then γ_k is close to t. With this assumption the expression for Δv reduces to

$$\Delta v = \frac{v_{c2} - v_{c1}}{1 - v_{c2}}$$

which is the same as was derived for the continuous time scaling case in the last section.

The expression for τ_c^k has a couple of differences from the expression for τ_c derived in the last section. Firstly, the denominator is $(1 - v_{c2})$ here rather than $(1 - v_{c1})$. Secondly, there is a τ_c^k estimate for each segment of the partition, $\mathcal{P}$. When the proposed Δv is equal to Δv_c (the optimal value for the contact), all the estimates τ_c^k will be nearly the same, namely τ_c. In that case the short-time correlations sum in a constructive fashion. Further, the time difference of arrival can be determined for any of the segment midpoints, γ_k. It is most useful to determine the most recent time difference—that at the midpoint of the last segment, γ_{N-1}:

$$\tau_c = \frac{D_{c1}^{N-1} - D_{c2}^{N-1} + (v_{c2} - v_{c1})\gamma_{N-1}}{1 - v_{c2}}. \tag{3.26}$$

The above discussion was developed ignoring any interferer. A similar set of expressions for Δv and τ can be developed considering only the interferer and ignoring

the contact. The actual expression for $\mathrm{E}[\hat{R}_{12}(\tau)]$ includes the sum of the autocorrelations for both the contact and the interferer, since cross correlations of independent signals are zero. As before, the scaling factor that yields a maximum for the contact is denoted Δv_c and that which yields a maximum for the interferer is Δv_i.

Each proposed Δv yields a correlogram which forms part of an ambiguity surface. The interferer and the contact each contribute a peak in that surface. To adequately resolve the two peaks, the total integration time must be such that

$$|\Delta v_c - \Delta v_i| \cdot T \gg \frac{1}{\text{bandwidth}}.$$

An additional condition that has been imposed is that each of the correlation times must be short. The short term correlations will not suffer significant degradation if, during the integration time, the delay does not vary by more than the correlation time of the source signal [5].

Selection of the total integration time, T, and of the segment correlation length, $[t_k, t_{k+1}]$, will be discussed at more length in Chapter 4.

3.3 Implementing the Select-Correlate-Sum Expressions

The mathematical basis for the Select-Correlate-Sum algorithm was developed in the preceding sections. This section introduces the application of those expressions in an algorithm to find the TDOA for a moving contact.

In a practical application, all that is available for processing are finite length sequences from the output of the two hydrophones. The pertinent equation for the implementation is (3.22) reproduced here for convenience:

$$\hat{R}_{12}(\tau) \approx \frac{1}{T} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} s_1(t) s_2(t + \gamma_k \Delta v - \tau) dt$$

where $s_1(t)$ and $s_2(t)$ are the signals received by Sensors 1 and 2 respectively; where $s_1(t)$ is partitioned in time with the partition $\mathcal{P} = \{0, t_1, t_2, t_3, \ldots, T\}$, with $t_0 = 0$

and $t_N = T$; and where the midpoints of the partition segments are given by

$$\gamma_k = \frac{t_k + t_{k+1}}{2}, \quad k = 0 \ldots N-1.$$

For each proposed scaling factor, Δv, the midpoint of each segment of $s_1(t)$ in the partition $\mathcal{P}$ is scaled to determine the midpoint of a matching segment from $s_2(t)$. Each pair of matching segments are cross correlated, and the N correlation results are averaged. Each Δv results in slightly different matching segments from $s_2(t)$ being selected and therefore yields a slightly different correlogram. The set of all correlograms forms the ambiguity surface, a plot of the correlation magnitude versus Δv versus τ. The two peaks of the ambiguity surface, interpolated if necessary, correspond to the contact and the interferer. Some *a priori* knowledge is required to identify the peaks. In the proposed scenario, it is known that Δv_i is zero since the interferer is collinear with the sensors. The delay for the interferer, $D_{i1} - D_{i2}$ is known to be the maximum possible given the separation of the two sensors. Finally, since the contact is assumed to be following the surface ship, it can be assumed that the contact's speed is close the the speed of the surface ship, which would be known.

3.3.1 Correcting Bias in the Estimates

The ambiguity surface parameters indicated by the location of the contact peak correspond to estimates of Δv_c and τ_c discussed in the preceding sections.

Δv_c was predicted to be

$$\Delta v_c = (v_{c2} - v_{c1})/(1 - v_{c2}).$$

The location of the peak found in the ambiguity surface indicates an estimate of Δv_c, denoted $\Delta \hat{v}_c$. Since the motion of the contact between the two sensors dictates that $v_{c1} = -v_{c2}$, and that the required compensating factor is $v_{c2} - v_{c1}$, $\Delta \hat{v}_c$ can be used

to provide estimates of v_{c1} and v_{c2} as follows:

$$\hat{v}_{c2} = -\hat{v}_c 1 = \frac{\Delta\hat{v}_c}{(2 + \Delta\hat{v}_c)}.$$

These estimates are then used to correct the bias in the estimate of the time difference, τ_c.

The value of the correlation variable corresponding to the location of the contact peak in the ambiguity surface is denoted $\hat{\tau}_c$ which is an estimate of τ_c predicted in (3.26). A number of operations are required in order to manipulate $\hat{\tau}_c$ into a useful estimate of the TDOA for the contact at the end of the observation period.

The estimate $\hat{\tau}_c$ is subject to several sources of error. It is a time average estimate. The presence of the interferer introduces a bias. The methods themselves introduce other biases.

In Section 3.2 the aggregate correlation peak was expected at

$$\tau_c = \frac{D_{c1}^{N-1} - D_{c2}^{N-1} + (v_{c2} - v_{c1})\gamma_{N-1}}{1 - v_{c2}}.$$

Even for long observation periods τ_c will be a biased estimator of $D_{c1} - D_{c2}$ due to the introduction by the Select-Correlate-Sum algorithm of the denominator term $(1 - v_{c2})$. This denominator is different than the expression derived for the fully scaled method (3.18) which is assumed to be more accurate.

A first bias correction towards getting a useful TDOA estimate from the peak location is to alleviate that difference by multiplying $\hat{\tau}_c$ by $(1-\hat{v}_{c1})/(1-\hat{v}_{c2})$. Equation (3.18) also includes a term to correct for the fact that the delay difference is desired for the end of the observation period, γ_{N-1}, but the correlation yields a delay difference for the midpoint of the correlation interval. Both of these corrections are contained in the expression:

$$\Delta D = \hat{\tau}_c \left(\frac{1 - \hat{v}_c}{1 - \hat{v}_{c2}} \right) (1 - \hat{v}_{c2}) - (\hat{v}_{c2} - \hat{v}_{c1})\gamma_{N-1}.$$

ΔD will generally be biased due to the presence of correlation lobes from the interferer. This is in fact the principal source of error. If, however, the interference is largely suppressed through the algorithm, then the bias will be small.

D_{c1} and D_{c2} were defined to be the delays experienced by the sound from the contact *received* at Sensor 1 and Sensor 2, respectively, at time $T/2$, the midpoint of the observation period. D_{c1}^{N-1} and D_{c2}^{N-1} were defined to be similar quantities, but for the time given by γ_{N-1}, the midpoint of the final segment in the partition of $s_1(t)$. ΔD is, therefore, an estimate of the difference in those delays. However, these signals have spent time travelling to the sensors and were emitted from the contact at separate times earlier than $T/2$ (or than γ_{N-1}). Since the contact is moving, it was in a different location at each of those earlier times.

What is really the desired quantity, however, is the time difference of arrival of sound *emitted* from the contact at the end of the observation period T (or the last segment midpoint γ_{N-1}).

The estimate of the desired TDOA will be denoted as $\Delta D'$. $\Delta D'$ can be calculated from the time difference ΔD. The derivation can be found in Appendix A, the result of which is used as follows:

$$\Delta D' = \Delta D \left(1 - \frac{\hat{s}^2}{c^2}\right) - \frac{\hat{s}l}{c^2} \tag{3.27}$$

where l is the distance between the two sensors, c is the speed of sound in the ocean, and where $\hat{s}$ is an estimate of the velocity of the contact. The velocity estimate is given by

$$\hat{s} = c\ \hat{v}_{c2}.$$

The result $\Delta D'$ is the estimate of the desired quantity, the time difference of arrival for sound emitted from the contact at the end of the observation period. This provides the most current clue as to the location of the contact.

A number of important issues still need to be examined: The selection of the

length of the segments in the partition $\mathcal{P}$; the length of the observation time, T; and the problem of aligning the correlations from each segment. These details will be each be examined in Chapter 4.

4. Selection of Key Parameters

In this chapter a number of issues, which were raised in the earlier chapters, are examined in more detail. Some issues have a bearing on the selection of key parameters used in the implementation. Other issues relate to choices that need to be made in the implementation. These issues include:

- The length of the partition segments necessary to avoid smearing effects.
- The length of the total observation time necessary to effectively suppress the interferer.
- The size of the change in scaling factor between successive guesses in the search for the proper scaling factor to compensate for contact motion.
- The alignment problem introduced when the extraction selected by a particular scaling factor does not align with existing data samples.

Each of these issues will be addressed later in this chapter. In the next chapter, the discussion of the simulation of the algorithm will rely upon the reasoning in this chapter for the selection of key simulation parameters.

4.1 Effect of Contact Motion

The object of this section is to show the effect on the cross correlation of contact motion during the observation period. It will show that the effect of contact motion can be expressed as the convolution of the autocorrelation function with a pulse of unit area and appropriate duration. The effect of contact motion can then be illustrated graphically by convolving an ideal correlation function with pulses of varying widths. The effect appears as a smearing of the correlogram, seen by a flattening of the

peak and a broadening of the main lobe. Smearing makes locating the peak of the correlogram more susceptible to noise.

Consider a very simple geometry of two sensors and a contact moving between them, away from Sensor 1 and towards Sensor 2. Since the effect of contact motion is desired, of course no scaling to compensate for contact motion is done. The signal received at Sensor 1 is $s_1(t)$ and that received at Sensor 2 is $s_2(t)$. Using $s_1(t)$ as a reference, $s_2(t)$ can be expressed as

$$s_2(t) = s_1(t(1+a) - D)$$

where a is the rate of change of the time difference of arrival (TDOA) and D is the initial TDOA. The rigour used in the last chapter for the various estimates of TDOA is not necessary in this discussion.

The time-average cross correlation over time T of the signals from Sensor 1 and Sensor 2 is given by

$$\begin{aligned} \hat{R}_{12}(\tau) &= \frac{1}{T}\int_0^T s_1(t)s_2(t-\tau)dt \\ &\quad \frac{1}{T}\int_0^T s_1(t)s_1((t-\tau)(1+a) - D)dt \\ &\quad \frac{1}{T}\int_0^T s_1(t)s_1(t+at-\tau(1+a) - D)dt. \end{aligned}$$

The expected value of this time-average cross correlation is

$$\begin{aligned} \mathrm{E}[\hat{R}_{12}(\tau)] &= \frac{1}{T}\int_0^T \mathrm{E}[s_1(t)s_1(t+at-\tau(1+a) - D)]dt \\ &= \frac{1}{T}\int_0^T R_{11}(\tau(1+a) - D - at)dt. \end{aligned}$$

Changing the variable of integration with $\lambda = at$ results in

$$\mathrm{E}[\hat{R}_{12}(\tau)] = \int_0^{aT} R_{11}(\tau(1+a) - D - \lambda)\frac{1}{aT}d\lambda. \tag{4.1}$$

Equation (4.1) can be expressed as

$$\mathrm{E}[\hat{R}_{12}(\tau)] = R_{11}(\tau(1+a) - D) \star p(\tau), \tag{4.2}$$

where $\star$ denotes the convolution operation and

$$p(\tau) = \begin{cases} \frac{1}{aT} & \text{for } 0 \leq \tau \leq aT \\ 0 & \text{otherwise.} \end{cases}$$

Equation (4.1) gives some insight into the effect of contact velocity on the shape of the cross correlation. In particular, it gives some insight into the shape of $\mathrm{E}[\hat{R}_{12}(\tau)]$ in the vicinity of its peak. In the absence of any contact velocity, the parameter a is zero, $p(\tau)$ becomes an impulse function, and the expected value of the cross correlation is simply the autocorrelation $R_{11}(\tau)$ shifted by the TDOA, D. When the contact is moving, the shape of $\mathrm{E}[\hat{R}_{12}(\tau)]$ is affected in two ways. Firstly, the velocity reshapes $R_{11}(\tau)$ by time compression or expansion by a factor of $1 + a$. Secondly, this time scaled $R_{11}(\tau)$ is then smeared through the convolution with a pulse whose width is proportional to the velocity.

Both the time scaling action and the smearing action can have the deteriorating effect of reducing the sharpness of the peak in $\mathrm{E}[\hat{R}_{12}(\tau)]$. The time scaling action will have a deteriorating effect when $1 + a > 1$, however, the smearing action of convolution with the pulse always causes deterioration by smoothing the correlation peak and making its detection more susceptible to noise. The relative effects of the time scaling and convolution depend on the observation period, T, relative to the bandwidth, B. For a large time bandwidth product, i.e., $BT \gg 1$, the smearing effect dominates.

To illustrate the effect of the convolution, an ideal low pass spectrum is used for the source. The corresponding autocorrelation function is a $sin(x)/x$ function. The $sin(x)/x$ function will be smeared by a pulse with unit area and width aT. The product aT is the amount of the change in the TDOA over the observation period.

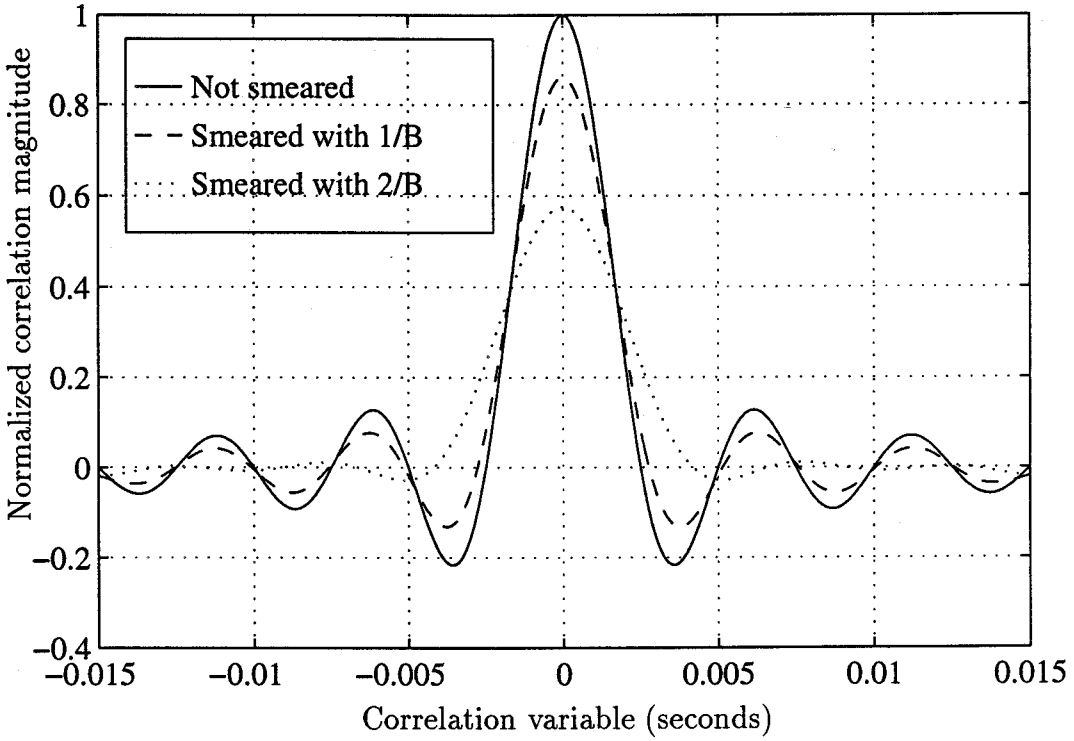

Figure 4.1 Autocorrelation of $\sin\pi Bt/\pi Bt$ convolved with rectangular pulse of width $1/B$ and $2/B$; $B =$ 400 Hz.

Kuhn et al. [5] say that the correlation will not be degraded significantly if the delay changes by less than the correlation time. The correlation time is approximately equal to $1/B$. Therefore the sharpness of the peak is not changed significantly if $aT < 1/B$.

Figure 4.1 shows the effect of smearing for aT equal to $1/B$ and $2/B$. The bandwidth of the source is $B = 400$ Hz and the autocorrelation function is $\sin\pi Bt/\pi Bt$. In agreement with Kuhn, a pulse width of $1/B$ shows only slight smearing of the correlation function as evidenced by an attenuation of the peak magnitude and broadening of the main lobe. When convolved with a pulse of width $2/B$, more significant smearing of the correlation occurs. [Also note the dramatic effect of smearing on the magnitude of side lobes. This effect is crucial to interferer suppression which will be discussed later.]

As would be expected, this figure is very similar to Figure 2.4 which was generated by cross correlating signals from a contact moving at several different speeds.

While Figure 4.1 illustrates in a general way the smearing effect of uncompensated contact motion on the shape of the correlogram, it does not reveal the specific implications of smearing on the selection of the extraction length and the observation period. Selection of these key parameters is examined in the next sections.

4.2 Selecting the Extraction Length

In order to avoid the effect of correlation smearing and make a good estimate of the time difference of arrival, compensation must be made for the motion of the contact. Two ways of compensating for contact motion have been discussed. In the first way, every point of $s_2(t)$ is scaled on the time axis. The scaling and resampling required in this method are computationally expensive. In the Select-Correlate-Sum method, a series of extractions from $s_2(t)$ is correlated with $s_1(t)$. The extractions are selected as follows: First, $s_1(t)$ is partitioned into N segments, and the midpoints of the segments are located. Then, using the scaling factor, these midpoints are scaled to another set of points that will be the midpoints of matching extractions from $s_2(t)$. The identified extractions from $s_2(t)$ are then individually correlated with $s_1(t)$. While the expense of scaling and resampling each point of $s_1(t)$ is avoided, compensation for contact motion is done through the selection of the extraction midpoints.

Because the points within the extractions have not been scaled, some smearing of the correlation is still expected. If, however, the length of each extraction is short, the degree of smearing will be small. This section will examine the selection of the length of extractions.

Following a development similar to that used in Section 3.2, ignoring interference and noise, and using loss coefficients equal to one, the signals from the hydrophone sensors can be expressed as

$$\begin{aligned} s_1(t) &= s_c(t - d_{c1}(t)) \\ s_2(t) &= s_c(t - d_{c2}(t)) \end{aligned}$$

where $d_{c1}(t)$ and $d_{c2}(t)$ are the time-varying time delays experienced by the signals in travelling from the contact to Sensor 1 and Sensor 2, respectively. The delay $d_{c1}(t)$ can be represented by a collection of Taylor series expansions. For each segment in the partition, $\mathcal{P}$, of $s_1(t)$, the delay $d_{c1}(t)$ is represented as a Taylor series expansion about the midpoint of the segment. For the purpose of analysis, the length of each segment is taken to be ΔT. Also for the purpose of analysis, and without loss of generality, only the mathematics of the first segment, which covers the time interval $t \in (0, \Delta T)$, will be shown. In doing this,

$$d_{c1}(t) \approx v_{c1} \cdot (t - \frac{\Delta T}{2}) + D_{c1}; \quad 0 \leq t < \Delta T$$

where

$$v_{c1} = \left. \frac{d\, d_{c1}(t)}{dt} \right|_{t=\Delta T/2}$$

and

$$D_{c1} = d_{c1}(\frac{\Delta T}{2}).$$

The delay $d_{c2}(t)$ can be similarly expanded.

The received signals are then

$$s_1(t) = s_c(t(1 - v_{c1}) - D_{c1} + v_{c1}\frac{\Delta T}{2}); \quad 0 \leq t < \Delta T$$

and

$$s_2(t) = s_c(t(1 - v_{c2}) - D_{c2} + v_{c2}\frac{\Delta T}{2}); \quad 0 \leq t < \Delta T.$$

The time-average cross correlation is given by:

$$\begin{aligned}
&\hat{R}_{12}(\tau) \\
&= \frac{1}{\Delta T}\int_{-\Delta T/2}^{\Delta T/2} s_1(t)s_2(t-\tau)dt \\
&= \frac{1}{\Delta T}\int_{-\Delta T/2}^{\Delta T/2} s_c(t(1-v_{c1}) - D_{c1} + v_{c1}\frac{\Delta T}{2})s_c((t-\tau)(1-v_{c2}) - D_{c2} + v_{c2}\frac{\Delta T}{2})dt,
\end{aligned}$$

and the expected value of the time-average cross correlation is

$$\begin{aligned}
&\mathrm{E}[\hat{R}_{12}(\tau)] \\
&= \frac{1}{\Delta T}\int_{-\Delta T/2}^{\Delta T/2} \mathrm{E}[s_c(t(1-v_{c1}) - D_{c1} + v_{c1}\frac{\Delta T}{2})s_c((t-\tau)(1-v_{c2}) - D_{c2} + v_{c2}\frac{\Delta T}{2})]dt \\
&= \frac{1}{\Delta T}\int_{-\Delta T/2}^{\Delta T/2} R_{cc}(t(v_{c2}-v_{c1}) + \tau(1-v_{c2}) - D_{c1} + D_{c2} + \frac{\Delta T}{2}(v_{c1}-v_{c2}))dt.
\end{aligned}$$

Changing the variable of integration with $\lambda = t(v_{c2} - v_{c1})$ results in

$$\begin{aligned}
&\mathrm{E}[\hat{R}_{12}(\tau)] = \\
&\quad \frac{1}{P}\int_{-P/2}^{P/2} R_{cc}(\lambda + \tau(1-v_{c2}) - D_{c1} + D_{c2} + \frac{\Delta T}{2}(v_{c1}-v_{c2}))d\lambda
\end{aligned} \tag{4.3}$$

where $P = \Delta T(v_{c2} - v_{c1})$. In this form (4.3) is recognizable as the expression obtained through the convolution of an autocorrelation with a pulse as follows:

$$\mathrm{E}[\hat{R}_{12}(\tau)] = R_{cc}(\tau(1-v_{c2}) - D_{c1} + D_{c2} + \frac{\Delta T}{2}(v_{c1}-v_{c2})) \star p(\tau)$$

where the operator $\star$ denotes convolution and

$$p(\tau) = \begin{cases} \frac{1}{\Delta T(v_{c2}-v_{c1})} & \text{for } -\frac{\Delta T}{2}(v_{c2}-v_{c1}) \le \tau \le \frac{\Delta T}{2}(v_{c2}-v_{c1}) \\ 0 & \text{otherwise.} \end{cases}$$

The smearing of the correlogram, resulting from contact motion during the integration time ΔT, can be effectively modelled with convolution of the autocorrelation term with the pulse $p(\tau)$. The short-term correlations will not suffer significant smearing degradation if the delay does not vary by more than the correlation time of the source signal during the correlation integration time ΔT. The change in delay is

given by the pulse width. The pulse has width $\Delta T(v_{c2} - v_{c1})$ and area 1. Therefore, to avoid serious smearing of the correlogram,

$$|\Delta T(v_{c2} - v_{c1})| < \frac{1}{B}. \tag{4.4}$$

Equation (4.4) governs the length of the segments in the partition $\mathcal{P}$. It can be seen that if the contact is not moving, then $(v_{c2} - v_{c1}) = 0$ and the length of ΔT is not an issue. If the speed of the contact is very fast, then the TDOA changes quickly and the segment length must be correspondingly shorter.

In the scenario described in Chapter 2, the bandwidth is about 400 Hz and the speed of the contact can be expected to vary between about 5 and 15 meters per second. This implies that ΔT must be less than about 370 ms and 120 ms, respectively. At a rate of 8000 samples per second, these times correspond to correlations of just 2960 samples and 960 samples.

Figure 4.2 shows an example of Select-Correlate-Sum processing on a simulated contact. For this example, there is no interferer, the contact is moving at -15 meters per second, and the observation period has been partitioned into twenty segments. Extractions of lengths 80, 100, and 150 ms were used and the cumulative correlations are plotted. Using a 100 ms extraction length provides additional peak amplitude over an 80 ms extraction length. An extraction length of 150 ms shows significant peak smearing. Figure 4.3 shows the same example with extraction lengths of 100 ms and 120 ms, at which smearing is becoming noticeable. This figure validates (4.4).

4.3 Selecting the Observation Period Length

While the SCS algorithm uses scaling to compensate for contact motion and avoid smearing of the correlation of the contact, the algorithm also exploits smearing to suppress the interferer. As was pointed out in Chapter 3, a proposed scaling factor Δv that maximizes the correlation for the contact will cause a poor correlation for the interferer, if the contact and the interferer have different velocities. This is because the

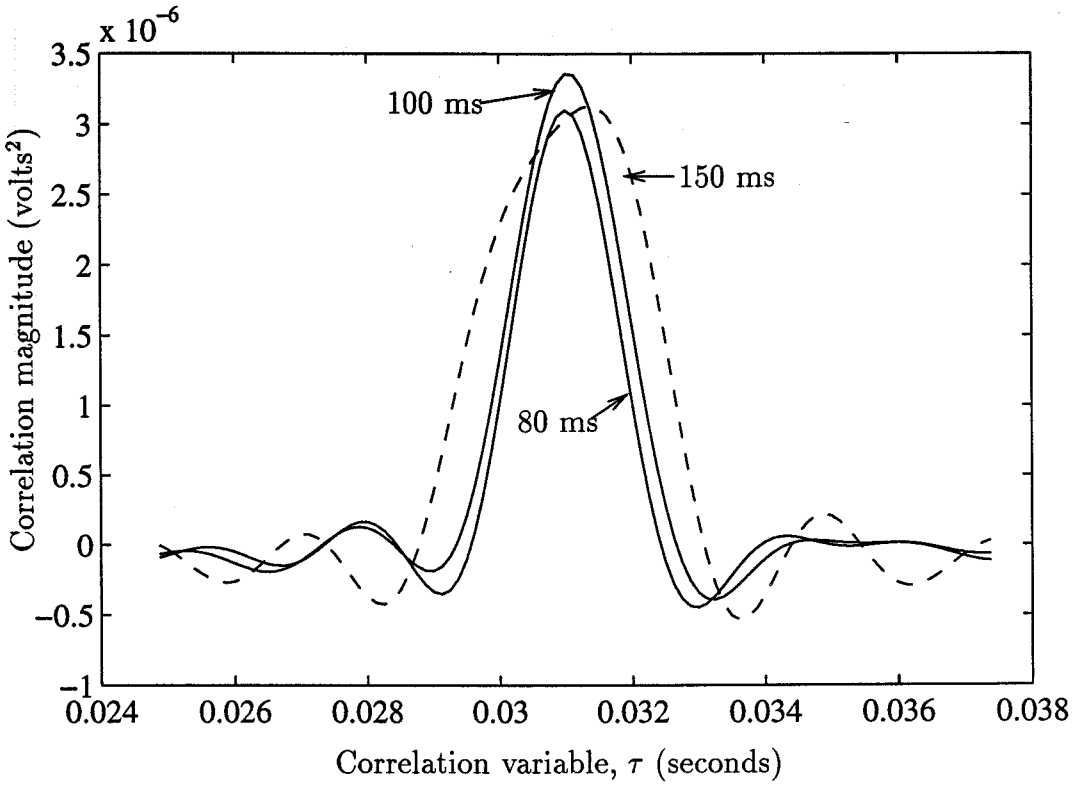

Figure 4.2 Select-Correlate-Sum method correlation of test signal using extraction lengths of 80 ms, 100 ms and 150 ms.

scaling that compensates for the contact motion at the same time causes a smearing of the correlation for the interferer. This section examines the relationship between the length of the observation period and the degree of suppression of the interferer.

The approach that will be used follows the development of Section 3.1. Using the expression for Δv_c which maximizes the correlation for the contact (in the absence of interference), the effect on the interferer will be examined.

Equation (3.17) indicates that the value Δv_c is one of the conditions which maximizes, for the contact, the expected value of the cross correlation of signals from Sensors 1 and 2. This Δv_c is now substituted into the integration term of the interferer in (3.16) as follows:

$$\begin{aligned} &\text{TERM 2 of (3.16)} \\ &= \frac{1}{T}\int_0^T R_{ii}(t(v_{i2}-v_{i1}-\Delta v_c(1-v_{i2}))+\tau(1+\Delta v_c)(1-v_{i2}) \end{aligned}$$

$$+ D_{i2} - D_{i1} + (v_{i1} - v_{i2})T/2)dt.$$

Change the variable of integration with $\lambda = t(v_{i2} - v_{i1} - \Delta v_c(1 - v_{i2}))$ to get

$$\text{TERM 2 of (3.16)} = \int_0^P R_{ii}(\lambda + \tau(1 + \Delta v_c)(1 - v_{i2}) + D_{i2} - D_{i1} + (v_{i1} - v_{i2})T/2)\frac{1}{P}d\lambda$$

where $P = T(v_{i2} - v_{i1} - \Delta v_c(1 - v_{i2}))$. This can be seen to be the expression for the convolution of two functions

$$\text{TERM 2 of (3.16)} = R_{ii}(\tau(1 + \Delta v_c)(1 - v_{i2}) + D_{i2} - D_{i1} + (v_{i1} - v_{i2})T/2) \star p(\tau)$$

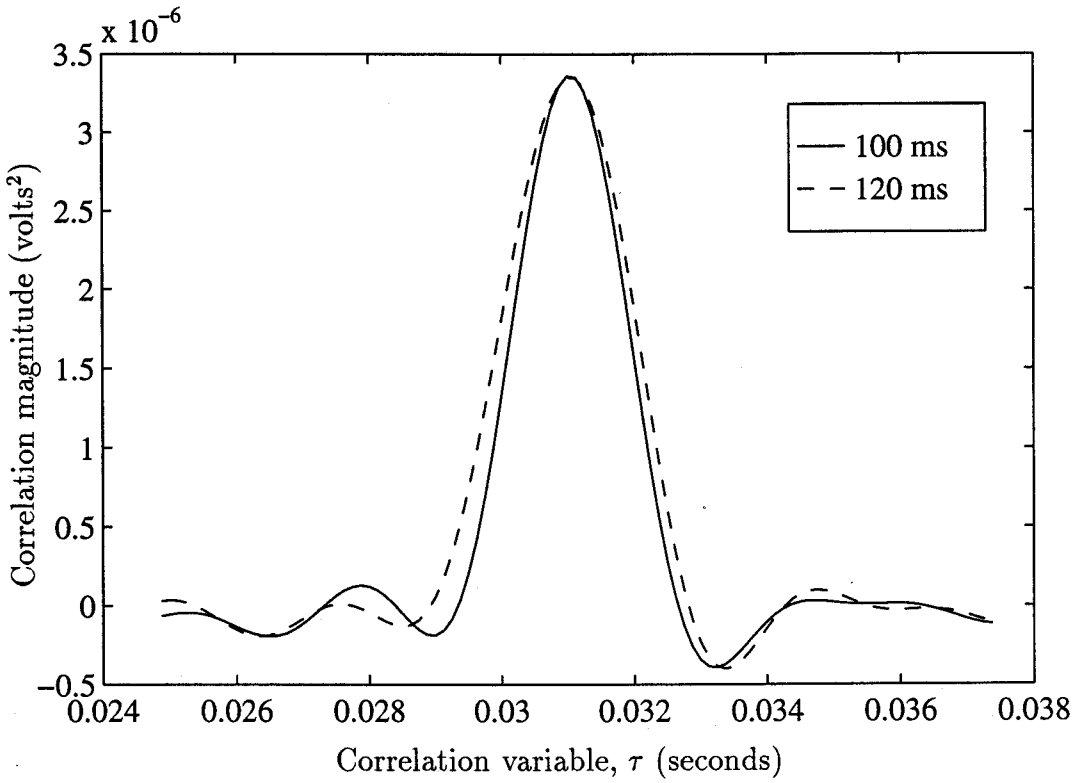

Figure 4.3 Select-Correlate-Sum method correlation of test signal using extraction lengths of 100 ms and 120 ms.

where

$$p(\tau) = \begin{cases} \frac{1}{P} = \frac{1}{T(v_{i2}-v_{i1}-\Delta v_c(1-v_{i2}))} & \text{for } 0 \leq \tau \leq T(v_{i2}-v_{i1}-\Delta v_c(1-v_{i2})) \\ 0 & \text{otherwise.} \end{cases}$$

As was shown in Section 4.1, the width of this unit area pulse is the amount that the TDOA for the interferer changes during the observation period T. To smear the interferer by a large amount, the pulse width should be much greater than $1/B$ where B is the bandwidth. Therefore,

$$\begin{aligned} \frac{1}{B} &\ll |T(v_{i2}-v_{i1}-\Delta v_c(1-v_{i2}))| \\ &\ll \left|T(\frac{v_{i2}-v_{i1}}{1-v_{i2}}(1-v_{i2})-\Delta v_c(1-v_{i2}))\right| \\ &\ll |T(\Delta v_i(1-v_{i2})-\Delta v_c(1-v_{i2}))| \\ &\ll |(1-v_{i2})T(\Delta v_i-\Delta v_c)|\,. \end{aligned}$$

Since v_{i2} is small (on the order of 1 percent) this expression can be simplified to

$$T \gg \frac{1}{B|\Delta v_i - \Delta v_c|}. \tag{4.5}$$

Equation (4.5) expresses, in terms of the rates of change of TDOA (or speeds) of the contact and interferer, how long an observation period is required to substantially smear the correlation of the interferer in an effort to suppress it, given a compensating scaling factor equal to Δv_c. The length of the observation period can be specified in multiples the value of T equal to the right had side of (4.5). Thus an observation period factor (OPF) of 10, represents the conditions where $TB|\Delta v_i - \Delta v_c| = 10$.

In the scenario of Chapter 2, Δv_i is zero (the TDOA for the interferer does not change) and Δv_c varies over a range of 0.007 to 0.02 for expected contact velocities of -5 to -15 meters per second. Since the bandwidth is 400 Hz the requirement for interferer suppression is that $T \gg 0.366$ seconds for -5 m/s and $T \gg 0.122$ seconds for -15 m/s.

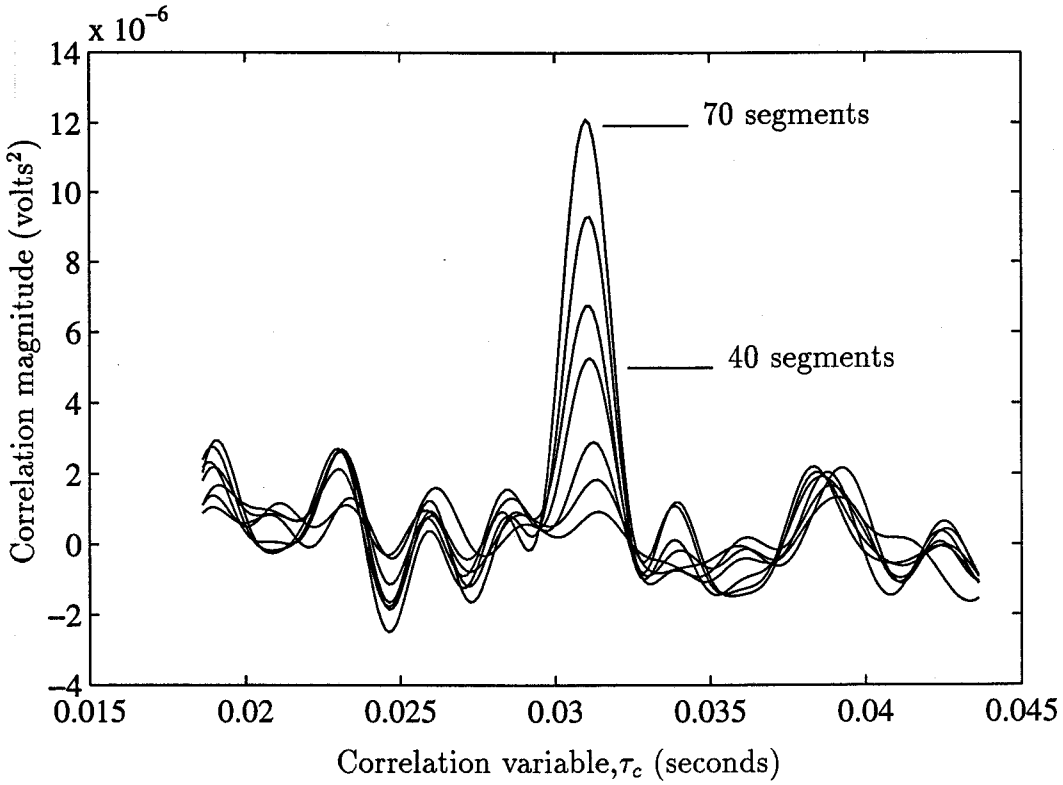

Figure 4.4 Select-Correlate-Sum method correlation of moving contact signal using from 10 to 70 extractions, all of length 100 ms.

Figure 4.4 illustrates an example of the suppression of the interferer for various observation period lengths, measured in the number of 100 ms segments included in the cumulative correlogram. For this example the contact is moving at -15 m/s and the interferer is stationary. The correlation peak for the contact is shown increasing monotonically for each accumulation of 10, 20, ..., 70 segments. With 40 segments accumulated (or an OPF of about 11) the correlation peak for the contact is distinguishable above the interference. With 70 segments (an OPF of about 20) the effect of the interferer has been well suppressed.

4.4 Permissible Delay Rate Mismatch

Using a scaling factor that is not correct has been shown to cause degradation (smearing) of the correlogram. This has a bearing on the size of the increment in the scaling factor used during the search for the optimum scaling factor.

In their study of compensating for contact motion for the purpose of time delay estimation, Kuhn et al. [5] derived equations and plotted the degradation in the cross correlation of received signals for varying mismatches between the scaling factor used for compensation and the scaling factor that would completely compensate for contact motion during the integration time. Their measure of correlation degradation was the decrease in the magnitude of the correlogram peak. Their results are useful in estimating the degree of delay rate mismatch that can be tolerated in the search for the scaling factors in this work. According to Kuhn, for a bandwidth of 350 Hz and a correlation time of about 5 seconds, a delay rate mismatch of 200 x 10^{-6} seconds per second would result in a 30% reduction in the correlogram peak magnitude. If, during the search for the optimum scaling factor, increments much larger than 200 x 10^{-6} were used, there would be a danger of missing the desired peak. None of the scaling factors in the search might closely match the required scaling factor, with the result that all of the correlogram peaks would be reduced due to delay rate mismatch. It might be, therefore, that none of the correlograms peaks would be prominent enough to be identifiable as the contact peak. On the other hand, small increment sizes increase the number of iterations required during the search for the optimum scaling factor. The value of 200 x 10^{-6} will be used as a guideline for increment size during the search.

4.5 Segment Alignment Problem

In using proposed scaling factors to calculate the midpoint of a segment from $s_2(t)$ (in an effort to account for contact motion), it is likely that the calculated segment midpoint is not coincident with an existing sample point in $s_2(t)$. This alignment problem, and how it will be alleviated, is discussed in this section.

As was introduced in Chapter 2, for efficiency only the midpoint of a segment from $s_2(t)$ is scaled to compensate for contact motion, and not each of the data points within the segment. In implementation, all that is required is to calculate which sample of $s_2(t)$ will be the midpoint based on the scaling factor and the midpoint of

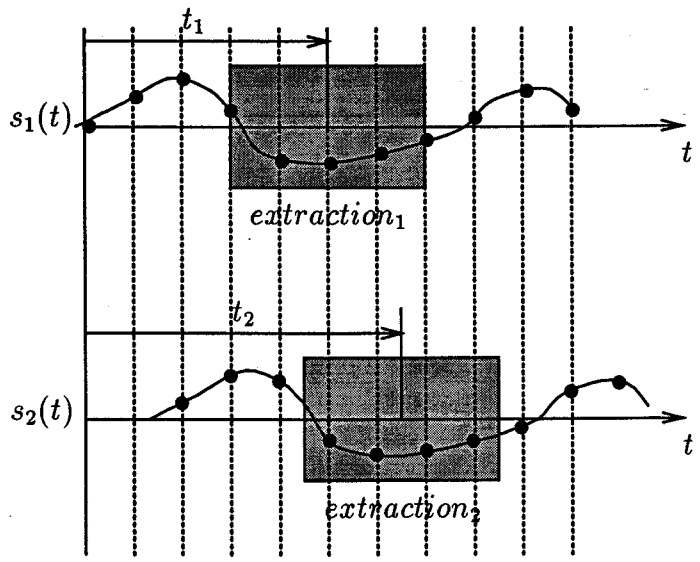

Figure 4.5 Illustration of alignment error as a result of scaling that computes a midpoint that is not coincident with any existing data point.

the segment from $s_1(t)$, denoted t_1. A problem arises when for a particular t_1, the proposed scaling results in the calculated midpoint for a matching segment from $s_2(t)$ (denoted t_2) that lies between sample points in $s_2(t)$. Refer to Figure 4.5.

If t_2 is not coincident with a sample from $s_2(t)$, then any segment chosen from actual sample points from $s_2(t)$ will not be exactly the segment called for by the scaling operation. When a segment sequence from $s_2(t)$ is correlated with the matching segment from $s_1(t)$, the data points in the segments will be offset in time and the peak of the correlogram will similarly be offset. The consequence is that as the short-time correlations are summed, the peaks will not be properly aligned, even if the optimum scaling factor is used.

This alignment error could be handled in several ways. The alignment error could be ignored. The magnitude of the error is half of an intersample interval or less. Alternatively, new data points in the segment from $s_2(t)$ could be interpolated at the calculated midpoint and at regular intervals from that midpoint. Or, after correlating each segment, the resulting correlation could be resampled prior to being combined with the others. Resampling is a computationally expensive operation and might not

be a practical option.

Another possibility is to choose an alternate t_1, which when scaled results in a t_2 that is (or very nearly is) coincident with a sample point. Since Δv is constant, it is possible to predict a t_1 that will scale to an appropriate t_2.

Selecting alternate t_1 values implies that the segments from $s_1(t)$ are different from those in the regular partition, $\mathcal{P}$, and that there is likely to be some overlap of the new segments. As long as it remains small relative to the length of the segments, overlap is not a serious concern.

The following is a description of an algorithm for selecting an alternate t_1 based on the difference $t_2 - t_n$ where t_n is an existing sample point in $s_2(t)$. The algorithm finds a new t_2 that is nearly coincident with a sample point. In the following, the time values in uppercase letters are sample sequence numbers.

Using T_1 suggested by the partition $\mathcal{P}$, scale to determine the proposed T_2 with

$$T_2 = T_1(1 + \Delta v).$$

Normally T_2 will not be an integer sample number. A corrective number of samples, denoted D, is calculated from the fractional part of T_2 as follows:

$$D = \text{round}(\frac{T_2 - \text{round}(T_2)}{\Delta v}),$$

then T_1 is corrected by D and scaled again

$$T_2' = (T_1 - D)(1 + \Delta v),$$

where T_2' will be very nearly an integer number. For Δv values corresponding to contact velocities from -5 to -15 meters per second, the maximum difference from an integer that T_2' will be is less than 0.01. Therefore, using this algorithm, the segments will be aligned to within one one-hundredth of a sample interval. The maximum

required correction to T_1 is 159 samples.

4.6 Accumulating Sparse Segments

Additional processing time could be saved by correlating and accumulating only some of the segments in a partition of $s_1(t)$. Clearly sufficient segments need to be included in order to provide enough data for a reliable time-average correlation. It is also clear that the segments which are used should span the entire observation period in order to maximize the amount of relative motion between the contact and the interferer. However, given those two conditions, all of the segments need not be accumulated.

In a case where the contact is moving slowly, the observation period needs to be quite long. With long extractions, it should not be necessary to accumulate all of the segments unless the interference is particularly intense.

Figure 4.6 shows the result of processing data from the same observation period (7.8 seconds) with 20, 40 and 60 segments of 100 ms. The speed of the contact is -15 m/s and the interferer is stationary. The amplitude of the correlogram peak, and the accuracy of the location of the peak, increases with additional processed segments.

Since the speed with which the TDOA can be established is a primary concern, it is possible that processing some segments while continuing to collect others will lead to an earlier solution. Depending on the speed of processors available, it could be that processing only sparse segments, at the risk of increased location estimate error, is a reasonable trade off.

This chapter has discussed the concerns around selecting values for key parameters for the SCS algorithm. Chapter 5 will describe the details of the simulation for generating and processing signals from the contact and interferer. Chapter 6 will present the results of the simulation trials.

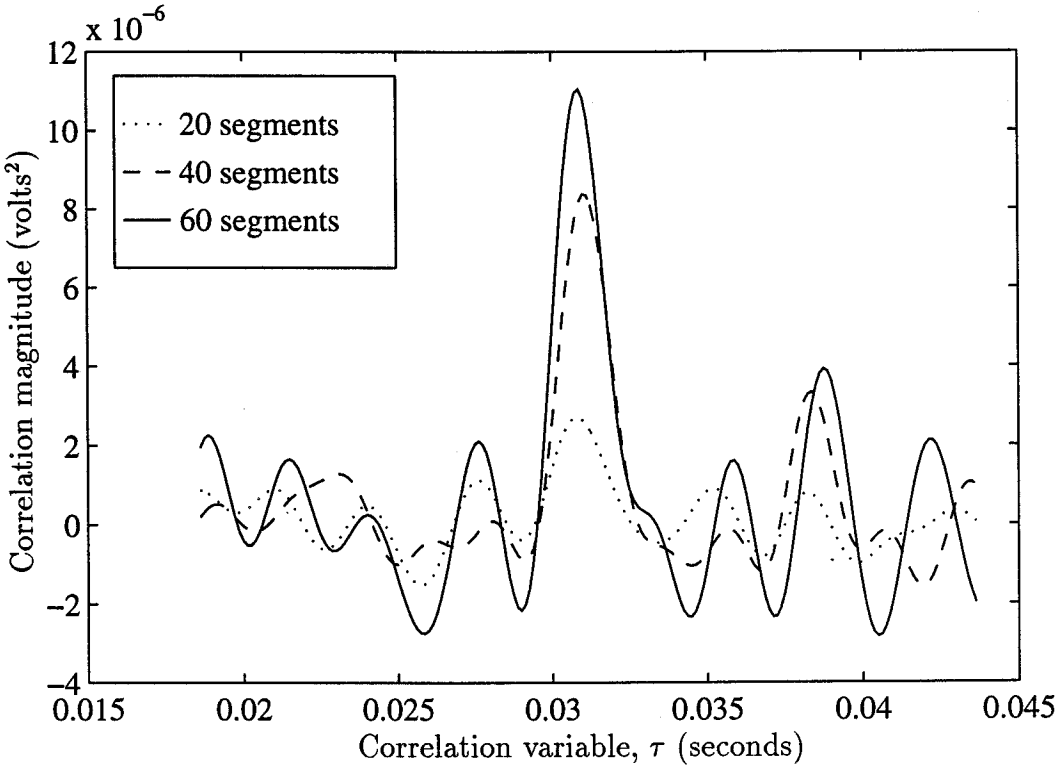

Figure 4.6 SCS correlation over an observation period of 7.8 seconds using only 20, 40 and 60 equally spaced 100 ms extractions.

5. Implementation of Simulation and Algorithms

This chapter contains a description of the procedures used in simulating the acoustic signals as they would be received by two hydrophone sensors, and a description of the implementation of the Select-Correlate-Sum method (hereafter referred to as the SCS method). It also contains some details for the full time-axis scaled (SCALED) method and for a modification of the SCS method in which only some of the partition segments are used (SPARSE method).

All synthesis and processing routines have been written as part of this project and are implemented on Sun workstations using the MATLAB programming environment from the MathWorks, Inc.

5.1 Signal Synthesis

Numeric sequences are synthesized and treated as the sampled voltage output from two hydrophone sensors. To do this, random sequences are generated, filtered, and treated as acoustic signals from underwater sources propagating through the ocean medium. The two sensors then convert all of the sound reaching them into a voltage signal. This procedure will now be described in more detail.

A normally distributed random sequence is generated and then filtered with a Butterworth filter. The filter was designed with specifications of a passband ripple of -1 dB and a minimum stopband attenuation of -15 dB. The passband cutoff was 400 Hz and the stopband began at 500 Hz. The sampling frequency is assumed to be 8000 samples per second. In the simulation, the filtering is done during generation rather than during receiving. It is assumed that the strong tonal components (from the propellers) are also filtered out. The data point amplitudes are multiplied after

filtering to portray acoustic sources of varying sound intensities.

The simplifying assumption is made that an acoustic source generating an intensity of 1 watt/m^2, measured at a distance of 1 meter, would cause a sensor to record a 1 volt RMS sequence. This corresponds to an acoustic pressure, $p = 1.22 \times 10^3$ μPa and a sensor conversion factor, $K = 8.16 \times 10^{-4}$ volts/μPa. The sound level of such a source is SL//μPa@1m = 181.76 dB. Since the bandwidth is flat to 400 Hz, the amount of power in each band of width 1 Hz is given by SL//μPa/Hz@1m = 155.74 dB. This source level is reasonable for surface ships according to Urick [9].

The interferer is always assumed to emit sound at this intensity. The intensity of the contact is varied for different experimental trials.

A channel vector is constructed to model the propagation of sound along each path from the source to each of the two sensors. The channel vector specifies an attenuation factor and delay value for each synthesized data point. Only spreading losses are implemented. Other losses are not nearly as significant over the short distances involved in this scenario. Each data point in the synthesized sequence is multiplied and delayed by the appropriate channel factors to simulate propagation through the ocean.

Because of the source motion, the arrival times of the propagated data points are not coincident, nor do they have the same rate. The sequence assumed to be arriving at Sensor 1 and the sequence assumed to be arriving at Sensor 2 are both resampled to give uniform and coincident samples. This is done using a sin x/x method involving the 200 adjacent data points [10].

Sequences for both the contact and the interferer are independently synthesized as just described. Next the sequences are summed, point for point, to give the final sensor output sequences.

The contact to interferer sound intensity ratio (CIR) is used to indicate the relative

magnitude of the two sources, each measured at 1 meter. That is,

$$\text{CIR} = 10 \log \frac{I_c}{I_i},$$

where I_c and I_i are the acoustic intensities of the contact and interferer respectively, both measured at 1 meter.

However, the CIR must be clearly differentiated from a resulting signal to noise ratio in the hydrophone output sequences. In the scenario, the proximity of the interferer to Sensor 2 is very significant. The surface ship (the interferer) wishes to rapidly determine the location of a following submarine (the contact) and also wishes to minimize the likelihood of the contact passing Sensor 2 and becoming undetectable between the ship and both sensors. Therefore, the ship must begin recording and processing sequences from the sensors soon after deploying the second sensor. The consequence of this is that the ship is then close to Sensor 2 and will be received very strongly. The distance of the interferer from Sensor 2 is more significant than the relative intensities of the contact and interferer. The trials used to produce the results in the next chapter are made with the interferer initially located 50 meters from Sensor 2 and moving away with the same speed and direction as the contact. The CIR should only be considered in a comparative sense.

The starting locations, speeds, and relative intensities are recorded for later reference to the estimates provided by the analysis methods.

5.2 SCALED Processing Method

This section describes some of the implementation details for the SCALED method, which uses full time-axis scaling of every data point to compensate for contact motion during the observation interval. The results and cost of this method are used as a basis for comparison with the results and costs of the SCS and SPARSE methods.

After loading the simulated data into buffers, a data point arrival time axis is regenerated. This is straightforward since the data was uniformly resampled at known

instances during its synthesis.

Time axis scaling, to compensate for contact motion, is done by altering each point's arrival time according to a current scaling factor and then resampling.

For each proposed Δv in the search space, the arrival time for each data point in the sequence from Sensor 2, $s_2(n)$, is scaled. The scaling operation is

$$s_{2ts}(n) = s_2(n(1 + \Delta v)),$$

which clearly establishes a relationship between the arrival times of data in $s_2(n)$, $t_n = n(1+\Delta v)$ and the arrival times of the same data in $s_{2ts}(n)$, t'_n. Given the arrival times for $s_2(n)$, the arrival times for the scaled version are then computed with

$$t'_n = t_n/(1 + \Delta v).$$

After scaling, the sequence is resampled at intervals spaced uniformly from the scaled arrival time of the midpoint of the observation period. The $\sin x/x$ method is used, involving the adjacent $\pm$ 50 points in the sum. The scaled and resampled sequence from Sensor 2 is then correlated with the unaltered sequence from Sensor 1.

The resulting ambiguity surface should contain two correlation peaks, one at the scaling factor and delay corresponding to the contact and another corresponding to the interferer. In most trials, however, the search space is limited and the ambiguity surface is generated only for the region in which the correlation peak for the contact is expected.

The location of the contact peak provides biased estimates of the compensating scaling factor and time difference of arrival of sound received at the end of the observation period. Some corrections of these biases are possible as described in Section 5.5. Finally, the contact speed estimate is calculated from the scaling factor coordinate at which the peak is found. The location estimate for the contact at the end of the observation period is calculated from the corrected delay coordinate for the peak.

5.3 SCS Processing Method

This section describes implementation details for the Select-Correlate-Sum (SCS) method. The principal difference between the SCS and the SCALED methods is that the SCS method avoids the computationally expensive scaling of every data point. Instead it correlates short extractions of the unscaled sequences.

The Sensor 1 sequence $s_1(n)$ is partitioned into N segments and the midpoint of each segment, m_k^1, is determined for $k = 0 \ldots N-1$. The sequence number of each point in $s_1(n)$ is used to identify the time at which the point is received.

Rather than scale the time axis of $s_2(n)$ so that it becomes a replica of $s_1(n)$, a set of points in $s_2(n)$ are located which would have scaled to correspond to the set of segment midpoints identified in $s_1(n)$.

In the example scenario, it is known that $s_2(n)$ will be a compressed version of the sound emitted from the moving contact and that $s_1(n)$ will be an expanded version. Rather than scale $s_2(n)$, points in $s_2(n)$, denoted m_k^2, are located in the compressed state which correspond to m_k^1. This is the opposite operation to scaling, accomplished with

$$m_k^2 = m_k^1(1 + \Delta v), \quad k = 0 \ldots \mathrm{N} - 1.$$

The length of the extractions is normally the same as the partition segment length. It is governed by the bandwidth, and Δv_c and Δv_i as specified in (4.4). For contact speeds of -15, -10 and -5 m/s, extraction lengths of 100 ms, 150 ms and 300 ms, respectively, are used.

Since Δv guesses are unrestricted, the scaled midpoints will in general not be coincident with existing samples in the Sensor 2 sequence. The sample nearest the calculated value is taken as the midpoint. This ignores the alignment problem and will result in some alignment error as was discussed in Chapter 4. However, this error is judged as acceptable for two reasons. Firstly, other error, such as that caused by the interferer is much larger and, secondly, the error introduced by not aligning the

midpoint with existing samples is at most one sample in error. This corresponds to a location error of less than 0.1 meter.

Each short extraction from Sequence 2 is then correlated with Sequence 1. A custom correlation procedure is used, since the edge effects of the MATLAB *xcorr* routine [11] are not tolerable using very short extractions. The custom procedure shifts the short extractions from Sequence 2 against all necessary data in Sequence 1. The amount of shifting is a parameter that restricts the region of the correlation variable to where the peak is expected. This greatly reduces the number of computations required. Each resulting correlation is summed with the others and the aggregate correlation is divided by the total integration time, N times the extraction length.

The set of aggregate correlations resulting from all of the Δv guesses form an ambiguity surface, which is restricted to the region where the peak corresponding to the contact is expected.

The peak of the ambiguity surface is interpolated using the *bi-cubic* option of the MATLAB *interp2* function. Twenty sub points are interpolated in each axis direction. The maximum of the interpolated space is taken as the correlation peak. A higher degree of precision is not necessary and is not valid given the error already admitted in not aligning the correlations.

The peak of the ambiguity surface indicates a biased estimate of the scaling factor required to compensate for contact motion, Δv_c and for the time difference of signals arriving at the time of the midpoint of the last segment in the partition. Some corrections of these biases are possible as described in Section 5.5. Finally, the contact speed estimate is calculated from the scaling factor coordinate at which the peak is found. The location estimate for the contact at the end of the observation period is calculated from the corrected delay coordinate for the peak.

Because of the random nature of the signals there will be some location error due to the presence of the interference and due to the alignment problem. Numerous trials are run to develop a value for the variance of the error.

5.4 SPARSE Processing Method

The SPARSE algorithm is identical to the SCS algorithm with the exception that only some of the partition segments are used for correlating. While maintaining the same observation period, and thereby capitalizing on all of the relative motion between the contact and interferer, only some of the available data is used. This reduces the number of computations at the expense of reduced interferer suppression. One half of the segments used for the SCS algorithm are used for the SPARSE algorithm throughout.

5.5 Validating Bias Correction

The expressions derived in Section 3.2 for the rate of change of delay and for the time delay using the SCS method both indicated biased estimates. Based on knowledge of the scenario, these biases are largely correctable, leading to more accurate estimates. Bias corrections were described in Section 3.3.1. This section verifies those bias corrections using an example from a special case scenario. In this special case scenario the interferer is stationary and located midway between the sensors. It is recognized that this is not a feasible case of the scenario that has been used throughout this thesis. However, this location provides a constant interference level throughout the observation period. The data being analyzed in this example is for a contact moving at -15 m/s and the CIR = 5 dB.

The known values for the speed and location of the contact at the end of the observation period will be compared with corresponding values determined from analysis of the received sequences. Starting with the coordinates of the contact peak in the ambiguity surface (namely $\Delta\hat{v}_c$ and $\hat{\tau}_c$) the bias corrections and transformations of Section 3.3.1 will lead to an estimate of the speed and location of the contact.

Figure 5.1 and Figure 5.2 show, in contour plots, close views of the contact peak in the ambiguity surface. The vertical axis is the scaling factor Δv and the horizontal axis is the correlation variable τ. The interpolated peak at coordinates $\Delta\hat{v}_c$ and $\hat{\tau}_c$,

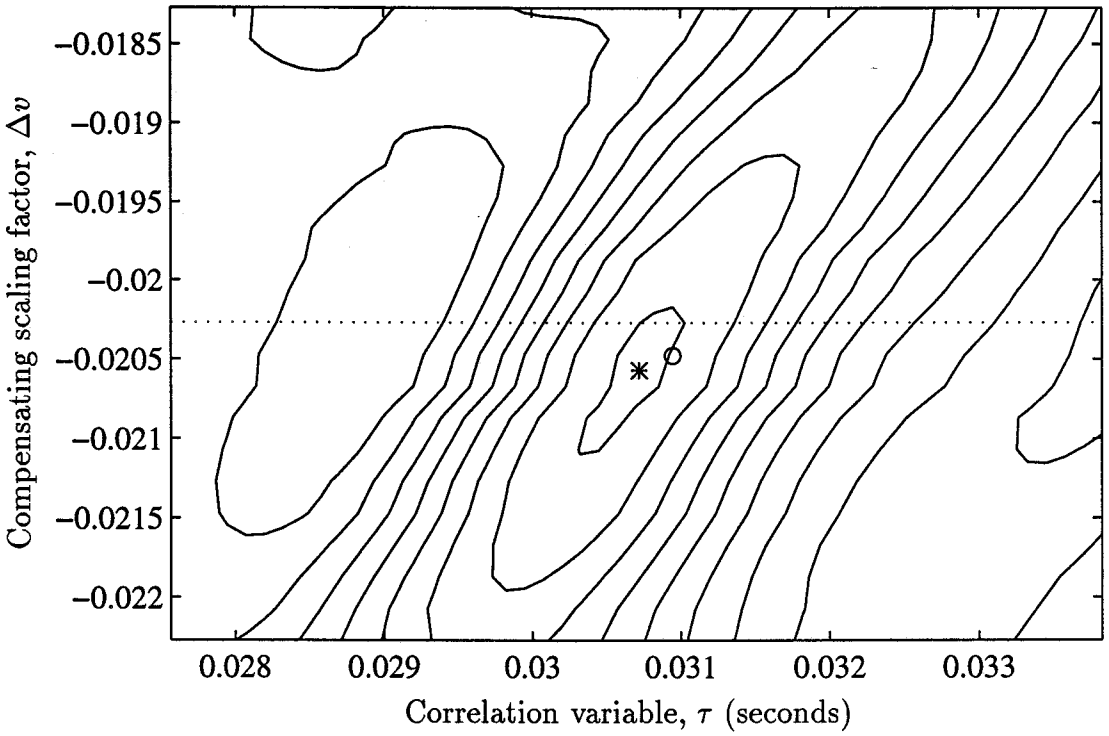

Figure 5.1 Contour of ambiguity surface in region of peak after an observation period factor of 10.

is marked with an asterisk. The dotted line indicates the value of Δv_c at which the peak would have been located if there were no interference.

The actual delay, which was recorded during data simulation, cannot be plotted directly. For long observation periods, τ_c (the expected delay coordinate of the contact peak) and $\Delta D'$ (the estimate of the time difference of arrival of interest, obtained by transformation of τ_c) can differ by a large amount. In order to map the actual value onto a plot of the ambiguity surface, the inverse of the transformations of Section 3.3.1 need to be applied. The result of applying those inverse transformations to the actual TDOA values is shown plotted as an 'o' in the figures.

Figure 5.1 shows the contact peak of the ambiguity surface after an observation period factor (OPF) of 10. Recall that the observation period factor is given by $\text{OPF} = TB|\Delta v_i - \Delta v_c|$. It can be seen that due to a short correlation time and due to the interference, the ambiguity surface peak is in error relative to the mapping

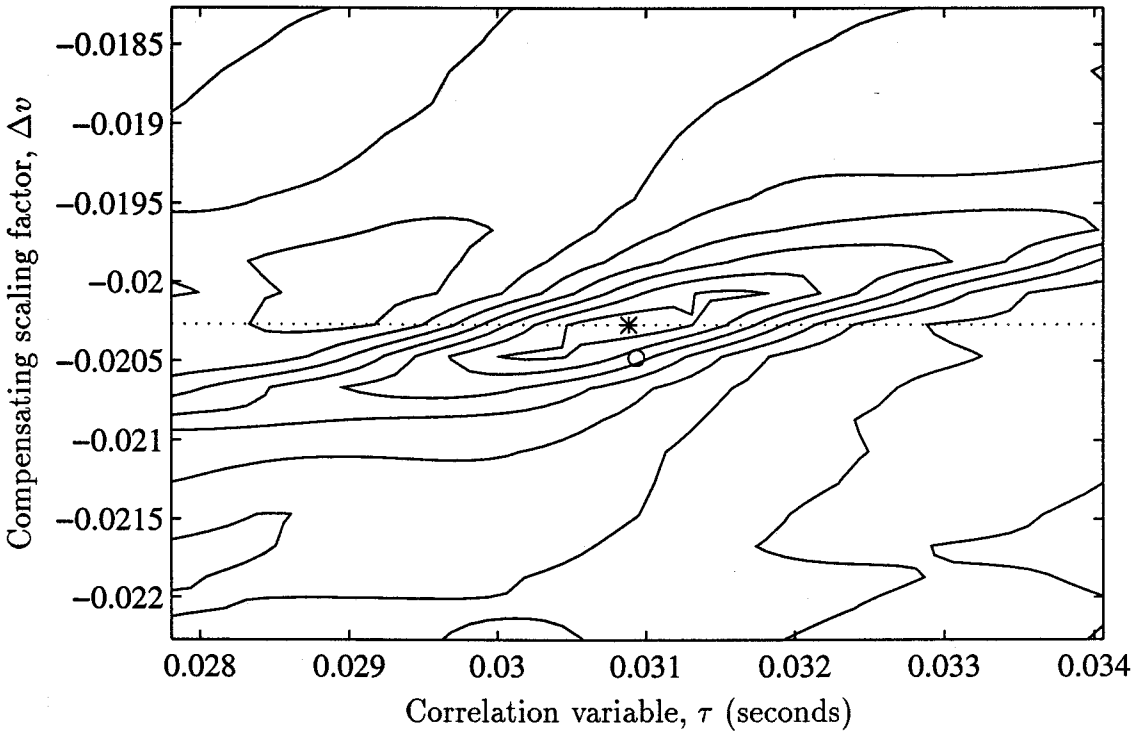

Figure 5.2 Contour of ambiguity surface in region of peak after an observation period factor of 50.

of the true delay and scaling factor, and the scaling factor is in error from what the method suggests it should be.

After an OPF of 50, however, as shown in Figure 5.2, the peak of the ambiguity surface is at the expected Δv scaling, and the estimated delay ('*') is very close to the actual value ('o'), as evidenced by the nearly vertical alignment of the two.

Another way to illustrate the accuracy of the SCS method is to compare the derived contact speed and location values to the true values. For the single example whose peak is plotted above, the time of the last segment midpoint is 3.6486 seconds. The location and speed of the contact at that time are shown in the following table. The error in the location of the contact is acceptable.

	Location in meters	Contact Speed in m/s
True values	-79.7294	-15.0000
Estimated	-79.7981	-14.9925
Error	-0.0688	-0.0075

Table 5.1 Comparison of true and estimated values for contact speed and location for the example of Section 5.5.

6. Results and Conclusions

This chapter presents an analysis of the performance of the Select-Correlate-Sum method for determining an estimate of the time difference of arrival for sound from a moving contact. It shows that the method can be successful in suppressing a strong interfering signal and obtaining an estimate of the TDOA that is sufficiently accurate for the scenario first discussed in Chapter 2. The performance of the SCS method is compared to that of the SPARSE and SCALED methods.

The chapter continues with a discussion of the possible use of the SCS method in practise, and the type of data that would need to be made available. Following that is a brief discussion of possible directions for continued research into the SCS method. The chapter ends with a short summary and concluding remarks.

6.1 Validation of the SCS Method

This section shows that the SCS method can be successful in accurately estimating the time difference of arrival (TDOA) of a moving contact at a particular point in time. To do this most effectively, the results of the method are applied to the special case scenario first used in the last chapter. In that special case, the interferer is located midway between the two sensors, and is stationary. While this is not a realistic case, in that the two sonobuoys were to have both been deployed behind the surface ship, this case causes the interferer to have a TDOA similar to that of the contact for at least part of the observation period.

Firstly, the ability to suppress the interferer will be shown with an example plot of the cross correlation of received signals. This is shown in Figure 6.1. The figure shows a plot of the cross correlation using the SCS method, with a segment length of 0.366 seconds and an observation period factor of 30. The CIR is 0 dB, which in this

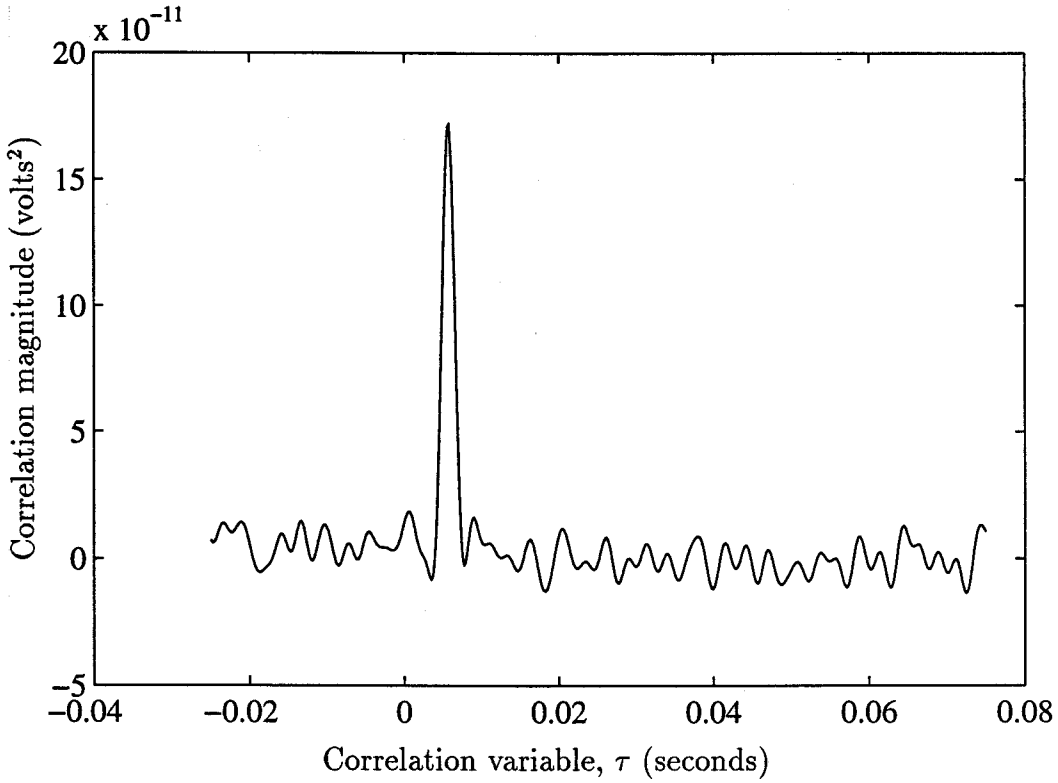

Figure 6.1 Correlogram for speed = -5 m/s, CIR = 0 dB, observation period factor = 30, scaling factor $\Delta v = \Delta v_c$. Interferer stationary and located midway between sensors.

case means a signal-to-noise ratio of about 0 dB, since the interferer and contact are approximately the same difference from either sensor. This correlogram was made using a compensating scaling factor equal to the optimum Δv_c for the contact moving at -5 m/s. A single correlogram peak is apparent in Figure 6.1, which is identifiable as that for the contact since it is found using a scaling factor corresponding to the speed of the contact. Further, the location of the peak, after proper transformation, gives an accurate estimate of the true location of the moving contact at the end of the observation period.

A correlogram peak representing the interferer is not visible, even though it would be expected to occur at $\tau = 0$ in the figure. It is therefore apparent that the interferer has been suppressed, and is only contributing as uncorrelated noise to the correlogram

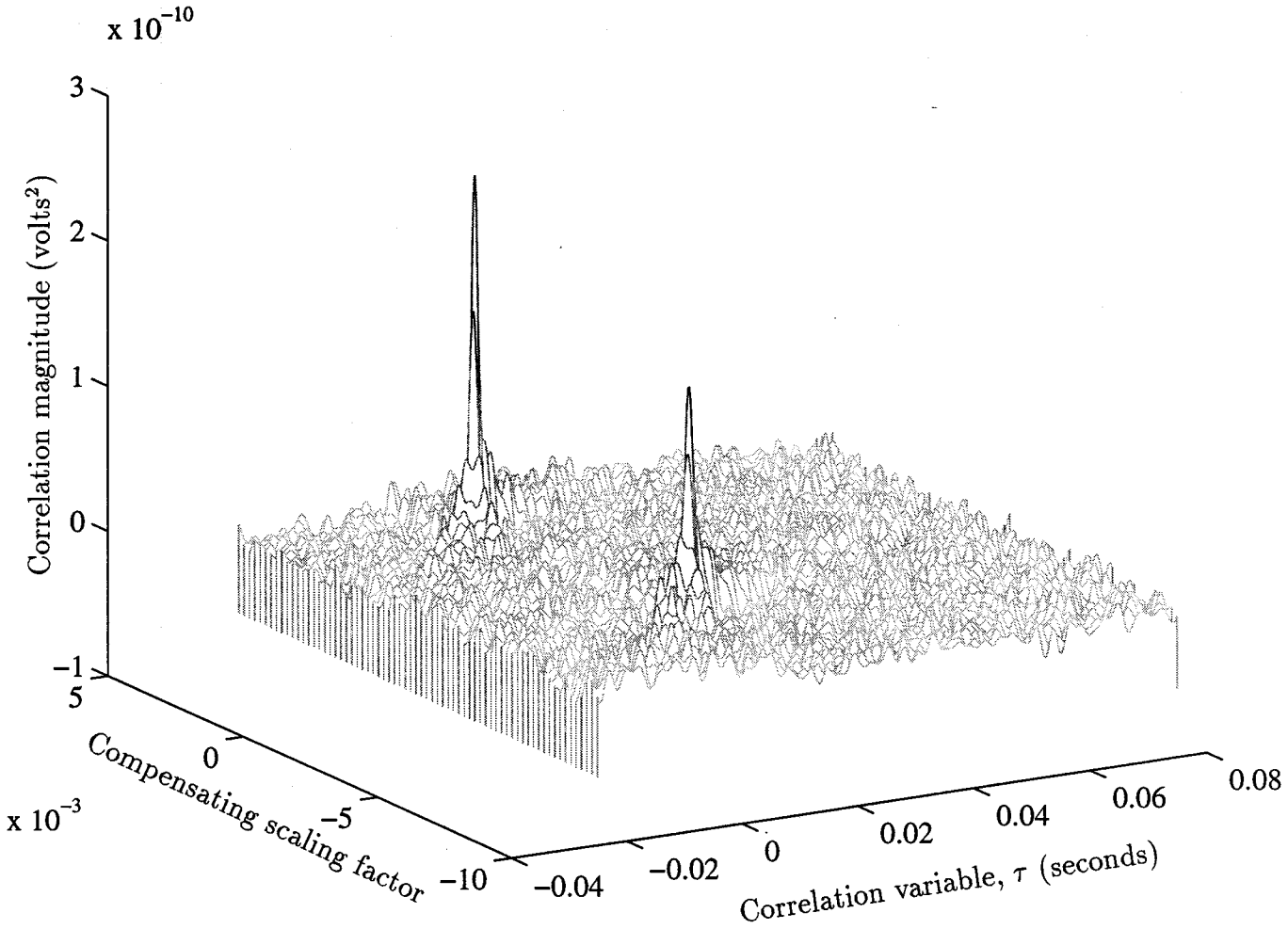

Figure 6.2 Ambiguity surface plot for speed = -5 m/s, CIR = 0 DB, OPF = 30, interferer stationary and located midway between sensors.

of Figure 6.1.

Figure 6.2 shows a waterfall plot for a relatively large patch of the ambiguity surface for the above scenario. The axis on the right base is the correlation variable τ in seconds. The axis on left base is the compensating scaling factor Δv and the z-axis, or height, represents the magnitude of the ambiguity surface. The observation period factor is 30. The contact speed is -5 m/s and the CIR = 0. The figure clearly shows exactly two peaks in this region of the possible search space.

The peak in the foreground, the smaller peak, is the peak due to the contact. The larger peak in the background is due to the interferer. The interferer peak is expected

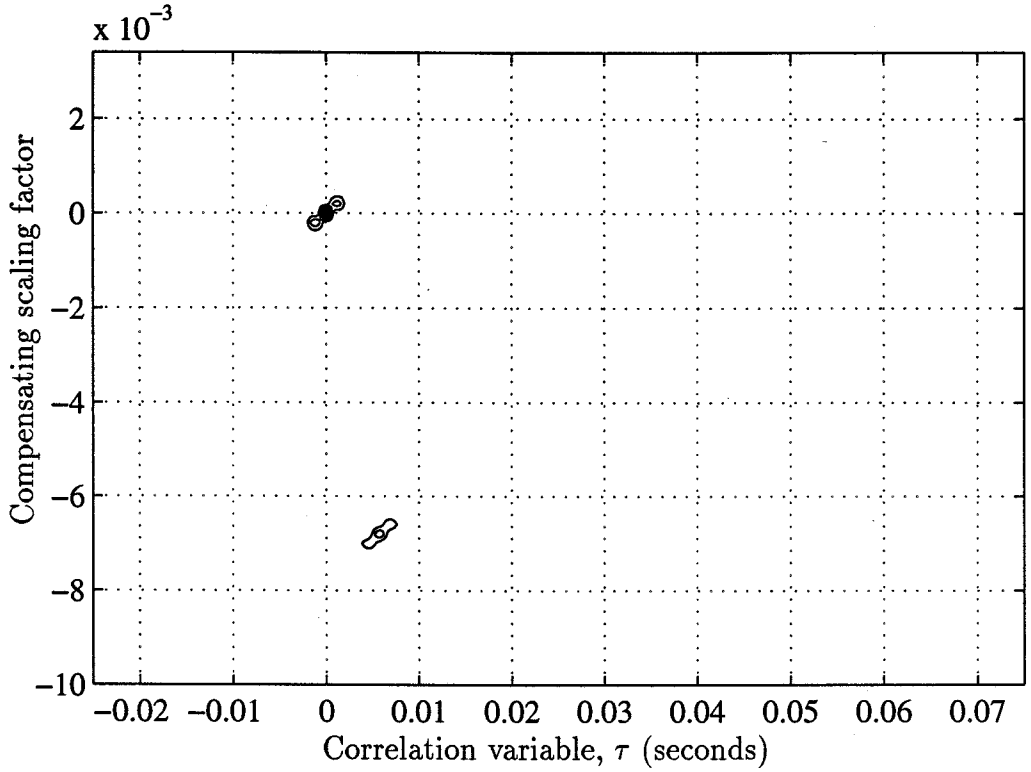

Figure 6.3 Contour plot of the square of the ambiguity surface for speed = -5 m/s, CIR = 0 DB, OPF = 30, interferer stationary and located midway between sensors.

to have a larger magnitude. The set of scaling factors used in this analysis includes the scaling factor 0 that matches exactly the optimum value for the stationary interferer. The set of scaling factors does not include the exact value required to compensate for contact motion. The coordinates of the peaks are easier to see in the next figure.

Figure 6.3 shows a contour plot for the same region of the ambiguity surface for this scenario. Note that for this contour plot the magnitudes of the surface have been squared in order to simplify the appearance of the plot; the squaring makes the peaks more prominent and reduces clutter. The two peaks are clearly visible. The interferer peak is clearly identifiable at $\tau = 0$ and $\Delta v = 0$, as expected since the interferer was located midway between the sensors and was not moving. The second peak can be identified as that of the contact. Its Δv coordinate of -6.8 x 10^{-3} maps to a contact speed of -4.9979 m/s (it is known to move at -5 m/s). The τ coordinate for the peak is 5.7125 x 10^{-3} seconds. This τ can be transformed to a TDOA at the end of the

observation period (10.8719 seconds) of -59.3613 meters. It is expected to be -59.3594 meters. [As before, the contact has a starting location of +5 meters, 2 seconds before the beginning of the observation period.]

Since the two peaks are clearly separated, for this one example it is clear that the SCS method has met the criteria of effectively suppressing a strong interferer. At the appropriate scaling factor, the TDOA for the contact has been correctly determined despite the presence of the strong interferer. The term "suppress" is used to indicate a reduction in the interfering effect of the interferer. None of the power in the interferer signal is lost. Instead, the interferer signal is decorrelated and rendered as noise of an equivalent power.

The variance of the location error will be used as a quantitative measure of the degree of interferer suppression.

The next section uses the location variance as a way to express some performance properties of the SCS method.

6.2 Performance Properties of the SCS Method

If the variance of the error in determining the location of the contact at the end of the observation period is to be used as a way to compare the three methods, then an explanation of some of the factors that influence the variance is required. The variance of the location error is affected by the contact to interferer ratio (CIR), by the speed of the contact, and by the length of the observation period. Each of these is examined.

Figure 6.4 shows the effect of different contact speeds on location variance. The figure shows the location error versus the observation period factor for contact speeds of -5, -10, and -15 m/s. The extraction lengths are 100, 200 and 300 ms respectively. The CIR is 5 dB. The observation period factor (OPF) is given by

$$\text{OPF} = TB|\Delta v_i - \Delta v_c|,$$

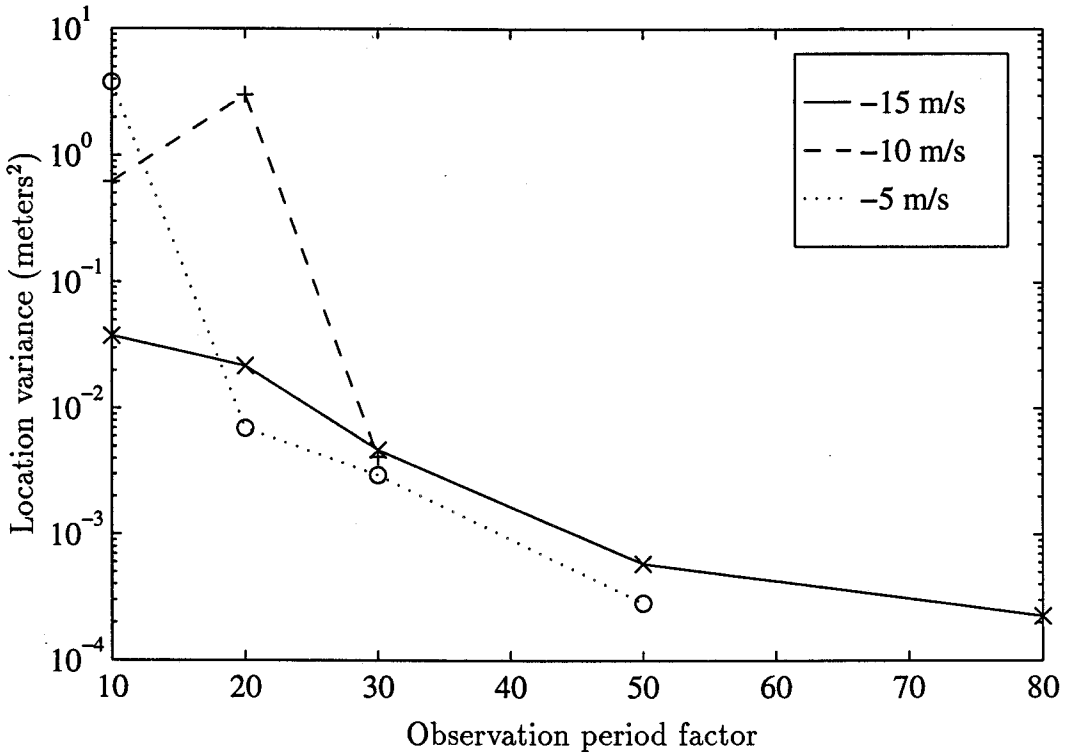

Figure 6.4 Variance in location error vs observation period factor for contact speeds of -5, -10 and -15 m/s. CIR is 5 dB.

where B is the bandwidth of the signal, and Δv_c and Δv_i are the optimum compensating scaling factors for the contact and interferer, respectively. This expression is derived from (4.5) in Section 4.3. The observation period must be large to ensure interferer suppression; values for the OPF ranging from 10 to 80 are used. Because of the interferer location in the scenario, Δv_i is zero.

The actual length of the observation period in seconds varies significantly for different contact speeds. Through using the OPF as the independent axis, the results for the three contact speeds can be more readily compared.

Figure 6.4 shows some erratic behaviour at low OPF values. Using OPF values of 10 or 20, the location of the ambiguity surface peak is quite variable. Actually, even higher variances should have been reported, were it not for a limitation imposed by

the implementation of the method (not by the method itself). The search space was limited to a small region in the vicinity of where the peak was expected. At low OPF values, sometimes the implementation did not detect a peak within the search space at all. Such cases would have contributed to high variances, but since no peaks were detected, such cases were omitted from the variation statistics.

However, the figure does show that the curves for the three speeds tend to converge for OPF values of 30 and more. Of course the length of the observation period in seconds required to attain similar levels of location variance is much longer for lower contact speeds than for higher speeds. For example, a contact speed of 5 m/s requires an observation level of three times the length of that required for a contact speed of 15 m/s. Conversely, for a given observation period length, a higher level of variance can be expected for a slower contact speed.

Figure 6.5 shows the effect of different contact to interferer ratios on the location variance. The figure plots the location variance versus the OPF for CIR of 5 dB, 0 dB, -5 dB and -10 dB. The contact speed is -15 m/s and the extraction length is 100 ms.

Some erratic behaviour is again evident at low OPF values. However, it is clear that increasing CIR causes a reduction in location error variance. The CIR values should only be used in a relative sense, and the CIR values should be carefully differentiated from signal to noise ratios (SNR). This is because in the scenario, the interferer is located quite close to Sensor 2 at the beginning of the observation period, and therefore has a very large impact at the beginning of the observation period.

The effect of the length of the observation period, measured in OPF, is evident in both Figure 6.5 and Figure 6.4. The variance decreases with increasing OPF in apparently exponential fashion. The decreasing variance reaches a limit at about 10^{-3} meters^2. This agrees with the error in location estimate expected as a result of not aligning the points of the correlations prior to summing, as discussed in Chapter 4. This degree of variance is more than acceptable for locating the contact between the two sensors.

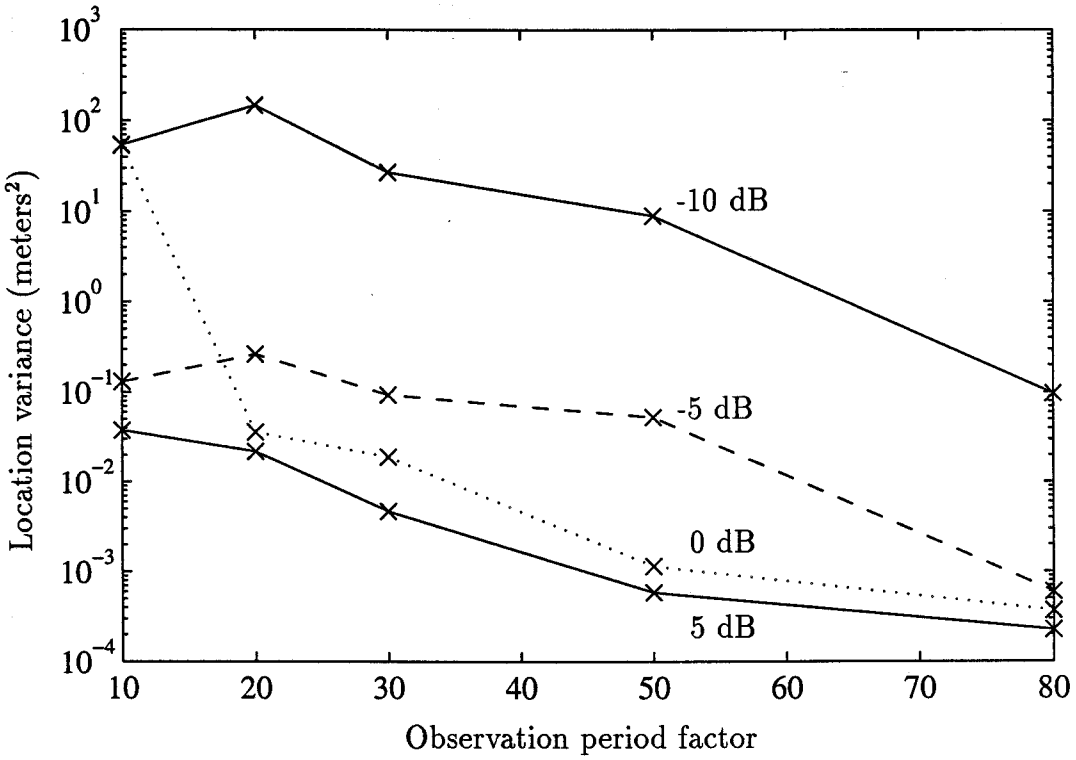

Figure 6.5 Variance in location error vs observation period factor for CIR ratios of 5 dB, 0 dB, -5 dB, and -10 dB. Contact speed is -15 m/s.

6.3 Comparison of Processing Methods

In addition to being able to determine an accurate estimate of the TDOA for a moving contact despite a strong interferer, an objective of the SCS method was to do so more efficiently than than the SCALED method, which uses time axis scaling of every data point in the observation period. This section will compare the performance and costs for the SCS, SPARSE and SCALED methods.

The three methods are first compared strictly on the basis of localization error variance. Then the computational costs are considered and the three methods are compared on the basis of cost for equivalent levels of location error variance.

The location error variance curves, versus OPF, for the three processing methods are shown in Figure 6.6. The speed of the contact is -15 m/s, the extraction length is 100 ms and the CIR is 5 dB. The figure clearly shows the relative performance of the

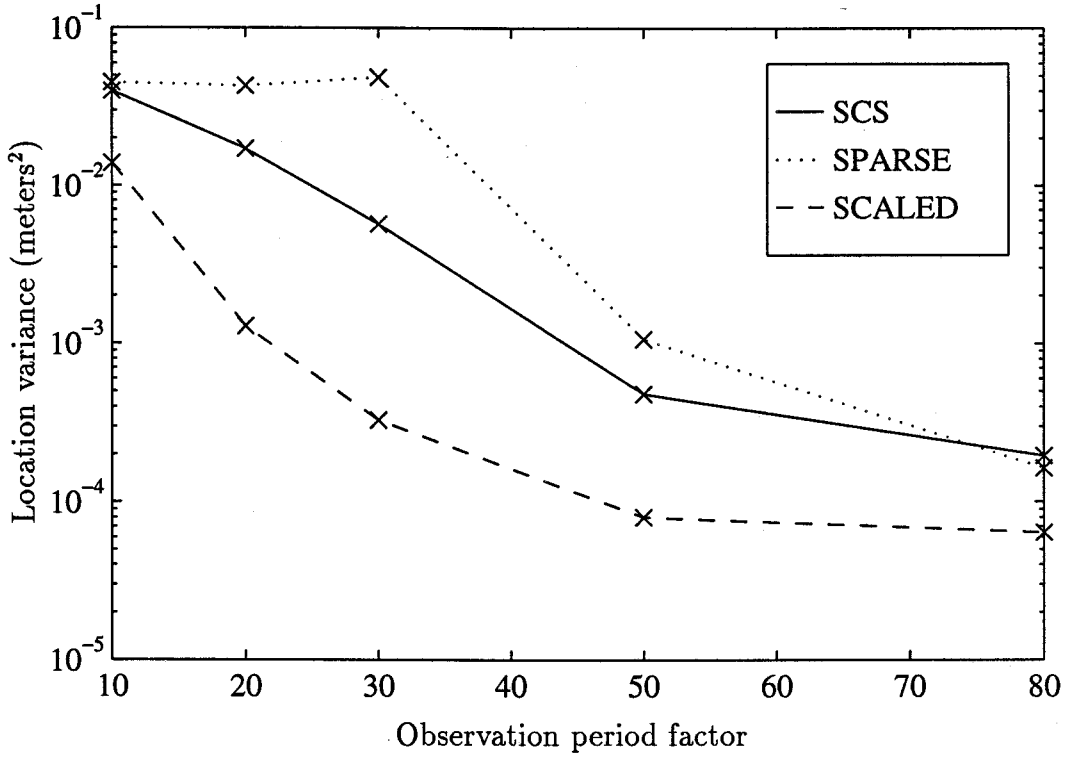

Figure 6.6 Variance in location error vs observation period factor for SCS, SPARSE and SCALED processing. CIR is 5dB and contact speed is -15 m/s.

three processing methods. It is expected that the SCALED method would have lower variance for two reasons. Firstly, the scaling in the SCALED method is accurate for all data points in the correlation time, while scaling is accurate only for the segment midpoints in the SCS method. This means that the magnitude of the correlation with the SCS method is not as large since the segments are not completely faithful replicas of one another. Secondly, since per-segment correlations are not completely aligned prior to summing them, some location error is introduced with the SCS method.

At a high OPF value of 80, the variance for the SPARSE method becomes effectively the same as the variance for the SCS method. If additional correlation data does not reduce the variance any further, then the limit for the processing method must have been reached.

At low OPF values, the SPARSE method has higher variances and is more susceptible to erratic behaviour. This is understandable because the total correlation time for the SPARSE method is one half that for the SCS methods, since one-half the number of segments were used as compared to the SCS method.

To include the processing costs of the three processing methods into the comparison, timings were made of the inner loops of the program code, thereby excluding from the timings more variable factors such as file reading and writing and network accesses. The MATLAB programming environment includes the *flops* routine to report the number of floating point operations between successive calls. This number remains consistent despite the speed of the computer processor being used, and is unaffected by processing loading. Therefore, this routine was used to produce the computational costs shown in Figure 6.7. The figure shows the cost in millions of flops (Mflops) of the three methods versus the OPF. This figure should only be construed to represent *relative* computational costs for the three methods.

The cost increases linearly with OPF in all three methods. As expected, the cost for the SPARSE method is half of that for the SCS algorithm, since only half of the segments are used in the SPARSE method. As expected as well, the cost for the SCALED method is significantly higher than the SCS method. Straight lines were fitted to the data as follows:

$$\begin{aligned} \text{SPARSE} &= 6.1036\ (\text{OPF}) + 0.8071 \\ \text{SCS} &= 12.4687\ (\text{OPF}) - 0.8431 \\ \text{SCALED} &= 33.0498\ (\text{OPF}) + 333.8278 \end{aligned}$$

The cost of the SCALED method is between 3 and 5 times the cost of the SCS method over the range of OPF values used.

In order to determine the relative cost of the three methods in obtaining the same level of location error variance, the results in Figure 6.6 and Figure 6.7 need to be combined. This is done as follows. From Figure 6.6, for a number of location

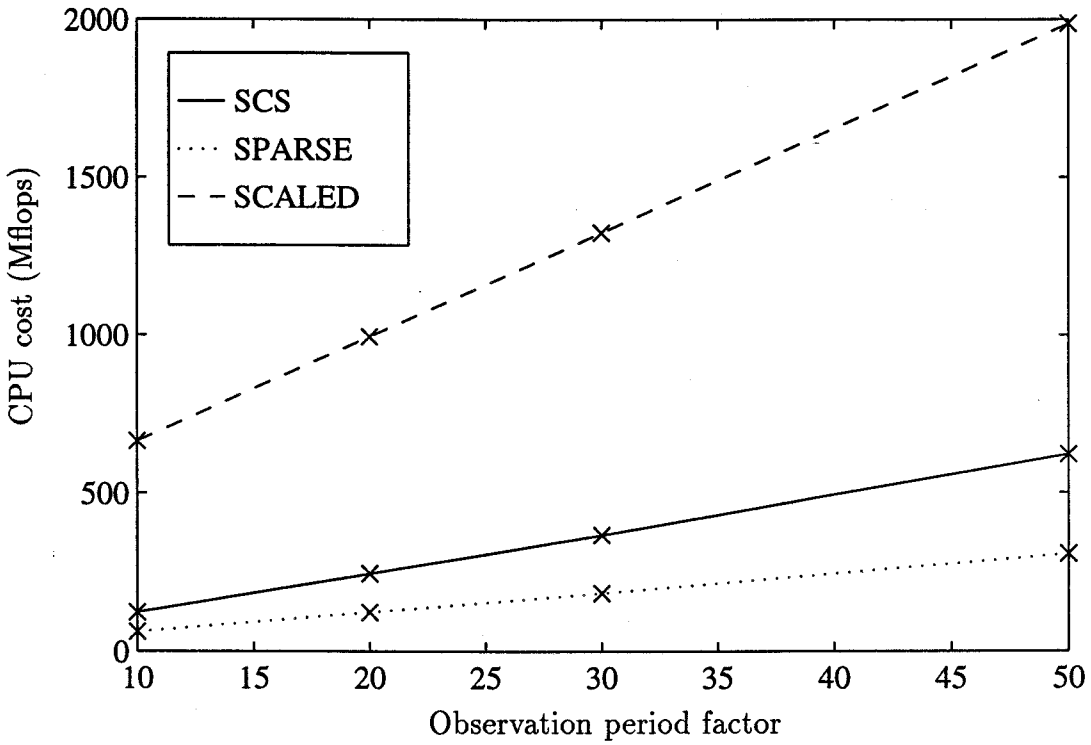

Figure 6.7 CPU cost in Mflops vs observation period factor for SCS, SPARSE and SCALED methods. CIR = 5 dB, contact speed = -15 m/s.

variance values, the corresponding OPF value is linearly interpolated for each of the three methods. Then the computational cost of each of those OPF values is computed using the linear equations derived from Figure 6.7. Then the cost determined for each method is plotted versus the location variance error in Figure 6.8.

Figure 6.8 shows the cost in Mflops versus location error variance for the three processing methods. For this figure the CIR is 5 dB, the contact speed is -15 m/s. The figure shows that for similar levels of location error variance, the SCS method has computationally lower cost than the SCALED method; it is more efficient. At lower variance levels, achieved at high OPF values, the SPARSE method is, however, even more efficient than the SCS method. At low OPF values, the high variance levels associated with the SPARSE method hurt its cost performance. It should be noted that the SCALED method can achieve location variance levels lower than are attainable with the SCS or SPARSE methods. It is also noteworthy that the costs of the SCS and SPARSE methods at their lowest variance level are lower than the costs

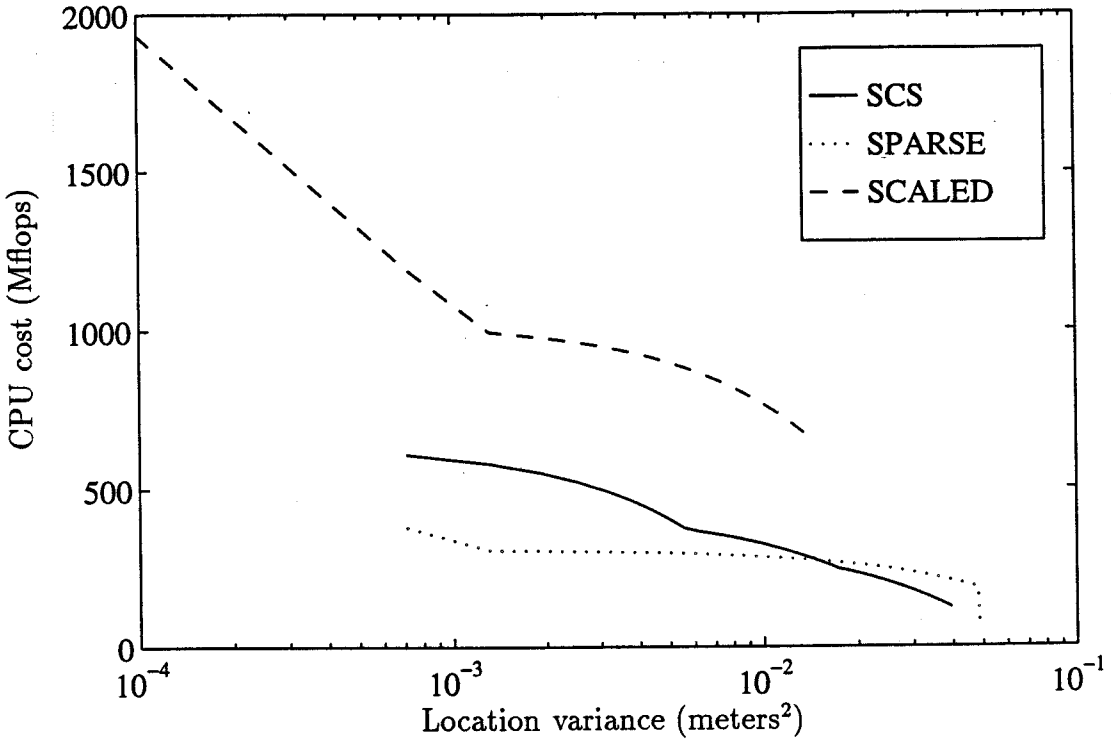

Figure 6.8 CPU cost in Mflops vs location variance for SCS, SPARSE, and SCALED methods. CIR = 5 dB, contact speed = -15 m/s.

for the SCALED method.

The computational costs are a significant consideration in the problem of locating the contact. Computational costs are directly related to the required processing time. A reduction in the processing time required means that the location estimate will be available sooner. A fast method also means that the search space can be feasibly made larger.

A comparison of the performance of the SCALED and SPARSE methods relative to the SCS method is shown in Figure 6.9. This figure shows the CPU cost relative to the SCS method of both the SPARSE and SCALED methods, for the levels of variance at which comparison is possible. For those variance levels, the SCS is roughly twice as efficient as the SCALED method. At high OPF values, and resulting lower variances, the SPARSE method is about twice as efficient as the SCS method.

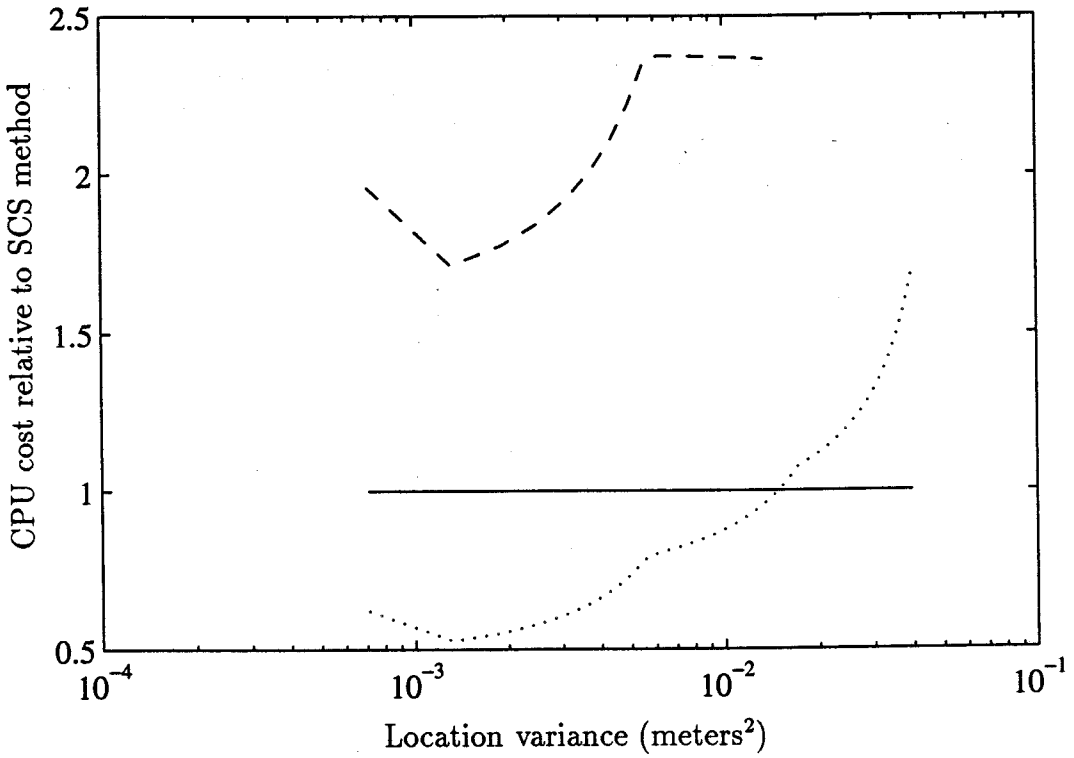

Figure 6.9 Plot of CPU cost relative to the SCS method vs location error variance. The solid line is the SCS method; the dotted line is the SPARSE method; and the dashed line is the SCALED method.

6.4 Desirable Data For SONAR Application

There are several related parameters involved in the performance of the SCS method. This section will explain the type of data that the surface ship in the scenario needs at hand during attempts to locate the following contact.

The most accurately known parameter will be the speed of the contact, since in this scenario, it is expected to have approximately the same speed as the interferer, which will be accurately known. While the interferer's Source Level will also be well known, the CIR cannot be accurately known due to the possible variations in range from the sensors of both the contact and the interferer.

Graphs of the form of Figure 6.5 for a range of ship's speeds would be most useful. For a particular speed, the independent axis can be the observation period in seconds rather than the observation period factor. The sonar operator can then

determine the level of location variance desired and can speculate on the likely CIR ratios, given some knowledge of the ship's distance from the sensors and the acoustic characteristics of the suspected following contact. The operator can then determine the required observation period.

6.5 Suggestions for Further Studies

Suggestions for further investigations of the SCS method will be discussed in four areas: Expressions for expected location variance, validation of the expected variance at a wider range of operating conditions, generalization of the SCS method to multipath sound paths and to contact motion not constrained to a path between the two sensors, and comparison of the SCS method with other efficient methods.

With the assumption that the SCS method makes a strong interference act in the same way as uncorrelated noise, an expression for the variance of time difference of arrival should bear a resemblance to the general case of a contact in the presence of only uncorrelated noise [12, 2]. However, the motion of the contact will complicate the derivation. The motion of the contact does create a Doppler frequency shift in the contact spectra received at the two sensors. Since the variance expressions are a function of those spectra, Doppler shifted versions of the contact's emitted spectrum need to be determined and used in the expression. Additionally, the nearness of both the contact and the interferer to the sensors, and their motion during the observation period, make the amplitude coefficients functions of time.

Once the variance expressions are determined, the SCS method variance results need to be validated over a wider range of operating conditions specified by contact speed, CIR levels and observation period lengths.

The scenario used throughout has omitted surface-reflected sound paths from the contact to the sensors, and has assumed that the contact was at the same depth as the sensors. Including surface reflections should be readily accommodated by SCS method as presented here, providing that contact stays collinear with the sensors.

The SCS method relies on the fact that the rate of change of delay from the contact to Sensor 1 is -1 times the rate change of delay from the contact to Sensor 2. That condition fails if more complex motion of the contact is allowed.

The surface reflections should show up as additional peaks in the ambiguity surface. Instead of one peak representing the contact, there would be four peaks corresponding to the cross correlations of the various combinations of the four separate sound paths [1]. Ideally the peaks would be well differentiated by their separate time differences and optimum scaling factors.

The ability of the method to suppress a strong interference makes it attractive in more general sonar localization problems, not just the scenario of the contact closely following the interferer. In the general case the contact and interferer are located anywhere in three dimensions. In addition to the multipath environment, the necessary size of the search space is very large. The extra efficiency provided by the SCS method would be valuable in this case.

Finally, the performance of the SCS method should be compared with alternative methods to the SCALED method, such as those proposed by Scarbrough [3], Kuhn et al. [5] and especially Betz [2].

6.6 Conclusion

For the scenario first described in Chapter 2, of a surface ship attempting to localize a moving submarine between two sonobuoys, the SCS method proposed in this work has been shown to be effective at providing accurate estimates of the actual location in at least certain operating conditions, and it has been shown to be more efficient than the SCALED method.

The methods were tested with contact to interferer ratios from +5 dB to -10 dB to show the increase in the variance of error related to the relative powers of the contact and interferer. The proximity of contact or interferer to either sensor is a more significant factor than the relative source strengths.

Speeds of -5, -10 and -15 m/s were used in tests to show the increase in the required observation period with faster speeds to show the same level of location error variance as at faster speeds. The method exploits the motion of the contact relative to the interferer, and faster contact speeds facilitate the method.

Observation periods, measured in observation period factors (OPF) ranged from 10 to 80. These factors correspond to times of from 1.2 to 9.76 seconds for a contact speed of -15 m/s and from 3.66 to 29.28 seconds for a contact speed of -5 m/s.

Especially for the SPARSE processing method, OPF values of 10 and 20 were not always successful in properly estimating the scaling factor and time difference. This was due to an implementation design to limit the computations required during testing of the methods. The problem was with the implementation and not the method. At OPF values 30 and above, the implementation provided consistent estimates of the scaling factor and time difference for all methods tested.

While the expressions developed for the required scaling factor and the time difference of arrival were biased, the bias was correctable. For CIR of 5 dB, 0 dB or -5 dB and for moderate OPF values, the variance in the location estimate was typically 0.1 meter or less.

The SCALED method was found to have the lowest location error variance for a particular observation period. The variance of the SPARSE method was slightly higher than for the SCS method.

The SCS method was shown to have twice the efficiency of the SCALED method. To achieve a given level of location error variance, the SCS method required about half the number of computations. At higher OPF values, the SPARSE method showed even greater efficiency. This improved efficiency could prove significant in implementing sonar methods to locate moving contacts.

References

[1] B. L. F. Daku, *The Accuracy in Locating an Underwater Acoustic Source in a Multipath Environment.* PhD thesis, University of Saskatchewan, 1990.

[2] J. W. Betz, "Comparison of the deskewed short-time correlator and the maximum likelihood correlator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 285–295, Apr. 1984.

[3] K. Scarbrough, *Analysis of Time Delay Estimator Performance.* PhD thesis, Kansas State University, 1984. Published as a Tech. Rep. of the Naval Underwater Systems Center, New London, CT, No. TR 7203, 19 June, 1984.

[4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[5] J. P. Kuhn, S. E. Robison, and D. W. Winfield, "Time delay detection and estimation techniques for moving sources," Tech. Rep., General Electric Company, Undersea Electronics Programs Department, Date Unknown.

[6] J. C. Rosenberger, "Passive localization," in *Underwater Acoustic Data Processing*, (Y. Chan, ed.), pp. 511–524, Kluwer Academic Publishers, 1989.

[7] L. Izzo, A. Napolitano, and L. Paura, "Interference-tolerant estimation of amplitude and time-delay parameters of a composite signal," in *Signal Processing V: Theories and Applications*, (L. Torres, E. Masgau, and M. Lagunas, eds.), Elsevier Science Publishers B.V., 1990.

[8] W. B. Adams, J. P. Kuhn, and W. P. Whyland, "Correlator compensation requirements for passive time-delay estimation with moving source or receivers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, pp. 158–168, Apr. 1980.

[9] R. J. Urick, *Principals of Underwater Sound.* McGraw-Hill, 3rd ed., 1983.

[10] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

[11] *MATLAB Reference Guide.* The MathWorks, Inc., Natick,MA, Aug. 1992.

[12] B. L. Daku and J. E. Salt, "Quality of underwater source localization in a multipath environment," *J. Acoust. Soc. Am.*, vol. 91, pp. 957–964, Feb. 1992.

Appendix A. Transforming Received TDOA to Emitted TDOA

Developing the expression for correcting difference in travel delay for sound signals received at time T to the time difference of arrival of sound signals emitted at time T was deferred to this appendix during Chapter 3. While the expression is necessary for determining the location of the contact, it is not closely related to the development of the time difference expressions.

An expression for ΔD was developed in Section 3.3.1. It represents the difference in delay for sound received at Sensor 1 and Sensor 2 at the end of the observation period, time T. Because the sound must have spent time travelling to the sensors from the contact, the sound must have been emitted earlier than time T. Generally, the time that the sound arriving at Sensor 1 spent travelling is different than the time that the sound arriving at Sensor 2 spent travelling. This implies that the sound received at Sensor 1 at time T was emitted at a different time than the sound received at Sensor 2 at time T.

Because the contact is moving, its location would have changed during the period between emitting the sounds that would eventually arrive at the two sensors at time T. Clearly, at different times the contact emits somewhat different sound, although the emitted sound is statistically stationary.

This is shown in Figure A.1. All lines are marked as distances—as the product of a time delay and the speed of either the contact or of sound.. The location of a moving contact is depicted for three instants in time, $p(t_1)$, $p(t_2)$, and $p(t_3)$. Delay $d_2(t_3)$ represents the delay already experienced by sound arriving at Sensor 2 at instant of time t_3. Since this delay is the longest, that sound was emitted earliest at time t_1. At t_1, the contact was located at position $p(t_1)$. Similarly sound arriving at Sensor 1 at

t_3, has experienced delay $d_1(t_3)$ and was emitted from the contact at time t_2 while it was at location $p(t_2)$. Delay $d'_2(t_3)$ is the delay that sound emitted from the contact at t_3 will yet experience until arriving at Sensor 2 in the future. Similarly $d'_1(t_3)$ is the delay that sound emitted at t_3 will experience travelling to Sensor 1.

The contact moves at a constant speed of s. The distance between the sensors is l.

The following relationships can be taken from the figure:

$$cd'_2(t_3) = cd_2(t_3) + sd_2(t_3) \quad \text{(A.1)}$$

$$cd'_1(t_3) = cd_1(t_3) - sd_1(t_3) \quad \text{(A.2)}$$

These two expressions are easily derived:

$$d'_2(t_3) = d_2(t_3) + \frac{s}{c}d_2(t_3) \quad \text{(A.3)}$$

$$d'_1(t_3) = d_1(t_3) - \frac{s}{c}d_1(t_3) \quad \text{(A.4)}$$

The desired quantity is $d'_1(t_3) - d'_2(t_3)$ so subtract (A.3) from (A.4):

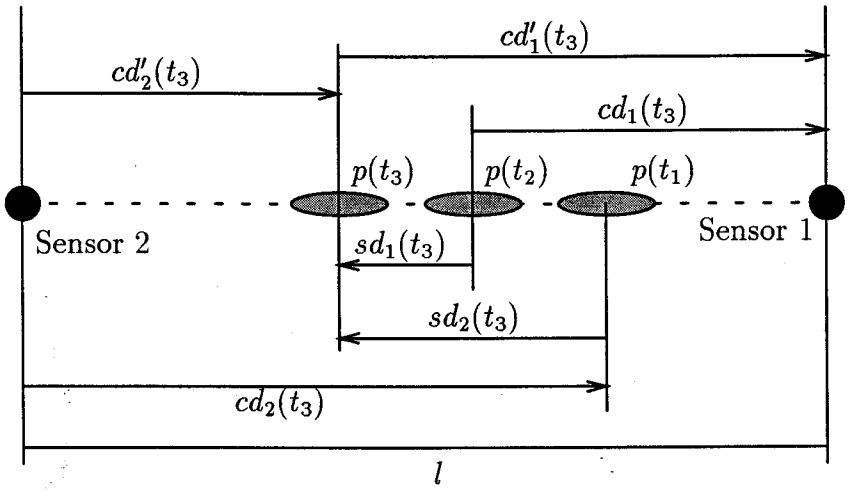

Figure A.1 Plan view of an ocean region showing the location of the contact at three instants in time.

$$d_1'(t_3) - d_2'(t_3) = d_1(t_3) - d_2(t_3) - \frac{s}{c}(d_1(t_3) + d_2(t_3)) \quad \text{(A.5)}$$

$$\Delta D' = \Delta D - \frac{s}{c}(d_1(t_3) + d_2(t_3)) \quad \text{(A.6)}$$

To determine the last term in (A.6), add (A.3) and (A.4):

$$d_1'(t_3) + d_2'(t_3) = d_1(t_3) + d_2(t_3) - \frac{s}{c}(d_1(t_3) - d_2(t_3)) \quad \text{(A.7)}$$

$$d_1(t_3) + d_2(t_3) = d_1'(t_3) + d_2'(t_3) + \frac{s}{c}(d_1(t_3) - d_2(t_3)) \quad \text{(A.8)}$$

$$d_1(t_3) + d_2(t_3) = \frac{l}{c} + \frac{s}{c}\Delta D \quad \text{(A.9)}$$

Substituting (A.9) into (A.6):

$$\Delta D' = \Delta D - \frac{s}{c}\left[\frac{l}{c} + \frac{s}{c}\Delta D\right] \quad \text{(A.10)}$$

$$\Delta D' = \Delta D\left(1 - \frac{s^2}{c^2}\right) - \frac{ls}{c^2} \quad \text{(A.11)}$$

Equation (A.11) gives the expression for calculating the TDOA of sound transmitted at some time t, given the difference in delay of sound arriving at time t. The expression uses the distance between the sensors and the speed of the contact (or at least an estimate of the contact speed).