

TRACKING MARINE MAMMALS USING PASSIVE ACOUSTICS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR IN PHILOSOPHY

IN

GEOLOGY AND GEOPHYSICS

DECEMBER 2007

By

Eva-Marie Nosal

Dissertation Committee:

L. Neil Frazer, Chairperson
Roy H. Wilkens
Fred K. Duennebier
Whitlow W.L. Au
Joseph R. Mobley

We certify that we have read this dissertation and that, in our opinion, it is satisfactory in scope and quality as a dissertation for the degree of Doctor of Philosophy in Geology and Geophysics.

DISSERTATION COMMITTEE

Chairperson

© 2007, Eva-Marie Nosal

ACKNOWLEDGEMENTS

First and foremost I thank L. Neil Frazer, my advisor and mentor, for his inspirational and intellectual guidance. I am very grateful to Roy Wilkens for his support and to my other committee members, Fred Duennebier, Whitlow Au, and Joe Mobley, for their interest and discussions. I am indebted to Greg Moore and Garrett Ito for providing access to their Linux cluster, to Torben Nielsen for my office Mac (especially the amazing display), and to the Maui High Performance Computing Center (MHPCC) for use of the Squall system. I thank Mike Porter for making his sound propagation code freely available and for his helpful suggestions. Ron Morrissey and Nancy DiMarzio at the Naval Undersea Warfare Center helped make the AUTEK datasets accessible. Funding was provided by a Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada, a Fellowship from the Research Corporation of the University of Hawaii, a Student Engagement Grant from the MHPCC (Sue Brown), and a Graduate Traineeship from the U.S. Office of Naval Research (Nicholas Chotiros and Ellen Livingston). On a personal note, I wish to express my deepest gratitude to my family for providing a solid and loving foundation and to Tom Fedenczuk for his unwavering support as my life partner.

ABSTRACT

It is difficult to study the behavior and physiology of marine mammals or to understand and mitigate human impact on them because much of their lives are spent underwater. Since sound propagates for long distances in the ocean and since many cetaceans are vocal, passive acoustics is a valuable tool for studying and monitoring their behavior. After a brief introduction to and review of passive acoustic tracking methods, this dissertation develops and applies two new methods. Both methods use widely-spaced (tens of kilometers) bottom-mounted hydrophone arrays, as well as propagation models that account for depth-dependent sound speed profiles.

The first passive acoustic tracking method relies on arrival times of direct and surface-reflected paths. It is used to track a sperm whale using 5 at the Atlantic Undersea Test and Evaluation Center (AUTEK) and gives position estimates that are accurate to within 10 meters. With such accuracy, the whale's pitch and yaw are estimated by assuming that its main axis (which points from the tail to the rostrum) is parallel to its velocity. Roll is found by fitting the details of the pulses within each sperm whale click to the so-called bent horn model of sperm whale sound production. Finally, given the position and orientation of the whale, its beam pattern is reconstructed and found to be highly directional with an intense forward directed component.

Pair-wise spectrogram (PWS) processing is the second passive acoustic tracking method developed in this dissertation. Although it is computationally more intensive,

PWS has several advantages over arrival-time tracking methods, especially in shallow water environments, for long duration calls, and for multiple-animal datasets, as is the case for humpback whales on Hawaiian breeding grounds. Results of simulations with realistic noise conditions and environmental mismatch are given and compared to other passive localization techniques. When applied to the AUTEK sperm whale dataset, PWS position estimates are within meters of those obtained using the time-of-arrival method.

TABLE OF CONTENTS

Acknowledgments	iii
Abstract	v
List of Tables	xi
List of Figures	xii
List of Abbreviations	xiv
Chapter 1. Introduction	1
1.1 Why track marine mammals with passive acoustics?	1
1.2 Overview of passive acoustic localization methods	4
1.3 Sound propagation modeling	10
Chapter 2. Direct-reflected time difference method	11
2.1 Paper 1:	12
Nosal E-M, LN Frazer (2006). Delays between direct and reflected arrivals used to track a single sperm whale. <i>Applied Acoustics</i> , 87 (11-12), 1187-1201.	
2.1.1 Introduction	12
2.1.2 Methods	14

2.1.2.1 Preliminaries	15
2.1.2.2 Detecting and classifying arrivals to establish DRTDs	17
2.1.2.3 Creating likelihood surfaces	22
2.1.3 Results	23
2.1.4 Error estimates	27
2.1.5 Discussion	31
2.1.6 Conclusions	33
2.2 Paper 2:	35
<p style="padding-left: 40px;">Nosal E-M, LN Frazer (2007). Sperm whale 3D track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones. <i>Journal of the Acoustical Society of America</i>, 122(4), 1969-1978.</p>	
2.2.1 Introduction	36
2.2.2 Data processing	39
2.2.2.1 Click detection, classification, and association	39
2.2.2.2 Level and pulse delay measurement	42
2.2.3 Localization	42
2.2.3.1 Methods	42
2.2.3.2 Error Estimates	45
2.2.3.3 Results	46
2.2.3 Swim Orientation	48
2.2.3.1 Conventions	48
2.2.3.2 Pitch and yaw	50

2.2.3.3 Roll	51
2.2.4 Beam pattern and directivity	54
2.2.4.1 Methods and results	54
2.2.4.2 Discussion	58
2.2.5 Click source levels	60
2.2.6 Concluding remarks	62
Chapter 3. Pair-wise spectrogram processing	63
3.1 Paper 3:	64
Nosal E-M, LN Frazer (2006). Pair-wise processing of spectrograms for localization of multiple broadband CW sources. <i>Newsletter of the IEEE Ocean Engineering Society</i> , Winter 2006.	
3.1.1 Introduction	64
3.1.2 Algorithms	67
3.1.3 Simulation Specifics	70
3.1.4 Discussion	74
3.1.5 Conclusion	76
3.2 Paper 4:	77
Nosal E-M, LN Frazer (in press). Modified pair-wise spectrogram processing for localization of unknown broadband sources. <i>IEEE Journal of Ocean Engineering</i> .	
3.2.1 Introduction	78
3.2.2 Overview of pair-wise waveform (PWW) processing	81

3.2.3 Relation between PWW, incoherent PWW (IPWW), and Bartlett processing	82
3.2.4 Reducing computational load for PWW processing	85
3.2.5 Overview of pair-wise spectrogram (PWS) processing	86
3.2.6 Reducing computational load for PWS processing	87
3.2.7 Simulated data	88
3.2.8 Localization method parameters and specifics	91
3.2.8.1 TOAD Method	92
3.2.8.2 Bartlett, PWW, and PWS processors compared	94
3.2.9 Results, discussion, and conclusion	95
3.3 PWS applied to AUTECH data	101
Chapter 4. Concluding remarks and future directions	102
Appendix	104
References	106

LIST OF TABLES

Table.2.1.1 Hydrophone positions	16
Table 3.2.1 Run times for the PWW and PWS processors	95
Table 3.2.2 Maximum values of surfaces in Figure 3.2.4	98

LIST OF FIGURES

2.1.1 Sound speed profile	16
2.1.2 Modeled DRTDs	17
2.1.3 Sperm whale waveform	18
2.1.4 Raw and processed waveforms	19
2.1.5 Spectrogram of the signal from Fig.2.1.4	20
2.1.6 Likelihood surfaces for the first 20 s of data	23
2.1.7 Estimated track (not smoothed)	25
2.1.8 Estimated x-y track	25
2.1.9 Smoothed three-dimensional track	26
2.1.10 Modeled DRTDs	29
2.1.11 Error contours	30
2.1.12 Error in x, y, and depth as a function of time	31
2.2.1 Paths taken by the p0 and p1 pulses	38
2.2.2 Estimated track (not smoothed)	47
2.2.3 Estimated two-dimensional track (not smoothed)	47
2.2.4 Half-widths of 95% confidence intervals for position	48
2.2.5 Angle approximation used in estimating the delay between p0 and p1	53
2.2.6 Estimated beam patterns	56
2.2.7 Estimated p1 beam pattern as a function of off axis angle	59

2.2.8 Source levels	61
2.2.9 Relative source level as a function of click number	62
3.1.1 Spectrogram of humpback whale signal used in simulations	70
3.1.2 Layout for 1 source simulation	71
3.1.3 Layout for 2 source simulation	72
3.1.4 Ambiguity surface with SNR 5 dB, 1 source, 3 receivers	72
3.1.5 Ambiguity surface with SNR 0 dB, 1 source, 3 receivers	73
3.1.6 Ambiguity surface with SNR -5 dB, 1 source, 3 receivers	73
3.1.7 Ambiguity surface SNR 0 dB, 2 sources, 4 receivers	75
3.2.1 Sound speed profile	89
3.2.2 Humpback whale signal used in simulations	89
3.2.3 Simulation configuration	90
3.2.4 TOAD method, Bartlett, PWW, and PWS ambiguity surfaces	97
3.3.1 Comparison DRTD/TOA method and PWS	101

LIST OF ABBREVIATIONS

AUTEC	Atlantic Test and Evaluation Center
CI	Confidence interval
CLF	Conditional likelihood function
DRTD	Direct-reflected time differences
FFT	Fast-Fourier transform
ICI	Inter-click interval
IPWW	Incoherent pair-wise waveform (processing)
MFP	Matched-field processing
MHPCC	Maui High Performance Computing Center
PAM	Passive acoustic monitoring
PWW	Pair-wise waveform (processing)
PWS	Pair-wise spectrogram (processing)
TOA	Time of (direct) arrival
TOAD	Time of (direct) arrival difference (between a pair of hydrophones)
NFFT	Number of points in the FFT window (used to create spectrograms)
NURC	Naval Undersea Warfare Center
PMRF	Pacific Missile Range Facility
SNR	Signal to noise ratio
SSP	Sound speed profile

Chapter 1

INTRODUCTION

1.1 Why track marine mammals with passive acoustics?

The past several decades have seen increased concern and controversy over the impact of human activity on marine mammal welfare [Green *et al.* 1994; Richardson *et al.* 1995; Malakoff 2001; NRC 2003; NRC 2005]. Anthropogenic noise in the ocean includes sound from shipping, naval operations, and seismic exploration. In response to noise, marine mammals change vocalizations rates, alter habitat use, move away from the source, lengthen songs, change respiration patterns, and possibly strand [Richardson *et al.* 1995; Frantzis 1998; Miller *et al.* 2000; Anon. 2001; Caldwell 2002; Jepson *et al.* 2003; Gordon *et al.* 2004; Taylor *et al.* 2004]. Among other things, the response is influenced by source level and frequency characteristics, sound propagation conditions, and the sensitivity of the animal [Richardson *et al.* 1995; DeRuiter *et al.* 2006; D'Spain *et al.* 2006;]. Since sound can propagate for long distances without suffering much attenuation, animals may be affected tens, hundreds, and even thousands of kilometers away from a source [Greene and Richardson 1988; Bowles *et al.* 1994; Nieuwkirk *et al.* 2004; Madsen *et al.* 2006]. In addition to short-term effects, long-term increases in ocean ambient noise [Curtis *et al.* 1999; Andrew *et al.* 2002; McDonald *et al.* 2006], potentially degrade

habitat by masking and interfering with acoustic signals that are used for communication, orientation, navigation, and detection of predators and prey [Payne and Webb 1971; Au 1993; Tyack and Clark 2000; Clark and Ellison 2003].

Methods used to study and monitor marine mammals in the wild include multi-sensor archival tags, visual surveys, and passive acoustics. Tags can provide detailed information on animal depth, orientation, physiology (including heart rate and breathing), and vocalizations [Schevill and Watkins 1966; Leatherwood and Evans 1979; Mate 1989; Goodyear 1993; Johnson and Tyack 2003]. They have facilitated several major advances in our understanding of the impact of noise on marine mammals as well as in behavioral studies and in bioacoustics [Watkins *et al.* 1993; Fletcher *et al.* 1996; Miller *et al.* 2004; Zimmer *et al.* 2005a; DeRuiter *et al.* 2006; Tyack *et al.* 2006; Watwood *et al.* 2006; Stimpert *et al.* 2007]. Disadvantages to tagging include logistical problems (tags are expensive and can be difficult to place on an animal), possibly altered behavior, and limited attachment time [Whitehead *et al.* 2000]. Furthermore, because they require that an animal be tagged, they cannot be used to detect animals for mitigation purposes. Most mitigation measures rely on trained visual observers aboard vessels who scan the sea surface for the presence of marine mammals [reviewed in Barlow and Gisinier 2006; Weir and Dolman 2007]. Visual methods play a key role in many cetacean studies and include photo identification [first described by Würsig and Würsig 1977; reviewed in Hammond *et al.* 1990], as well as aerial [e.g. Watkins and Schevill 1979; Würsig *et al.* 1984, 1993; Scott and Perryman 1991; Mobley *et al.* 1999; Mobley 2005, 2006], ship/boat-based [e.g. Johnston *et al.* 2007; Williams and Thomas 2007], and ground based studies [e.g. Würsig and Würsig 1979, 1980; Clark and Clark 1980; Tyack 1981; Noad and Cato 2007].

Limitations include sea state, daylight, and the amount of time an animal spends near the surface (as little as 5% of the time for some deep diving species [Barlow 1999]). Passive acoustic monitoring (PAM) methods can be used to detect animals that are submerged at any time of day, in poor weather conditions, and they are used extensively in studies of marine mammal behavior and movement [e.g. Leaper *et al.* 1992; McDonald *et al.* 1995; Stafford *et al.* 1998; Au *et al.* 2000; VanParijs *et al.* 2002]. PAM is limited to vocalizing animals and although most cetaceans are vocal, they may be silent for long periods of time and may silence in response to noise [Watkins and Schevill 1975]. PAM is also hindered by the incomplete repertoire representations for some species [Dawbin and Cato 1992; Mellinger *et al.* 2000].

The complementary nature of tagging, visual, and acoustic methods means that they can be especially useful in combination. For example, Zimmer *et al.* [2005a] used visual sightings and a towed hydrophone system to estimate echolocation source characteristics from a tagged sperm whale. Vocalizations recorded from tagged whales improve PAM capabilities by adding to the known repertoire [Johnson *et al.* 2004; Stimpert *et al.* 2007]. Combining visual and acoustic detection methods can improve tracks and increase the detection probability, although methods to relate acoustic and visual detection statistics to the true population need further development [Ko *et al.* 1986; Frankel *et al.* 1995; Noad and Cato 2001, 2007; Clark and Fistrup 1997; Tiemann *et al.* 2006].

PAM is useful on its own for census efforts and behavioral studies, particularly for continuous, long-term monitoring [Clark and Ellison 1988; McDonald *et al.* 1995; Stafford *et al.* 1998; Norris *et al.* 1999] and in hostile or inaccessible areas [Wartzok *et*

al. 1992]. PAM can also be used for bioacoustics research on free-ranging animals [Au *et al.* 1974, 1986, 1987, 2002, 2006; Møhl *et al.* 1990; Thode *et al.* 2002; Wahlberg 2002; Au and Benoit-Bird 2003; Au and Herzing 2003; Rasmussen *et al.* 2004; Madsen and Wahlberg 2007]. Critical information regarding the biosonar of marine mammals has been derived from tests with trained or captive animals [reviewed in Au 1993; Thomas *et al.* 2004], but such studies are limited to smaller species, and biosonar performance might differ for free-ranging animals [Au *et al.* 1974, 2004; Au and Herzing 2003; Madsen *et al.* 2004a,b].

1.2 Overview of passive acoustic localization methods

Passive acoustic monitoring refers to the use of acoustic signals to detect, classify, and localize calling animals. Detection and classification often require sophisticated techniques, and the development of these methods is an active area of research [Altes 1980; Mellinger and Clark 1993; Potter *et al.* 1994; Mellinger 2000, 2004; Chesmore 2001; Oswald *et al.* 2003, 2007; Gillespie 2004; Roch *et al.* 2007]. Except for the detection algorithm used in the first two papers of this dissertation, detection and classification are not dealt with here. Passive acoustic localization (estimation of position) and tracking (taking positions estimates over time) are also active areas of research, and they are the topic of this dissertation.

The use of acoustics to track marine life was pioneered in the 1960s and 1970s. Walker [1963] tracked sources of 20-Hz pulses, apparently from a whale, using three

hydrophones. Cummings *et al.* [1964] found the positions of fish and invertebrates, also using three hydrophones. Watkins and Schevill [1972, 1977] used a three-dimensional array with four hydrophones to track the movements of individual whales. Since that time, many studies have used a variety of passive acoustic localization methods and hydrophone configurations to track marine mammals [e.g. Cummings and Holliday 1985; Møhl *et al.* 1990; Freitag and Tyack 1993; Clark *et al.* 1994; Stafford *et al.* 1998].

A commonly used method of passive acoustic localization is known as the time-difference of arrival (TOAD) method (also known as multilateration or hyperbolic positioning/fixing). TOAD methods are useful in a broad range of applications: civilian and military applications to locate aircraft, submarines, ground vehicles, and stationary sources such as explosions, geophysical applications to monitor seismicity [e.g. Fox *et al.* 1995], terrestrial biological applications to estimate animal positions [e.g. Mennill *et al.* 2006]. TOAD methods have been used to track almost every imaginable source of sound, from human speakers to gunshots [Lahr and Fischer 1993; Vermaak and Blake 2001; Bucher and Misra 2002].

In the TOAD method, a signal reaches two spatially separated receivers at different times because of different propagation path lengths from the source to the receivers. For known receiver positions, the locus of possible source locations is a hyperboloid. A third receiver provides another TOAD measurement, which defines a second hyperboloid and a line of possible source locations is defined by the intersection of these two hyperboloids. A fourth receiver defines a third hyperboloid, with the intersection of all three hyperboloids defining a point, which is the estimated source location. Note that each additional receiver actually adds as many TOADs (hence

hyperboloids) as there were receivers, but only one new TOAD is unique. Also, depending on the receiver configuration, the intersection of the three hyperboloids may be two points (or infinitely many points in a degenerate case), in which case a fifth hydrophone is required to localize a source in three dimensions [Tyrrell 1964; Spiesberger 2001].

The first step in TOAD methods is to estimate the signal time delay between each pair of hydrophones. The most commonly used method is correlation, in which the estimate is the time-lag that maximizes the cross-correlations between the received signals [Knapp and Carter 1976]. For marine mammal applications, both filtered waveforms [Clark *et al.* 1986; Spiesberger and Fristrup 1990; Mitchell and Bower 1995; Janik *et al.* 2000; Tiemann *et al.* 2004] and spectrograms [Altes 1980; Clark *et al.* 1986; Frankel *et al.* 1995; Janik *et al.* 2000; Clark and Ellison 2003; Tiemann *et al.* 2004] of the recorded signals have been used in the cross correlation. TOADs can also be estimated by using a matched-filter approach if a template of the call can be estimated [Stafford *et al.* 1998]. The best method to use to estimate TOAD varies depending on the signal, noise, and propagation characteristics of the problem.

The second step in the TOAD method is to find the point of intersection (or the closest such point if intersection is imperfect) of the hyperboloids. Assuming constant speed of sound propagation, the problem can be expressed as a system of linear equations. For a well-defined problem (not underdetermined/overdetermined by too few/many receivers), a closed form solution to this system gives the source location [e.g., Schmidt 1972; Watkins and Schevill 1971; Schau and Robinson 1987; Delsome *et al.* 1980; Spiesberger and Fristrup 1990; Wahlberg *et al.* 2001]. For overdetermined systems,

a least-squares approach can be used to give the best source position [Spiesberger and Fristrup 1990; Wahlberg *et al.* 2001]. Reflections from the bottom and surface can be treated as recordings made by virtual receivers and incorporated into the solution [Urick 1983; Møhl *et al.* 1990; Aubauer *et al.* 2000; Wahlberg *et al.* 2001]. Using reflections improves the accuracy of estimated source positions [Møhl *et al.* 1990; Wahlberg *et al.* 2001; Thode *et al.* 2002] and fewer real receivers (as few as one) are required for localization [Aubauer *et al.* 2000; Tiemann *et al.* 2006; Laplace 2007].

Error analysis [reviewed in Taylor 1997] can be approached by comparing locations obtained by different receiver subsets, or by comparison with positions determined visually [Cleator and Dueck 1995; Smith *et al.* 1998; Aubauer *et al.* 2000; Janik *et al.* 2000]. Another approach involves linear error propagation and considers uncertainties in sound speed, receiver position, ray bending, and TOAD measurement [Watkins and Schevill 1971; Spiesberger and Fristrup 1990; Wahlberg *et al.* 2001].

The TOAD method can be used with different hydrophone configurations, each with its own strengths and weaknesses, to meet various monitoring and research requirements. For example, a closely spaced planar array is better for echolocation research [Au *et al.* 2002; Rasmussen *et al.* 2002; Au and Herzing 2003], but it is limited to short distances and animal positions that are perpendicular to the plane of the array. A widely spaced array is ideal for tracking in three dimensions over long distances [Stafford *et al.* 1998; Tiemann *et al.* 2004], but applications are limited to sufficiently loud calls (so that they can be heard on several hydrophones) and for logistical reasons (e.g. cost, clock synchronization, and hydrophone position uncertainties). As refraction becomes significant at long distances, widely spaced arrays are also more sensitive to assumptions

of straight-line sound propagation, and sound propagation models may be required to obtain accurate position estimates [Chapman 2004; Tiemann *et al.* 2004].

A somewhat different approach, which is a version of beamforming [Johnson and DeGraag 1982], is required to localize a source using a towed line array [Leaper *et al.* 1992; Gillespie 1997; Barlow and Taylor 2005]. This method estimates the bearing to a sound relative to the tow cable axis from the TOAD measured between two hydrophones spaced a few meters apart by assuming plane wave propagation. Range is estimated via a time-motion analysis of the changes in estimated bearing as the platform moves. The method requires that the speed of the vessel be much greater than the speed of the vocalizing animal, that the animal vocalizes continuously for several minutes, and that individuals vocalizing simultaneously can be distinguished – conditions that are not always met [Thode 2005]. Furthermore, this method does not distinguish between horizontal and vertical range.

Time of arrival is not the only component of a recorded call that contains information about animal location; phase and sound pressure levels can also be useful. Directional hydrophones can be used to estimate the direction to high-frequency vocalizations [Whitehead and Gordon 1986] and for hydrophones configured less than a wavelength apart, differences in phase can be used to estimate the bearing to animal calls [Clark 1980]. Cato [1998] proposed a method using only the differences in received levels to estimate position, although we note here that this method might not be applicable for animals with high directionality. If the source level of the call is known, then it may be possible to estimate the distance to an animal from the received level.

However, data on source levels are scarce, individuals can vary their source levels, and directionality will again confound the problem.

Matched-field processing (MFP) uses all available information (timing, sound pressure level, and phase) to estimate source position (among other applications). The underwater acoustics community developed MFP in the 1980s and 1990s for naval purposes. Possibly because implementation can be costly and computationally demanding, MFP has seen only limited application in marine mammal localization problems [Thode *et al.* 2000; Tiemann *et al.* 2004; Thode 2005]. Very briefly, localizing a source via MFP involves predicting the receiver response given a source at some candidate source position and then comparing the predicted response with the measured response. This process is repeated for each point in a grid of candidate source positions, and the candidate position giving the best agreement between predictions and measurements is chosen as the estimated source position. Details can be found in reviews by Tolstoy [1993] and Baggeroer *et al.* [1993] and the references therein. Since MFP was designed to exploit the phase information, it traditionally requires large line arrays and low frequency sources (higher frequencies become incoherent for long distance propagation).

Although passive acoustic methods for marine mammal localization yielded much useful information over the past 30 years, there is still a need for improvement. Most effort has focused on short duration calls (such as clicks) because arrival time estimates are relatively simple in these cases, while longer duration calls have reflections overlapping direct arrivals. There is a need to improve methods for localizing long duration calls made by many species (baleen whales in particular). Also, methods that

increase the accuracy of estimated animal positions, for example by using more realistic sound propagation models, will facilitate behavioral studies by providing more detailed information on animal movement. Another important goal of passive acoustics is to track multiple individuals simultaneously for insight into marine mammal communication and interaction. Real-time methods that enable continuous, long-term monitoring are the long-term goal.

1.3 Sound propagation modeling

Marine mammal localization methods are usually implemented with an isospeed (constant sound speed) assumption. In many cases, particularly those involving propagation over long distances, this can lead to inaccurate position estimates [Chapman 2004; Tiemann *et al.* 2004; Nosal and Frazer 2006a]. To overcome this problem, this dissertation uses the Gaussian beam acoustic propagation model BELLHOP [Porter and Bucker 1987; Porter and Liu 1994; Porter 2005] to model sound propagation through the ocean. For a given frequency ω and fixed source/receiver positions, BELLHOP uses environmental inputs (sound speed profiles, surface and bottom properties, and so forth) to calculate the arrival time, t_k , and complex amplitude, a_k , of the k^{th} beam at the receiver. These Green's function is then a sum of beams:

$$G(\omega) = \sum_k e^{i\omega t_k(\omega)} a_k(\omega). \quad (1.1)$$

Chapter 2

DIRECT-REFLECTED TIME DIFFERENCE METHOD

The first paper (Section 2.1) of this Chapter introduces and applies a method that relies strictly on the differences in the time of arrival of direct and surface reflected paths (DRTD). Unlike the TOAD method, the DRTD method is insensitive to receiver timing offset, which makes it ideal for problems where receivers cannot be synchronized. It is potentially useful to estimate and correct the timing offset between receivers using a source of opportunity (such as a whale call). The second paper (Section 2.2) combines the DRTD method with a TOAD method to obtain the track of a sperm whale with accuracy to within 10 m. A method to estimate the orientation of the whale is developed and used to recover the beam pattern of sperm whale clicks.

Both the DRTD and the combined TOAD/DRTD methods are implemented using a simplified MFP approach; times of arrivals are predicted for every point on a grid of candidate source positions, then predictions are compared with measured values to find the whale position. Probability density functions are used to quantify error caused by uncertainties in receiver position, sound-speed profiles, and TOA measurements. An acoustic sound propagation model is used to account for depth-dependent sound speed variations. This is important for widely spaced hydrophone arrays because refraction can be significant for propagation over long distances.

2.1 Paper 1

Nosal E-M, LN Frazer (2006). Delays between direct and reflected arrivals used to track a single sperm whale. *Applied Acoustics*, 87 (11-12), 1187-1201.

Abstract – One dataset made available for the 2nd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics featured a single sperm whale recorded for 25 minutes on 5 widely spaced, bottom-mounted hydrophones in the Tongue of the Ocean. In this paper, we track the whale using a model-based method that relies on the difference in arrival times along direct and surface-reflected propagation paths. Four receivers are required to estimate positions in three dimensions. Details of the method are presented, and tracks are estimated using an isospeed and a depth-dependent sound speed profile (SSP). Depth estimates for the isospeed SSP are about 50–100 m shallower than for the depth-dependent SSP, and horizontal positions are similar. Performance estimates indicate that the depth-dependent SSP results are more accurate, with estimated depths of 650–760 m and average vertical and horizontal swim speeds (not accounting for current) of 0.30 m/s and 2.40 m/s, respectively.

2.1.1 Introduction

Recordings of sperm whale vocalizations in the Tongue of the Ocean were made available to participants of the 2nd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics. The datasets are from March 23 and 30, 2002 and were prepared by the Naval Undersea Warfare Center (NUWC). In the March

23 dataset a single sperm whale is clearly heard on all 5 bottom-mounted hydrophones for the full 25 minutes of recordings.

As they are the predominant vocalization present in the recordings, this work concentrates on the so-called regular (or usual) clicks emitted by sperm whales during deep dives [Whitehead and Weilgart 1991]. Regular clicks have inter-click intervals (ICIs) of 0.5–2 s [Goold and Jones 1995; Jaquet *et al.* 2001], duration of about 10–20 ms [Goold and Jones 1995], and energy from below 100 Hz to above 20 kHz [Watkins *et al.* 1993; Zimmer *et al.* 2005a]. Due to these characteristics, as well as the deep-water environment and bottom-mounted hydrophones used for the recordings, direct and surface-reflected arrivals can be easily identified. Specifically, reflected arrivals come shortly after their associated direct arrivals and they have lower amplitude, less power at high frequencies, and slightly longer duration. Moreover, the effect of source-receiver spacing on the time delay between corresponding direct and surface-reflected rays (DRTD) is clearly audible; DRTDs decrease with increasing source/receiver separation. For a moving sperm whale DRTDs vary between receivers as well as with time on a single receiver. Cato [1998] and Aubauer *et al.* [2000] explain this effect for an isospeed SSP.

Motivated by this dataset, we developed and implemented a tracking method that relies entirely on DRTDs. DRTDs have previously been exploited for localization [Cato 1998; Aubauer *et al.* 2000; Thode 2005; Skarsoulis and Kalogerakis 2005], but they have typically been used with isospeed SSPs to establish range, and not (to our knowledge) to estimate a three dimensional track using widely spaced receivers, as is done here. A ray-tracing model that accounts for the depth-dependent sound-speed SSP (SSP) gives

different, and presumably more accurate, estimates than an isospeed SSP. We also give approximate error maps for depth and x-y coordinates of location. Unfortunately, no visual or tagging data is available to verify the estimated track of the sperm whale. Nevertheless, the estimated track is consistent with other observations of sperm whale behavior, which lends confidence to our predictions.

2.1.2 Methods

Before giving the details of the method, we first provide a general overview, noting that at least four receivers are required for the localizations. Signals are subdivided into short time intervals, and a list of candidate source depths is created. Each time interval and depth is processed separately. At each receiver, the DRTD is established by click detection scheme. These measured DRTDs are compared to modeled DRTDs to estimate the horizontal separation of source and receiver. This separation defines (the radius of) a circle centered at the receiver. If the search has been conducted at the correct source depth, the arrival times have been accurately determined, and the environment has been perfectly modeled, all receiver circles intersect at a single point, which is the position of the source. In most cases, however, the circles do not intersect at a single point, and a point of best agreement must be determined. This is accomplished by creating a likelihood surface (a probabilistic indicator of source location sometimes referred to as an ambiguity surface) for each receiver; the surfaces are assigned value 1 along the circles (highest probability) and decay according to a Gaussian weighting away from the circle, both inward and outward. Likelihood surfaces for each receiver are averaged to create a total likelihood surface at the current search depth. This process is

repeated for all candidate source depths, and the point with maximum likelihood is declared the overall estimated source position at that time. Only positions at times that give sufficiently large likelihood are retained.

2.1.2.1 Preliminaries

The signal at each receiver is sub-divided into short time intervals, typically several tens of seconds long, which can overlap. Two factors are considered when choosing interval lengths. First, longer intervals contain more clicks, which helps to reduce errors in estimated DRTDs. For intervals that are too long, however, movement of the whale may result in significant variation of the DRTDs within the interval. Through trial and error, 20 s intervals (which typically contained between 10 and 25 clicks) were chosen as a good compromise for the workshop dataset. A 15 s overlap was used since it gave good time resolution for the track while keeping run-times reasonably low.

Next, a look-up table of predicted DRTDs as a function of range for all receivers and candidate source depths is created. Hydrophone positions (Table 2.1.1) were provided by NUWC. All phones were 17 feet off the bottom except K, which was 18 feet off the bottom. In this work, the Gaussian beam acoustic propagation model BELLHOP [Porter and Liu 1994] was used to model the environment and create this table. The range list varied from 5 m to 10 km in 5 m increments. Since the hydrophones were all within 7.5 km of one another, this allowed for searches several kilometers beyond the boundary of the receiver array. Candidate depths covered the entire water column at 10 m resolution. Two sound speed profiles (SSP) were used. One was an isospeed SSP with a sound-speed of 1510 m/s; the other was a depth-dependent SSP, the average historical

SSP from the Tongue of the Ocean for March, taken from the Generalized Digital Environment Model [GDEM] (Fig.2.1.1). Figure 2.1.2 shows modeled DRTDs as a function of horizontal separation for one hydrophone and three candidate source depths.

Table.2.1.1. Hydrophone positions provided by NUWC

Hydrophone	x-position (m)	y-position (m)	depth (m)
G	10658.04	-14953.63	1530.55
H	12788.99	-11897.12	1556.14
I	14318.86	-16189.18	1553.58
J	8672.59	-18064.35	1361.93
K	12007.50	-19238.87	1522.54

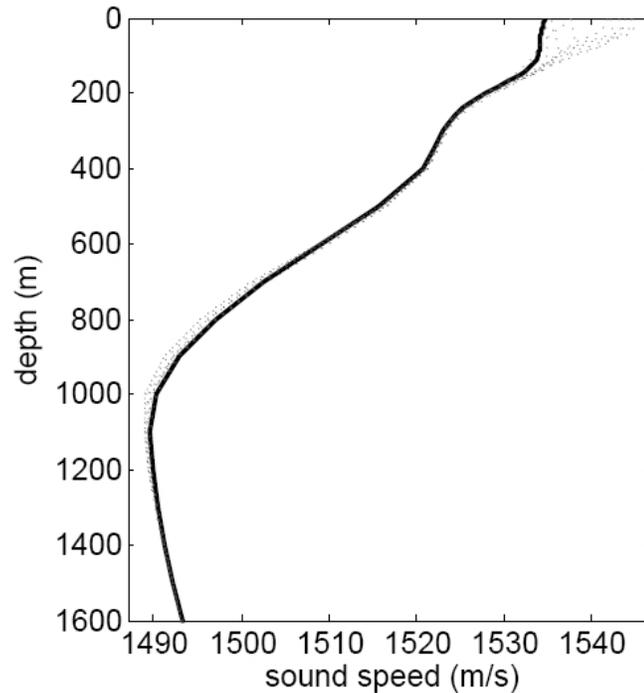


Fig 2.1.1 Historical SSPs for the region. The SSP for March, which is when the data was collected, is shown in bold and was used to model DRTDs.

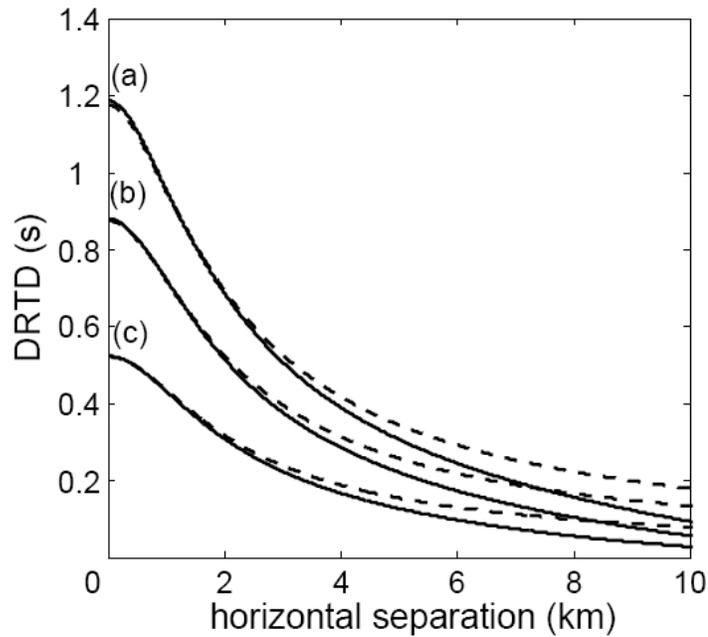


Fig 2.1.2 Modeled DRTDs for hydrophone H as a function of horizontal separation for source depths of (a) 890 m, (b) 670 m, and (c) 400 m. Solid lines are for the depth-dependent SSP for March; dashed lines are for an isospeed SSP of 1510 m/s.

2.1.2.2 Detecting and classifying arrivals to establish DRTDs

As mentioned in the introduction, the characteristics of the source and the environment make it easy to identify direct and surface-reflected arrivals in the time series. Short-duration calls result in no overlap between associated direct and surface-reflected arrivals, and their broadband nature can be used to reduce noise (see next paragraph). The ICI is usually long enough so that a reflected arrival precedes the direct arrival from the next click. The deep-water environment reduces complications from multiple arrivals, and bottom-mounted hydrophones mean that bottom reflections arrive immediately after direct signals, so they are not confused with surface reflections. Refer to Figure 2.1.3 for a waveform of a typical sequence of direct and reflected arrivals.

Direct arrivals are high amplitude and quite evenly spaced, with reflected arrivals between them. In some cases the direct and reflected arrivals are not so clear (see Fig.2.1.4 (a)).

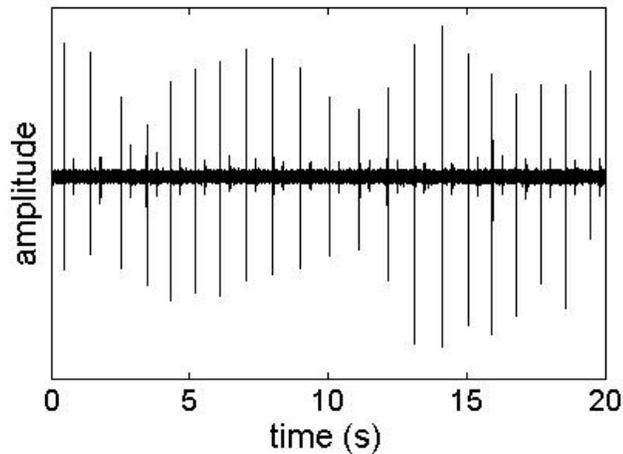


Fig 2.1.3 Waveform at hydrophone J for the data segment beginning at 170 s. Direct arrivals have large relative amplitude. Surface reflections come between the direct arrivals and have smaller relative amplitude.

To improve the detection process for such difficult cases, a spectrogram method was employed. The spectrogram method exploits the broadband nature of the sperm whale clicks. First, a complex spectrogram is created from the hydrophone signal, which has a sampling frequency of 48 kHz, via a short-time Fourier transform. A high-pass filter is applied to the time series of each frequency channel in the spectrogram. Since sperm whale clicks are less than 25 ms in duration, the filter cutoff is set at 40 Hz. Filtering is done in the frequency domain using half of a Hanning window to roll-off with an 80 Hz transition bandwidth. This reduces slowly varying sounds, such as tonal noise from equipment or boats. After filtering, magnitudes are taken of the filtered spectrogram points, and each frequency channel is divided by the mean of the entire channel; this de-

emphasizes the lower frequencies that have more background noise. Finally, the frequency channels are summed to give a time signal with lower background noise and coarser time resolution than the original signal. The time resolution of the channel sum depends on the duration and overlap of the windows used in the discrete FFT transforms. Figure 2.1.5 shows the spectrogram and the processed spectrogram for the signal in Fig.2.1.4(a) using 512-point Hanning windows with 256-point overlap. The channel sum (hereinafter referred to as the filtered signal) is shown in Figure 2.1.4(b).

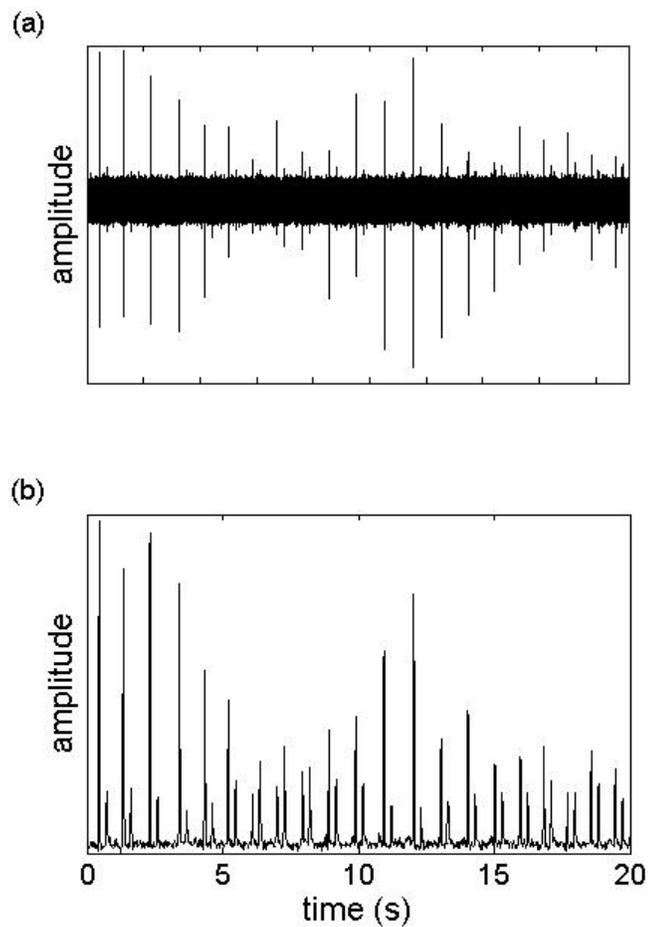


Fig 2.1.4(a) Waveform at hydrophone K for the data segment beginning at 170 s. Noise completely covers the reflected arrivals. (b) After applying the spectrogram method, noise is significantly reduced and reflected arrivals are clearly identifiable.

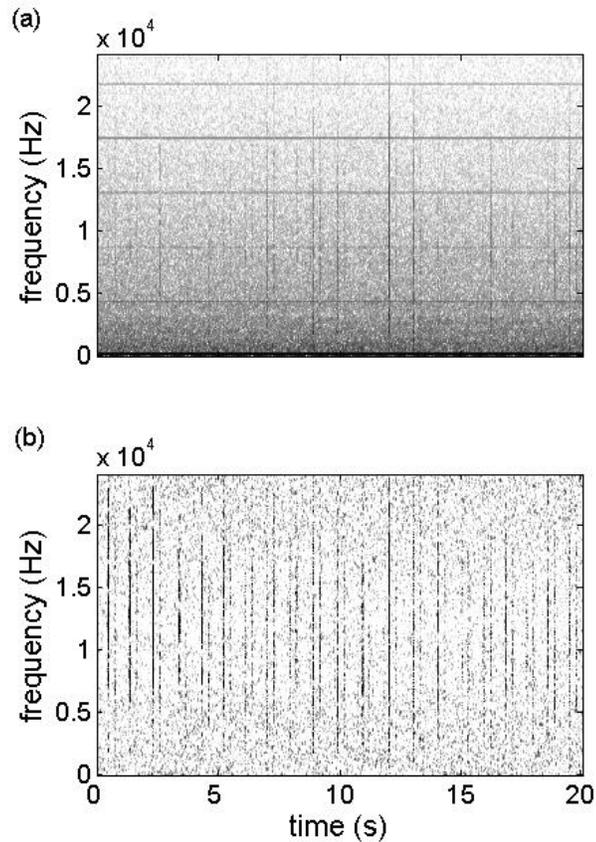


Fig 2.1.5 Spectrogram of the signal from Fig.2.1.4 (a) created using 512-point Hanning windows with 256-point overlap. (a) Original spectrogram; and (b) after high-pass filtering and dividing each frequency channel by its mean. Note in (b) that the tonals have been removed, and the low frequencies (with significant background noise) have been de-emphasized.

Peaks in the filtered signal are classified as direct arrivals if they have amplitude greater than some threshold percent of the mean of the three largest amplitudes in the current time interval. For the results presented here, a 50 % threshold was used. This was chosen by trial and error via visual inspection of signals and detected arrivals in numerous cases. Too high of a threshold caused many direct arrivals to be missed; too low a threshold caused some reflected arrivals to be incorrectly classified as direct arrivals.

The peak with maximum amplitude arriving between two classified direct arrivals was classified as the surface-reflected arrival corresponding to the direct arrival preceding it. Since direct arrivals were not always entirely impulsive (*i.e.*, they had non-zero duration), and since the bottom-reflected arrival could sometimes be seen immediately following the direct arrival, care was taken not to look for the surface reflection too soon after the direct arrival; arrivals within 100 ms of the direct arrival were not included in the search for surface reflections. Further, since the surface-reflected arrivals were expected to have amplitudes between about 10% and 75% of the amplitude of the corresponding direct arrival, pairs that did not meet these expectations were discarded. Having classified direct and surface-reflected arrivals pairs, DRTDs were computed from their relative arrival times. The median of all resulting times was declared the representative DRTD for the current time window and receiver. A median was taken instead of a mean, since misclassifications can result in large outliers that significantly affect the mean.

In some cases, DRTDs could not be established on all receivers. This occurred at times when the whale was silent, as well as for various cases when the detection scheme failed. This included cases with very poor signal-to-noise ratios and those in which a surface-reflected signal arrived immediately before, at the same time as, or after the subsequent direct arrival (due to short ICIs). Time intervals in which a DRTD could not be established on at least four hydrophones were discarded.

2.1.2.3 Creating likelihood surfaces

The following steps are repeated for all candidate source depths. The horizontal separation with modeled DRTD closest to the measured DRTD is found for each receiver. To create a likelihood surface, a grid is created that covers the horizontal plane of interest. For the workshop dataset, the grid range used was 7000 m to 16000 m N-S and -21000 m to -10000 m E-W, with 10 m resolution in both directions. For each receiver, the likelihood value is a function of each grid point's radial distance from the perimeter of a circle centered at the receiver with radius given by the horizontal separation corresponding to that receiver. A Gaussian weighting function, with standard deviation 500 m, was found to work well for the current dataset. Likelihood surfaces for all receivers on which DRTD could be established are averaged to give the total likelihood surface (with value between 0 and 1) at the current candidate source depth. Figure 2.1.6 shows the likelihood surface at three different depths. When the candidate source depth is too shallow or too deep, the receiver circles do not intersect closely, resulting in lower maxima. The positions and values of the maxima are stored.

After this process has been completed for all candidate source depths, the point with the maximum likelihood value is chosen as the estimated source position. Smaller maximum likelihood values indicate greater uncertainty in the source position. Times with likelihood below a preset level are discarded as having too much error. For the workshop data, a threshold of 0.850 was used.

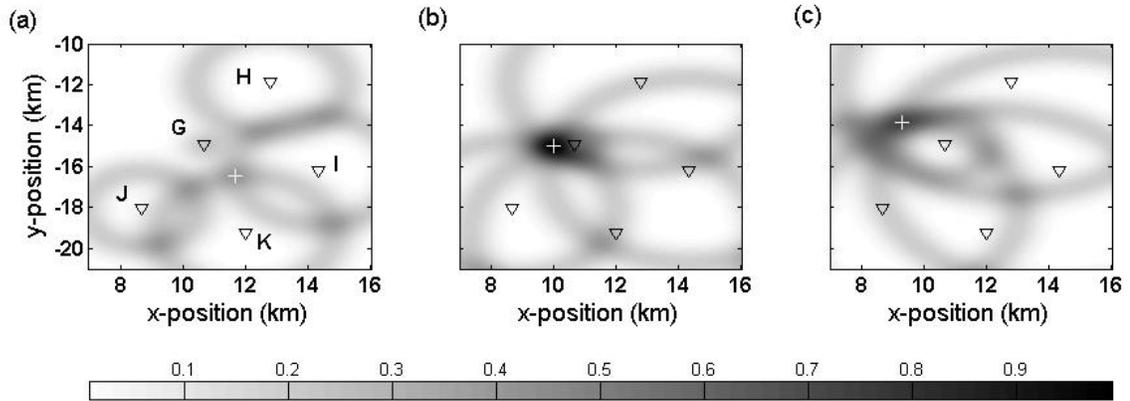


Fig 2.1.6 Likelihood surfaces for the first 20 s of data at depths of (a) 400 m (b) 670 m (c) 890 m. Triangles indicate receiver locations and are labeled in (a). White crosses mark position estimates, with surface values (a) 0.401 (b) 0.997 (c) 0.729. The estimated source position, chosen from (b), is [10010 m, -15020 m, 670 m].

2.1.3 Results

The process was automated by a collection of MATLAB codes. No effort was made to optimize the code for efficiency. Modeling of the environment (calculation of predicted DRTDs) took less than a minute and was done once for each SSP. After this, run times were about three times real-time on a 2.8 MHz Pentium IV; 5 minutes of data took about 17 minutes to process. Using coarser time and/or space resolution can reduce run times. Also, a more intelligent search can increase the efficiency of the algorithm (*e.g.*, the swim speed of the whale is limited so it is not necessary to search the entire water column if the position of the whale is established for previous time intervals).

Results are shown in Figure 2.1.7 as estimated position (x-, y-, and depth) versus time. In these results, 13% of all time intervals were eliminated because DRTDs could not be established on 4 or more receivers. Of the remaining time intervals, 24% were eliminated in the isospeed SSP case and 20% were eliminated in the depth-dependent

SSP case because maximum likelihood values were less than 0.850. The mean of the maximum likelihood values for the remaining points is 0.920 for the isospeed SSP and 0.998 for the depth-dependent SSP, suggesting that the depth-dependent SSP results are more accurate. The x- and y- tracks for the isospeed SSP and the depth-dependent SSP are quite similar. The depth tracks are also similar, with the depth-dependent SSP track between 650 m and 760 m and about 100 m deeper than the isospeed track. This difference reiterates [Aubauer *et al.* 2000; Thode 2005; Chapman 2004] the importance of incorporating the effects of a depth-dependent SSP into methods for tracking marine mammals over long ranges using passive acoustics. In Figure 2.1.8, the track from the depth-dependent SSP is plotted in the x-y plane with the hydrophone positions. Figure 2.1.9 shows the smoothed track from the depth-dependent SSP in three dimensions. A five-point moving average filter was used along each direction to accomplish the smoothing.

The average swim speed was estimated from the smoothed track by calculating the velocity for each time step and taking the mean over all times. The vertical swim speed was 0.30 m/s and the horizontal swim speed was 2.40 m/s. It is not known what the current was in the area at the time, so the horizontal swim speed relative to the current cannot be estimated.

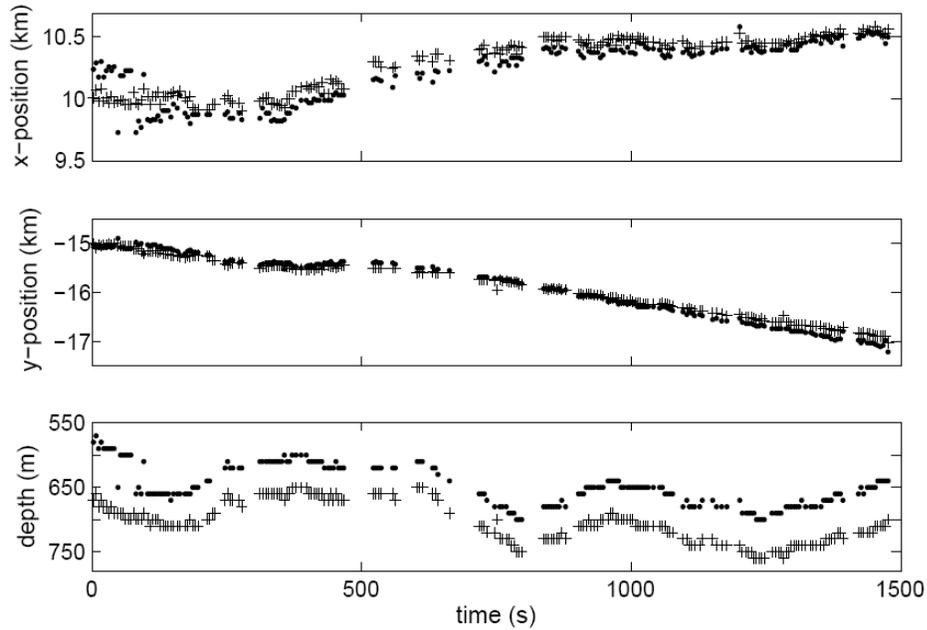


Fig 2.1.7 Estimated track (not smoothed) of the sperm whale with the isospeed SSP (dots) and the depth-dependent SSP (crosses). Times with performance values less than 0.850, and those for which DRTDs could be established on only 3 (or fewer) receivers, have been eliminated.

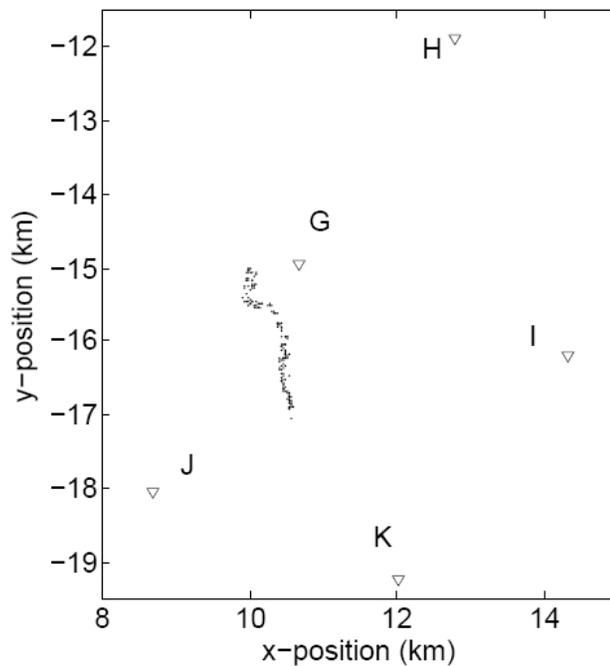


Fig 2.1.8 Estimated two-dimensional track (not smoothed) of the sperm whale with the depth-dependent SSP. Triangles indicate receiver positions.

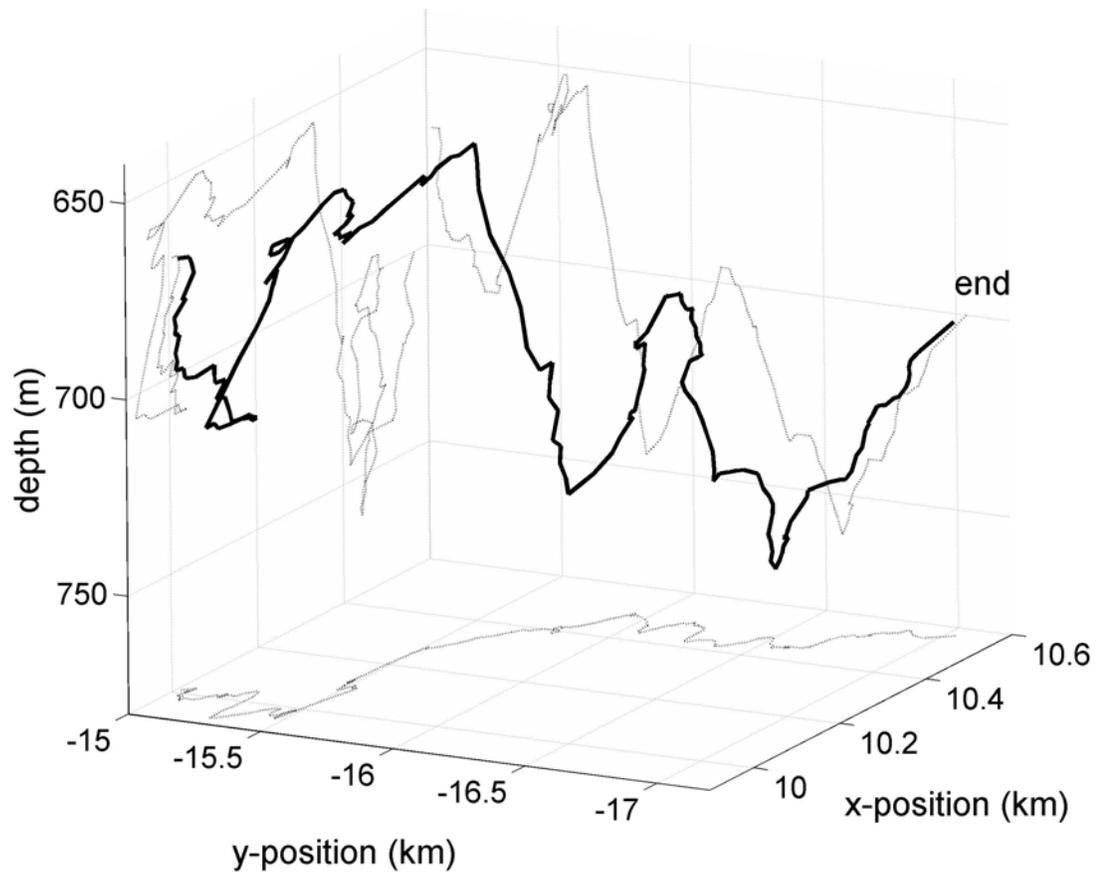


Fig 2.1.9 Smoothed three-dimensional track (solid line) estimated using the depth-dependent SSP. Projections onto the three planes are shown with dashed lines.

2.1.4 Error estimates

To estimate the error in source position associated with the method presented here, error maps of the array for x- y- and z- directions were created under the assumptions that sources of error are independent, error in one direction is independent of error in the other directions, and errors are normally distributed. As in time-difference of arrival methods [Wahlberg 2004], there are three main input variables (hence sources of error) associated with the DRTD method: sound speed SSP, measured DRTD, and receiver position. Since no measurement of receiver position error was available, we deal only with errors due to SSP and DRTD.

First, consider errors in the x- and y- directions. For receiver i and position $\bar{p} = (x_p, y_p, z_p)$, depth is fixed to find $\sigma_{r,i}$, the standard deviation in range for receiver i (see Fig. 2). This is a combination of the error due to the SSP, $\sigma_{ssp,r,i}$, and the error due to DRTD, $\sigma_{drt,d,r,i}$. Letting t_d be the modeled DRTD between receiver i and \bar{p} for a reference SSP, we approximate $\sigma_{ssp,r,i}$ as the difference between the minimum and maximum ranges corresponding to t_d over all possible SSPs, and $\sigma_{drt,d,r,i}$ as one fourth of the difference between the ranges corresponding to $t_d \pm 2\sigma_{drt,d}$ for the reference SSP, where $\sigma_{drt,d}$ is the estimated standard deviation of DRTD measurements. To the circle passing through \bar{p} , centered on receiver i , we attach a Gaussian PDF with standard deviation

$$\sigma_{r,i} = \sqrt{\sigma_{ssp,r,i}^2 + \sigma_{drt,d,r,i}^2}.$$

To estimate error in the x-direction from the combined PDFs at all receivers, each PDF is approximated as locally linear. In other words, the PDF for each receiver is

approximated by a ridge whose axis is the line tangent to the corresponding circle at \bar{p} .

Let θ_i denote the angle from the x -axis to the i^{th} hydrophone, then the combined PDF at point (x, y_p, z_p) is:

$$\begin{aligned} p(x) &\propto \prod_i \exp \left[-\frac{(x-x_p)^2 \cos^2(\theta_i)}{2 \sigma_{r,i}^2} \right] \\ &= \exp \left[-\frac{(x-x_p)^2}{2} \sum_i \frac{\cos^2(\theta_i)}{\sigma_{r,i}^2} \right]. \end{aligned} \quad (2.1.1)$$

Normalization is automatic as:

$$p(x) = \frac{1}{\sqrt{2\pi} \sigma_x} \exp \left[-\frac{(x-x_p)^2}{2\sigma_x^2} \right]; \quad \text{in which } \sigma_x = \left(\sum_i \frac{\cos^2(\theta_i)}{\sigma_{r,i}^2} \right)^{-1/2}. \quad (2.1.2)$$

Similarly for error in the y -direction, the combined PDF at point (x_p, y, z_p) is:

$$p(y) = \frac{1}{\sqrt{2\pi} \sigma_y} \exp \left[-\frac{(y-y_p)^2}{2\sigma_y^2} \right]; \quad \text{in which } \sigma_y = \left(\sum_i \frac{\sin^2(\theta_i)}{\sigma_{r,i}^2} \right)^{-1/2}. \quad (2.1.3)$$

Error in the z -direction is handled in a similar manner. For receiver i and position

$\bar{p} = (x_p, y_p, z_p)$, range is fixed to find the standard deviations in depth for receiver i :

$\sigma_{ssp,d,i}$, $\sigma_{drt,d,i}$, and $\sigma_{d,i} = \sqrt{\sigma_{ssp,d,i}^2 + \sigma_{drt,d,i}^2}$ (see Fig. 2.1.10). The combined PDF at

point (x_p, y_p, z) is simply:

$$p(z) = \frac{1}{\sqrt{2\pi} \sigma_z} \exp \left[-\frac{(z-z_p)^2}{2\sigma_z^2} \right]; \quad \text{in which } \sigma_z = \left(\sum_i \sigma_{d,i}^{-2} \right)^{-1/2}. \quad (2.1.4)$$

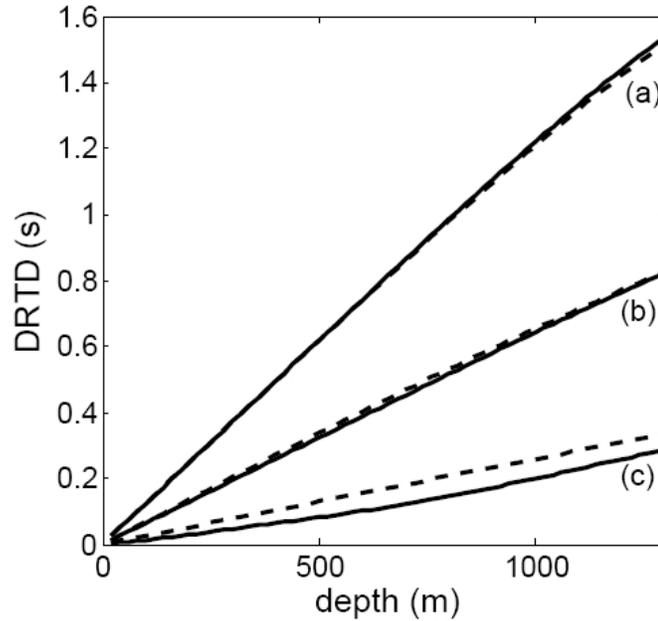


Fig 2.1.10 Modeled DRTDs for hydrophone H as a function of depth for source ranges of (a) 500 m, (b) 2500 m, and (c) 7500 m. Solid lines are for the depth-dependent SSP for March; dashed lines are for an isospeed SSP of 1510 m/s.

Based on the width of the processed clicks (about 10 ms), the standard deviation of the DRTD measurements, σ_{drtid} , was set to 5 ms. Figure 2.1.11 (a)-(c) shows location error maps (at 700 m depth) obtained using the SSPs from all twelve months (see Fig.2.1.1) as the collection of possible SSPs. Error maps for different depths are similar. For the horizontal directions, error from DRTD measurement only is about double the error from SSP only. In the vertical direction, DRTD measurement errors are similar to SSP errors. Figure 2.1.11 (d)-(e) shows error maps obtained when the isospeed SSP is added to the collection of possible SSPs. In this case, errors associated with SSP only are an order of magnitude greater than errors associated with DRTD only. In Figure 2.1.12, the errors associated with the actual tracks are plotted, with the errors calculated for (a) the depth-dependent case using the monthly SSPs and (b) the isospeed case using the

isosped SSP in addition to the monthly SSPs. In all instances, error in the vertical direction is less than error in the horizontal directions; this is expected because DRTD changes more rapidly with depth (Fig.2.1.10) than with range (Fig.2.1.2).

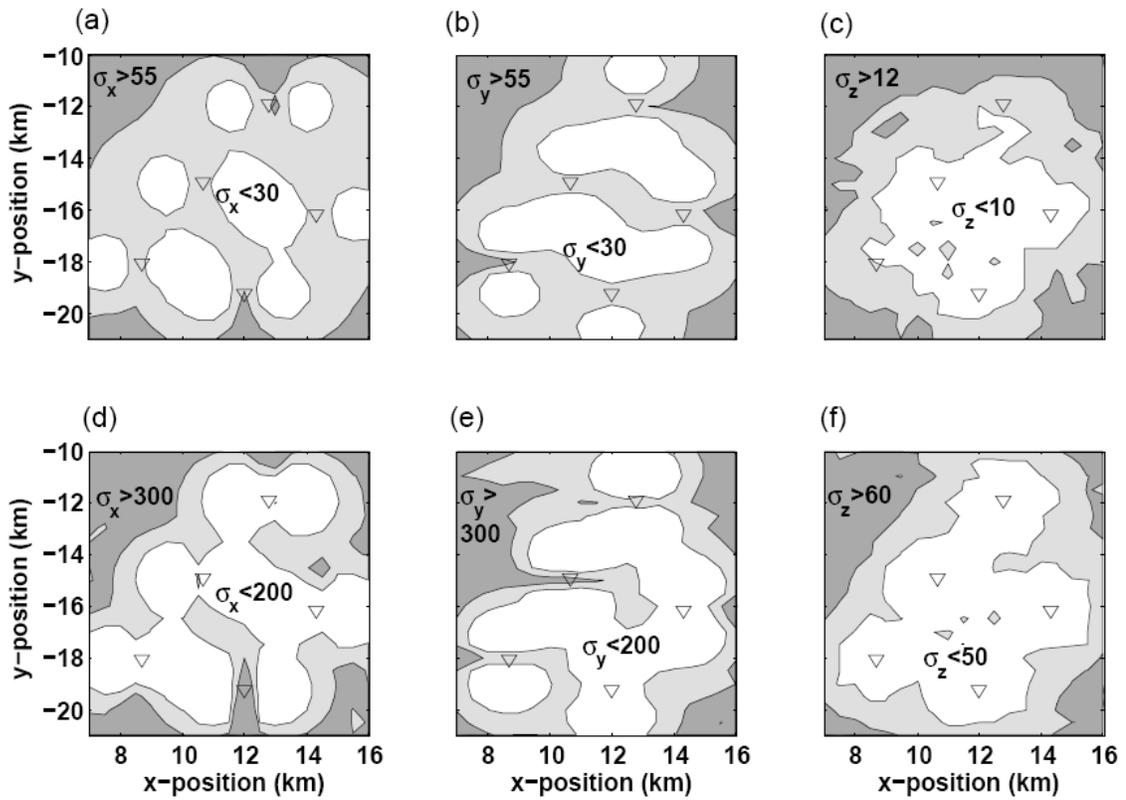


Fig 2.1.11 Contour intervals (values as indicated) of one standard deviation in x-, y-, and z- source position at 700 m depth due to uncertainties in SSPs and DRTDs. (a)-(c) Depth-dependent SSP; (d)-(e) Isosped SSP. Triangles indicate receiver positions.

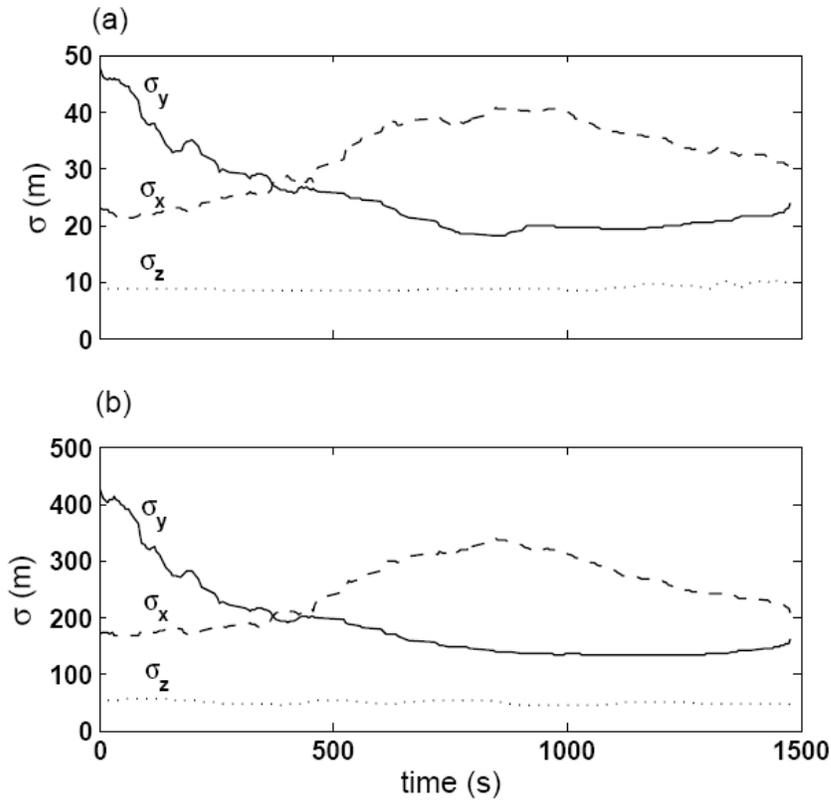


Fig 2.1.12 One standard deviation in x-, y-, and z- source position (as a function of time) along estimated tracks due to uncertainties in SSPs and measurement of DRTDs. (a) Depth-dependent SSP; (b) Isospeed SSP.

2.1.5 Discussion

Although it cannot be confirmed by tags or sightings, the estimated track is consistent with what is expected for a sperm whale. In particular, sperm whale dives are typically many hundreds of meters deep [Papastavrou *et al.* 1989; Watkins *et al.* 1993; Clarke *et al.* 1993; Zimmer *et al.* 2003] with reports of dives in excess of 1000m[Loyker 1977; Wahlberg 2002]. Dives may last up to 90 minutes [Goold and Jones 1995], but are more commonly between about 25 and 50 minutes [Watkins *et al.* 1993; Papastavrou *et al.* 1989; Gordon and Steiner 1992; Jaquet *et al.* 2000]. Also, the estimated swim speeds

agree with those observed in previous studies [Watkins *et al.* 1993; Papastavrou *et al.* 1989; Wahlberg 2002; Watkins *et al.* 2002]. Our error estimates suggest that the track using the depth-dependent SSP is correct to about 100 m in horizontal position and 20 m in depth.

An important advantage of the DRTD method over arrival time difference methods commonly used for marine mammal localization is that it is much less sensitive to synchronization errors in timing between receivers. This is because DRTD measurements are estimated for individual hydrophones, rather than between pairs of hydrophones. Although a comprehensive study of synchronization error was not performed, it is worth noting that a 2.34 s offset between two of the five hydrophones that was (unknown and) present in the original version of the distributed dataset did not significantly affect our predicted track.

Several problems are associated with the DRTD method. First, for near-surface sources, direct and surface-reflected clicks are difficult to distinguish. Shadow zones present another problem for near-surface sources, although this effect would likely occur on only one receiver, and can be overcome for sufficiently large (> 5 receiver) arrays. Furthermore, surface roughness associated with gravity waves may have a significant effect on reflected arrival times [Skarsoulis and Kalogerakis 2005; Godin and Fuks 1989], and hence on estimated DRTDs. Uncertainty in receiver location is also a problem, as it is with all localization techniques. Methods to locate the receivers more accurately [Wahlberg *et al.* 2001], or to include variable receiver position in the modeling are useful for this [Michalopoulou and Ma 2005]. Our detection scheme is for a single animal, but improved schemes that can distinguish calls of individuals [Thode *et al.*

2002; Mellinger 2002] might extend its applicability to multiple whales. Finally, it would be prudent to compare and combine the DRTD method with other localization techniques to give more accurate track estimates.

The reader is advised that some of the methods presented here were done with a somewhat “quick and dirty” mentality. This approach was taken because we wanted to test the feasibility of using DRTDs for 3-D localization without getting tangled in detail, and there is certainly much room for improvement. For example, the method would benefit from a more sophisticated (and objective) detection and classification scheme [Zimmer *et al.* 2005a]. Also, likelihood surfaces should incorporate errors in measurement and modeling instead of using the (empirical and somewhat arbitrary) standard deviation of 500 m. Rather than searching over candidate depths, likelihood volumes could be created in three dimensions. Further, the receiver log likelihoods (rather than the likelihoods themselves) should be averaged to create overall likelihood surfaces. Among other possible improvements, these things would reduce errors and allow for more accurate error estimates.

2.1.6 Conclusions

Recordings of a single sperm whale on 5 bottom-mounted hydrophones in a deep-water environment were used to track the animal in three dimensions for 25 minutes. A model-based method based on the arrival time difference between direct and surface-reflected clicks was used in the tracking and described in detail. All 5 hydrophones were used and at least four hydrophones are needed to apply the method. A depth-dependent SSP led to better performance estimates than an isospeed SSP. Run times were about

three times longer than real-time, but can be reduced to real-time by decreasing resolution or by using a faster machine. Although we did not have data to verify the track visually or otherwise, it is consistent with sperm whale behavior. Estimated horizontal positions were similar for both SSPs, but depth for the isospeed SSP was about 50–100 m shallower than for the depth-dependent SSP. The estimated depth of the whale varied between 650 m and 760 m for the depth-dependent SSP. The average vertical and horizontal swim speeds were 0.30 m/s and 2.40 m/s, respectively.

2.2 Paper 2

Nosal E-M, LN Frazer (2007). Sperm whale 3D track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones. *Journal of the Acoustical Society of America*, 122(4), 1969-1978.

Abstract – In an earlier paper [Nosal and Frazer 2006a], a sperm whale was tracked in 3D using direct and surface-reflected time differences (DRTD) of clicks recorded on five bottom-mounted hydrophones, a passive method that is robust to timing errors between hydrophones. This paper refines the DRTD method and combines it with a time of (direct) arrival (TOA) method to improve the accuracy of the track. Knowing the position and origin time of each click, pitch and yaw are obtained by assuming the main axis of the whale is tangent to the track. Roll is then found by applying the bent horn model of sperm whale phonation, in which each click is composed of two pulses, p0 and p1, that exit the whale at different points. With instantaneous pitch, roll and yaw estimated from time differences, amplitudes are then used to estimate the beam patterns of the p0 and p1 pulses. The resulting beam patterns independently confirm those obtained by Zimmer *et al.* [2005a] with a very different experimental set-up. A method for estimating relative click levels is presented and used to find that click levels decrease toward the end of a click series, prior to the “creak” associated with prey capture.

2.2.1 Introduction

The main purpose of this paper is to demonstrate and progress the use of passive acoustic methods for studying marine mammals in the wild, especially odontocetes. In a recent paper [Nosal and Frazer 2006a], we studied the improvement in ray-based tracking that occurs when a realistic sound speed profile is used instead of an assumed isospeed profile. We tracked a sperm whale using the difference between direct and surface-reflected click arrival times (DRTD), a method that is robust to time-origin errors on different hydrophones. Here we refine the DRTD method and combine it with a time of (direct) arrival (TOA) method to get a combined method that is more accurate than either method separately. The time and position estimates are precise enough that we can approximate swim velocity and orientation at each click, which we then use to estimate click beam patterns and levels.

This paper focuses on clicks with regular inter-click intervals (ICIs) of 0.45-2s, called ‘usual’ clicks by Whitehead and Weilgart [1990], emitted by sperm whales while foraging at depth. A typical foraging dive lasts about 45 minutes, and begins with a steep, steady descent to a depth of several hundred meters [Watkins *et al.* 2002, Watwood *et al.* 2006], followed by a period of searching for prey and then a steep, steady ascent. Series of regular clicks emitted at foraging depth are often terminated by a ‘creak’ of clicks with high repetition rate followed by several seconds of silence [Gordon 1987; Mullins *et al.* 1988; Goold and Jones 1995]. The regular clicks are likely used for echolocation [Norris and Harvey 1972; Whitehead and Weilgart 1991; Goold and Jones 1995; Møhl *et al.* 2000; Jaquet *et al.* 2001; Madsen *et al.* 2002 *a*; Whitehead 2003; Møhl *et al.* 2003] while the creaks mark the terminal phase of prey capture [Miller *et al.* 2004]. Regular clicks are

short in duration, broadband (100Hz – over 20 kHz), and have a powerful forward directed beam [Møhl *et al.* 2000; Madsen *et al.* 2002*a,b*]. Regular clicks are often heard on hydrophones several kilometers from the vocalizing animal. Being of short-duration, direct and reflected arrivals can often be distinguished, making clicks ideal for passive localization.

We track a single sperm whale from its regular clicks for 23 minutes using recordings on 5 bottom-mounted hydrophones. The data were recorded at the Atlantic Undersea Test and Evaluation Center (AUTEC) located in the Tongue of the Ocean (off Andros Island, Bahamas). They were provided by the Naval Undersea Warfare Center for the 2nd International Workshop of Detection and Localization of Marine Mammals using Passive Acoustics. The sampling rate was 48 kHz and the hydrophone positions are listed in Table 2.1.1. Some further details can be found in Adam *et al.* [2006], but unfortunately, the anti-alias filter, frequency response, sensitivity, and directionality of the sensors were not available. Accordingly, our results are limited by the assumption of an omni-directional and flat frequency response, and absolute sound pressure levels cannot be found.

The dataset used here is the same dataset that we used to develop the DRTD method [Nosal and Frazer 2006*a*]. The track of the sperm whale in this dataset has also been obtained using time of arrival differences between pairs of receivers [Giraudet *et al.* 2006; Morrissey *et al.* 2006; White *et al.* 2006]. The improved accuracy of the combined DRTD/TOA method used in this paper allows us to estimate the velocity of the sperm whale from position and time differences between successive clicks. The pitch and yaw

of the whale then follow from the assumption that the main axis of the whale is parallel to its velocity vector.

In order to find roll, we apply the bent-horn model of sperm whale phonation, which was proposed to explain the multi-pulse structure of sperm whale clicks [Norris and Harvey 1972; Møhl 2001] and is supported by recent studies [Madsen *et al.* 2003; Møhl *et al.* 2003; Zimmer *et al.* 2005a]. In the bent-horn model, a single sound is generated at the phonic lips (Figure 2.2.1). Some energy leaks directly into the water as a p_0 pulse. Most of the energy transmits back through the spermaceti organ, reflects off the frontal sac in front of the skull, transmits forward into the “junk,” and exits into the water as the p_1 pulse. Since the p_1 pulse follows a longer path than p_0 , it arrives later, giving the click a multi-pulse structure. Other click components (resulting from further reflections in the head and other exit points) are also present [Møhl 2001], but they are not required or considered here. Since the measured delay between the p_0 and p_1 pulses depends on the orientation of the whale [Zimmer *et al.* 2005b; Laplanche *et al.* 2006], it can be used to recover roll.

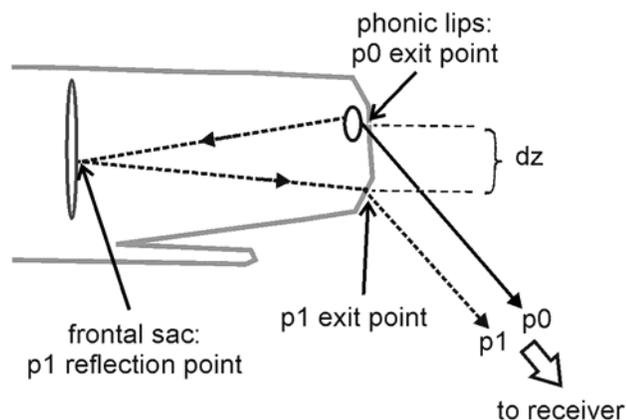


Fig. 2.2.1 Diagram illustrating the paths taken by the p_0 and p_1 pulses according to the modified bent-horn model of sound production in sperm whales.

With position, velocity, pitch, roll and yaw obtained solely from travel time differences, we then use relative amplitudes to estimate the beam patterns and directivity indices of the p0 and p1 pulses that comprise a click. Our results agree with previous studies [Møhl 2003; Zimmer *et al.* 2005a] which found that the p1 pulse has a narrowly focused, forward-direct beam, that the p0 pulse is slightly weaker and more broadly backward-directed, and that a low-frequency, nearly omni-directional component is characteristic of all clicks. Finally, we correct click amplitudes for beam pattern and propagation loss to estimate relative click levels within each click sequence, finding that click source levels decrease toward the end of a click series.

2.2.2 Data processing

2.2.2.1 Click detection, classification, and association

The beginnings and ends of the clicks were detected using an automated transient detector [Page 1954; Wald 1947; Abraham 2000; Zimmer *et al.* 2003; 2005a]. To reduce noise, each time series was filtered using a second-order, high-pass Butterworth filter with a 300 Hz low cut. The envelope of each filtered time series was calculated as the magnitude of the corresponding analytic signal, where the analytic signal has real and imaginary parts consisting of the original time series and its Hilbert transform, respectively [e.g., Bracewell 2000]. Given the instantaneous signal amplitude (envelope) e_n , a test variable V_n was calculated as:

$$V_n = \frac{e_n^2}{N_n} \quad (2.2.1)$$

where N_n is the noise estimate. For the first noise estimate, N_1 is taken as the mean square envelope value (over all samples for that hydrophone). Detection decisions and updates for subsequent noise estimates are made according to the value of V_n in the following algorithm:

$$\text{If } \begin{cases} V_n > T_0 \\ V_n < T_x \\ \text{other} \end{cases} \text{ then } \begin{cases} \text{decide detection} \\ \text{set } T_x = T_1 \\ \text{set } N_{n+1} = N_n \\ \\ \text{decide no detection} \\ \text{set } T_x = T_2 \\ \text{set } N_{n+1} = (1 - \alpha)N_n + \alpha e_n^2 \\ \\ \text{no decision} \\ \text{keep current } T_x \\ \text{set } N_{n+1} = N_n \end{cases} \quad (2.2.2)$$

where T_0 , T_1 , and T_2 are the thresholds for decision of detection, end of detection, and noise, respectively; $T_x \in \{T_1, T_2\}$ is the current threshold; and α is the exponential weighting on the power estimate when no signal is detected. For the first sample, T_x is set equal to T_1 . Threshold and weighting values that performed well were $T_0 = 25$ (13.98 dB), $T_1 = 9$ (9.54 dB), $T_2 = 4$ (6.02 dB), and $\alpha = 1/100$.

At each time step, this algorithm decides if there is a signal present (detection) or not (no detection). The algorithm operates in two modes: signal and noise. In the signal mode, signal present is decided while the value of the test variable is greater than the detection threshold ($V_n > T_0$). No decision is made if the value of the test variable is less

than the threshold for detection but greater than the threshold for the end of detection ($T_1 < V_n < T_0$). Signal not present is decided once the value of the test variable drops below the end of detection threshold ($V_n < T_1$). Here, the algorithm switches to noise mode. In this mode, signal not present is decided while the test variable remains below the noise threshold ($V_n < T_2$) and the noise variable is updated at each time step according to an exponential weighting (more weight toward recent values). No decision is made if the value of the test variable is less than the threshold for detection but greater than the threshold for noise ($T_2 < V_n < T_0$). Signal present is decided once the value of the test variable jumps above the detection threshold ($V_n > T_0$). Here, the algorithm switches to signal mode.

Each click resulted in a direct arrival, usually followed by a lower-amplitude surface reflection. Direct-reflected pairs were classified according to the following criteria: (1) the amplitude of the direct arrival varies slowly; (2) the inter-click interval between successive direct arrivals varies slowly; (3) the reflected arrival has lower amplitude than the direct arrival; and (4) the time between the direct and reflected arrival (DRTD) is similar to that of the preceding direct-reflected pair.

Each click detected on two or more receivers was numbered sequentially. Clicks on different receivers were associated by comparing intervals between clicks, which should be nearly identical on all receivers. To eliminate incorrect associations due to click time measurement error, this comparison included intervals between all clicks in a series, not only those immediately preceding or following a given click. In total, 1324 clicks were numbered, with 1102, 913, 868, 1163, and 1035 clicks detected on receivers 1, 2, 3, 4, and 5 respectively. The number of clicks detected on a total of 2, 3, 4, and 5

receivers was 137, 324, 480, and 383. Only clicks recorded on 3 or more hydrophones (total of 1187 clicks) were used for localization.

2.2.2.2 Level and pulse delay measurement

The maximum of the envelope was used to estimate the received peak pressures of each direct click. Using frequency and time windows, peak pressures of the p0, p1, and low-frequency (LF) components were also obtained. Following Zimmer *et al.* [2005a], the p0 and p1 pulses were defined to fall in time windows from, respectively, -2 to 3 ms and 3 to 10 ms, relative to detection of the start of the click. They were both defined to fall in a frequency window of 3 to 22 kHz. The identified p0 and p1 times and amplitudes corresponded to maxima of the envelope of the filtered signal. The LF component was defined with a time window of -2 to 10 ms and a frequency window of 300 Hz to 3 kHz. The delay between the p0 and p1 pulses, τ , was estimated by subtracting p0 arrival time from p1 arrival time.

2.2.3 Localization

2.2.3.1 Methods

For each receiver, the acoustic propagation model BELLHOP [Porter 2005] was used to create a lookup table of TOAs, DRTDs, takeoff beam angles, and transmission losses for a list of candidate source ranges and depths. The historic depth-dependent sound speed profile for the area (24° 45' N, 77° 45' W) in March was taken from the Generalized Digital Environment Model [GDEM] and is the same as the profile used in Nosal and Frazer [2006a]. The depth list varied from 5 m to 1550 m with 5 m increments,

and the range list varied from 5 m to 20 km with 5 m increments. Since arrival times varied smoothly for the depths and ranges of interest, all required TOAs and DRTDs were interpolated from the values in the lookup table using cubic splines.

To determine the time and position of each click, we first created a 4-D grid of candidate source points (one dimension for time and three for position). Errors in DRTD and TOA were assumed to be normally distributed. Ideally, DRTD and TOA should be regarded as functions, not just of source position, but also of sound speed and receiver positions, and likelihood surfaces should be maximized over this much larger parameter space. However, to reduce computational requirements, we incorporated the uncertainties in sound speed profile and receiver positions into the standard deviations for DRTD and TOA in a “worst case” manner. Standard deviations σ_{drtd} and σ_{toa} were calculated as:

$$\sigma_{\text{drtd}} = \sqrt{2\sigma_{\text{meas}}^2 + \sigma_{\text{rp}}^2 + \sigma_{\text{drtd}}^2} \quad (2.2.3)$$

$$\sigma_{\text{toa}} = \sqrt{\sigma_{\text{meas}}^2 + \sigma_{\text{rp}}^2 + \sigma_{\text{toa}}^2} \quad (2.2.4)$$

where σ_{meas} is the standard deviation (std) in the measured click times, σ_{rp} is the std due to uncertainty in receiver position, and σ_{drtd} and σ_{toa} are the std (due to sound speed uncertainty) in modeled DRTDs and TOAs. We used $\sigma_{\text{meas}} = 5$ ms based on the widths of the clicks (about 10 ms) and $\sigma_{\text{rp}} = 2$ ms corresponding approximately to a best-guess receiver position uncertainty of 3 m (actual position uncertainty is unknown). To determine σ_{drtd} and σ_{toa} , the DRTD and TOA lookup tables were recalculated for all 12 months using historic sound speed profiles (also from the GDEM). This gave 12 possible TOAs and DRTDs for each range and depth. The difference between the minimum and the maximum of these 12 values approximates the width of the uncertainty curves. The

maximum such width over all ranges and depths (“worst-case”) was taken as two std, giving $\sigma_{\text{drttd}} = \sigma_{\text{toa}} = 3$ ms. Using the maximum width simplifies the calculations, by allowing one std to be used for all candidate points, and it over-estimates final errors.

For candidate whale position \mathbf{s} and click time t , the DRTD and TOA likelihood functions were computed as:

$$L_{\text{drttd}}(\mathbf{s}) = \frac{1}{(2\pi\sigma_{\text{drttd}}^2)^{N/2}} \exp\left[\frac{-1}{2\sigma_{\text{drttd}}^2} \sum_j (\overline{DRTD}_j(\mathbf{s}) - DRTD_j)^2\right] \quad (2.2.5)$$

$$L_{\text{toa}}(\mathbf{s}, t) = \frac{1}{(2\pi\sigma_{\text{toa}}^2)^{N/2}} \exp\left[\frac{-1}{2\sigma_{\text{toa}}^2} \sum_j (\overline{TOA}_j(\mathbf{s}, t) - TOA_j)^2\right] \quad (2.2.6)$$

where the sums are over all receivers that heard the click, N is the number of receivers that heard the click, $DRTD_j$ and TOA_j are the measured values on receiver j and \overline{DRTD}_j and \overline{TOA}_j are the modeled values. The total likelihood value is the product of these:

$$L(\mathbf{s}, t) = L_{\text{drttd}}(\mathbf{s})L_{\text{toa}}(\mathbf{s}, t). \quad (2.2.7)$$

The point (\mathbf{s}, t) that maximizes L is the estimated source position and time. An advantage of distinct likelihood surfaces is that they can be examined separately as a diagnostic, since persistent differences between locations from the two methods are an indication that hydrophone time origins may be different (degrading TOA), or that the sound speed profile in the upper part of the water column is inaccurate (degrading DRTD).

For computational efficiency, two passes were made. The first pass was coarsely sampled in space (10 m grid spacing) and time (10 ms time spacing). For the first click, the spatial search volume covered the full water column in depth and extended 5 km past the boundary defined by the receivers. Time was searched from 0 to 20 s. For the other

clicks the boundary of the search volume was based on the time, Δt , between the current click and the preceding localized click. This was estimated from the measured time between these two clicks on a single phone that heard both clicks. The search volume was centered on the position estimate of the previous click, and bounded in all three directions by double the maximum possible swim distance in Δt , i.e. $8\Delta t$ (assuming a swim speed of at most 4 m/s). Time was searched from the previously localized click until $2\Delta t$ after it.

The second pass refined the position and time estimate from the first pass by searching a smaller, more finely sampled, volume centered on the position and time found in the first pass. The search volume for this pass was sampled at intervals of 1 m in space and 1 ms in time. It was bounded in space by the coarsely determined source location, plus or minus 20 m in all directions, and in time by 200 ms before and after the coarsely determined click time.

2.2.3.2 Error estimates

The literature on bioacoustic localization arrays contains various approaches to quantify error [e.g., Whalberg *et al.* 2001; Spiesberger and Wahlberg 2002]. Since the complete likelihood surfaces were already calculated in the localization step above, we applied a somewhat different approach, using the likelihood surfaces to give error estimates.

95% confidence intervals (CIs) were estimated from conditional likelihood functions (CLFs) by identifying the interval containing 2.5% to 97.5% of the cumulative likelihood for the parameter of interest. For example, to find the confidence interval in the x -position for a single click, all other parameters (y -position, z -position, and click

time) were fixed to their values (y_0, z_0, t_0) at the estimated source position and time. The corresponding CLF, $L(x_j | y_0, z_0, t_0)$, was calculated according to Eq. (2.2.7) for a list of possible x -positions, x_j . The cumulative CLF was then calculated as

$$C(x) \approx \frac{\sum_{x_j < x} L(x_j)}{\sum_{x_j} L(x_j)}. \quad (2.2.8)$$

The denominator normalizes the distribution, and the equality is approximate because of the discrete sampling of x . The list of x -positions ranged from $x_0 - 1$ km to $x_0 + 1$ km (since the CLF was very close to 0 at 1 km away from x_0), with 1 m resolution. Then the 95% CI is $[x_{2.5\%}, x_{97.5\%}]$, where $x_{2.5\%}$ and $x_{97.5\%}$ are such that $C(x_{2.5\%}) = 0.025$ and $C(x_{97.5\%}) = 0.975$. CIs for y -position, z -position, and time were computed similarly.

2.2.3.3 Results

Figure 2.2.2 shows the resulting x - y - and z - positions obtained for clicks heard on three or more receivers. Figure 2.2.3 shows the positions in the x - y plane. The click time list ranged from $t_0 - 1$ s to $t_0 + 1$ s with 0.1 ms resolution. The resulting CI half-widths for position are shown in Figure 2.2.4. The half-widths for time were less than 4.5 ms, 5 ms, and 5.8 ms for clicks heard on 5, 4, and 3 receivers, respectively.

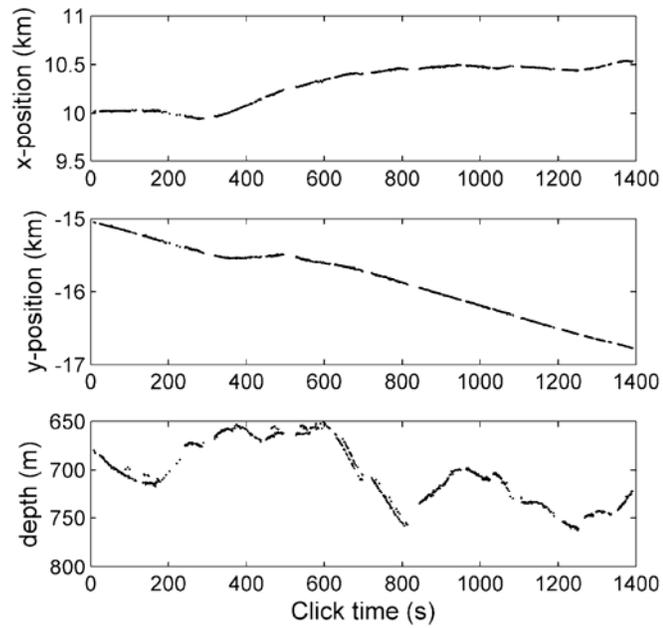


Fig. 2.2.2 Estimated track (not smoothed) of the sperm whale. Positions of clicks detected on 3 or more receivers are plotted as dots against the time of the click.

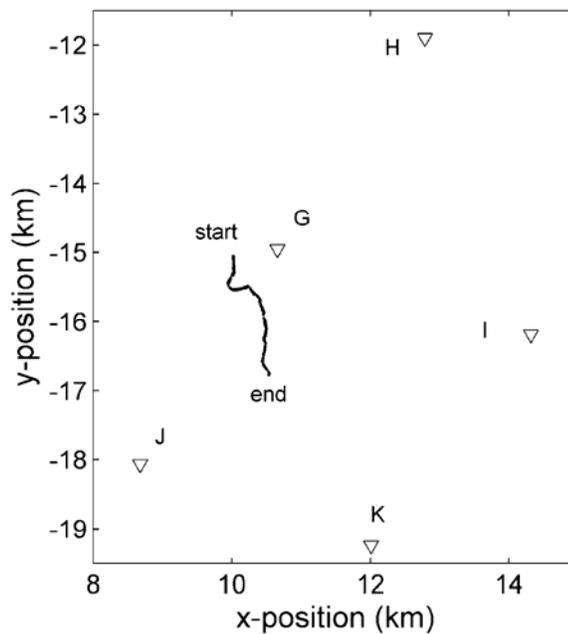


Fig. 2.2.3 Estimated two-dimensional track (not smoothed) of the sperm whale. Positions are plotted as dots to form the track, as in Fig. 2.2.2, and triangles indicate receiver positions.

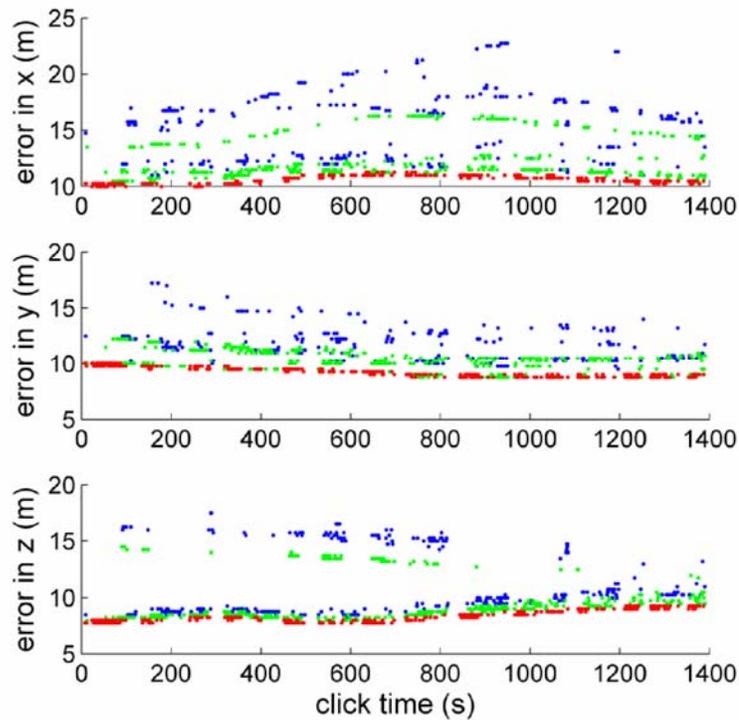


Fig. 2.2.4 Half-widths of 95% confidence intervals for position. Red, green, and blue indicate clicks detected on 5, 4, and 3 receivers, respectively.

2.2.3 Swim Orientation

2.2.3.1 Conventions

This section outlines the orientation conventions used below. Two reference frames are required: the earth frame and the whale frame, notated as unprimed and primed, respectively. In the earth frame, positive x , y , and z are directed east, north, and upward, respectively. In the whale frame, positive x' , y' , and z' point forward (rostrally along the whale's long axis), left, and dorsally, respectively. The two frames coincide when the whale is traveling due east, horizontal and upright.

Three angles are required to transform between the whale and earth frames: yaw (θ_z), pitch (θ_y), and roll (θ_x), which are rotations about the z , y , and x axes, respectively. For yaw and roll, positive values correspond to a coordinate system rotation in a clockwise direction when looking away from the origin along the axis of rotation. For consistency with conventions used by Johnson and Tyack [2003], and so that a positive pitch corresponds to a nose-upward orientation, positive pitch corresponds to a counter-clockwise rotation when looking away from the origin along the y -axis. Note that this convention differs from standard Euler and pitch-roll-heading convention [Goldstein 1980]. Thus in our convention a whale with zero yaw, pitch, and roll is swimming eastward, horizontally, and upright. From this $\theta_x = \theta_y = \theta_z = 0$ orientation, the whale turns left to increase yaw, toward the surface to increase pitch, and clockwise to increase roll. To make them unique, θ_x , θ_y , and θ_z are constrained to the intervals $(-180^\circ, 180^\circ]$, $[-90^\circ, 90^\circ]$, and $(-180^\circ, 180^\circ]$, respectively.

A vector in the earth (unprimed) frame is expressed in whale (primed) frame coordinates via three matrices that commute only in the limit of very small angles (so the order of multiplication is important):

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_x(\theta_x) R_y(\theta_y) R_z(\theta_z) \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2.2.9)$$

with

$$R_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & \sin \theta_x \\ 0 & -\sin \theta_x & \cos \theta_x \end{bmatrix} \quad (2.2.10)$$

$$\mathbf{R}_y(\theta_y) = \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \quad (2.2.11)$$

$$\mathbf{R}_z(\theta_z) = \begin{bmatrix} \cos \theta_z & \sin \theta_z & 0 \\ -\sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.2.12)$$

2.2.3.2 Pitch and yaw

The first step in recovering swim attitude is to approximate the velocity of the whale at each click. To do this, a vector-valued position function $\mathbf{f}(t) = (f_x(t), f_y(t), f_z(t))$ was fit to the calculated click positions and times by minimizing a weighted sum of squared position error and acceleration:

$$E(\mathbf{f}) = a \sum_{j=1}^{N_c} |\mathbf{s}_j - \mathbf{f}(t_j)|^2 + (1-a) \int_{t_1}^{t_{N_c}} \left| \frac{d^2 \mathbf{f}}{dt^2} \right|^2 dt, \quad (2.2.13)$$

where the sum is over all localized clicks j , N_c is the total number of localized clicks, \mathbf{s}_j and t_j are the estimated position and time of click j , and $a = 0.7$ is a smoothing parameter. Velocity in the earth frame, $\mathbf{v}(t) = (v_x(t), v_y(t), v_z(t))$, is found by taking the first derivative of \mathbf{f} .

To recover pitch and yaw, we assume that the whale's main axis is parallel to its velocity. This assumption neglects the effects of current and the ability of the whale to move laterally or vertically, as well as any scanning movements of the head, so the goodness of the approximation increases with the forward speed of the whale. Pitch and yaw can then be computed as:

$$\theta_y = \sin^{-1}(v_z / |\mathbf{v}|) \quad (2.2.14)$$

$$\theta_z = \tan^{-1}(v_y / v_x) \quad (2.2.15)$$

where \tan^{-1} is the four-quadrant, inverse tangent.

2.2.3.3 Roll

Once position, pitch, and yaw are known, roll is estimated from the delay τ between the p0 and p1 pulses (Fig. 2.2.1). To do this, we build on methods introduced by Zimmer *et al.* [2005b] and Laplanche *et al.* [2006]. The modeled delay is split into two components: $\tau = t_c + t_{\Delta p}$. The constant component, t_c , is the time required for sound to travel from the phonic lips to the frontal sac, where it reflects, and thence forward to the p1 exit point. It is assumed fixed for a given animal. The second component, $t_{\Delta p}$, is the difference between the travel time from the p1 exit point to the receiver and the travel time from the p0 exit point to the receiver. It depends on the exit point of the p1 pulse relative to the phonic lips—we assume this exit point is fixed for a given animal—and on whale position, receiver position, and roll. The exit point of the p1 pulse is located at the junk and is directly ventral to the phonic lips [Madsen 2002; Zimmer *et al.* 2005b]. Hence, if we take the phonic lips to be at point (0,0,0) in the whale frame, the p1 exit point can be approximated as (0,0,-dz), with dz fixed for the individual whale (Fig. 2.2.1).

Given the position of the whale, \mathbf{s} , and the receiver, \mathbf{r} , in the earth frame, as well as the pitch and yaw angles previously determined, τ can be modeled (for various t_c , dz, and roll angles) as follows. For each click and receiver, we find the takeoff direction of the ray that connects the receiver and localized source positions. For a constant sound-speed profile, this direction in the earth frame would simply be $\mathbf{r} - \mathbf{s}$ with source position \mathbf{s} and receiver position \mathbf{r} . To find this direction for our depth-dependent sound speed

profile, the list of takeoff angles from BELLHOP is interpolated to get the vertical angle (or elevation), ϕ , of the ray in the earth frame. The ray direction vector in the earth frame is then:

$$\mathbf{b} = \left[r_x - s_x, \quad r_y - s_y, \quad \tan(\phi) \sqrt{(r_x - s_x)^2 + (r_y - s_y)^2} \right]^T. \quad (2.2.16)$$

This ray direction is transformed into the whale frame direction vector, \mathbf{b}' by applying Eq. (2.2.9) with the calculated yaw, pitch, and candidate roll angle for the current click. The azimuth $\alpha' \in [-180^\circ, 180^\circ]$ and elevation $\phi' \in [-90^\circ, 90^\circ]$ of the ray in whale coordinates are then calculated as:

$$\alpha' = \tan^{-1} \left(\frac{b'_y}{b'_x} \right) \quad \text{and} \quad \phi' = \tan^{-1} \left(\frac{b'_z}{\sqrt{(b'_x)^2 + (b'_y)^2}} \right), \quad (2.2.17)$$

where \tan^{-1} is the four-quadrant inverse tangent. Here, positive/negative azimuth corresponds to a leftward/rightward directed beam, and positive/negative elevation corresponds to an upward/downward directed beam. Elevation and azimuth of 0° correspond to a beam directed along the whale's main axis, x' .

Since the distance from the whale to the receiver is much greater than dz , the vertical takeoff angles of the p0 pulse and the p1 pulse from the junk exit point are well approximated by ϕ' . Then $t_{\Delta p}$ is approximated as (Figure 2.2.5):

$$t_{\Delta p} \cong \sin(\phi') dz / c \quad (2.2.18)$$

in which c is the speed of sound through water (c_w), for $-90^\circ \leq \alpha' \leq 90^\circ$, but c is the variable speed of sound through whale tissue (c_t), for $\alpha' > 90^\circ$ or $\alpha' < -90^\circ$. The change

in sound speed is necessary because for clicks propagating forward ($-90^\circ \leq \alpha' \leq 90^\circ$) both the p0 path, and the p1 path after exiting the junk, pass primarily through the water, while backward propagating pulses pass through whale tissue. For each click, c_w is found from the value of the sound speed profile interpolated to the depth of the whale, while c_t is a function of temperature, pressure, ray elevation, and ray azimuth (different angles mean that sound passes through different tissues). For simplicity, the unknown value of c_t is assumed here to be constant and is estimated in the following optimization step.

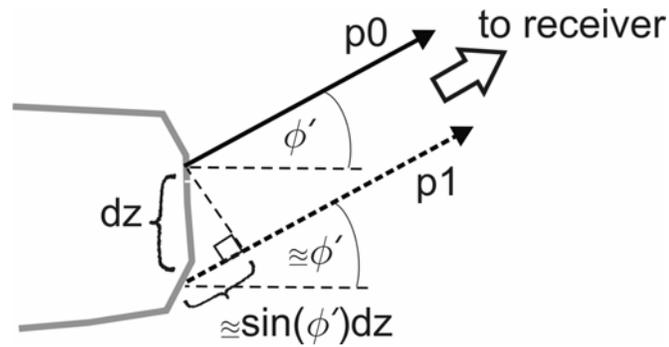


Fig. 2.2.5 Angle approximation used in estimating the delay between p0 and p1.

The constants t_c , dz , and c_t , and the roll for each click are found as follows. With t_c , dz , and c_t fixed over all clicks, we find the roll for each click that minimizes the difference between measured and modeled τ in a least squares sense over all receivers. Summing over all clicks gives the total squared error associated with the current values of t_c and dz . This total error is minimized over t_c , dz , and c_t .

The best fit values were $t_c = 6.6$ ms, $dz = 1.30$ m, and $c_t = 1540$ m/s. The value $t_c = 6.6$ ms corresponds to a whale length of 14.42 m using the formulas of Gordon [1991],

and to a whale length of 13.61 m using the formula of Rhineland and Dawson [2004]. The estimated p1 exit point located 1.30 m ventral of the phonic lips for a whale over 13.5 m makes sense anatomically assuming that the exit point is on the junk [Møhl 2001]. It is also consistent with the results of Zimmer *et al.* [2005b], who found the p1 exit point to be 1.10 m ventral of the phonic lips for a 12 m whale. The derived value $c_t = 1540$ m/s is high compared to the value of 1370 m/s found by Flewellen and Morris [1978] for the speed of sound through spermaceti oil at 33 °C at 1 atm. It is more consistent with (although still on the high end of) values for more similar conditions given by Goold *et al.* [1996], who found that sound speed in spermaceti oil increased from 1390 to 1540 m/s with increasing pressure (from 0 to 90 atm) and decreasing temperature (from 38 to 22 °C). The seemingly high value for c_t found here possibly stems from the fact that our animal is alive, and that the p1 pulse passes through other whale tissue (not only spermaceti oil) to get to the receiver; however, we have made numerous assumptions and approximations that invariably introduce error, and our estimate will need to be examined in future work.

2.2.4 Beam pattern and directivity

2.2.4.1 Methods and results

The azimuth and elevation of each click to each receiver were found for the calculated position and orientation data as outlined in the previous section. The received levels obtained in Section 2.2.2.2 were corrected for transmission loss using the values in the lookup table from Section 2.2.3.1 to get click levels. Since hydrophone sensitivity (or clipping level) was unavailable, click levels could only be found as values relative to

some arbitrary level, chosen such that the weakest click level corresponded to 0 dB.

Hence, we report only “relative click levels”, by which we mean the difference between the current click level and the minimum click level (over all clicks).

Relative click levels are plotted as a function of azimuth and elevation in Figure 2.2.6. Since 324, 480, and 383 clicks were localized on 3, 4, and 5 receivers, respectively, a total of $324 \times 3 + 480 \times 4 + 383 \times 5 = 4807$ points are plotted. Multiple clicks with similar azimuth and elevation were measured, and the figures show higher levels overlapping lower levels, which helps to reduce the effect of variation in click source levels by approximating the maximum level in each direction. The resulting beam patterns are similar, although with somewhat broader peaks, to the patterns found by Zimmer *et al.* [2005a] who used a similar approach.

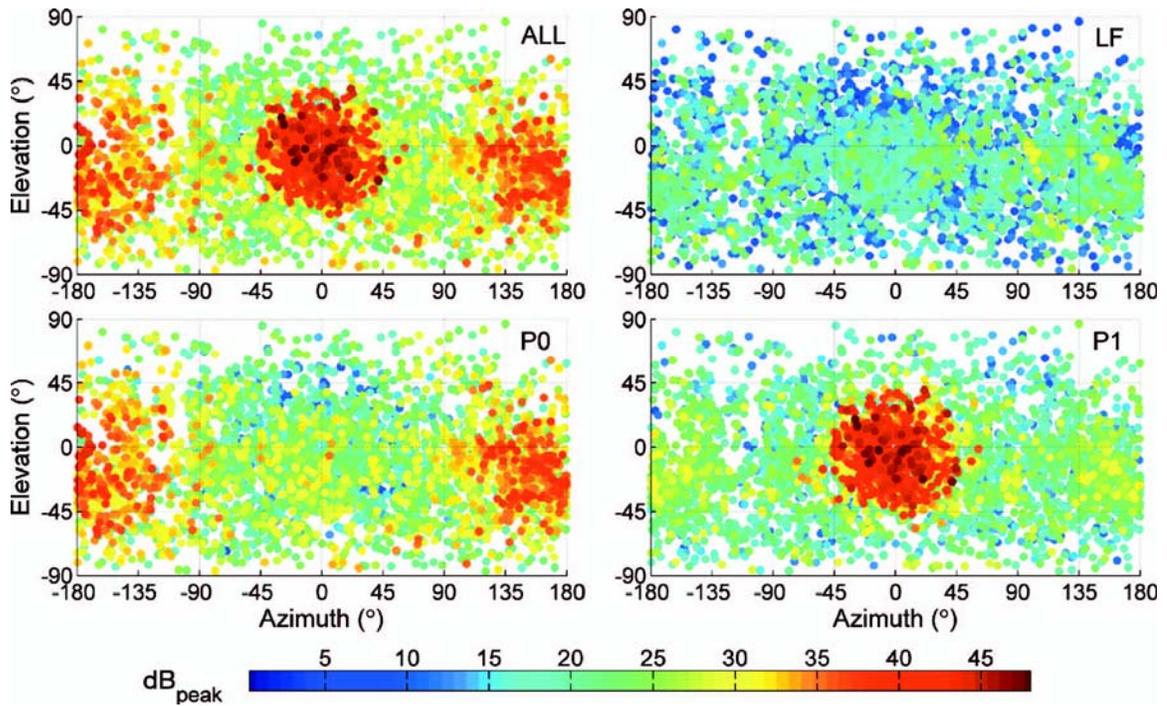


Fig. 2.2.6 Estimated beam patterns from 4807 recorded clicks. Since receiver sensitivities were not available, these are not absolute click levels but relative levels, such that 0 dB corresponds to the weakest recorded click. For these figures, recorded levels were corrected for transmission loss and were plotted as colored dots on the appropriate elevation/azimuth position. Higher levels are shown overlapping lower levels to minimize the effect of a variable source level. Results are shown for the full click, p0 pulse, p1 pulse, and LF components.

Although absolute source levels cannot be estimated, source level differences can be found: the maximum source level of the p1 pulse measured here was 8.8 dB higher than the maximum source level of the p0 pulse and 19.4 dB higher than the maximum source level of the LF component. These values are consistent with estimates reported by Zimmer *et al.* [2005a] of 210 dB_{peak} for the p1 pulse, 200 dB_{peak} for the p0 pulse, and 190 dB_{peak} for the LF component, all re: 1μPa at 1m.

Directivity indices were estimated according to a discretized version of Eq. (3-10) of Au [1993]:

$$DI = 10 \log \frac{4\pi}{\sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\phi} \left(\frac{p(\alpha'_i, \phi'_j)}{P_{\max}} \right)^2 \cos \alpha' \Delta \alpha' \Delta \theta'}$$

in which N_α is the number of azimuth steps of width $\Delta\alpha'$, N_ϕ is the number of elevation steps of width $\Delta\phi'$, $p(\alpha'_i, \phi'_j)$ is the received pressure corrected for transmission loss for the bin corresponding to azimuth step α'_i and elevation step ϕ'_j (recall that the primes denote whale frame coordinates), and p_{\max} is the maximum received pressure over all angles. Step widths of 2.5° were used for both azimuth and elevation, and the maximum pressure over each bin was used for $p(\alpha'_i, \phi'_j)$. Estimated directivity indices were 21.8 dB for the p1 pulse, 9.4 dB for the p0 pulse, and 5.2 dB for the LF component. In comparison, Møhl *et al.* [2003] reported a p1 directivity index of 27 dB and Zimmer *et al.* [2005a] reported a p1 directivity index of 26.7 dB and a p0 directivity index of 7.4 dB.

2.2.4.2 Discussion

Similarly to the case discussed in Zimmer *et al.* [2005a], it is likely that the maximum source level of the p1 pulse is underestimated here due to clipping of the high-intensity arrivals (197 out of all 4807 signals used reached clipping amplitude), limited sampling bandwidth, and a small sample size of on-axis clicks (only 8 clicks within 5° of the main beam axis). This explains in part why our beam pattern for the p1 pulse has a broader peak and lower directivity than reported by Zimmer *et al.* [2005a] and Møhl *et al.* [2003].

To deal with clipping, we follow Zimmer *et al.* [2005a], in which a model broadband beam pattern is fit to the 90th percentile for off-axis angles between 20 and 90 degrees (binned here into 2 degree intervals). In Fig.2.2.7, which is similar to Fig. 9 of Zimmer *et al.* [2005a], all measured (relative) levels are plotted with gray dots as a function of off-axis angle. The 90th percentile for each (2 degree) off-axis bin is plotted in black and the best-fit modeled beam pattern is plotted in red. The p1 pulse was modeled as a Gabor function emitted from a circular piston [Au 1993]. The parameters that give the best fit (in a least squares sense) are peak frequency 15 kHz, signal duration 0.60 ms, and piston radius 0.40 m. These correspond to a p1 directivity index of 25.2 dB. This is much closer to, although still considerably less than, the values of 27 dB and 26.7 dB reported by Møhl *et al.* [2003] and Zimmer *et al.* [2005a], respectively. Unfortunately, maximum source level could not be estimated here because hydrophone sensitivity was unknown.

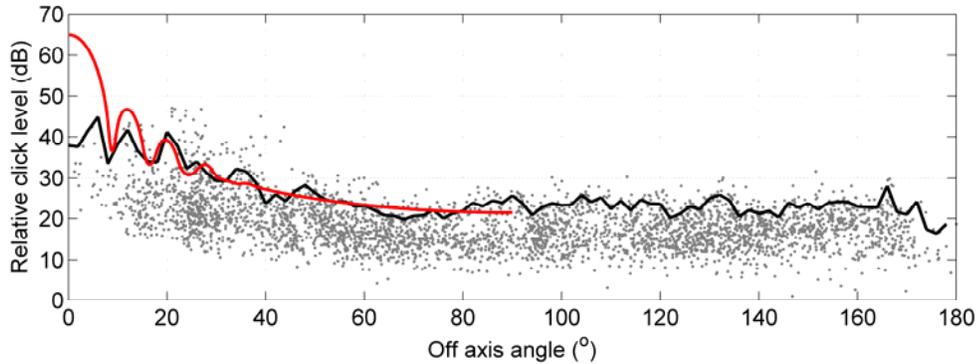


Fig. 2.2.7 Scatter plot of estimated p1 beam pattern as a function of off axis angle. Levels are relative, as in Fig. 2.2.6. The black line represents the 90th percentile for each off axis angle bin (bin size 2 degrees). The red line represents the beam pattern predicted for a circular piston with parameters fitted to measured values (black line) for off axis angles between 20 and 90 degrees. Results and best-fit parameters are given in the text.

Other sources of error in our beam pattern are from errors in estimated pitch, yaw, and roll. These stem from uncertainties in source location and click time, from the assumption that the whale's main axis coincides with the velocity vector, and from approximations made in application of the bent-horn model (constant sound speed through whale tissue, for example). Further, our results are limited by the assumption that all clicks have the same beam pattern and by the assumptions on receiver response (flat-frequency, omni-directional, and the same for all receivers). The remarkable agreement of our beam patterns with those of Zimmer *et al.* [2005a] and Møhl *et al.* [2003], both of which used very different experimental set-ups, suggests that either our assumptions are at least approximately satisfied or that the errors caused by the assumptions tend to cancel over multiple hydrophones.

2.2.5 Click source levels

Click source levels were estimated from the measured beam patterns by finding best-fit levels to a “model” beam pattern. For a given direction, the model beam pattern was assigned the maximum received level, corrected for transmission loss, over all directions within 5° . This approach was preferred over binning the received levels into discretized azimuth and elevation steps, which would have resulted in a non-uniform weighting of the received levels since different elevation bins subtend different solid angles (this also explains why the clicks in Figure 2.2.6 are more densely populated at elevations closer to 0°). For a sufficient number of clicks, this should eliminate variations due to click source level, giving a model that well approximates the true beam pattern [Zimmer *et al.* 2005a].

Relative click source levels (at a distance of 1m on the acoustic axis) were estimated by minimizing the misfit between the model levels and the received levels corrected for transmission loss and source level. Minimization was done in a least squares sense over all receivers that heard the click. Again, because receiver sensitivity was unknown, only relative click source levels could be found; the resulting relative click source levels are shown in Figure 2.2.8. A total of 14 complete click series and 2 incomplete click series (at the beginning and the end of the dataset) were recorded, where a series is defined as ending in a creak or at least 5 seconds of silence. Click levels vary by about 20 dB (in agreement with the dynamic range reported by Madsen *et al.* [2002]) and tend to steadily decrease toward the end of each click series. There were no apparent correlations between the inter-click intervals and source levels, or between whale depth (or orientation) and source level. However, as shown in Figure 2.2.9, there is a significant

relationship between click level and the order of the click within its series. This suggests that the variation in click level may be a consequence of the click production mechanism, whereby a click series begins at some constant level and decreases with each subsequent click. However, other explanations are possible. For example, since we do not have target range information it is not possible to determine if level is controlled by some automatic gain control mechanism, as might be employed by dolphins [Au and Benoit-Bird 2003; Au and Herzing 2003] but possibly not by beaked whales [Madsen *et al.* 2005].

Re-calculating the directivity indices after correcting with these click source levels gave indices of 22.9 dB, 9 dB, and 5 dB for the p1, p0, and LF components, respectively. Since corrected beam patterns were very similar to those in Figure 2.2.6 they are not presented.

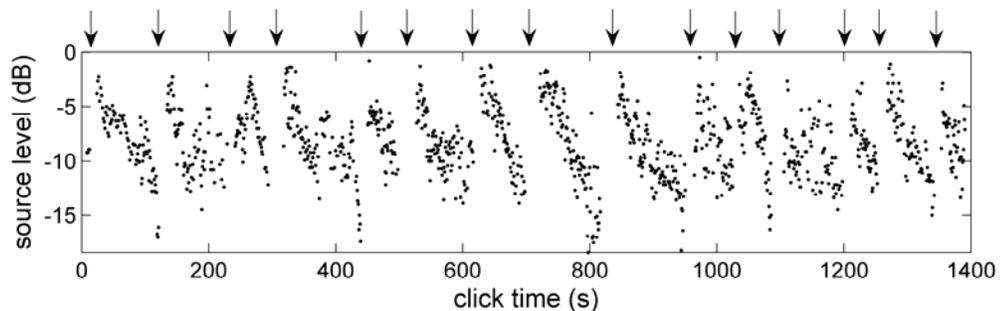


Fig. 2.2.8 Source levels relative to the strongest recorded source level as a function of time. Click levels decline by 10–15 dB from the start to the end of most click series. The beginning of each series is indicated by an arrow at the top of the figure.

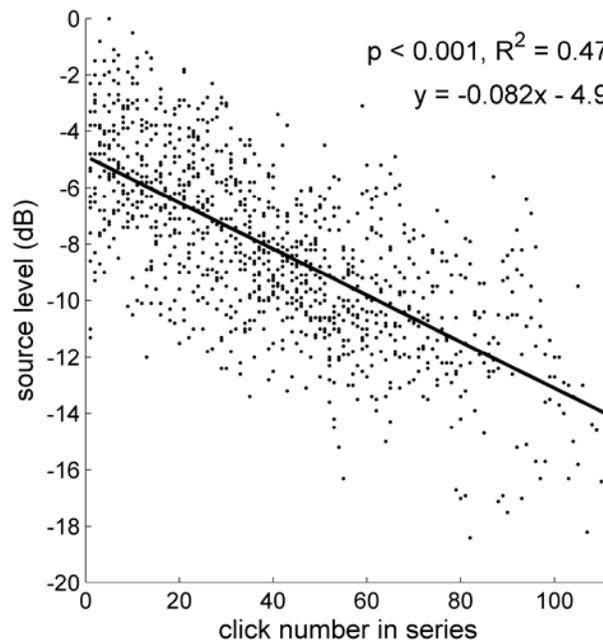


Fig. 2.2.9 Relative source level as a function of click number within its series. Data are pooled from all 14 complete click series. The significant correlation and the negative slope of the regression line suggest that click level decreases with click number within a series.

2.2.6 Concluding remarks

Although our method to recover roll is specific to sperm whales, our estimation of pitch and yaw is applicable to any clicking marine mammal. Since beam patterns exhibit rotational symmetry, at least to a first approximation – as for sperm whales [Zimmer *et al.* 2005a; this paper] and for bottlenose dolphins [Au 1993] – it may be useful to estimate beam patterns as a function of off-axis angle only. In that case, roll is not needed, so the methods developed here can be used to obtain directivity indices and estimates of click level for any clicking marine mammal recorded on bottom-mounted hydrophones in the wild.

Chapter 3

PAIR-WISE SPECTROGRAM PROCESSING

It is not always possible to extract time-of-arrival information as easily as in the single sperm whale dataset considered in Chapter 2. Difficulties arise in shallow water environments that have many reflections, in datasets with multiple calling animals, in noisy environments, and with long duration calls (typical of baleen whales) for which direct and reflected arrivals overlap and interfere. Pair-wise waveform (PWW) and pair-wise spectrogram (PWS) processing were developed to deal with these situations. PWW and PWS extend MFP methods to localize unknown sources of any frequency using widely spaced arrays. As in the DRTD/TOAD method, PWW/PWS use acoustic propagation models that account for depth-dependent sound-speed profiles and probability density functions to quantify error.

Section 3.1 (Paper 3) develops PWW and PWS processing and tests them using a simulated dataset. Section 3.2 introduces a modified version of PWS processing that is computationally less demanding. It also explores the effects of certain parameters in PWS, such as the number of points used to create spectrograms. PWS is applied to another simulated dataset, which covers greater ranges and higher frequencies than in Section 3.1, and it is compared to the TOAD method.

3.1 Paper 3

Nosal E-M, LN Frazer (2006). Pair-wise processing of spectrograms for localization of multiple broadband CW sources. *Newsletter of the IEEE Ocean Engineering Society*, Winter 2006.

Abstract - A pair-wise processing algorithm has been developed to localize broadband sources in shallow water. A simple sparse hydrophone array with number of elements roughly equal to the maximum number of sources is used. The sources can be continuous-wave (i.e. no onset times), and no previous knowledge of source signatures is required. The processor is spatially coherent and partially frequency coherent. Simulations show considerable improvement over conventional (i.e. frequency incoherent) matched field techniques under realistic noise conditions, with environmental mismatch and multiple sources. Spectrograms have been incorporated into the algorithm to make use of higher frequencies at greater ranges. Our work is motivated by the problem of localizing multiple singing humpback whales.

3.1.1 Introduction

The goal of our research is to extend and implement passive acoustic localization algorithms for use in tracking vocalizing humpback whales on winter breeding grounds. Acoustical techniques have advantages over visual and tagging techniques since they are non-invasive and unobtrusive, they are not interrupted by poor weather conditions or lack

of daylight, they enable continuous and remote sensing, and they are cost and time efficient.

Although localization methods for underwater sources have made great progress over the last 25 years [reviewed in Tolstoy *et al.* 1993; Baggeroer *et al.* 1993; Hursky *et al.* 2004], their application to humpback whale localization is problematic because of their need for large numbers of hydrophones, e.g. vertical line arrays, assets that are seldom available to scientists studying whales. Accordingly, acoustical methods for locating whales have often relied on simple assumptions, such as constant sound-speeds and straight-line propagation, that are not satisfied by the shallow water environments in which humpbacks are usually found [Chapman 2004]. Model-based methods (i.e. those that use computer models of acoustic propagation) are desirable in this problem, but available model-based methods can be difficult to apply, mainly due to the characteristics of humpback vocalizations. These include:

- Unknown waveforms: the whale's song is not known (or, technically, how far the whale is into the song is unknown).
- Continuous waveforms: song units typically consist of up-sweeps, down-sweeps, and constant-frequency contours [Payne and McVay 1971].
- Multiple sources: singers tend to space themselves about 4-6 km apart [Frankel 1994], although the spacing becomes tighter with increasing density of whales.
- Broadband, mid-frequencies: 30 Hz – 8kHz [Winn and Winn 1978].

Since model-based algorithms depend on the agreement of measured signals with synthetic signals, they have difficulty with high-frequency sources, i.e. sources located many wavelengths from receivers. For source-receiver offsets many wavelengths long,

fluctuations and uncertainties in sound speed profile and bathymetry distort actual signals to the point where they no longer agree with signals synthesized under the assumption of a constant environment. Thus popular model-based techniques have been limited to low frequencies (well below 1 kHz) where such environmental mismatch is less harmful. Only recently have mid and high frequencies started being explored for use in source localization [Hodgkiss *et al.* 1997; Hursky *et al.* 2004].

Another limit to existing techniques is that most require additional assumptions about the source. For example it is often assumed that there is only one source, or that the source waveform is known, impulsive, or narrow-band. In addition, as noted above, some techniques rely on line arrays. We consider arrays with a few hydrophones separated by many source wavelengths (sparse arrays) because they are often the only type of array available to whale researchers, and they are usually the simplest and least costly type of array [Hayes *et al.* 2000]. Various sparse arrays currently in operation (e.g. AUTECH, PMRF, and the Southern California Offshore Acoustic Range) can be used to gather marine mammal data.

In this work we address the problem of low spatial resolution (of arrays with relatively few hydrophones) by utilizing the frequency coherence of the source signal as well as its spatial coherence; we do this without the usual requirement that the source signal be known. We address the problem of lowered coherence at high frequencies by processing spectrograms instead of waveforms.

3.1.2. Algorithms

To deal with unknown, continuous-wave sources, a pair-wise waveform (PWW) processor is used. It is an extension of the pair-wise inversion technique of Frazer and Sun [1998], with application of ideas from Westwood's broadband processor [1992]. Here we assume that all hydrophones have the same unknown transfer function; arrays in which different hydrophones have different transfer functions can be accommodated by four-wise processors [Frazer and Sun 1998] with lower resolution.

To understand the PWW processor, consider the received signals at two hydrophones, $R_1(\omega)$ and $R_2(\omega)$. Let $\hat{G}_1(\omega)$ and $\hat{G}_2(\omega)$ denote the channel Green's functions from the source to the first and second hydrophones, respectively. The received spectra (measured) are the products of the source spectrum, $W(\omega)$, with the impulse responses, i.e.

$$R_n(\omega) = W(\omega)\hat{G}_n(\omega) \quad n = 1, 2.$$

Now, let $G_n(\bar{x}, \omega)$ denote the modeled Green's functions between receiver n and candidate source location \bar{x} . We introduce the following two products:

$$H_{12}(\bar{x}, \omega) = R_1(\omega)G_2(\bar{x}, \omega)$$

$$H_{21}(\bar{x}, \omega) = R_2(\omega)G_1(\bar{x}, \omega).$$

Denote the correct source location by \bar{s} . Then $G_n(\bar{s}, \omega) \approx \hat{G}_n(\omega)$ (approximately since the propagation model cannot be perfect). This leads to:

$$H_{12}(\bar{s}, \omega) \approx W(\omega)\hat{G}_1(\omega)\hat{G}_2(\omega) \approx H_{21}(\bar{s}, \omega).$$

For a single pair of receivers our PWW processor (a probabilistic indicator of source location) is given by

$$\varphi(\bar{x}) = \frac{\sum_{\omega} \left[(H_{12})^* H_{21} + (H_{21})^* H_{12} \right]}{\sum_{\omega} |H_{12}|^2 + \sum_{\omega} |H_{21}|^2} \quad (3.1.1)$$

where * denotes conjugation. The reason for this definition of $\varphi(\bar{x})$ can be understood as follows. Think of H_{12} and H_{21} as two complex column vectors with frequency as the row index. Concatenate them twice, once with H_{12} above H_{21} , then vice-versa, to make two longer vectors. Then $\varphi(\bar{x})$ is just the normalized inner product of these two longer vectors. The definition above is preferable to just taking the inner product between H_{12} and H_{12} directly because it adds symmetry to the algorithm; it does not matter which receiver is named 1 and which is named 2. By the Cauchy-Schwartz inequality, the processor reaches its maximum value (unity) when $H_{12} = H_{21}$. In particular, $\varphi(\bar{x})$ is maximized at the true source location $\bar{x} = \bar{s}$.

To reduce computational requirements, note that $H_{12}^* H_{21} = (H_{21}^* H_{12})^*$.

Consequently, the PWW processor can be written as:

$$\varphi(\bar{x}) = \frac{2 \operatorname{Re} \left(\sum_{\omega} H_{12}^* H_{21} \right)}{\sum_{\omega} |H_{12}|^2 + \sum_{\omega} |H_{21}|^2} \quad (3.1.2)$$

The single-pair PWW processor can be generalized to $N > 2$ receivers by summing coherently over receiver pairs:

$$\varphi_{pww}(\bar{x}) = \frac{2 \operatorname{Re} \left(\sum_{\omega} \sum_{i=1}^{N-1} \sum_{j>i}^N H_{ij}^* H_{ji} \right)}{\sum_{\omega} \sum_{i=1}^N \sum_{j \neq i}^N |H_{ij}|^2} . \quad (3.1.3)$$

To address the problem of incoherence at long ranges, we process spectrograms instead of waveforms. We call this the pair-wise spectrogram (PWS) processor.

Spectrograms are less sensitive to mismatch and fluctuations in the ocean wave-guide, particularly at higher frequencies. Our use of spectrograms is in the spirit of envelope processing [Hursky *et al.* 2004] in which signal envelopes are processed instead of waveforms. In contrast to envelope processing, however, PWS processing retains both time and frequency characteristics, and can still benefit from coherence at low frequencies.

Let $S_{ij}(\bar{x}, t, f)$ denote the spectrogram formed from $H_{ij}(\bar{x}, \omega)$, where t and f are time and frequency steps respectively. Above the crossover frequency fc (to be determined), only the envelope of each channel is processed, and the mean is removed from each envelope because a constant offset holds no information. The formula for the PWS processor is analogous to that of the PWW processor:

$$\varphi_{pws}(\bar{x}) = \frac{2 \operatorname{Re} \left(\sum_t \sum_f \sum_{i=1}^{N-1} \sum_{j>i}^N S_{ij}^* S_{ji} \right)}{\sum_t \sum_f \sum_{i=1}^N \sum_{j \neq i}^N |S_{ij}|^2} \quad (3.1.4)$$

Once again, intuition into this processor is gained by thinking of the spectrograms as vectors and taking normalized inner products. Of course, a weighting over frequencies and/or times may be introduced in either of the processors above to emphasize or de-

emphasize certain aspects of the signal. For a slowly drifting source, for example, it may be advantageous to put more weight on more recent times.

3.1.3. Simulation Specifics

The Bartlett, PWW, and PWS processors were implemented in MATLAB. Bellhop [Porter and Bucker 1987] was used to model impulse responses. Simulations were run for a 700 m by 700 m by 200 m (constant depth) area. The sound speed profile used was typical of that seen in Hawaiian winter waters. It was based on historical values taken from the Generalized Digital Environmental Model. A humpback whale signal, 40 s long and sampled at 2 kHz, was propagated (by convolution with modeled impulse responses) from several source locations to several receiver locations within the range. A spectrogram of the signal used is shown in Figure 3.1.1.

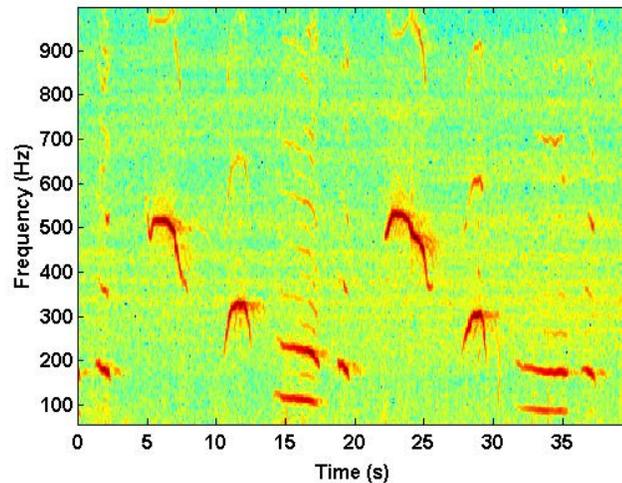


Fig. 3.1.1 Spectrogram of humpback whale signal used in simulations.

Simulations shown here are for a case with 1 source and three receiver, and another case with two sources and four receivers. Figures 3.1.2 and 3.1.3 show the source/receiver layouts as well as the search areas from a top down perspective. In the first case the source was at (252 m, 304 m, 60 m) and the three receivers were at (47 m, 102 m, 60 m), (175 m, 647 m, 60 m), and (603 m, 200 m, 60 m). In the 2 source case, the first source and three receivers were in the same positions as in the 1 source case. The second source was at (452 m, 573 m, 60 m) and the fourth receiver was at (677 m, 697 m, 60 m). The same signal was used for the second source as for the first source, with the first 20 s and the last 20 s swapped in the time domain.

Simulated noise was of the worst-case type: many noise sources with source signatures identical to that of the actual source, except for their randomized strengths and start times. These “noise whales” were placed at every grid point in the search area. The signals from the “noise whales” were propagated to the receivers and summed in time to give the background noise. The power of the “noise whales” was adjusted to give a specified average signal-to-noise ratio (SNR) over all receivers.

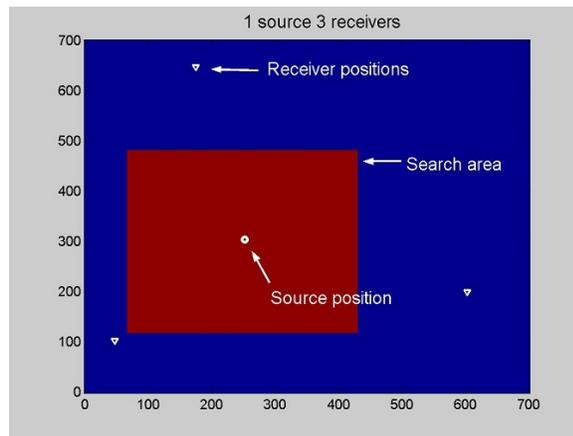


Fig. 3.1.2. Simulation layout for 1 source simulation. Vertical and horizontal distances are in m.

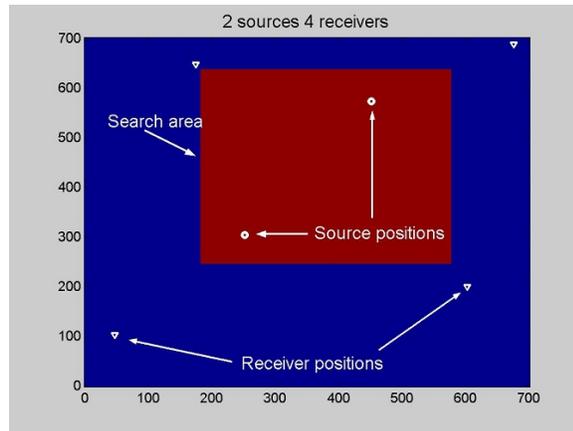


Fig. 3.1.3. Simulation layout for 2 source simulation. Vertical and horizontal distances are in m.

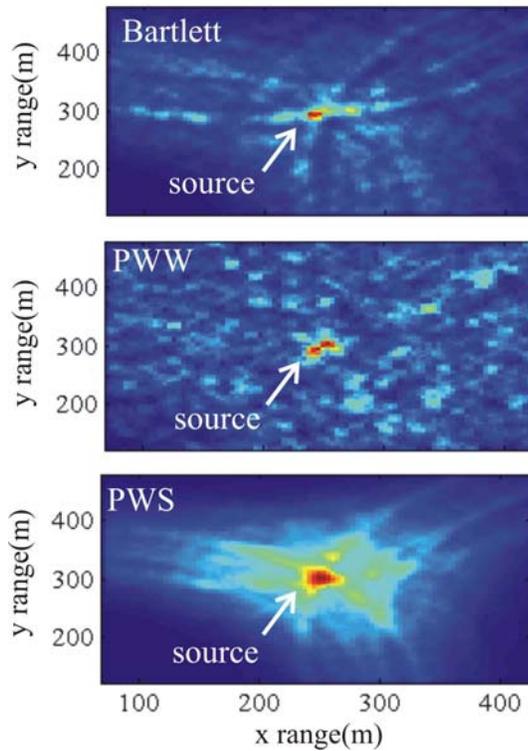


Fig. 3.1.4. SNR 5 dB. 1 source, 3 receivers. All three processors successfully localize the source

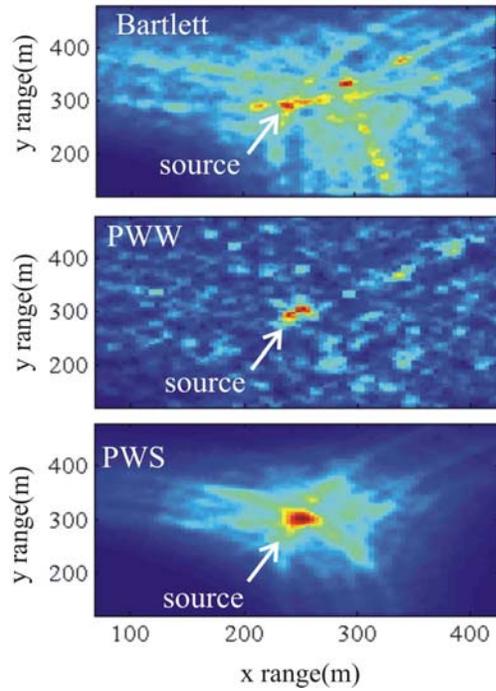


Fig. 3.1.5. SNR 0 dB. 1 source, 3 receivers. The PWW and PWS processors successfully localize the source which is lost to the Bartlett processor.

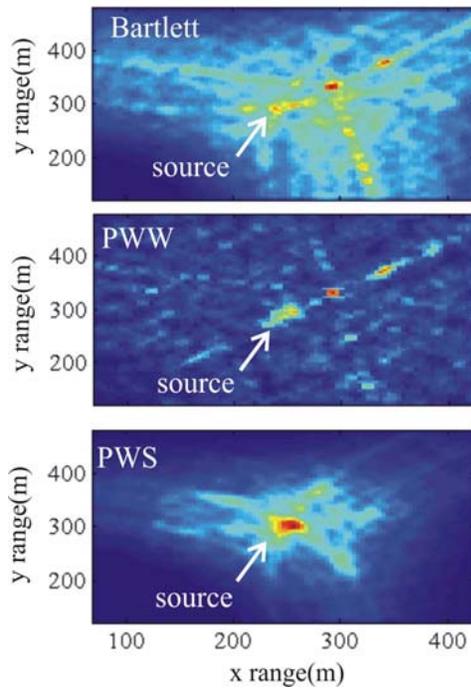


Fig. 3.1.6. SNR -5 dB. 1 source, 3 receivers. Only the PWS processor successfully localizes the source.

After generation of the noisy, synthetic data, the PWW, PWS, and Bartlett algorithms were used to try to locate the whales. The grid used in the localization was at a single depth (60 m) and grid spacing was 4 m (ultimately, searches will be conducted over several depths). Spectrograms were generated using a 256 point FFT. Signals were Hanning windowed prior to computing each spectrum, and there was a 128 point overlap between successive time windows of the spectrogram. Environmental mismatch was introduced in the form of incorrect water depth; all inversions shown are for a depth of 204 m rather than of 200 m. SNRs were gradually decreased to explore the localization error due to noise. Only frequencies up to 200 Hz were used in the Bartlett and PWW processors since higher frequencies became too incoherent to add useful information. For the PWS processor, the crossover frequency f_c was 100 Hz.

3.1.4. Discussion

Ambiguity surfaces for the three processors for increasing levels of noise are shown in Figures 3.1.4-3.1.6. The images have been individually scaled so that red and blue correspond to the maximum and minimum surface values, respectively. At 5 dB SNR, all three processors localize the source accurately. At 0 dB SNR, however, spurious sources begin to appear with the Bartlett processor, while the pair-wise processors still find the source. Increasing the noise to -5 dB, only the PWS processor correctly localizes the source. Only the PWS processor is able to localize both sources in the 2 source case with 0 dB SNR (Fig. 3.1.7).

It is evident from these simulations that the PWS is more robust with respect to noise and mismatch than the other two processors, but that this benefit is gained by loss

of resolution. Indeed, the localized sources in the PWS surface are “smeared” over a larger area. This effect is minimized by raising the cutoff frequency, f_c , and/or by reducing the number of points in the spectrogram FFT window.

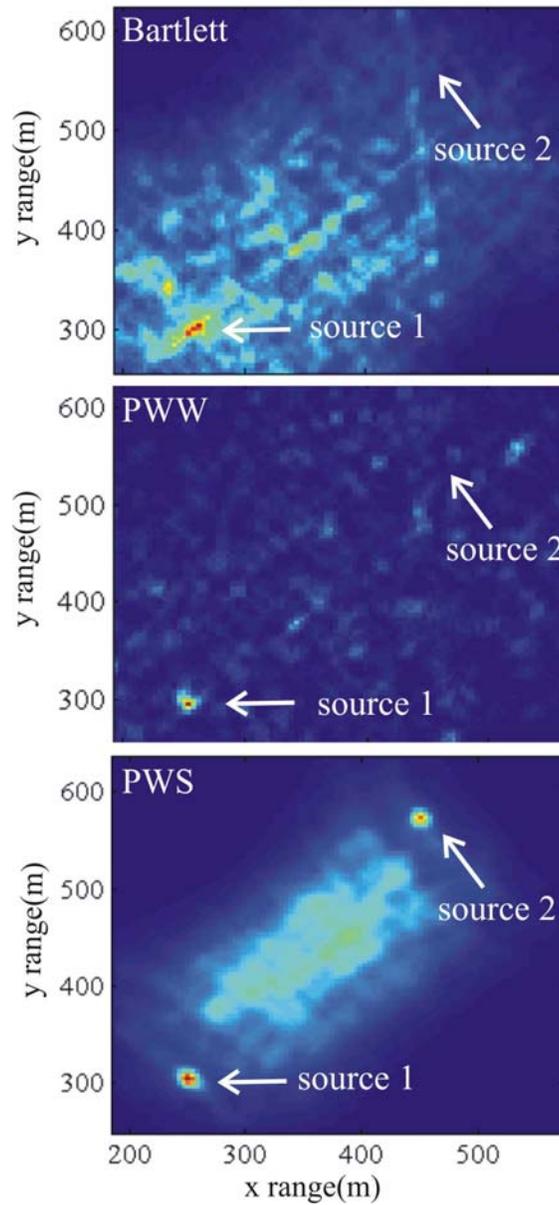


Fig. 3.1.7 SNR 0 dB. 2 sources, 4 receivers. Only the PWS processor localizes both sources (4 receivers).

3.1.5. Conclusion

A pair-wise spectrogram (PWS) processor has been proposed for the localization of multiple broad-band unknown continuous-wave sources in shallow water. It appears robust with respect to mismatch and noise; with only three receivers, a single source could be localized under conditions of both environmental mismatch and signal to noise ratios worse than -5 dB. Two identical (but unknown) sources could be localized in mismatch and 0 dB SNR. Many aspects of these processors remain to be explored through simulations and analysis of real data.

Of interest for future work is the use of cochleagrams instead of spectrograms in the PWS processor. These are auditory representations of sound (as heard by whales) whereby a cochlear filter-bank [Helweg *et al.* 2000; Moore 2003] is applied to the signal. Being biologically more relevant than spectrogram processing, cochleagrams may prove to aid in localization – perhaps the whales know something we don't.

Currently, efforts are being made to modify the PWS processor to further reduce computational requirements, which will allow us to work at longer ranges and at higher sampling frequencies.

3.2 Paper 4

Nosal E-M, LN Frazer (in press). Modified pair-wise spectrogram processing for localization of unknown broadband sources. *IEEE Journal of Ocean Engineering*.

Abstract—Pair-wise waveform (PWW) and pair-wise spectrogram (PWS) processors for 3-D localization of unknown, continuous-wave, broadband sources in shallow water have been developed and implemented [Nosal and Frazer 2006b]. The processors use sparse hydrophone arrays and are applicable to multiple sources, which can be unknown, continuous-wave, and broadband. Here, we give new formulas for these two processors that significantly reduce computational requirements, making localization at longer ranges and higher frequencies feasible. The new processors are motivated by a demonstration that an incoherent version of the PWW processor (in which processor outputs at different frequencies are averaged after being processed independently) is the Bartlett processor without auto-receiver terms. The new PWW processor is mathematically equivalent to the original version, though much faster. The new PWS processor is mathematically equivalent to the original version only in the limit of infinite spectrogram window length, but for window lengths that are optimal with the old PWS processor, the new PWS processor gives essentially the same results with much greater speed. Simulations comparing PWS processing to Bartlett, PWW processing, and a time difference of arrival method indicate that the main advantage of PWS processing is for multiple sources in uncertain, high noise environments at ranges many wavelengths long. With PWS, increased robustness with respect to mismatch is obtained at the expense of

reduced resolution; varying PWS processor parameters (such as the size of windows used to create spectrograms) optimizes this tradeoff. This work is motivated by the problem of localizing singing humpback whales, and simulation results use whale sources.

3.2.1 Introduction

Acoustic techniques for localization of marine mammals are an important addition to visual and tagging techniques. Despite recent advances in localization methods for underwater sources [Baggeroer *et al.* 1993; Tolstoy 1993; Hodgkiss *et al.* 1997; Hursky *et al.* 2004], acoustic techniques are seldom used to track singing humpback whales. This is because the problem is complicated by limited resources, by the source characteristics (unknown, continuous-wave [Payne and McVay 1971], broadband [Winn and Winn 1978], multiple sources [Frankel 1994]), and the by the environment (shallow water, sound-speed ducts, long ranges, and sound-speed uncertainties).

The pair-wise waveform (PWW) and pair-wise spectrogram (PWS) processors [Nosal and Frazer 2006b] were developed with the humpback localization problem in mind. Both processors are in the spirit of matched field processing (MFP). That is, the measured sound field is compared to the predicted sound field for a source at a candidate position. A grid of candidate source positions is examined to find the point of best agreement. In contrast to most MFP methods, which typically rely on line arrays [Baggeroer *et al.* 1993; Tolstoy 1993; Hodgkiss *et al.* 1997], the PWW and PWS processors use (fixed or floating) sparse hydrophone arrays (few hydrophones separated by many source wavelengths), as these are more available to marine mammal researchers due to their simplicity and lower cost [Hayes *et al.* 2000]. Moreover, data may be

available from large arrays currently in operation (e.g. AUTEK, PMRF, and the Southern California Offshore Acoustic Range).

For unknown, continuous-wave source signals, pair-wise processing [Westwood 1992; Frazer and Sun 1998] is used to retain partial frequency (as well as spatial) coherence. This improves the spatial resolution of arrays with relatively few hydrophones and strengthens performance for multiple sources. At high frequencies (i.e. sources many wavelengths from receivers), measured signals lose coherence due to environment fluctuations and uncertainties, which limits the applicability of the PWW and other waveform based MFP methods [Baggeroer *et al.* 1993; Tolstoy 1993; Westwood 1992] to lower frequencies (typically well below 1kHz). In the PWS processor, this problem is addressed by processing spectrograms instead of waveforms.

PWS processing is related to envelope processing [Hursky *et al.* 2004], which has successfully been used to extend MFP methods to high frequencies. To some extent, spectrograms can be thought of as generalized envelopes since taking the magnitude of (individual points of) a spectrogram channel at one frequency is similar to taking the envelope of the signal after applying a band-pass filter centered at that frequency. In this light, the reason that spectrogram processing is expected to work better than waveform processing at high frequencies (or in cases with significant environmental mismatch and noise) is intuitive; a small change in a waveform that causes measured and synthetic waveforms to disagree has a much smaller effect on the spectrograms, so that measured and synthetic spectrograms can still agree quite closely. PWS processing has the advantage over envelope processing of retaining both time and frequency information and permitting high and low frequencies to be processed differently.

Spectrograms have previously been used to estimate time of arrival differences in marine mammal detection methods [Mellinger and Clark 2000] and hyperbolic fixing localization methods [Clark *et al.* 1986; Janik *et al.* 2000; Clark and Ellison 2000; Tiemann *et al.* 2004]. Low frequency spectrograms have also been used to determine group and phase velocity relationships for long-range localization using a single hydrophone [Kuperman *et al.* 2001]. The PWS processor is the first (to our knowledge) to apply spectrograms in a MFP approach.

In their original forms [Nosal and Frazer 2006*b*] the PWW and PWS processors were computationally inefficient, which greatly constrained their usefulness. In this paper we demonstrate the equivalence of Bartlett (linear) MFP, with auto-receiver terms removed, to an incoherent version of PWW processing in which frequencies are separately processed and then averaged. The demonstration leads, very usefully, to a form of the PWW processor that is less intuitive but computationally much more efficient than the original form. After a review of the PWS processor, a similar computationally efficient form of the PWS processor is developed.

Using these new, more efficient forms of the PWW and PWS processors, we were able to run simulations for much longer ranges and higher frequencies. The simulations showed that larger FFT windows (than were previously [Nosal and Frazer 2006*b*] used) to create spectrograms make the PWS processor more successful at localizing multiple sources at long ranges (on the order of several kilometers at 2 kHz). Specifically, for simulations with significant environmental mismatch we are able to localize 2 sources with only 3 receivers in signal-to-noise ratios (SNRs) as poor as -10 dB; by comparison, 4 receivers and 0 dB SNR were previously [Nosal and Frazer 2006*b*] required to localize

2 sources. Our simulations compare the PWS processor with a time of arrival difference (TOAD) method, the Bartlett processor and the PWW processor.

3.2.2 Overview of pair-wise waveform (PWW) processing

The PWW processor [Westwood 1992; Frazer and Sun 1998; Nosal and Frazer 2006b], is used to deal with unknown, continuous-wave sources. Let $R_i(\omega)$ denote the Fourier-transformed received signal at the i^{th} hydrophone (N hydrophones in total), let $\hat{G}_i(\omega)$ denote the unknown true Green's function between the unknown actual source location \bar{s} and the i^{th} hydrophone, let $G_i(\bar{x}, \omega)$ denote the modeled Green's function between the i^{th} hydrophone and a candidate source location \bar{x} , and let $W(\omega)$ denote the source spectrum. For all receiver pairs, define the following products:

$H_{ij}(\bar{x}, \omega) = R_i(\omega)G_j(\bar{x}, \omega)$. The PWW processor is then given by:

$$\begin{aligned} \varphi_{pww}(\bar{x}) &= \frac{\sum_{\omega} \sum_{i=1}^N \sum_{j \neq i}^N H_{ij}^* H_{ji}}{\sum_{\omega} \sum_{i=1}^N \sum_{j \neq i}^N |H_{ij}|^2} \\ &= \frac{2 \operatorname{Re} \left(\sum_{\omega} \sum_{i=1}^{N-1} \sum_{j>i}^N H_{ij}^* H_{ji} \right)}{\sum_{\omega} \sum_{i=1}^N \sum_{j \neq i}^N |H_{ij}|^2} \end{aligned} \quad (3.2.1)$$

where $*$ denotes conjugation. The first line shows that PWW is just a normalized inner product, symmetric over receivers. The second line holds because $H_{ij}^* H_{ji} = (H_{ji}^* H_{ij})^*$.

Since $R_i(\omega) = W(\omega)\hat{G}_i(\omega)$ and $G_i(\bar{s}, \omega) \approx \hat{G}_i(\omega)$ for all receivers (approximately since the propagation model cannot be perfect), it follows that

$$H_{ij}(\bar{s}, \omega) \approx W(\omega)\hat{G}_i(\omega)\hat{G}_j(\omega) \approx H_{ji}(\bar{s}, \omega).$$

Thus, by the Cauchy-Schwartz inequality, the processor reaches its maximum (ideally unity) at the true source location $\bar{x} = \bar{s}$.

A weighting over frequencies and/or times may be introduced to emphasize aspects of the signal. For a slowly drifting source, for example, it may be advantageous to put more weight on more recent times. Also, it is important to window time and frequency channels with windows tapered at both ends to avoid ringing that manifests as spurious sources.

Our PWW processor differs from Westwood's [1992] processor in two respects. First, in the PWW processor the order of receivers does not matter (evident from the first equality in Eq.(3.2.1)), while in Westwood's processor it does (see the Appendix). This receiver symmetry is at the expense of losing the imaginary parts in the numerator products (and hence some frequency coherence). The second difference is in how the products are normalized.

3.2.3 Relation between PWW, incoherent PWW (IPWW), and Bartlett processing

We first show that the Bartlett linear processor [Bucker 1976; Tolstoy 1993] with the auto-receiver terms removed is equivalent to a frequency-incoherent version of the PWW processor, i.e., a PWW processor in which frequencies are processed individually

and then an average is taken over the outputs at different frequencies. We refer to this as the incoherent PWW, or IPWW, processor. To see the equivalence, first define vectors

$$\bar{R}(\omega) = \begin{bmatrix} R_1(\omega) \\ R_2(\omega) \\ \vdots \\ R_N(\omega) \end{bmatrix} \quad \text{and} \quad \bar{G}(\bar{x}, \omega) = \begin{bmatrix} G_1(\bar{x}, \omega) \\ G_2(\bar{x}, \omega) \\ \vdots \\ G_N(\bar{x}, \omega) \end{bmatrix}.$$

The Bartlett processor is given by [Bucker 1976; Tolstoy 1993]:

$$\varphi(\bar{x}) = \frac{1}{N_\omega} \sum_{\omega} \frac{|\bar{R}^* \bar{G}|^2}{|\bar{R}|^2 |\bar{G}|^2} \quad (3.2.2)$$

where N_ω is the number of frequencies. Expanding terms in the numerator gives

$$\begin{aligned} |\bar{R}^* \bar{G}|^2 &= \left(\sum_i R_i^* G_i \right)^* \left(\sum_j R_j^* G_j \right) \\ &= \sum_i \sum_j (R_i G_j) (R_j^* G_i^*) \\ &= \sum_i \sum_j H_{ij} H_{ji}^* \end{aligned}$$

and in the denominator,

$$\begin{aligned} |\bar{R}|^2 |\bar{G}|^2 &= \sum_i |R_i|^2 \sum_j |G_j|^2 \\ &= \sum_i \sum_j |R_i G_j|^2 \\ &= \sum_i \sum_j |H_{ij}|^2. \end{aligned}$$

After removal of the auto-receiver ($i = j$) terms from the numerator and denominator, the

Bartlett processor becomes

$$\begin{aligned}
\varphi(\bar{x}) &= \frac{1}{N_\omega} \sum_\omega \frac{|\bar{R}^* \bar{G}|^2 - \sum_i |R_i^* G_i|^2}{|\bar{R}|^2 |\bar{G}|^2 - \sum_i |R_i|^2 |G_i|^2} \\
&= \frac{1}{N_\omega} \sum_\omega \frac{\sum_i \sum_{j \neq i} H_{ij} H_{ji}^*}{\sum_i \sum_{j \neq i} |H_{ij}|^2} \\
&= \frac{1}{N_\omega} \sum_\omega \frac{2 \operatorname{Re} \left(\sum_{i=1}^{N-1} \sum_{j>i}^N H_{ij}^* H_{ji} \right)}{\sum_{i=1}^N \sum_{j \neq i}^N |H_{ij}|^2}. \tag{3.2.3}
\end{aligned}$$

This is just Eq. (3.2.1) applied to individual frequencies, then averaged, which is our IPWW processor.

It is evident from this derivation that the both the IPWW and PWW processors differ from the Bartlett processor since they do not include the auto-receiver ($i=j$) terms. The auto-receiver terms are constants and hold no useful information for localization, so processor resolution must increase when these terms are removed. Normalization in the PWW processor is done after all frequency contributions are summed, so each frequency is weighted proportionally to the spectral power of the received signals, while the Bartlett and IPWW processors weight all frequencies equally. Because of this, the PWW processor puts more weight on frequencies with more energy, and since these are usually the frequencies with higher signal-to-noise ratios, resolution is increased in the PWW processor. Overall, the order of processor resolution from higher to lower is PWW, IPWW, and Bartlett. Because of this, PWW is expected to outperform IPWW and IPWW is expected to outperform Bartlett in cases without much noise or mismatch and in

cases where localization includes optimization of parameters in a well-chosen environmental model. For the same reason, however, Bartlett can outperform IPWW and IPWW can outperform PWW in cases with significant mismatch, if environmental parameters are not searched. Since relative performance of the processors changes depending on the amount and nature of noise and mismatch, testing is always advisable in particular cases.

3.2.4 Reducing computational load for PWW processing

The relation between IPWW processing and Bartlett processing suggests an alternative formula for the PWW processor. Indeed, a derivation similar to that in Section 3.2.3, but in the opposite direction, gives

$$\varphi_{pww}(\bar{x}) = \frac{\sum_{\omega} \left(|\bar{R}^* \bar{G}|^2 - \sum_i |R_i^* G_i|^2 \right)}{\sum_{\omega} \left(|\bar{R}|^2 |\bar{G}|^2 - \sum_i |R_i^* G_i|^2 \right)}. \quad (3.2.4)$$

This formula for the PWW processor is less intuitive, but it is mathematically equivalent to the first version and requires much less computational effort to implement. In the first version, H_{ij} is calculated for each receiver *pair*, while in the second version, only $R_i^* G_i$ is computed for each receiver. This reduces the number of operations from $O(N^2)$ to $O(N)$, where N is the number of receivers; thus, run times are greatly reduced.

3.2.5 Overview of pair-wise spectrogram (PWS) processing

As discussed in the introduction, PWS processing [Nosal and Frazer 2006b] is used to address the problem of incoherence at long ranges. It is similar to PWW processing, except that spectrograms (which are less sensitive to environmental mismatch) are processed instead of waveforms.

Let $S_{ij}(\bar{x}, \tau, f)$ denote the complex spectrogram formed from $H_{ij}(\bar{x}, \omega)$, where τ and f are time and frequency steps, respectively. The PWS processor is defined as

$$\varphi_{pws}(\bar{x}) = \frac{2 \operatorname{Re} \left(\sum_{\tau} \sum_f \sum_{i=1}^{N-1} \sum_{j>i}^N S_{ij}^* S_{ji} \right)}{\sum_{\tau} \sum_f \sum_{i=1}^N \sum_{j \neq i}^N |S_{ij}|^2}. \quad (3.2.5)$$

As waveforms may still agree at lower frequencies, it is often advantageous to run the PWS processor twice: once with complex spectrograms for frequencies less than some crossover frequency f_c and once with magnitudes (or envelopes) for frequency channels above f_c . The results are then averaged. In the spectrogram envelope version, the mean is removed from each frequency channel because a constant offset holds no information. Envelope processing reduces error and the effects of mismatch but at the cost of lower resolution. The optimum crossover frequency f_c is determined during processing by the agreement of the complex version and the envelope version of the processor. Roughly speaking, the use of complex spectrograms in PWS gives results similar to PWW. Increases in noise level, source range, and environmental mismatch all favor a lower crossover frequency.

3.2.6 Reducing computational load for PWS processing

Surprisingly, the mathematical trick used to get the computationally efficient version of the PWW processor (Eq. (3.2.4)) can also be applied to the PWS processor. For a given candidate source location \bar{x} , let t_i denote the time of the first arrival in the modeled Green's function (between \bar{x} and receiver i). Let $g_i(\bar{x}, t)$ be the modeled time-domain Green's function advanced by t_i so that the first arrival is at time 0. Similarly, let $r_i(t)$ be the signal at receiver i in the time domain, advanced by t_i . Denote the spectrogram of $r_i(t)$ by $\tilde{R}_i = \tilde{R}_i(f, \tau)$, where $nfft$ points are used to create the spectrograms. For all time steps in $\tilde{R}_i(f, \tau)$, let $\tilde{G}_i = \tilde{G}_i(\bar{x}, f, \tau)$ be the FFT of the first $nfft$ points in $g_i(\bar{x}, t)$. If the time corresponding to $nfft$ is considerably longer than the delay time between the first and last arrivals in $g_i(\bar{x}, t)$, then $\tilde{S}_{ij} \approx \tilde{R}_i \tilde{G}_j$, in which \tilde{S}_{ij} is the spectrogram of $r_i(t) * g_j(\bar{x}, t)$ (where $*$ denotes convolution). Make the following definitions

$$\tilde{\tilde{R}}(f, \tau) = \begin{bmatrix} \tilde{R}_1(f, \tau) \\ \tilde{R}_2(f, \tau) \\ \vdots \\ \tilde{R}_N(f, \tau) \end{bmatrix} \quad \text{and} \quad \tilde{\tilde{G}}(\bar{x}, f, \tau) = \begin{bmatrix} \tilde{G}_1(\bar{x}, f, \tau) \\ \tilde{G}_2(\bar{x}, f, \tau) \\ \vdots \\ \tilde{G}_N(\bar{x}, f, \tau) \end{bmatrix}.$$

A development similar to that for the PWW processor then gives a computationally simpler version of the PWS processor:

$$\varphi_{pws}(\bar{x}) \approx \frac{2 \operatorname{Re} \left(\sum_{\tau} \sum_f \sum_{i=1}^{N-1} \sum_{j>i}^N (\tilde{R}_i \tilde{G}_j)^* \tilde{R}_j \tilde{G}_i \right)}{\sum_{\tau} \sum_f \sum_{i=1}^N \sum_{j \neq i}^N |\tilde{R}_i \tilde{G}_j|^2}$$

$$= \frac{\sum_{\tau} \sum_f \left(\left| \tilde{R}^* \tilde{G} \right|^2 - \sum_i \left| \tilde{R}_i^* \tilde{G}_i \right|^2 \right)}{\sum_{\tau} \sum_f \left(\left| \tilde{R} \right|^2 \left| \tilde{G} \right|^2 - \sum_i \left| \tilde{R}_i^* \tilde{G}_i \right|^2 \right)}. \quad (3.2.6)$$

As with the new version of the PWW processor, this form of the PWS processor reduces run times by a factor of N . The two forms of the PWS processor are not mathematically equivalent, because of the approximation $\tilde{S}_{ij} \approx \tilde{R}_i \tilde{G}_j$, which is exact only in the limit as $nfft$ becomes infinite. Nevertheless, they give indistinguishable results for the $nfft$ values thought to be optimal for processing with the exact formula. Thus, in our experience there is no reason to use the exact formula.

3.2.7 Simulated data

All simulations and localization techniques discussed were implemented in MATLAB. Simulations were run for a 10 km by 10 km by 200 m (constant depth) area with a sound speed profile typical of Hawaiian winter waters. The sound speed profile was based on historical values taken from the Generalized Digital Environmental Model [GEDM] and is shown in Figure 3.2.1. Geoacoustic properties of the seafloor were taken from typical values for sand [Fu *et al.* 2004]: density 1.86 g/cm³, compressional wave speed 1620 m/s, compressional wave attenuation 0.83 dB/wavelength. Three receivers were used in the simulations, with (x, y, depth) coordinates (1150 m, 1080 m, 30 m), (2380 m, 8620 m, 30 m), and (9340 m, 3630 m, 30 m). Figure 3.2.3 shows the receiver configuration in plan view. The grid used in the localization was at a single depth (60 m) with 200 m grid spacing. Given the sound-speed profile, bathymetry, and geoacoustic properties of the bottom, the Gaussian beam tracing model BELLHOP [Porter and

Bucker 1987; Porter and Liu 1994; Porter 2005] was used to calculate magnitudes, phases, and travel times for each source-receiver pair and each receiver-grid point pair. Green's functions were computed from these multi-path parameters.

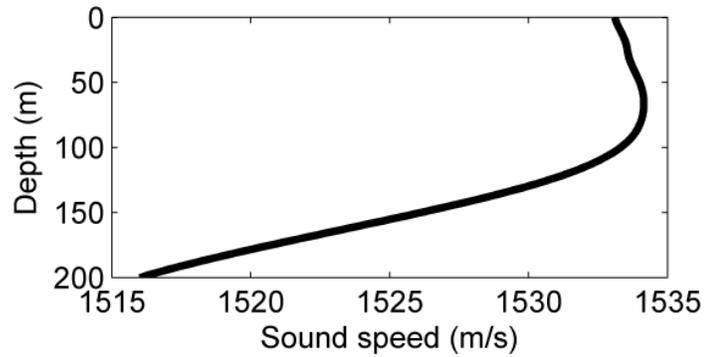


Fig. 3.2.1 Sound speed profile used in forward model; inversion assumed a homogeneous SSP with sound speed 1530 m/s.

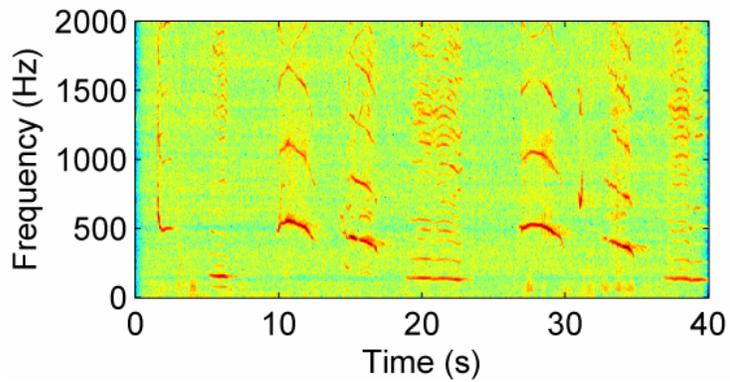


Fig. 3.2.2 Humpback whale signal used in simulations.

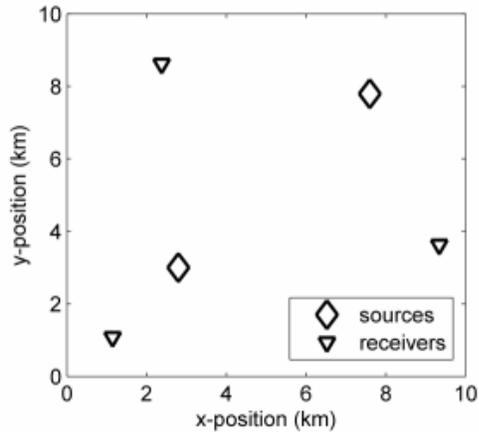


Fig 3.2.3 Simulation configuration

To create the simulated signal, a humpback whale signal (shown in Figure 3.2.2), 40 s long and sampled at 4 kHz, was propagated (by convolving the source signal with the appropriate Green's function) from two source positions with (x, y, depth) coordinates (2800 m, 3000 m, 60 m) and (7600 m, 7800 m, 60 m). Source positions are shown in Figure 3.2.3. The signal, recorded approximately half a mile from the singer, was taken from 4 minutes and 10 seconds into track 1 of the Audio CD, Rapture of the Deep [Knapp 2001]. The first 20 s and the last 20 s of the humpback signal at the second source position were swapped, so that the two sources were similar but not identical. For sources identical in signature and timing, the received signals from individual whales are nearly indistinguishable, and all processors perform poorly. Moreover, the case of identical sources is unrealistic; humpbacks sing very similar songs, but do not sing the same part of the song at precisely the same time.

The source positions were intentionally placed on grid points. This was done in fairness to the Bartlett and PWW processors. A coarse grid spacing of 200 m was chosen

for reduced run times, but it can disproportionately degrade the performance of the Bartlett and PWW processors; in these wave-based processors, sidelobes can overpower true sources when grid spacing is not sufficiently fine [Tolstoy 1993]. Placing the sources on grid points helped to minimize this effect.

Noise was of the worst-case type: many noise sources with source signatures identical to that of the sources, except for their randomized strengths and phase shifts. These “noise whales” were placed at every grid point in the search area. The signals from the “noise whales” were propagated to the receivers and summed in time to give the background noise. The power of the “noise whales” was adjusted to give a specified average SNR over all receivers. In other simulations, different types of noise (such as ambient noise recorded in Hawaii) were used, instead of and in addition to, the “noise whales”, but the “noise whales” consistently resulted in worse performance of the processors. This is likely because, in the “noise whale” case, the noise levels are highest at the same frequencies where the source levels are highest.

3.2.8 Localization method parameters and specifics

After generation of the noisy, synthetic data a time difference of arrival (TOAD) method [Tiemann *et al.* 2004], the Bartlett processor, the PWW processor and the PWS processor were used to try to locate the simulated singers for varying SNRs.

Environmental mismatch was introduced in the form of incorrect sound speed profile and water depth; all inversions assumed homogeneous sound speed (1530 m/s) and water depth 204 m (rather than 200 m). The processors were used to create ambiguity surfaces

(probabilistic indicators of source position) in which the value for the ambiguity surface at each grid point is the processor value at that point.

3.2.8.1 TOAD Method

The TOAD method is from Tiemann *et al.* [2004], and further details can be found in their paper. For each receiver pair, the 40 s signals were windowed into 10 s frames with 2 s overlap. For each frame and receiver pair, a digitized-spectrogram correlation method [Tiemann *et al.* 2004] was used to estimate time-lags between whale calls. Spectrograms were generated using 512-point Hanning FFT windows with 90% overlap. Since there was more than one source, the two time-lag bins that gave the highest spectral correlation scores were chosen provided they exceeded a 100-point threshold. This differs from Tiemann *et al.* [2004], in which only one peak was taken at a time since only one source was being localized. Here, taking fewer or greater than two time-lags reduced performance. If only one time-lag was chosen, fewer correct sources were found, and if more than 2 time-lags were chosen, more spurious sources (and no new correct sources) were found. This optimization of the TOAD method for two sources used information that might not be available in a real application.

It is expected that a group of time-lags corresponding to the same whale call satisfies the following:

$$\left| \hat{T}_{12}^{mg} + \hat{T}_{23}^{mg} - \hat{T}_{13}^{mg} \right| \leq \varepsilon \quad (3.2.7)$$

where \hat{T}_{ij}^{mg} is the measured time-lag between hydrophones i and j for time frame m and time-lag group g , and ε is a tolerance factor to account for the lack of precision in time-

lag measurements. Only groups of lags satisfying Eq. (3.2.7) with $\varepsilon = 0.1$ s were used (this is the value that gave best results). More than one surface was formed for each time frame when more than one group of lags satisfied Eq. (3.2.7). Surfaces for each time frame m , lag group g , receiver pair ij , and grid point \bar{x} were created according to Eq. (20) of Tiemann *et al.* [2004]:

$$L_{ij}^{mg}(\bar{x}) = \sqrt{\frac{|T_{ij}(\bar{x}) - \hat{T}_{ij}^{mg}|}{d_{ij}/c}}. \quad (3.2.8)$$

In the above, $T_{ij}(\bar{x})$ is the modeled time lag between receivers i and j for a source at grid point \bar{x} , d_{ij} is the distance between receiver pair ij , and c is the minimum possible sound speed. The denominator of Eq.(3.2.8) is a normalization by the maximum possible time lag between receiver pair ij [Tiemann *et al.* 2004]. Surfaces A_g^m for each time frame m and lag group g were generated according to an exponentiated version of Eq. (22) of Tiemann *et al.* [2004]:

$$A_g^m(\bar{x}) = \exp\left(\sum_{i=1}^N \sum_{j>i} L_{ij}^{mg}(\bar{x}) \alpha_{ij}(\bar{x})\right) \quad (3.2.9)$$

where $\alpha_{ij}(\bar{x})$ is the total predicted transmission loss in dB for the two paths between each of the receivers i and j and grid point \bar{x} . Exponentiation was used so that surfaces have maxima (and not minima) in the most likely whale positions, as is the case for the other processors. The surfaces A_g^m were combined for the m^{th} frame by taking the maximum surface values over all groups at each grid point.

$$S^m(\bar{x}) = \max_g A_g^m(\bar{x}). \quad (3.2.10)$$

Since the sources are stationary, surfaces for all frames were combined (again via maximums) to give an overall ambiguity surface

$$S_{total}(\bar{x}) = \max_m S^m(\bar{x}). \quad (3.2.11)$$

Maximums were used to combine the surfaces (rather than summation, for example) because they gave the best results. By taking maximums, correctly localized sources that were found in only a few surfaces A_g^m still appeared in the overall surface S_{total} . Similarly, a spurious source that appeared in several individual surfaces A_g^m was not over-emphasized.

3.2.8.2 Bartlett, PWW, and PWS processors compared

The Bartlett and PWW processors used frequencies up to 200 Hz only. Simulations using higher frequencies gave worse performance, indicating that higher frequencies were too incoherent to add useful information. The PWS processor used frequencies up to 2 kHz.

For the PWS processor, spectrograms were generated using 2 s long Hanning FFT windows with 50% overlap. A 2 s FFT window is significantly longer than the 0.128 s window used in our previous work [Nosal and Frazer 2006*b*]. By running numerous simulations, we found that a longer FFT window makes the PWS processor more robust with respect to mismatch and noise. Also, when a longer FFT window is used, the processor is less sensitive to changes in candidate source position and the resulting ambiguity surface has much broader peaks. Although this does not allow the source to be localized as precisely in space, a longer FFT window permits a much coarser grid (here

we used a grid spacing of 200 m compared to 4 m [Nosal and Frazer 2006b]). Ideally, the PWS could be used in a first sweep of a broad area with coarse grid spacing and a long FFT window, then more locally with finer grid spacing and a shorter FFT window.

As phase is retained in PWS processing of frequencies below the crossover frequency f_c (to make use of coherence), these lower frequencies serve to sharpen the ambiguity surface peaks. In the simulations presented here, trial and error was used to find an optimal crossover frequency f_c of 100 Hz.

3.2.9 Results, discussion, and conclusion

Table 3.2.1 gives the run times on a 2 GHz Pentium 4 PC for the old (Eq. (3.2.1)) and new (Eq. (3.2.4)) forms of the PWW processor. There is a 3 times reduction in run time (approximately), consistent with the observation from Section IV that the new form requires N (number of receivers) times fewer operations than the old form. Also shown are run times for the old (Eq. (3.2.5)) and new (Eq. (3.2.6)) forms of the PWS processor, in which run speeds are again reduced by a factor of 3.

Table 3.2.1 Run times for the original and modified forms of the PWW and PWS processors.

	Original forms Eqs. (3.2.1) & (3.2.5)	Modified forms Eqs. (3.2.4) & (3.2.6)
PWW run time (minutes)	32	10
PWS run time (minutes)	61	22

Figure 3.2.4 shows ambiguity surfaces with each grid point colored according to the value of the ambiguity surface at that point. The ambiguity surfaces have been squared and blurred with a 2-dimensional, 3 by 3 Gaussian lowpass filter with standard deviation 0.5. The images have been scaled to use the full colormap, with red corresponding to the maximum value attained by the surface, and blue corresponding to the minimum. As a result, scales for each image are different. Colorbars are not shown since only relative levels within each surface are of interest, with maxima (red) corresponding to source position estimates. The maximum values for each of the surfaces (before squaring and blurring) are given in Table 3.2.2. These numbers are not particularly significant since they are not probabilities and they should not be used to compare processor performance. They are provided for reference purposes only. For all processors, localized source positions are slightly off, which is due to the introduced environmental mismatch. As might be expected, the first source lost to all processors as noise increases is the source outside of the receiver array.

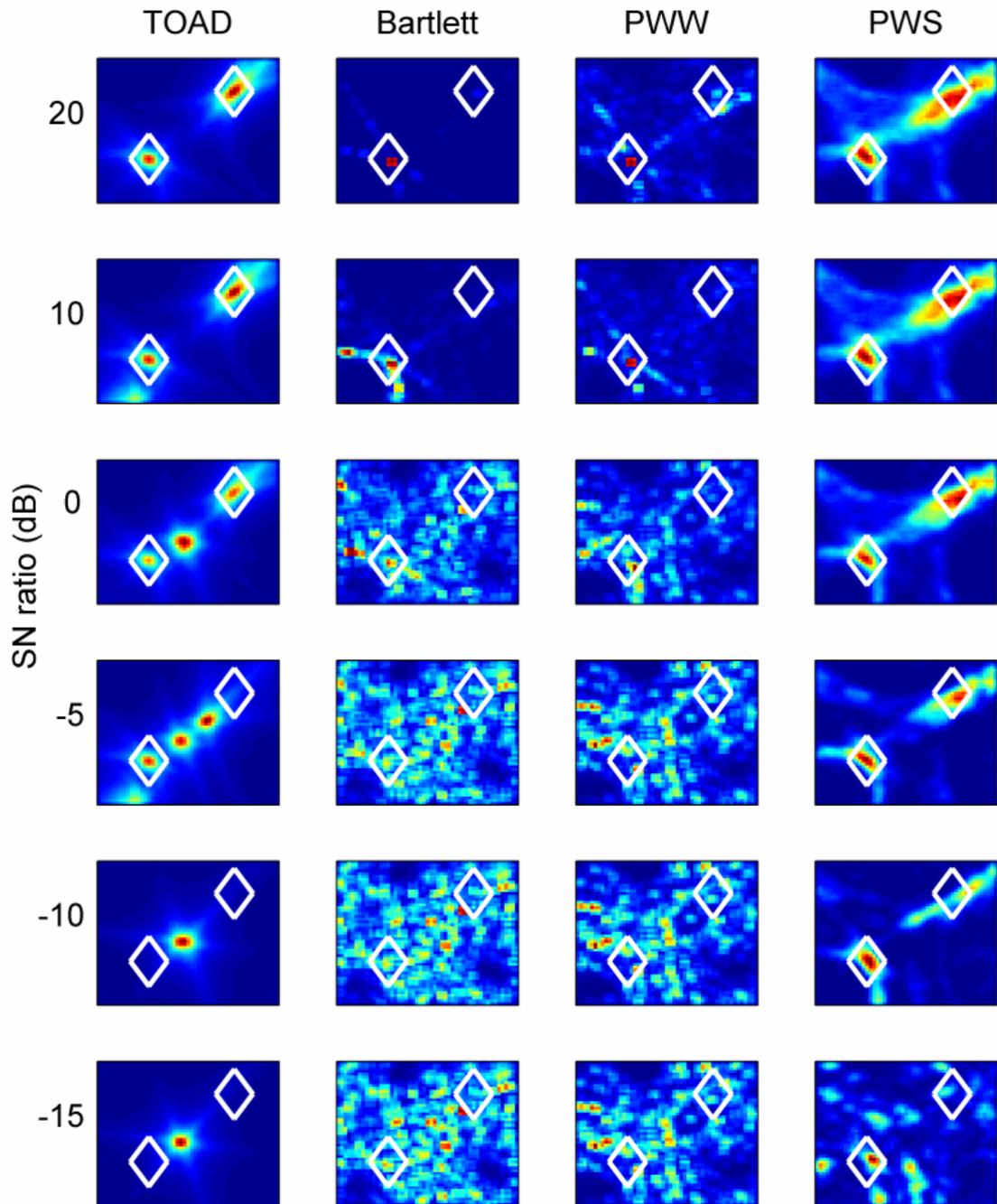


Fig. 3.2.4 TOAD method, Bartlett, PWW, and PWS ambiguity surfaces for various SNRs. Correct source positions are centered in the white diamond markers. Colorbars are individually scaled, and maximum values for each surface are given in Table 3.2.2.

Table 3.2.2 Maximum values of surfaces in Figure 3.2.4. These are provided for reference purposes only; they are not probability values and cannot be used to compare processor performance.

SN ratio (dB)	TOAD	Bartlett	PWW	PWS
20	0.854	0.233	0.204	0.356
10	0.787	0.069	0.175	0.333
0	0.715	0.032	0.082	0.233
-5	0.702	0.031	0.081	0.099
-10	0.683	0.030	0.080	0.016
-15	0.652	0.029	0.075	0.006

The TOAD method (optimized for two sources) is more successful than either Bartlett or the PWW processor at SNRs of 20 and 10 dB. At 0 dB both sources are still localized, but a spurious source appears that is stronger than either of the real sources. In contrast, neither the Bartlett nor the PWW processors localize both sources. At SNR -5 dB, one source is still found but two spurious sources appear (in Bartlett and PWW, both sources have been lost completely), and at -15 dB, only one spurious source remains.

In the simulation for SNR 20 dB, the PWW processor exhibits minor improvement over the Bartlett processor in that both sources are localized (although the source outside of the receiver array is weak), while the Bartlett processor finds only one source. Both processors localize one source at 10 dB SNR, but the Bartlett processor gives spurious sources. Improvement of the PWW over the Bartlett processor is also seen at 0 dB SNR, where the PWW processor localizes one source (albeit quite weakly), while the Bartlett processor does not localize either source. In the simulations for this paper, the noise was always in exactly the same band as the signal, so the advantage of PWW over Bartlett was reduced. As discussed in Section 3.2.3, we did not expect an

obvious advantage of either the PWW or the Bartlett processor over the other, and in other cases with noise and mismatch, the Bartlett processor may outperform the PWW processor.

At high levels of noise it is clear that the PWS processor is the best of the four; it localizes both sources correctly for SNR as low as -10 dB. Even at -15 dB SNR, one source is still localized, although spurious sources begin to appear. It is apparent from the simulations that the PWS processor has lower resolution and lower error than the other processors. As discussed in Section 3.2.8, this is due to the long (2 s) FFT window. If higher resolution is desired, a shorter FFT window should be used, for which a finer grid may be required. The performance of the PWS processor shows improvement from previous simulations [Nosal and Frazer 2006*b*] in which 4 receivers and 0 dB SNR were required to localize two sources. This is made possible by the use of longer FFT windows (2 s compared to 0.128 s).

Several points regarding the PWS processor should be mentioned. First, when working with real data, it will be necessary to process for an array of candidate source depths to create ambiguity volumes rather than surfaces. Second, processing longer lengths of signal will improve results by reducing the effects of noise, provided that the sources remain stationary (as in the case of humpback whale singers). Also, more receivers will improve predictions; in our experience the PWS processor can only localize as many sources as there are receivers. As with other processors, different source characteristics and power levels will give varying degrees of success. The PWS processor seems to perform remarkably better when the sources have different frequency bands (simulations not included). Since the PWS processor does not discriminate

between source characteristics, directional noise sources (such as ships and other whales) may be localized as actual sources. Indeed, when localizing one humpback, other humpbacks in the area are effectively noise sources themselves. Problems arising from constant tone sources and tonals from recording equipment are eliminated through the mean removal process in PWS processing. This step may be replaced with a high-pass filtering of envelopes of spectrogram frequency channels.

In summary, a new version of the PWS processor [Nosal and Frazer 2006*b*] reduces computational requirements of the processors by a factor of N , where N is the number of receivers. In simulations with environmental mismatch, three receivers were used to localize two sources in SNRs down to -10 dB. For -15 dB SNR, one of two sources was found. In these simulations, the PWS processor outperformed a TOAD method, the Bartlett processor, and the PWW processor. Compared with the other methods, the PWS processor sacrifices spatial resolution in order to localize higher frequency signals at greater ranges on a coarser computational grid. This tradeoff may be adjusted by changing the length of FFT windows used to create the spectrograms.

3.3 PWS applied to AUTEK data

PWS was applied to the same single sperm whale dataset that was used in Chapter 2 with similar results. Figure 3.3.1 shows the resulting track compared to the track obtained using the combined DRTD/TOAD method in Section 2.2. These results used a grid spacing of 10 m and 0.05 s long FFT window to create spectrograms. The predicted positions agree quite well and lend confidence to the PWS method and implementation. It is not obvious why the depths disagree. Possible causes include grid spacing, the high directionality of sperm whale clicks, or uncertainties in receiver positions and sound speed profiles.

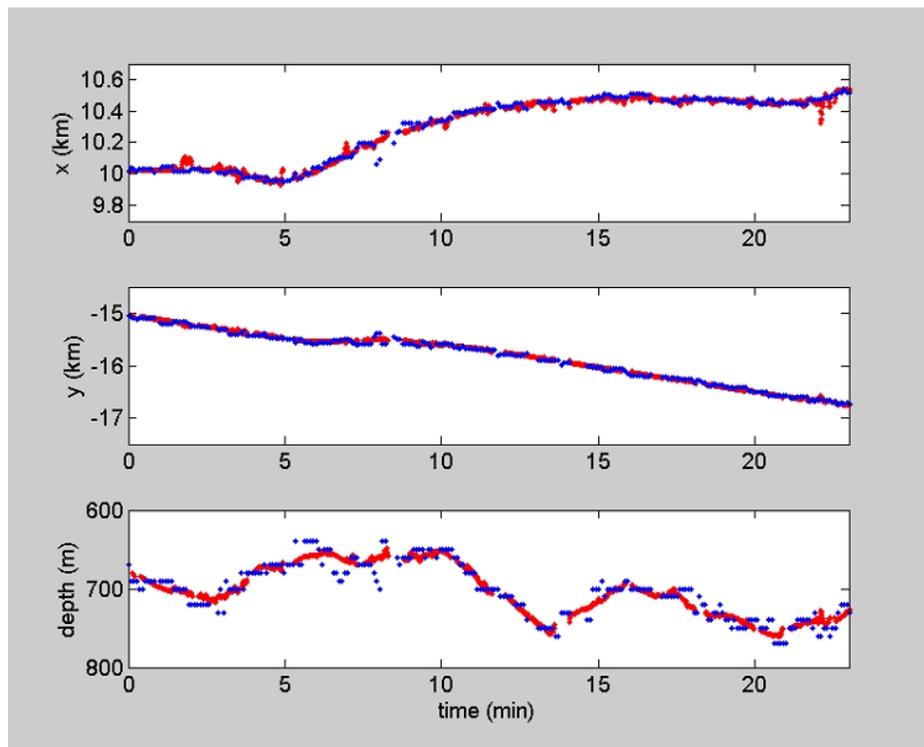


Fig. 3.3.1 Comparison of sperm whale tracks obtained using the DRTD/TOA method (red) and PWS processing (blue).

Chapter 4

CONCLUDING REMARKS AND FUTURE DIRECTIONS

All of the methods described and developed in this dissertation are really a variation on the same theme of localization by matching predicted fields with measured fields. Indeed, even the simplest TOAD method matches (a version of) the predicted field with (a version of) the measured field. In this case, only the time-of-arrival of the direct signal is kept for the measured and predicted fields, and the field prediction is simplified by straight-line assumptions. The TOAD/DRTD does the same thing but increases the accuracy of estimated positions by including surface reflections in the measured and predicted fields and by using a propagation model. PWW/PWS processing goes one step further by including phase/frequency and sound pressure level information.

When deciding between methods it is important to consider the tradeoff between the accuracy and power of the processor on one hand, and the computational demands and modeling complexity on the other hand.

The next step for the DRTD/TOAD method is to generalize it to multiple animal tracking. One approach for this is to develop a method to separate click trains from individual animals, which should be possible given the relatively stable inter-click intervals of many odontocetes (e.g. sperm whales [Baggenstoss 2007]). It may be possible to separate individuals based on spectral components for certain species, or to separate calls using the feature-based extraction methods that are being developed for

passive acoustic detection problems [e.g. Kim *et al.* 2006; Oswald *et al.* 2007]. Future developments and improvements to the PWS processor will be prompted by application to real datasets. Of immediate interest is to use PWS to track multiple singing humpback whales simultaneously to study interactions between singers.

The coming years will undoubtedly see significant advances in the techniques available to study marine mammals. For example, recent advances in piezoelectric materials have led to the development of vector sensors. In addition to pressure (measured by hydrophones), vector sensors can measure particle velocity, which makes them useful in compact array configurations [Nehorai and Paldi 1994; Hawkes and Nehorai 2003]. Beyond acoustics, marine mammal monitoring capabilities will likely benefit from advances in autonomous and remotely controlled vehicles, including surface vessels, aircraft, and underwater vehicles. Various other technologies hold promise for marine mammal studies, including radar, infrared and hyper-spectral imagery, satellite imagery, and LIDAR (Light and Detection Ranging) [Barlow and Gisiner 2006]. Integration of tagging, visual, and acoustic data with these new approaches is one of the most exciting and potentially fruitful challenges.

APPENDIX

This appendix demonstrates that Westwood's processor [Westwood 1992] is dependent on receiver order. For receiver i , denote the received signal by $R_i(\omega)$ and the modeled Green's function by $G_i(\bar{x}, \omega)$ as in Section 3.2. Westwood [1992] defines the "data" cross spectrum as:

$$D_{ij}(\omega) = R_i(\omega)R_j^*(\omega) \quad (\text{A1})$$

and the modeled cross spectrum as

$$M_{ij}(\omega) = G_i(\bar{x}, \omega)G_j^*(\bar{x}, \omega) . \quad (\text{A2})$$

His processor is given by [1992]:

$$\varphi_W = \frac{\left| \sum_{i=1}^N \sum_{j=i+1}^N \sum_B D_{ij}(\omega)M_{ij}^*(\omega) \right|}{\left(\sum_{i=1}^N \sum_{j=i+1}^N \sum_B |D_{ij}(\omega)|^2 \right)^{1/2} \left(\sum_{i=1}^N \sum_{j=i+1}^N \sum_B |M_{ij}(\omega)|^2 \right)^{1/2}} \quad (\text{A3})$$

where B denotes a frequency band chosen to maximize SNR. To see the dependence on receiver order, apply this processor with three receivers at a single frequency:

$$\varphi_{W1} = \frac{|D_{12}M_{12}^* + D_{13}M_{13}^* + D_{23}M_{23}^*|}{\left(|D_{12}|^2 + |D_{13}|^2 + |D_{23}|^2 \right)^{1/2} \left(|M_{12}|^2 + |M_{13}|^2 + |M_{23}|^2 \right)^{1/2}} \quad (\text{A4})$$

where the dependence on frequency is dropped in the notation for simplicity. Now, suppose that the first and second receivers are switched in order. Then the processor becomes:

$$\varphi_{W2} = \frac{|D_{21}M_{21}^* + D_{23}M_{23}^* + D_{13}M_{13}^*|}{\left(|D_{21}|^2 + |D_{23}|^2 + |D_{13}|^2\right)^{1/2} \left(|M_{21}|^2 + |M_{23}|^2 + |M_{13}|^2\right)^{1/2}} . \quad (\text{A5})$$

Since $|D_{ij}| = |D_{ji}|$ and $|M_{ij}| = |M_{ji}|$, the two denominators are the same. In the numerators the sum of the second and third terms is the same; however, the first term in the numerator of Eq. (A4) is the conjugate of the first term in the numerator of Eq. (A5), i.e.

$$D_{12}M_{12}^* = \left(D_{21}M_{21}^*\right)^* . \quad (\text{A6})$$

It follows that the numerators are not equal, so that Eqs. (A4) and (A5) give different values; hence the processor is not symmetric with respect to receiver order.

REFERENCES

- Abraham DA (2000). *Passive acoustic detection of marine mammals*. SM-351
SACLANT Undersea Research Centre, La Spezia, Italy.
- Adam O, J-F Motsch, F Desharnais, NA DiMarzio, D Gillespie, RC Gisiner (2006).
Overview of the 205 workshop on detection and localization of marine mammals
using passive acoustics. *App. Acoust.* 67, 1061-1070.
- Altes R (1980). Detection, estimation, and classification with spectrograms. *J. Acoust.
Soc. Am.* 67, 1232-1246.
- Andrew RK, BM Howe, JA Mercer (2002). Ocean ambient sound: comparing the 1960s
with the 1990s for a receiver off the California coast. *Acoust. Res. Lett. Online* 3, 65-
70.
- Anon. (2001). *Joint interim report on the Bahamas marine mammal stranding even of
15-16 March 2000 (December 2001)*. NOAA unpublished report, 59 pp.
- Au WWL (1993). *Sonar of Dolphins*. Springer, New York.
- _____ (2004). Echolocation signals of wild dolphins. *Acoust. Phys.* 50(4), 454-462.
- Au WWL, KJ Benoit-Bird (2003). Automatic gain control in the echolocation system of
dolphins. *Nature* 423, 861-863.
- Au WWL, DL Herzing (2003). Echolocation signals of wild Atlantic spotted dolphin
(*Stenella frontalis*). *J. Acoust. Soc. Am.* 113, 598-604.

- Au WWL, RW Floyd, RH Penner, AE Murchison (1974). Measurement of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncatus*, Montagu in open waters. *J. Acoust. Soc. Am.* 56(4), 1280-1290.
- Au WWL, PWB Moore, D Pawloski (1986). Echolocation transmitting beam of the Atlantic bottlenose dolphin. *J. Acoust. Soc. Am.* 80(2), 688-691.
- Au WWL, RH Penner, CW Turl (1987). Propagation of beluga echolocation signals. *J. Acoust. Soc. Am.* 82(3), 807-813.
- Au WWL, J Mobley, WC Burgess, MO Lammers (2000). Seasonal and diurnal trends of chorusing humpback whales wintering in waters off western Maui. *Mar. Mamm. Sci.* 16(3), 530-544.
- Au WWL, JKB Ford, KA Allman (2002). Echolocation signals of foraging killer whales (*Orcinus orca*). *J. Acoust. Soc. Am.* 111, 2343-2344.
- Au WWL, AA Pack, MO Lammers, LM Herman, MH Deakos, K Andrews (2006). Acoustic properties of humpback whale song. *J. Acoust. Soc. Am.* 120(2), 1103-1110.
- Aubauer R, MO Lammers, WWL Au (2000). One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water. *J. Acoust. Soc. Am.* 107(1), 2744-2749.
- Baggeroer B, WA Kuperman, PN Mikhalevsky (1993). An overview of matched field methods in ocean acoustics. *IEEE J. Ocean. Eng.* 18, 401-424.
- Barlow J (1999). Trackline detection probability for long-diving whales. pp. 209-21 in GW Garner, SC Amstrup, JL Laake, BFJ Manly, LL McDonald, DG Robertson (Eds.) *Marine mammal survey and assessment methods*. Balkema Press, Rotterdam, Netherlands.

- Barlow J, R Gisiner (2006). Mitigation, monitoring and assessing the effects of anthropogenic sound on beaked whales. *J. Cet. Rec. Mgmt.* 7(3), 239-249.
- Barlow J, BL Taylor (2005). Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. *Mar. Mamm. Sci.* 21(3), 429-445.
- Bowles AE, M Smultea, B Wursig, DP McMaster, D Palaka (1994). Relative abundance and behavior of marine mammals exposed to transmission from the Heard Island Feasibility Test. *J. Acoust. Soc. Am.* 96(4), 2469-2484.
- Bracewell RN (2000). The Hilbert Transform, pp. 359-364 in *The Fourier transform and its applications*, 3rd ed. McGraw-Hill Companies, Inc., Singapore.
- Bucker H (1976). Use of calculated sound fields and matched-field detection to locate sound sources in shallow water. *J. Acoust. Soc. Am.* 59, 368-373.
- Bucher R, D Misra (2002). A synthesizable VHDL model of the exact solution for three-dimensional hyperbolic positions system. *VLSI Design* 15(2), 507-520.
- Caldwell J (2002). Does air-gun noise harm marine mammals? *The Leading Edge* 21, 75-78.
- Cato DH (1998). Simple methods of estimating source levels and locations of marine animal sounds. *J. Acoust. Soc. Am.* 104(3), 1667-1678.
- Chapman DMF (2004). You can't get there from here: Shallow water sound propagation and whale localization. *Can. Acoust.* 32(2), 167-171.
- Chesmore ED (2001). Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Appl. Acoust.* 62, 1359-1374.

- Clark CW (1980). A real-time direction finding device for determining the bearing to the underwater sounds of southern right whales, *Eubalaena australis*. *J. Acoust. Soc. Am.* 68, 508-11.
- Clark CW, JM Clark (1980). Sound playback experiments with southern right whales (*Eubalarna australis*). *Science* 207, 663-664.
- Clark CW, WT Ellison (2000). Calibration and comparison of acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pr. Barrow, Alaska, 1984-1993. *J. Acoust. Soc. Am.* 107, 3509-3517.
- _____ (2003). Potential use of low-frequency sound by baleen whales for probing the environment: Evidence from models and empirical measurements. pp. 547-589 in J Thomas, CF Moss, M Valter (Eds.). *Echolocation in bats and dolphins*, University of Chicago Press, Chicago.
- Clark CW, KM Firstrup (1997). Whales 95: A combined visual and acoustic survey of blue and fin whales off Southern California. pp. 583-600 in *Rep. Int. Whal. Comm.* vol. 47.
- Clark CW, WT Ellison, K Beeman (1986). Acoustic tracking of migrating bowhead whales. *Proceedings of IEEE Oceans 1986*, 341-346.
- Clarke MR, HR Martins HR, P Pascoe (1993). The diet of sperm whales, *Physeter macrocephalus* Linnaeus 1758, off the Azores. *Philos. Trans. R. Soc. London Ser. B* 339, 67-82.
- Cleator H, L Dueck (1995). Three-dimensional accuracy of a four-hydrophone array. *11th Biennial Conf. Bio. Mar. Mamm.* 14-18 Dec. 1995, Orlando.

- Cummings WC, DV Holliday (1985). Passive acoustic location of bowhead whales in a population census off Point Barrow, Alaska. *J. Acoust. Soc. Am.* 78, 1163-1169.
- Curtis KR, BM Howe, JA Mercer (1999). Low frequency ambient sound in the North Pacific: long time series observations. *J. Acoust. Soc. Am.* 196, 3189-3200.
- Dawbin WH, DH Cato (1992). Sounds of pygmy right whale (*Caperea marginata*). *Mar. Mamm. Sci.* 8, 213-219.
- Delsom JM, M Morf, B Friedlander (1980). Source location from time differences of arrival: Identifiability and estimation. *IEEE Int. Conf. Acoust., Speech, Sig. Process.*, 818-824.
- DeRuiter SL, PL Tyack, Y-T Lin, AE Newhall, JF Lynch, PJO Miller (2006). Modeling acoustic propagation of airgun array pulses recorded on tagged sperm whales (*Physeter macrocephalus*). *J. Acoust. Soc. Am.* 120(6), 4100-4114.
- D'Spain GL, A D'Amico, DM Fromm (2006). Properties of the underwater sound fields during some well documented beaked whale mass stranding events. *J. Cet. Rec. Manage.* 7(3), 223-238.
- Fletcher S, BJ LeBoeuf, DP Costa, PL Tyack, SB Blackwell (1996). Onboard acoustic recordings from diving northern elephant seals. *J. Acoust. Soc. Am.* 100, 2531-39.
- Flewellen CG, RJ Morris (1978). Sound velocity measurements on samples from the spermaceti organ of the sperm whale *Physeter catodon*. *Deep-Sea Res.* 25, 269-277.
- Fox C, E Radford, RP Dziak, T-K Lau, H Matsumoto, AE Schreiner (1995). Acoustic detection of a seafloor spreading episode on the Juan de Fuca Ridge using military hydrophone arrays. *Geophys. Res. Lett.* 22(2), 131-134.

- Frankel AS (1994). *Acoustic and visual tracking reveals distribution, song variability and social roles of humpback whales in Hawaiian waters*. Ph. D. dissertation, University of Hawaii at Manoa, Honolulu, HI.
- Frankel AS, CW Clark, LM Herman, CM Gabriel (1995). Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawaii determined using acoustic and visual techniques. *Can. J. Zool.* 73, 1134-1146.
- Frantzis A (1998). Does acoustic testing strand whales? *Nature* 392(6671), 29.
- Frazer LN, X Sun (1998). New objective functions for waveform inversion. *Geophysics*, 63, 213-222.
- Freitag LE, PL Tyack (1993). Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echo location clicks. *J. Acoust. Soc. Am.* 93, 2197-2205.
- Fu SS, C Tao, M Prasad, RH Wilkens, LN Frazer (2004). Acoustic properties of coral sands, Waikiki, Hawaii. *J. Acoust. Soc. Am.* 115(5), 2013-2020.
- GDEM (Generalized Digital Environment Model) website (last accessed on May 1, 2007): <<https://128.160.23.42/gdemv/gdemv.html>>
- Gillespie D (1997). An acoustic survey for sperm whales in the Southern Ocean Sanctuary conducted from the RSV Aurora Australis. *Rep. Int. Whal. Comm.* 47, 897-907.
- Gillespie D (2004). Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. *Can. Acoust.* 32(2), 39-47.
- Giraudet P, H Glotin H (2006). Real-time 3D tracking of whales by precise and echo-robust TDOAs of clicks extracted from 5 bottom-mounted hydrophones records of the AUTECH. *Appl. Acoust.* 67, 1106-1117.

- Godin OA, IM Fuks (1989). Travel-time statistics for signals scattered at a rough surface. *J. Acoust. Soc. Am.* 85, 83-89.
- Goldstein H (1980). The Euler Angles, Section 4-4, pp. 143-148, and Euler Angles in Alternate Conventions, Appendix B, pp. 606-610 in *Classical Mechanics*, 2nd ed., Addison-Wesley, Reading, MA.
- Goodyear JD (1993). A sonic/radio tag for monitoring dive depths and underwater movements of whales. *J. Wildl. Mgmt.* 57, 503-13.
- Goold JC, SE Jones (1995). Time and frequency domain characteristics of sperm whale clicks. *J. Acoust. Soc. Am.* 98(3), 1279-1291.
- Goold JC, JD Bennell, SE Jones (1996). Sound velocity measurements in spermaceti oil under the combined influences of temperature and pressure. *Deep-Sea Res. I* 43(7), 961-969.
- Gordon JCD (1987). *The behaviour and ecology of sperm whales off Sri Lanka*. Ph.D. dissertation, University of Cambridge, UK.
- Gordon JCD (1991). Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations. *J. Zool. Lond.* 224, 301-314.
- Gordon JCD, L Steiner (1992). Ventilation and dive patterns in sperm whales, *Physeter macrocephalus*, in the Azores. *Rep. Int. Whal. Commn.* 42, 561-565.
- Gordon J, D Gillespie, J Potter, A Frantzis, MP Simmonds, R Swift, D Thompson (2004). A review of the effects of seismic surveys on marine mammals. *Mar. Technol. Soc. J.* 37, 16-34.
- Greene CR Jr., WJ Richardson (1988). Characteristics of marine seismic survey sounds in the Beaufort Sea. *J. Acoust. Soc. Am.* 83(6), 2246-2254.

- Green DM, HA DeFerrari, D McFadden, JS Pearse, AN Popper, WJ Richardson, SH Ridgway, PL Tyack (1994). *Low-frequency sound and marine mammals: Current knowledge and research needs*. National Academy Press, DC, NRC Rep.
- Hammond PS, SA Mizroch, GP Donovan (Eds.) (1990). Individual recognition of cetaceans: Use of photo-identification and other techniques to estimate population parameters. *Rep. Int. Whal. Comm.* 12, Cambridge.
- Hawkes M, A Nehorai (2003). Wideband source localization using a distributed acoustic vector-sensor array. *IEEE Trans. Signal Process.* 51, 1479-1491.
- Hayes SA, DK Mellinger, DA Croll, DP Costa, JF Borsani (2000). An inexpensive passive acoustic system for recording and localizing wild animal sounds. *J. Acoust. Soc. Am.* 107, 3552-3555.
- Helweg DA, DS Houser, PWB Moore (2000). An integrated approach to the creation of a humpback whale hearing model. San Diego, Technical 1835.
- Hodgkiss WS, WA Kuperman, JJ Murray, GL D'Spain, and LP Berger (1997). High frequency matched field processing. pp. 229-234 in AP Lyons (Ed.). *High frequency acoustics in shallow water*, NATO SACLANT Undersea Research Center, La Spezia.
- Hursky P, MB Porter, M Siderius (2004). High-frequency (8-16 kHz) model-based source localization. *J. Acoust. Soc. Am.* 115, 3021-3032.
- Jaquet N, S Dawson, L Douglas (2001). Vocal behavior of male sperm whales: Why do they click? *J. Acoust. Soc. Am.* 109, 2254-2259.
- Jaquet N, S Dawson, E Slooten (2000). Seasonal distribution and diving behavior of male sperm whales off Kaikoura: Foraging implications. *Can. J. Zool.* 78, 407-419.

- Janik VM, SM VanParijs, PM Thompson (2000). A two-dimensional acoustic localization system for marine mammals. *Mar. Mamm. Sci.* 16, 437-447.
- Jepson, PD, M Arbelo, R Deavilee, IAP Patterson, P Castro, JR Baker, E Degollada, HM Ross, P Herraез, AM Pocknell, F Rodriguez, FE Howiell, A Espinosa, RJ Reid, JR Jaber, V Martin, AA Cunningham, A Fernandez (2003). Gas-bubble lesions in stranded animals: Was sonar responsible for a spate of whale deaths after an Atlantic military exercise? *Nature* 425(6958), 575-76.
- Johnson HJ, SR DeGraaf (1982). Improving the resolution of bearing in passive sonar arrays by eigenvector analysis. *IEEE Trans. Acoust. Speech Sig. Process.* 30, 638-647.
- Johnson MP, PL Tyack (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Ocean Eng.* 28, 3-12.
- Johnston DW, ME Chapla, LE Williams, CK Mattila DK (2007). Identification of humpback whale *Megaptera novaeangliae* wintering habitat in the Northwestern Hawaiian Islands using spatial habitat modeling. *Endang. Spec. Res.* 3, 249-257.
- Kim KH, P Hursky, MB Porter, JA Hildebrand, EE Henderson, SM Wiggins(2006). Automated passive acoustic tracking of dolphins in free-ranging pods. In SM Jesus, OC Rodriguez (Eds.), *Proc. 8th Eurp. Conf. Underwater Acoust* 12-15 June, Carvoeiro, Portugal.
- Knapp CH, GC Carter (1976). The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech, Sig. Process.* 24, 320-327.
- Knapp Jr. P (2001). *Rapture of the Deep – Humpback whale singing*. Compass Recordings.

- Ko D, JE Zeh, CW Clark, WT Ellison, BD Krogman, RM Sonntag (1986). Utilisation of acoustic location data in determining a minimum number of spring migrating bowhead whales, unaccounted for by the icebased visual census. *Rep. Int. Whal. Commn.* 36, 325-38.
- Kuperman WA, GL D'Spain, KD Heaney (2001). Long range localization from single hydrophone spectrograms. *J. Acoust. Soc. Am.* 106(5), 1935-1943.
- Lahr JC, FG Fischer (1993). Location of acoustic sources using seismological techniques and software. *USGS Open-file Report 93-221*, Menlo Park.
- Laplace C (2007). A Bayesian method to estimate the depth and the range of phonating sperm whales using a single hydrophone. *J. Acoust. Soc. Am.* 121(3), 1519-1528.
- Laplanche C, O Adam, M Lopatka, J-F Motsch (2006). Measuring the off-axis angle and the rotational movements of phonating sperm whales using a single hydrophone. *J. Acoust. Soc. Am.* 119(6), 4074-4082.
- Leeper R, O Chappell, J Gordon (1992). The Development of practical techniques for surveying sperm whale populations acoustically. *Rep. Int. Whal. Commn.* 42, 549-60.
- Leatherwood S, WE Evans (1979). Some recent uses and potentials of radiotelemetry in field studies of cetaceans. In HE Winn, BL Olla, *Behavior of marine mammals: Current perspectives in research* vol. 3, *Cetaceans*, Plenum Press, New York.
- Lockyer C (1977). Observations on diving behaviour of the sperm whale, *Physeter catodon*. pp. 591-609 in M Angel (Ed.) *A Voyage of Discovery*, Pergamon, Oxford.
- Madsen PT (2002). Morphology of the sperm whale head: A review and some new findings. In *Sperm whale sound production*, Ph.D. dissertation, University of Aarhus, Denmark.

- Madsen PT, M Wahlberg (2007). Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales. *Deep-Sea Res. I* 54, 1421-1444.
- Madsen PT, R Payne, NU Kristiansen, M Wahlberg, I Kerr, B Møhl (2002b). Sperm whale sound production studied with ultrasound time/depth-recording tags. *J. Exp. Biol.* 205, 1899-1906.
- Madsen PT, M Wahlberg, B Møhl (2002a). Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: implications for echolocation and communication. *Behav. Ecol. Sociobiol.* 53, 31-41.
- Madsen PT, DA Carder, WWL Au, PE Nachtigall, B Møhl, SH Ridgway (2003). Sound production in neonate sperm whale. *J. Acoust. Soc. Am.* 113, 2988-2991.
- Madsen PT, I Kerr, R Payne R (2004a). Echolocation clicks of two free-ranging delphinids with different food preferences: false killer whales (*Pseudorca crassidens*) and Risso's dolphin (*Grampus griseus*). *J. Exp. Bio.* 207, 1811-1823.
- _____ (2004b). Source parameter estimates of echolocation clicks from wild pygmy killer whales (*Feresa attenuata*). *J. Acoust. Soc. Am.* 116, 1909-1912.
- Madsen PT, M Johnson, N Aguilar deSoto, WMX Zimmer, P Tyack (2005). Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*). *J. Exp. Bio.* 208, 181-194.
- Madsen PT, M Johnson, PJO Miller, N Aguilar Soto, J Lynch, P Tyack (2006). Quantitative measures of air-gun pulses recorded on sperm whales (*Physeter macrocephalus*) using acoustic tags during controlled exposure experiments. *J. Acoust. Soc. Am.* 120(4), 2366-2379.
- Malakoff, DA (2001). Roaring debate over ocean noise. *Science* 291, 576-8.

- Mate B (1989). Satellite-monitored radio tracking as a method for studying cetacean movements and behaviour. *Rep. Int. Whal. Commn.* 39, 389-91.
- MATLAB, The MathWorks, Inc.
- McDonald MA, JA Hildebrand, SC Webb (1995). Blue and fin whales observed on a seafloor array in the Northeast Pacific. *J. Acoust. Soc. Am.* 98, 712-721.
- McDonald MA, JA Hildebrand, SM Wiggins (2006). Increases in deep ocean ambient noise in the northeast pacific west of San Nicolas Island, California. *J. Acoust. Soc. Am.* 120(2), 711-718.
- Mellinger D (2002). *ISHMAEL 1.0 User's Guide*. NOAA/PMEL.
- _____ (2004). A comparison of methods for detecting right whale calls. *Can. Acoust.* 35(2), 55-65.
- Mellinger D, CW Clark (1993). Bioacoustic transient detection by image convolution. *J. Acoust. Soc. Am.* 93, 2358.
- _____ (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.* 107(6), 3518-3529.
- Mellinger D, CD Carson, CW Clark (2000). Characteristics of minke whales (*Balaenoptera acutorostrata*) pulse trains recorded near Puerto Rico. *Mar. Mamm. Sci.* 16, 739-756.
- Mennill DJ, JM Burt, KM Fristrup, SL Vehrencamp (2006). Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest. *J. Acoust. Soc. Am.* 119(5), 2832-2839.

- Michalopoulou Z-H, X Ma X (2005). Source localization in the Haro Strait primer experiment using arrival time estimation and linearization. *J. Acoust. Soc. Am.* 118(5), 2924-2933.
- Miller PJO, N Biassoni, A Samuels, PL Tyack (2000). Whale songs lengthen in response to sonar. *Nature* 405, 903.
- Miller PJO, MP Johnson, PL Tyack (2004). Sperm whale behaviour indicates the use of scholocation click buzzes 'creaks' in prey capture. *Proc. R. Soc. Lond. B* 271, 2239-2247.
- Mitchell S, J Bower (1995). Localization of animal calls via hyperbolic methods. *J. Acoust. Soc. Am.* 97, 3352-3353.
- Mobley Jr. JR (2005). Assessing responses of humpback whales to North Pacific Acoustic Laboratory (NPAL) transmissions: Results of 2001-2003 aerial surveys north of Kauai. *J. Acoust. Soc. Am.* 117, 1666-1673.
- Mobley Jr. JR (2006). *Results of 2006 aerial surveys of humpback whales north of Kauai*. Annual report submitted to the North Pacific Acoustic Laboratory (NPAL), Scripps Ocean. Inst., 19 pp.
- Mobley Jr. JR, GB Bauer, LM Herman (1999). Changes over a ten-year interval in the distribution and relative abundance of humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters. *Aq. Mamm.* 25(2), 63-72.
- Moore BCJ (2003). *An introduction to the psychology of hearing*, 5th ed. Academic Press, San Diego.
- Møhl B (2001). Sound transmission in the nose of the sperm whale *Physeter catodon*. A postmortem study. *J. Comp. Physiol. A* 187, 335-340.

- Møhl B, A Surlykke, LA Miller (1990). High intensity narwhal clicks. pp. 295-303 in J Thomas, R Kastelein (Eds.). *Sensory ability of cetaceans*, Plenum Press, New York.
- Møhl B, M Wahlberg, PT Madsen, A Heerfordt, A Lund (2000). Sperm whale clicks: Directionality and source level revisited. *J. Acoust. Soc. Am.* 114, 1143-1154.
- Møhl B, M Wahlberg, PT Madsen, A Heerfordt, A Lund (2003). The monopulsed nature of sperm whale clicks. *J. Acoust. Soc. Am.* 114, 1143-1154.
- Morrissey RP, J Ward, N DiMarzio, S Jarvis, DJ Moretti (2006). Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean. *Appl. Acoust.* 67, 1091-1105.
- Mullins J, H Whitehead, L Weilgart (1988). Behaviour and vocalizations of two sperm whales, *Physeter macrocephalus*, off Nova Scotia. *Can. J. Fish. Aquat. Sci.* 45, 1736-1743.
- Nehorai A, E Paldi (1994). Acoustic vector-sensor array processing. *IEEE Trans. Signal Process.* 42, 2481-2491.
- Nieukirk SL, KM Stafford, DK Mellinger, RP Dziak, CG Fox, Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean, *J. Acoust. Soc. Am.* 115, 1832-1843.
- Noad MJ, DH Cato (2001). A combined acoustic and visual survey of humpback whales off southeast Queensland. *Memoirs of the Queensland Museum* 47, 507-523.
- _____ (2007). Swimming speeds of singing and non-singing humpback whales during migration. *Mar. Mamm. Sci.* 23(3), 418-495.
- Nosal E-M, LN Frazer (2006a). Delays between direct and reflected arrivals used to track a single sperm whale. *Applied Acoustics*, 87 (11-12), 1187-1201.

- _____ (2006b). Pair-wise processing of spectrograms for localization of multiple broadband CW sources. *Newsletter of the IEEE Ocean Engineering Society*, Winter.
- Norris KS, GW Harvey (1972). A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon L.*). pp. 397-417 in SR Galler, K Schmidt-Koenig, GJ Jacobs, RE Belleville (Eds.) *Animal orientation and navigation*, NASA Special Publication 262.
- Norris TF, M McDonald, J Barlow (1999). Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration. *J. Acoust. Soc. Am.* 106, 506-514.
- NRC, National Research Council (2003). *Ocean noise and marine mammals*. National Academies, Washington, D.C.
- _____ (2005). *Marine mammal populations and ocean noise*. National Academies, Washington, D.C.
- Oswald JN, J Barlow, TF Norris (2003). Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean. *Mar. Mamm. Sci.* 19, 20-37.
- Oswald JN, S Ranking, J Barlow, MO Lammers (2007). A tool for real-time acoustic species identification of delphinid whistles. *J. Acoust. Soc. Am.* 122(1), 587-595.
- Page SE (1954). Continuous inspection schemes. *Biometrika* 41, 100-115.
- Papastavrou V, SC Smith, H Whitehead (1989). Diving behavior of the sperm whale, *Physeter macrocephalus*, off the Galapagos Islands. *Can. J. Zool.* 67, 839-846.
- Payne RS, S McVay (1971). Songs of humpback whales. *Science* 173, 585-597.
- Payne RS, D Webb (1971). Orientation by means of long range acoustic signaling in baleen whales. *Ann. NY Acad. Sci.* 188, 110-41.

- Porter MB (2005). *BELLHOP* (a gaussian beam/finite element beam code). Available in the *Acoustics Toolbox* at <<http://www.hlsresearch.com/oalib/Modes/AcousticsToolbox/>>
- Porter MB, HP Bucker (1987). Gaussian beam tracing for computing ocean acoustic fields. *J. Acoust. Soc. Am.* 82(4), 1349-1359.
- Porter MB, YC Liu (1994). Finite-element ray tracing. pp. 947-956 in D Lee, MH Schulz (Eds.). *Proc. Int. Conf. Theor. Comp. Acoust.* World Scientific, Singapore.
- Potter JR, DK Mellinger, CW Clark (1994). Marine mammal call discrimination using artificial neural networks. *J. Acoust. Soc. Am.* 96, 1255-62.
- Rasmussen MH, LA Miller, WWL Au (2002). Source levels of clicks from free-ranging white beaked dolphins (*Lagenorhynchus albirostris* Gray 1846) recorded in Icelandic waters. *J. Acoust. Soc. Am.* 111, 1122-1125.
- Rasmussen MH, M Wahlberg, LA Miller (2004). Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*). *J. Acoust. Soc. Am.* 116(3), 1826-1831.
- Rhinelander MQ, SM Dawson (2004). Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length. *J. Acoust. Soc. Am.* 115, 1826-1831.
- Richardson WJ, CR Greene Jr., CI Malme, DH Thomson (1995). *Marine mammals and noise*. San Diego: Academic Press.
- Roch MA, MS Soldevilla, JC Burtenshaw, EE Henderson, JA Hildebrand (2007). Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. *J. Acoust. Soc. Am.* 121(3), 1737-1748.

- Schau HC and Robinson AZ (1987). Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Trans. Acoust., Speech, Sig. Process.* 35, 1223-1225.
- Schevill WE, WA Watkins (1966). *Radio-tagging of whales*. Woods Hole Technical Report, MA.
- Smith GW, GG Urquhart, DN MacLennan, B Sarno (1998). A comparison of theoretical estimates of the errors associated with ultrasonic tracking using a fixed hydrophone array and field measurements. *Hydrobiologia* 371/372, 9-17.
- Schmidt RO (1972). A new approach to geometry of range difference location. *IEEE Trans. Aerosp. Electron. Syst.* 8, 821-835.
- Scott MD, WL Perryman (1991). Using aerial photogrammetry to study dolphin school structure, In K Pryor, KS Norris (Eds.) *Dolphin societies: Discoveries and puzzles*, University of California Press, Berkeley.
- Skarsoulis EK, MA Kalogerakis (2005). Ray-theoretic localization of an impulsive source in a stratified ocean using two hydrophones. *J. Acoust. Soc. Am.* 118(5), 2934-2943.
- Spiesberger JL (2001). Hyperbolic location errors due to insufficient numbers of receivers. *J. Acoust. Soc. Am.* 109(6), 3076-3079.
- Spiesberger JL, KM Fristrup (1990). Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. *Am. Nat.* 135, 107-153.
- Spiesberger JL, M Wahlberg (2002). Probability density functions for hyperbolic and isodiachronic locations. *J. Acoust. Soc. Am.* 112(6), 3046-3052.

- Stafford KM, CG Fox, DS Clark (1998). Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean. *J. Acoust. Soc. Am.* 104, 3616-3625.
- Stimpert A, DN Wiley, WWL Au, MP Johnson, R Arsenault (2007). ‘Megapclicks’: acoustic click trains and buzzes produced during night-time foraging of humpback whales (*Megaptera novaeangliae*). *Bio. Lett.* Online Aug 8, 2007, 1-4.
- Taylor B, J Barlow, R Pitman, L Balance, T Klinger, D DeMaster, J Mildebrand, J Urban, D Palacios, J Mead (2004). *A call for research to assess risk of acoustic impact on beaked whale populations*, Paper SC/56/E36 presented to the IWC Scientific Committee, July 2004, Sorrento Italy, 4pp.
- Taylor JR (1997). *An introduction to Error Analysis*, 2nd ed. Chap. 3. University Science, Sausalito, CA.
- Thode A (2005). Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments. *J. Acoust. Soc. Am.* 18(6), 3575-3584.
- Thode A, GL D.Spain, WA Kuperman (2000). Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations. *J. Acoust. Soc. Am.* 107, 1286-1300.
- Thode A, DK Mellinger, S Stienessen, A Martinez, K Mullin (2002). Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico. *J. Acoust. Soc. Am.* 112(1), 308-321.
- Thomas J, CF Moss, M Valter (Eds.) (2004). *Echolocation in bats and dolphins*, University of Chicago Press, Chicago.

- Tiemann CO, MB Porter, LN Frazer (2004). Localization of marine mammals near Hawaii using an acoustic propagation model. *J. Acoust. Soc. Am.* 115(6), 2834-2843.
- Tiemann CO, SW Martin, JR Mobley (2006). Aerial and acoustic marine mammal detection and localization on Navy ranges. *IEEE J. Ocean. Eng.* 31(1), 107-119.
- Tiemann CO, AM Thode, J Straley, V O'Connell, K Folkert (2007). Three-dimensional localization of sperm whales using a single hydrophone. *J. Acoust. Soc. Am.* 120(4), 2355-2365.
- Tolstoy A (1993). *Matched field processing*. World Scientific, Singapore.
- Tyack PL (1981). Interactions between singing Hawaiian humpback whales and conspecifics nearby. *Behav. Ecol. Sociobiol.* 8, 105-116.
- Tyack PL, CW Clark (2000). Communication and acoustic behavior of dolphins and whales. pp.156-244 in WWL Au, AN Popper, RR Fay (eds.) *Hearing by whales and dolphins*. Springer, New York.
- Tyack PL, M Johnson, NA Soto, A Sturlese, PT Madsen (2006). Extreme diving of beaked whales. *J. Exp. Bio.* 209, 4238-4253.
- Tyrrell WA, (1964). Design of acoustic systems. pp. 65-86 in WN Tavolga (Ed.), *Marine Bioacoustics*. Pergamon, Oxford, UK.
- Urick RJ (1983). *Principles of underwater sound*, 3rd ed., Peninsula, Los Altos, CA.
- VanParijs SM, J Smith, PJ Corkeron (2002). Using calls to estimate the abundance of inshore dolphins: A case study with Pacific humpback dolphins *Sousa chinensis*. *J. Appl. Scol.* 39, 853-864.

- Vermaak J, A Blake (2001). Nonlinear filtering for speaker tracking in noisy and reverberant environments. *Proc. IEEE Conf. on Acoust., Speech, Signal Proc.* 5., 3021-3024.
- Wahlberg M (2002). The acoustic behaviour of diving sperm whales observed with a hydrophone array. *J. Exp. Mar. Bio. and Ecol.* 281, 53-62.
- ____ (2004). Comparing a linear with a non-linear method for acoustic localization. *Can. Acoust.* 32(2):125-131.
- Wahlberg M, B Møhl, PT Madsen (2001). Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics. *J. Acoust. Soc. Am.* 109(1), 397-406.
- Wald A (1947). *Sequential analysis*. Wiley, New York.
- Walker RA (1963). Some intense, low-frequency, underwater sounds of wide geographic distribution, apparently of biological origin. *J. Acoust. Soc. Am.* 35, 1816-1824.
- Wartzok D, S Sayegh, H Stone, J Barchak, W Barnes (1992). Acoustic tracking system for monitoring under-ice movements of polar seals. *J. Acoust. Soc. Am.* 92(2), 682-687.
- Watkins WA, WE Schevill (1971). Four-hydrophone array for acoustic three-dimensional localization. *Woods Hole Oceanographic Institution, Technical Report*, Ref. No. 71-60.
- ____ (1972). Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. *Deep-Sea Res.* 19, 691-706.
- ____ (1975). Sperm whales (*Physeter catodon*) react to pingers. *Deep-Sea Res.* 22, 123-129.

- _____ (1977). Spatial distribution of *Physeter catodon* (sperm whales) underwater. *Deep-sea Res.* 24, 693-699.
- _____ (1979). Aerial observation of feeding behavior in the four baleen whales: *Eubalaena glacialis*, *Balaenoptera borealis*, *Megaptera novaeangliae*, and *Balaenoptera physalus*. *J. Mammal.* 60, 155-63.
- Watkins WA, MA Daher, KM Fristrup, TJ Howald, GN DiSciara (1993). Sperm whales tagged with transponders and tracked underwater by sonar. *Mar. Mamm. Sci.* 9, 55-67.
- Watkins WA, MA Daher, NA DiMarzio, A Samuels, D Wartzok, KM Fristrup, PW Howey, RR Maiefski RR (2002). Sperm whale dives tracked by radio tag telemetry. *Mar. Mamm. Sci.* 18(1), 55-68.
- Watwood SL, PJO Miller, M Johnson, PT Madsen, PL Tyack (2006). Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). *J. Animal Ecol.* 75, 814-825.
- Weir CR, SJ Dolman (2007). Comparative review of the regional marine mammal guidelines implemented during industrial seismic surveys, and guidance towards a worldwide standard. *J. Int. Wildl. Law Policy* 10, 1-27.
- Westwood EK (1992). Broadband matched-field source localization. *J. Acoust. Soc. Am.* 91, 2777-2789.
- White PR, TG Leighton, DC Finfer, C Prowles, O Baumann (2006). Localisation of sperm whales using bottom-mounted sensors. *Appl. Acoust.* 67, 1074-1090.
- Whitehead H (2003). *Sperm whales: social evolution in the ocean*. Chicago University Press.

- Whitehead H, L Weilgart (1990). Click rates from sperm whales. *J. Acoust. Soc. Am.* 87, 1798-1906.
- _____ (1991). Patterns of visually observable behavior and vocalizations in groups of female sperm whales. *Behavior* 118, 275-296.
- Whitehead H, J Christal, PL Tyack (2000). Studying cetacean social structure in space and time: Innovative techniques. pp. 65-87 in J Mann, RC Connor, PL Tyack, H Whitehead (Eds.) *Cetacean Societies*, University of Chicago Press, Chicago.
- Williams R, L Thomas (2007). Distribution and abundance of marine mammals in coastal waters of British Columbia, Canada. *J. Cet. Res. Mangmt.* (in press)
- Winn HE, LK Winn (1978). The song of the humpback whale *Megaptera novaeangliae* in the West Indies. *Mar. Biol.* 47, 97-114.
- Würsig B, M Würsig (1977). The photographic determination of group size, composition and stability of coastal porpoises (*Tursiops truncatus*). *Science* 198, 755-756.
- _____ (1979). Behavior and ecology of the bottlenose dolphin, *Tursiops truncatus*, in the South Atlantic. *Fish. Bull.* 77, 399-412.
- _____ (1980). Behavior and ecology of the dusky dolphin, *Lagenorhynchus obscurus*, in the South Atlantic. *Fish. Bull.* 77, 871-890.
- Würsig B, EM Dorsy, MA Fraker, RS Payne, WJ Wilson, RS Wells (1984). Behavior of bowhead whales, *Balaena mysticetus*, summering in the Beaufort Sea: Surfaceing, respiration, and dive characteristics. *Can. J. Zool.* 62, 1910-1921.
- Würsig B, J Guerrero, GK Silber (1993). Social and sexual behavior of bowhead whales in fall in the Western Arctic: A re-examination of seasonal trends. *Mar. Mamm. Sci.* 9, 103-110.

Zimmer WMX, MP Johnson, A D'Amico, P Tyack (2003). Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*). *IEEE Jour. Oceanic Eng.* 28(1), 13-28.

Zimmer WMX, PL Tyack, MP Johnson, PT Madsen (2005a). Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. *J. Acoust. Soc. Am.* 117(3), 1473-1485.

Zimmer WMX, PT Madsen PT, V Teloni, MP Johnson, PL and Tyack (2005b). Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production. *J. Acoust. Soc. Am.* 118(5), 3337-3345.