PAIR-WISE SPECTROGRAM PROCESSING USED TO TRACK A SPERM WHALE

Eva-Marie Nosal and L. Neil Frazer
School of Ocean and Earth Science and Technology, 1680 East-West Road, Honolulu HI 96822

ABSTRACT

We developed a model-based localization method called pair-wise spectrogram (PWS) processing to track marine mammals using widely spaced hydrophone arrays [Nosal & Frazer, IEEE J. Ocean. Eng., in press]. Here we use PWS to track the sperm whale from the dataset provided for the *2005 Workshop on detection and localization of marine mammals using passive acoustics* (*Monaco*). This dataset provides a good opportunity to validate and explore the properties of the PWS processor. We demonstrate the relationship between pair-wise processing and a time-of-arrival method for a simple case. We also show how varying the size of windows used to create spectrograms optimizes the tradeoff between processor resolution and robustness, and how these parameters can be adjusted according to grid spacing. PWS position estimates are within tens of meters of those obtained using a careful time-of-arrival method applied to individual clicks.

RÉSUMÉ

Nous avons développé une méthode de localisation basée sur un modèle de propagation du son, appelée traitement de spectrogramme par paires (PWS : pair-wise spectrogram), pour déterminer les trajectoires de mammifères marins à l'aide de réseaux d'hydrophones très espacés [Nosal & Frazer, IEEE J. Ocean. Eng., in press]. Nous présentons une application de cette méthode visant à déterminer la trajectoire d'un cachalot à partir du jeu de données fourni par *l'Atelier 2005 sur la détection et la localisation des mammifères marines à l'aide du repérage acoustique passif* (*Monaco*). Ce jeu constitue une bonne occasion de valider et d'explorer les propriétés du PWS. Nous démontrons la relation entre le traitement par paire et la méthode des temps d'arrivée pour ce cas simple. Nous montrons aussi comment la modification de la taille des fenêtres utilisées pour créer les spectrogrammes permet d'optimiser le compromis entre la résolution et la robustesse du traitement et comment ces paramètre peuvent être ajustés en fonction de la taille de la grille. La différence entre les positions estimées avec PWS et celles obtenues en utilisant la méthode des temps d'arrivée appliquée aux clicks individuels est de l'ordre d'une dizaine de mètres.

1 INTRODUCTION

The most commonly used methods for tracking marine mammals are time difference of arrival (TDOA) methods [*e.g.* Watkins and Schevill 1972; Clark *et al.* 1986; Spiesberger and Fristrup 1990; Janik *et al.* 2000]. In TDOA methods, the difference in time of arrival between pairs of hydrophones is estimated, usually via cross-correlation of waveforms or spectrograms. Each receiver pair defines a hyperboloid, and the intersection of hyperboloids (from various receiver pairs) defines the position of the source. Depending on the receiver geometry, four or five receivers are required to localize the source in three dimensions [Spiesberger 2001]. Reflections from the bottom and surface can be treated as recordings made by virtual receivers [Urick 1983]. Using reflections improves the accuracy of estimated source positions [Møhl et al 1990; Wahlberg *et al.* 2001; Thode *et al.* 2002] and reduces the number of required receivers [Aubauer *et al.* 2000; Tiemann *et al.* 2007; Laplance 2007].

TDOA methods are usually implemented with an isospeed assumption. This has the advantage of providing closed-form solutions and rapid run-times. It is acceptable in many cases (*e.g.* nearly isospeed conditions, relatively short distance propagation), particularly when care is taken to account for the resulting errors [Wahlberg *et al.* 2001; Speisberger and Wahlberg 2002]. In other cases, a depth-dependent sound speed profile can significantly improve position estimates [Chapman 2004; Tiemann *et al.* 2004; Nosal and Frazer 2006]. To remove isospeed assumptions, model-based TDOA methods can be implemented using a matched field approach [Tiemann *et al* 2004; Nosal and Frazer 2006, 2007]. TDOAs are estimated (as before), a 3D grid of candidate source location is created, and TDOAs are modeled repeatedly for a source at each of the grid points. The modeled

TDOAs are then compared to the measured TDOAs, and the estimated source location is the one that gives the best agreement. Therefore, we can regard the TDOA method as a matched-field method in which arrival time is the only part of the field being "matched."

Pair-wise spectrogram (PWS) processing [Nosal and Frazer in press] extends model-based TDOA methods by matching amplitude and phase information in addition to arrival times. Simulations have shown [Nosal and Frazer in press] that pair-wise processing is a promising passive acoustic localization method, but it has yet to be tested on real data. In this paper, a single sperm whale dataset is used to validate and explore the processor. Since TDOA methods give very good position estimates for this dataset, they are used to "ground-truth" the PWS position estimates.

2 DATA

The dataset was made available by Naval Undersea Warfare Center for the 2nd International Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics [Adam *et al.* 2006]. It features a single sperm whale producing regular clicks (on average 1.06 clicks/s) for 25 minutes. Recordings are from 5 widely-spaced bottom-mounted hydrophones at the Atlantic Undersea Test and Evaluation Center in the Tongue of the Ocean (off Andros Island, Bahamas). Signal to noise ratios vary between receivers, and clicks, and are typically between 2 and 30 dB. The sampling rate is 48 kHz. Filtering, phone sensitivity, and directivity are unknown. Hydrophone positions are given in Table 1. The track of this sperm whale has been found using various TDOA methods [Nosal and Frazer 2006, 2007; Giraudet *et al.* 2006; Morrissey *et al.* 2006; White *et al.* 2006].

Table.1. Hydrophone positions provided by NUWC

Phone	x-pos (m)	y-pos (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

3 PWW & PWS processing

3.1 Symbols and notation

s source waveform (time domain)
S source waveform (frequency domain)
r_i received signal at phone i (time)
R_i received signal at phone i (frequency)
$\hat{g}_i$ modeled impulse response at phone i (time)
$\hat{G}_i$ modeled impulse response at phone i (frequency)
N number of samples in a signal
N_r number of receivers
$*$ complex conjugate
ω radian temporal frequency

3.2 Overview

This section gives a brief overview of pair-wise processing. Complete details can be found in [Nosal and Frazer in press]. A 3D grid of candidate source location is created, and at a given candidate source location, the pair-wise waveform (PWW) processor is given by:

$$\rho_{pww} = \frac{\sum_{n=1}^{N}\sum_{i=1}^{N_r}\sum_{j\neq i}^{N_r} H_{ij}^{*}(\omega_n)H_{ji}(\omega_n)}{\sum_{n=1}^{N}\sum_{i=1}^{N_r}\sum_{j\neq i}^{N_r}\left|H_{ij}(\omega_n)\right|^2} \quad (1)$$

where $H_{ij}(\omega_n) = R_i(\omega_n)\hat{G}_j(\omega_n)$. This processor is maximized at the correct source location, since there the modeled impulse response is approximately equal to the true impulse response, *i.e.* $\hat{G}_i \approx G_i$ (not exactly equal because the model is never perfect), so that $H_{ij} = R_i\hat{G}_j = SG_i\hat{G}_j \approx SG_j\hat{G}_i = R_j\hat{G}_i = H_{ji}$.

Appendix A.3 shows that if only arrival time is used, pair-wise waveform (PWW) processing is equivalent to TDOA methods in a limiting case. Pair-wise processing assumes that all hydrophones have the same impulse response, but it makes no assumption about the spectrum of the source [Frazer and Sun 1998].

PWS processing is similar to PWW processing, except that spectrograms are processed instead of waveforms. This is useful because spectrograms are less sensitive to imperfections in the model due to uncertainties in sound speed profiles, receiver positions, and so on (see Appendix A.4 for an explanation). The PWS processor is:

$$\rho_{pws} = \frac{\sum_{m}\sum_{n}\sum_{i=1}^{N_r}\sum_{j\neq i}^{N_r} \tilde{H}_{ij}^{*}(T_m,f_n)\tilde{H}_{ji}(T_m,f_n)}{\sum_{m}\sum_{n}\sum_{i=1}^{N_r}\sum_{j\neq i}^{N_r}\left|\tilde{H}_{ij}(T_m,f_n)\right|^2} \quad (2)$$

where $\tilde{H}_{ij}$ denotes the spectrogram of H_{ij} and the sums are over all spectrogram time (T_m) and frequency (f_n) channels.

In nearly all problems of interest the number of sources is unknown, hence the output of ρ_{pww} and ρ_{pws} is potentially

multimodal and should be generated on a grid of candidate locations.

3.3 Processing of data

The PWS results presented here used the raw data only (no pre-processing) and the Gaussian beam acoustic propagation model BELLHOP [Porter and Bucker 1987; Porter 2005] was used to model Green's functions. Received signals were split into 30 s segments that overlapped by 20 s and each segment was processed separately. A 30 s segment length was chosen as an optimum after trials with different length segments. Shorter segments gave less consistent position estimates, possibly because some segments didn't contain enough clicks, and longer segments reduced performance, probably due to movement of the animal.

Within each segment, an important parameter is the window length used to create spectrograms. Simulations have shown [Nosal and Frazer in press] that longer window lengths give position estimates that are not as sensitive to uncertainties in the bathymetry, sound speed profile, receiver positions, and so on. With longer windows, the peaks of the likelihood surfaces are broader, which means that coarser grid spacing can be used. Figure 1 illustrates this using the first 30 s of the dataset. Spectrograms were all created using Hanning windows with 50% overlap. With a grid spacing of 200 m, a window length of 256 ms gives a good first estimate of the animal's position, while a window length of 32 ms does not (the animal is lost somewhere between grid points). With finer grid spacing (Fig. 2), spectrogram windows can be shortened to get increasingly precise position estimates. Trial runs indicated that a good window length is the travel time of sound between two grid points, *i.e.* window length should be roughly $2\Delta d/c$ for a grid spacing of Δd and speed of sound $c \approx 1500$ m/s.

Larger grid spacing gives faster run-times (with fewer points to process) so the tradeoff between robustness and resolution can be used to reduce computational requirements. Rough position estimates were made by running the processor for a large area with 200 m grid spacing and spectrogram window lengths of 256 ms.

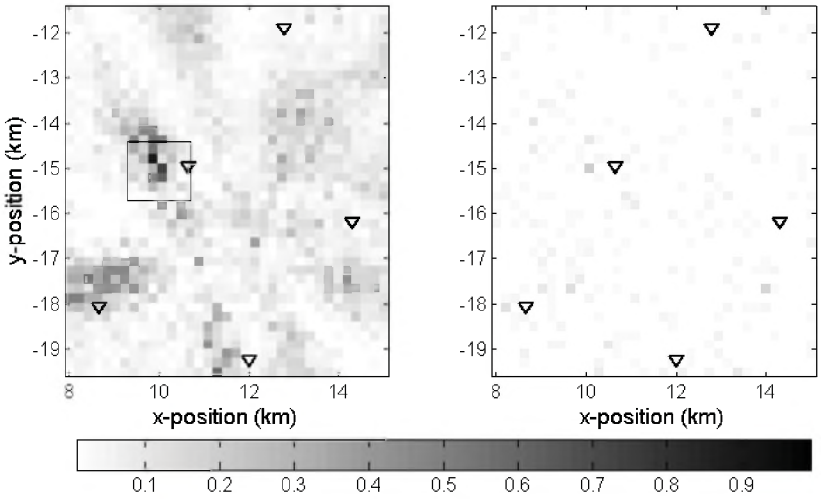

Figure 1. Plan view of the likelihood surface at a single depth of 685 m (approximately the correct depth of the animal) for the first 30 s of the dataset and a 200 m grid spacing. Receiver positions are indicated by triangles. Spectrograms use Hanning windows with 50% overlap and window lengths of (a) 256 ms, and (b) 32 ms. With such coarse grid spacing, 32 ms windows are too short to give a position estimate (unless the animal is near a grid point). The box in (a) indicates the area shown in Figure 2.

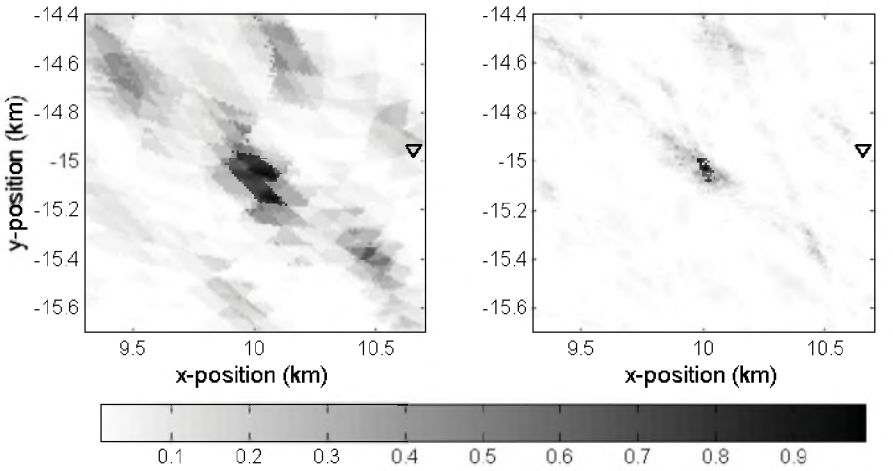

Figure 2. The boxed area shown in Fig 1 (a) processed here with 10 m grid spacing. With such fine grid spacing, 32 ms windows (b) give a more precise position estimate than 256 ms windows (a).

Rough position estimates were refined by processing smaller areas (surrounding the most promising grid point only) with 50 m grid spacing and spectrogram window lengths of 64 ms. Final estimates were obtained by further refining these with 10 m grid spacing and 16 ms windows. Resulting position estimates are shown in Figure 3. They are all within 40 m (and usually within 10 m) of those obtained with the TDOA method by Nosal and Frazer [2007], which have estimated 95% confidence interval half-widths of less than 25 m (details of this TDOA method are provided in Nosal and Frazer [2007]).

Error estimates were made as in Nosal and Frazer [2007] using conditional likelihood functions (CLFs). To get error in x at a given time step, y and z were fixed at the estimated source location and the resulting CLF was summed across x position to get the cumulative CLF for x (this can be thought of as summing along the horizontal strip that passes through the brightest point in Figure 2). 95% confidence intervals (CI) correspond to the distance between the x-positions at which the conditional CLF attains 2.5% and 97.5%, respectively. CIs for y and z were obtained analogously. For all dimensions, and over the whole track, CI half widths were within 45 m.

Note that error estimates cannot be directly compared because the TDOA method gave position estimates for every click while the PWS estimates are for 30 s intervals. It is evident, however, that the TDOA position estimates are more accurate for this dataset. On average, the whale moved 12, 40, and 9 m in the x, y, and z directions, respectively, in 30 s so that CI half widths within 45 m for the PWS methods are reasonable and may be partially explained by animal movement within a segment.

4 DISCUSSION

When deciding between localization 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. For the dataset considered here, with a single animal and very clear arrivals, the TDOA method excels since it gives better position estimates and is faster and easier to implement (to process 23 minutes of data, the TDOA method took 20 min while PWS processing took 43 minutes). Nevertheless, the PWS did successfully track the animal, which validates the PWS processor for this simple case. Future work will deal with more complicated datasets (more noise, multiple animals, shallower water, and long-duration calls) for which the PWS processor was developed [Nosal and Frazer in press].

One consideration that has not been addressed in PWS processing is source directivity. To use amplitudes, PWS assumes an omni-directional source; this does not hold for most marine mammals. Even though sperm whales clicks are highly directional [Møhl *et al.* 2003], the problem did not affect our results for this dataset, probably because there was enough information in the arrival times to overcome it. One solution for cases where directionality is a problem may be to use lower frequencies only, which are not as directional as higher frequencies. High frequencies might still contribute to the position estimates if spectrograms are thresholded to contribute only time of arrival information. Another (more difficult, but possibly more useful) approach is to include directivity in the source model and animal orientation in the search space.

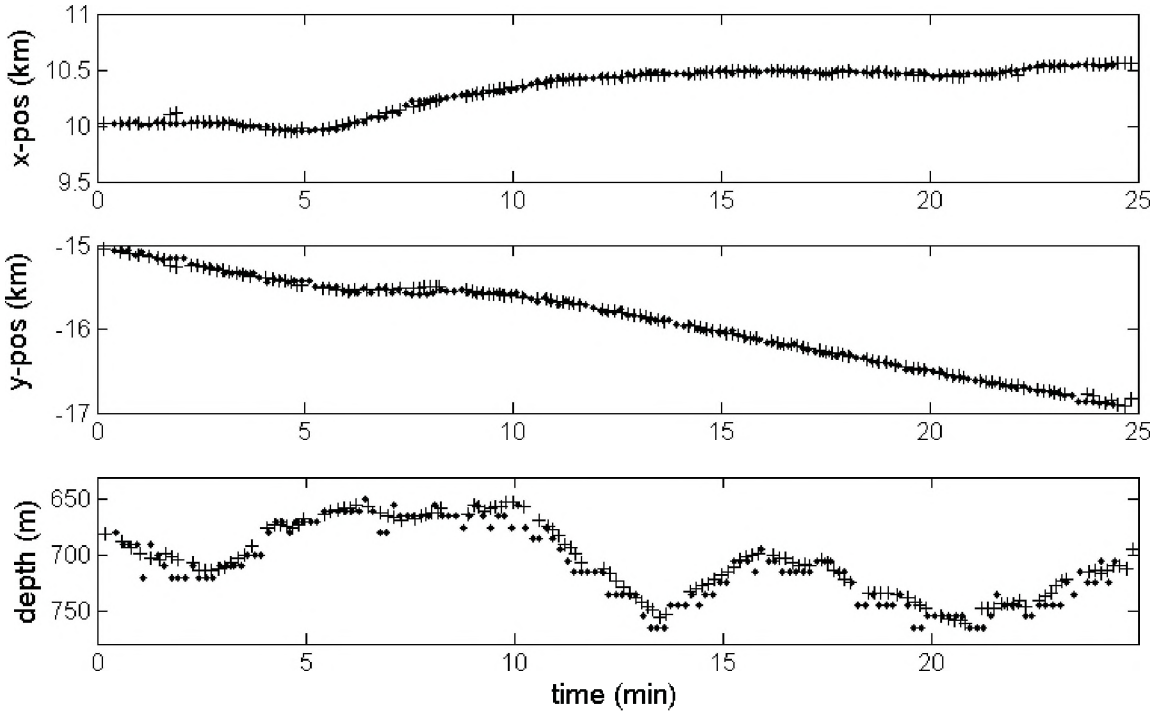

Figure 3. Whale positions estimated using PWS processing (dots) and TDOA processing (crosses).

APPENDIX

A.1 Notation

The appendix uses the same symbols and notation as in Section 3.1. The following are added:

$\otimes$ convolution:

$$(f \otimes g)(t_m) = \sum_{n=1}^{N} f(t_n) g(t_m - t_n)$$

$\circ$ cross-correlation:

$$(f \circ g)(t_m) = \sum_{n=1}^{N} f(t_n) g(t_m + t_n)$$

$\delta(t_n - \tau)$ unit impulse.

$$\delta(t_n - \tau) = \begin{cases} 1 \text{ when } t_n = \tau \\ 0 \text{ otherwise} \end{cases}$$

Discrete fourier transform of f, $DFT(f)$, is denoted using upper case:

$$f(t_n) \xleftrightarrow{DFT} F(\omega_n)$$

Inner product: $\langle f, g \rangle = \sum_{n=1}^{N} f(t_n) g^*(t_n)$

A.2 Required properties/relationships:

$$\delta(t_n - \tau) \otimes f(t_n) = f(t_n - \tau) \quad \text{(P1)}$$

$$\delta(t_n - \tau) \circ f(t_n) = f(t_n + \tau) \quad \text{(P2)}$$

$$f \circ g \xleftrightarrow{DFT} F^* G \quad \text{(P3)}$$

Power theorem: $\langle f, g^* \rangle = \frac{1}{N} \langle F, G^* \rangle$ (P4)

A.3 TDOA vs PWW for a simple case

This appendix aims to provide some intuition about the PWW processor and its relationship to the TDOA method. Consider the limiting case of infinitely fine grid spacing. Also consider unit impulses for the whale signal and the true and modeled impulse responses: respectively, $s(t_n) = \delta(t_n - \tau_0)$; $g_i(t_n) = \delta(t_n - \tau_i)$; and $\hat{g}_i(t_n) = \delta(t_n - \hat{\tau}_i)$, where τ_0 is the time at which the whale produces the impulse, τ_i is the true direct arrival time at receiver *i*, and $\hat{\tau}_i$ is the modeled direct arrival time at receiver *i*. A delayed impulse is an unrealistic Green's function, but it is useful for gaining insight.

By (P1) the received signal at receiver *i* is:

$$r_i(t_n) = s \otimes g(t_n) = \delta(t_n - \tau_0 - \tau_i). \quad \text{(A1)}$$

The cross-correlation of the received signals r_i and r_j at hydrophone pair *i-j* is then (by P2):

$$\begin{aligned} r_i \circ r_j(t_n) &= \delta(t_n - \tau_0 - \tau_i) \circ \delta(t_n - \tau_0 - \tau_j) \\ &= \delta(t_n + \tau_0 + \tau_i - \tau_0 - \tau_j) \\ &= \delta(t_n + \tau_i - \tau_j) \end{aligned} \quad \text{(A2)}$$

Similarly, the cross-correlation of the modeled impulse responses $\hat{g}_i(t)$ and $\hat{g}_j(t)$ at hydrophone pair *i-j* is

$$\hat{g}_i \circ \hat{g}_j(t_n) = \delta(t_n + \hat{\tau}_i - \hat{\tau}_j) \quad \text{(A3)}$$

In the TDOA method, $\tau_i - \tau_j$ is found as the time that maximizes $r_i \circ r_j$. The best candidate source location is the one that minimizes the difference between $\tau_i - \tau_j$ and $\hat{\tau}_i - \hat{\tau}_j$. In our unit impulse case, this is equivalent to finding the source location that maximizes the inner product between $r_i \circ r_j$ and $\hat{g}_i \circ \hat{g}_j$ since

$$\begin{aligned} \left\langle r_i \circ r_j, \hat{g}_i \circ \hat{g}_j \right\rangle &= \sum_{n=1}^{N} \delta(t_n + \tau_i - \tau_j)\delta(t_n + \hat{\tau}_i - \hat{\tau}_j) \\ &= \begin{cases} 1 \text{ if } \tau_i - \tau_j = \hat{\tau}_i - \hat{\tau}_j \\ 0 \text{ otherwise} \end{cases} \end{aligned} \quad \text{(A4)}$$

Since the signals are impulses, the denominator in the PWW processor, Eq. (1), for a single receiver pair is N. Accordingly, for receiver pair *i-j* the PWW processor is (by P3 and P4):

$$\begin{aligned} \rho_{pww} &= \frac{1}{N} \sum_{n=1}^{N} \left(R_i(\omega_n) \hat{G}_j(\omega_n) \right)^* R_j(\omega_n) \hat{G}_i(\omega_n) \\ &= \frac{1}{N} \sum_{n=1}^{N} \left(R^*_i(\omega_n) R_j(\omega_n) \right) \left(\hat{G}_i(\omega_n) \hat{G}^*_j(\omega_n) \right) \\ &= \frac{1}{N} \left\langle DFT(r_i \circ r_j), \left(DFT(\hat{g}_i \circ \hat{g}_j) \right)^* \right\rangle \\ &= \left\langle r_i \circ r_j, \hat{g}_i \circ \hat{g}_j \right\rangle \end{aligned} \quad \text{(A5)}$$

Compare this to (A4) to see that the TDOA method and the PWW processor are equivalent for a single receiver pair *i-j*. In principle, the argument can be extended to include surface and bottom reflections and to show that TDOA methods that use cross-correlation of spectrograms to find TDOAs are equivalent to PWS processing, within the limits of the geometrical acoustics approximation to a wavefield.

In general, differences between the methods include the way in which different receiver pairs are combined. Moreover, PWW retains amplitude and phase information from the signals, while the TDOA method keeps only time of arrival information.

A.4 PWW vs PWS processing

The simple example in A.3 illustrates the motivation for using spectrograms in pair-wise processing instead of waveforms. With finite grid spacing, $\tau_i - \tau_j$ is never exactly equal to $\hat{\tau}_i - \hat{\tau}_j$. Even for a candidate source location exactly at the correct source location, the environment and model are imperfect, *i.e.* $\tau_i - \tau_j \neq \hat{\tau}_i - \hat{\tau}_j$, so by (A4) the processor is always zero. In a real case with non-unit impulses, the processor is not zero but attains only low values. By taking spectrograms, arrivals are smeared in time, so they don't need to match as precisely. This gives poorer resolution (positions are not found as precisely) but allows for the use of coarser grids (see Section 3.3) and makes the PWS processor more robust with respect to environmental and modeling uncertainties than the PWW processor (demonstrated using simulations in Nosal and Frazer [in press]).

ACKNOWLEDGMENT

This work was funded by the Office of Naval Research. The authors thank Mélanie Abecassis for help in translating the abstract.

REFERENCES

Adam O, J-F Motsch, F Desharnais, NA DiMarzio, D Gillespie, RC Gisiner (2006). Overview of the 2005 workshop on detection and localization of marine mammals using passive acoustics. *Appl. Acoust.* 67, 1061-1070.

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.

Chapman DMF (2004). You can't get there from here: Shallow water sound propagation and whale localization. *Can. Acoust.* 32(2), 167-171.

Clark CW, WT Ellison, K Beeman (1986). Acoustic tracking of migrating bowhead whales. *Proceedings of IEEE Oceans* 1986, 341-346.

Frazer LN, Sun X (1998). New objective functions for waveform inversion. *Geophysics* 63, 1-10.

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 AUTEC. *Appl. Acoust.* 67, 1106-1117.

Janik VM, SM VanParijs, PM Thompson (2000). A two-dimensional acoustic localization system for marine mammals. *Mar. Mamm. Sci.* 16, 437-447.

Laplance 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.

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 (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.

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

Nosal E-M, LN Frazer (2007). Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones. *J. Acoust. Soc. Am.* 122(4), 1969-1978.

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

Porter MB, HP Bucker (1987). Gaussian beam tracing for computing ocean acoustic fields. *J. Acoust. Soc. Am.* 82(4), 1349-1359.

Porter MB (2005). *BELLHOP*. Last accessed on Dec 9, 2007 at < http://oalib.hlsresearch.com/>.

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.

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.

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, 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.

Urick RJ (1983). *Principles of underwater sound*, 3rd ed. Peninsula, Los Altos, CA.

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.

White PR, TG Leighton, DC Finfer, C Prowles, C Baumann (2006). Localisation of sperm whales using bottom-mounted sensors. *Appl. Acoust.* 67, 1074-1090.

Photo Credit:Greg Schorr. Copyright: Cascadia Research, Olympia, WA, USA