CHIRP CORP.
8248 SUGARMAN DRIVE LA JOLLA CA 92037 (619) 453-4406

Final Report

Contract No. N00014-95-C-0297

Report Title:

Signal Processing and Acoustic Imaging Models for Animal Sonar

Author:

Richard A. Altes, Ph.D.
Chirp Corporation
8248 Sugarman Drive
La Jolla, CA 92037

Funding Agency:

Office of Naval Research
Cognitive and Neural S&T Division, Code 342
800 North Quincy Street
Arlington, VA 22217-5660

Date:

June 20, 1999

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	20 June 1999	Final Report: 01 Sept 95 - 31 Aug 98

4. TITLE AND SUBTITLE

Signal Processing and Acoustic Imaging Models for Animal Sonar

5. FUNDING NUMBERS

N00014-95-C-0297

6. AUTHOR(S)

Richard A. Altes, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Chirp Corporation
8248 Sugarman Drive
La Jolla, CA 92037

8. PERFORMING ORGANIZATION REPORT NUMBER

N00014-95-C-0297
Final Report

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research, Code 342
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A biologically inspired acoustic imaging process is obtained by considering the constraints (e.g., limited array/aperture size), and advantages (e.g., unrestricted motion, wide bandwidth) of biological sonar systems. The resulting system is easy to implement compared with conventional synthetic aperture sonar. Images are formed sequentially; an image is updated on the basis of each new echo. The processor is Doppler tolerant and is tolerant of target-induced phase shifts that occur as the result of aspect changes. The simplicity of the process leads to generalized images that can distinguish between smooth and rough surfaces and that can detect and classify low-reflectivity targets in strong volume clutter. Accuracy is improved via an associative gradient descent algorithm that can converge rapidly to a global minimum by utilizing prior information, extra acoustic information that is not directly used for imaging (e.g., resonance phenomena), and non-acoustic cues. An image-based tracker exploits sequential image formation and parallel processing to compensate for motion perturbations and to model simultaneous imaging and pursuit of prey in animal sonar systems. The imaging system is remarkably robust with respect to sparse angular sampling; image quality degrades gracefully as angular sampling intervals are increased well beyond conventional limits. Insights into neurophysiology and cognition include the utilization of top-down, bottom-up processing to obtain high resolution images from low resolution representations, as in the superior colliculus.

14. SUBJECT TERMS

Broadband sonar, Biomimetics, Animal echolocation, Synthetic aperture sonar, Tracking, Acoustic imaging, Motion compensation, Adaptive focusing, Mine Countermeasures

15. NUMBER OF PAGES

48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited (UL)

Abstract (long version)

The azimuth/elevation (cross-range) resolution of a bat or dolphin sonar is theoretically much worse than its range resolution at ranges of one-half meter or more. Synthetic aperture sonar (SAS) processing can be used to improve cross-range resolution by summing echoes from the same target point as seen from different relative positions of the sonar and target. Dolphins and bats that transmit broadband, frequency modulated signals use Doppler tolerant waveforms and may be incapable of pulse-to-pulse coherent processing. A relevant form of SAS involves Doppler tolerant, noncoherent delay-and-sum processing, and it is similar to back projection tomography.

Another technique for improving cross-range resolution is to scan the environment with multiple, overlapping beams from a single sensor position, and to solve the resulting set of simultaneous equations for the reflectivity of each point (beam pattern deconvolution). Both beam scanning and translational motion are observed in animals. Beam deconvolution and SAS can be combined in a sonar processor that forms an internal model of the environment and that updates this model by comparing it with echo data. This updating method is known as the algebraic reconstruction technique (ART) in medical imaging applications, and it is equivalent to gradient descent for solving deconvolution (sharpening) problems. Simple, biologically feasible versions of back projection SAS and ART-SAS can form images of reflectivity and other target features as a function of range, azimuth, and elevation. The resulting reflectivity and feature images are used to explain how dolphins find buried fish.

ART-SAS can be implemented with a gradient descent optimization process that uses top-down, bottom-up processing. This processor obtains high resolution target representations from low resolution, nonlinear signal representations, and it can be used to generalize spectrogram correlation models to explain bats' apparent pulse compression capability (matched-filter equivalent processing). Gradient descent parameter estimation can be combined with a target classifier that uses associative memory. The resulting associative gradient descent process has fast convergence and avoids spurious local minima. Gradient descent top-down, bottom-up processing compares low-resolution versions of sensor images (as in the superior colliculus) with low-resolution (deliberately degraded) versions of a high-resolution image model. The high-resolution model is updated via the error generated by the low-resolution comparison. Although conventional image sharpening operations (Laplacian, 2-D high-pass, lateral inhibition) operators are not precluded by this process, they seem to be unnecessary.

A software implementation of the biologically inspired acoustic imaging system is used to form conventional images and feature images of mines that are ensonified with a dolphin-like pulse. Some of these images are formed with very sparse aspect sampling, corresponding to area coverage rates that are at least an order of magnitude larger than the area coverage rates of conventional SAS systems. This improvement is associated with bionic, nonlinear suppression of artifacts associated with SAS point spread function sidelobes, and with tolerance to aspect dependent phase changes induced by physical scattering mechanisms.

Motion compensation and adaptive focusing are obtained with an image-based tracker, which is a requirement for a biological system that pursues and images prey simultaneously. Image-based tracking accomplishes motion compensation and adaptive focusing by utilizing sequentially formed test images together with an image evaluation criterion. The tracker can incorporate both translational and rotational motion.

Table of Contents

List of Figures

1. Introduction

Animal sonar systems typically are characterized by large bandwidths, motion of the transmitter/receiver, and small aperture (array size) relative to man-made sonars. Such systems have cross-range (azimuth and/or elevation) resolution that is much worse than their range resolution. For bats and dolphins, the theoretical disparity between range and cross-range resolution becomes large for ranges in excess of half a meter. This disparity can be mitigated by using synthetic aperture sonar (SAS) processing. SAS forms an image in which cross-range resolution is commensurate with range resolution. With the possible exception of single-pulse Doppler processing, however, conventional SAS appears to be much too complicated for implementation by biological systems.

Recent results indicate that synthetic aperture processing can be greatly simplified. By considering simplifications that allow for biological implementation and generalizations that can emulate animal echolocation capabilities, SAS processing has actually been advanced beyond the previous state-of-the-art for man-made systems. These advances involve high resolution feature images, image-based tracking for motion compensation, sparse angular sampling, and the utilization of all available knowledge for acoustic imaging. All available knowledge includes prior expectations, non-acoustic sensory information, and acoustic information that is not explicitly associated with imaging, such as resonances.

Experimentally derived images that are shown in this report are from targets that were suspended in lake water. The targets were completely contained within the transmitter/receiver sonar beam width and echoes were obtained as the targets were rotated. The transmitted signal was a dolphin-like pulse with 10 dB bandwidth between 50 and 150 kHz. The echoes were recorded by ARL, Univ. of Texas at Austin and were put into PC format and furnished to Chirp Corp. by P. Moore, D. Helweg, and J. Sigurdson, Code D351, SPAWAR Systems Center, San Diego.

2. Doppler-based SAS

Some synthetic aperture systems depend upon Doppler sensitivity while others are Doppler tolerant. A Doppler-based system utilizes the angle dependence of range-rate for a moving sonar and a stationary scattering point. A relatively large range-rate is observed along the path of the sonar platform motion, and zero range-rate is observed orthogonal to the path of motion. A mapping thus exists between azimuth angle and range-rate. For side-looking systems, an object appears to rotate relative to the sonar, and this rotational motion can be used to form an image of the object.

Range-rate can be measured with a single pulse if the pulse has sufficient time-bandwidth product to estimate relevant Doppler-induced time compressions of the signal [1,2]. If a single pulse has insufficient time-bandwidth product for Doppler-based angle measurements, then a fully coherent system can use multiple pulses for range rate estimation. In the multi-pulse case, many echoes are stored and processed as though they are all obtained from a long-duration transmitted signal that is composed of many transmitted pulses. Even if a new pulse is transmitted only after the echo from the previous pulse is received, the new pulse is part of the composite signal. This kind of "pulse-Doppler" processing requires a coherent integration time that extends over multiple transmissions and receptions. Different scattering points correspond to different range-rate vs time

histories, and a coherent pulse-Doppler processor is used to form a matched filter for each image point [3-5].

3. Doppler tolerant, tomographic SAS

It is uncertain whether echolocating animals are capable of multi-pulse, coherent processing, although sensitivity to a phase shift of a *single echo* has been demonstrated in bats [6-8]. A conservative model assumes that *multi-pulse* coherent processing capability is lacking. Even if bats or dolphins have the ability to coherently sum echoes from different transmissions, a noncoherent SAS processor or semi-coherent SAS (summation of envelope detected matched filter outputs) is advantageous because:

1. The SAS processor is much easier to implement and is more tolerant of small errors between predicted and actual ranges and range-rates,
2. Little resolution is lost when wideband, biological signals are used, since the envelope of the signal auto-correlation function contains only a few of the fine-structure oscillations that are used for coherent processing;
3. The aspect sampling constraints that are associated with spatial aliasing (the effect of synthetic array element locations that are separated by more than half a wavelength) are removed, and are replaced by graceful degradation for sparse angular sampling (degradation that can be predicted from the peak-to-sidelobe ratio of the SAS point spread function, which is the same as the range, cross-range ambiguity function of the synthetic array processor);
4. The image is tolerant of aspect-dependent phase changes that are introduced by viewing the same target point from different directions.

A wideband, Doppler tolerant, tomographic synthetic aperture processor can be simplified to remove the requirement for fully coherent processing. In the case of dolphins, even the single-pulse matched filter assumption (or a process that is equivalent to matched filtering) is not required. This type of SAS processor is described in the following paragraphs.

A moving sonar transmits signals and receives echoes from a sequence of points along its path of motion. The receiving points along the sonar's path are regarded as the locations of elements that are part of a large, synthetic array. This synthetic array can focus on a particular point by delay-and-sum beam forming. A compensatory delay is inserted at the output of each element, such that all the echoes from a given scattering point occur at the same time. The resulting time-registered echoes are then added. *The delay-and-sum process is equivalent to forming a spatial matched filter for echoes from the chosen scattering point.*

The delay-and-sum process is also equivalent to reconstructing an image from its projections [9]. Projections occur naturally in radar/sonar data. All scattering points that are within the physical beam width of the sonar and that are at the same range (i.e., that lie along the same constant-range surface) contribute to the same echo sample. The sequence of echo samples on an A-scan (matched filter response vs range) represents a projection of the scatterer reflectivity distribution along the range axis. Different transmitter/receiver locations correspond to different propagation directions, and thus to different projections of the scatterer reflectivity distribution as shown in Figure 1.

Several methods can be used to reconstruct the reflectivity distribution from its projections [10]. The back projection algorithm is nearly identical to delay-and-sum beam forming, as demonstrated in Appendix A [11]. Back projection or delay-and-sum beam forming can be implemented

sequentially, such that the reflectivity estimate of each pixel is updated with each new echo. The echo sample that corresponds to a given target point is added to the sum of previous samples (one from each previous echo) that correspond to the same point. At a given sonar location, the echo sample corresponding to a chosen target point also corresponds to all the other target and clutter points at the same range. At a new location, the constant-range surface is rotated, and the echo sample for the chosen target point corresponds to other points that are located on a *different* constant-range surface, as in Figure 1.

Delay-and-sum synthetic aperture processing does not require Doppler information. In fact, processing would be simplified if the sonar were to stop at each synthetic array element location, transmit a signal, receive the resulting echoes, and then move to the next transmit/receive location. To take advantage of this simplification without stopping, the system can use Doppler tolerant signals, such that the matched filter response is not sensitive to range-rate.

Another simplification is to use noncoherent delay-and-sum beamforming, such that matched filter envelopes are used and phase is discarded. This simplification is feasible with wide band signals, since the envelope of the matched filter response contains comparatively few oscillatory "fine structure" peaks, which correspond to phase information. The resulting processor is semicoherent; a matched filter is used for each echo, but different echoes are noncoherently combined by summing the envelope detected matched filter responses.

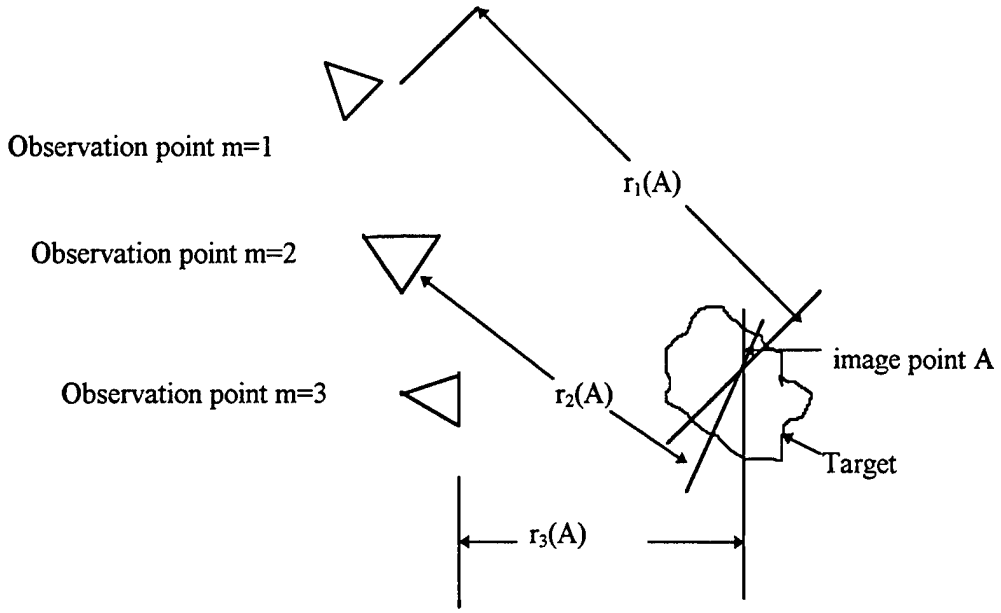


Figure 1. As a sensor moves, it creates a synthetic array with elements m=1,2,3,.... The range of point A is different for each of these elements. To focus the synthetic array on point A, the corresponding delays are compensated by a delay-and-sum operation. Each line through point A represents part of a constant-range surface that lies within the beam width of the transmitter. All the scattering points that are on a line contribute to the m^{th} echo at range $r_m(A)$.

Wide-band, short duration pulses such as those used by dolphins are Doppler tolerant with or without semicoherent processing. Long duration wide band signals with hyperbolic frequency modulation (linear period modulation) such as those used (or approximated) by many FM bats are Doppler tolerant when they are processed by a semicoherent receiver [12-14]. In the case of dolphins, a matched filter may be approximated by the band-pass operation of the receiver, since the signal has very small time-bandwidth product. A tomographic SAS model for dolphins can thus use noncoherent processing without a matched filter assumption.

A tomographic SAS model for bats requires pulse compression via matched filtering, inverse filtering, or an equivalent process, together with noncoherent pulse-to-pulse summation capability. A process that is equivalent to matched filtering or inverse filtering may be synthesized by spectrogram correlation [15,16] or by a time-frequency plane version of the top-down, bottom-up gradient descent process to be discussed in the Section 9.

4. The range, cross-range ambiguity function (SAS point spread function)

The delay-and-sum receiver response to a point scatterer is the sum of the rotated constant-range curves in Figure 1. This sum is shaped like an asterisk. For M different sensor positions, the center point of the asterisk is M times larger than an individual line, as shown in Figure 2.

A sampled version of a two-dimensional reflectivity distribution is an array of sample points with different reflectivities. The back projection SAS image of the sampled reflectivity distribution is a superposition of weighted, shifted versions of the function in Figure 2, where the weights correspond to the sample point reflectivities and the shifts correspond to the locations of the sample points. This weighted sum is a discrete convolution operation. The image of the reflectivity distribution is the convolution of the function in Figure 2 with the actual distribution (or a real, non-negative version of the actual distribution).

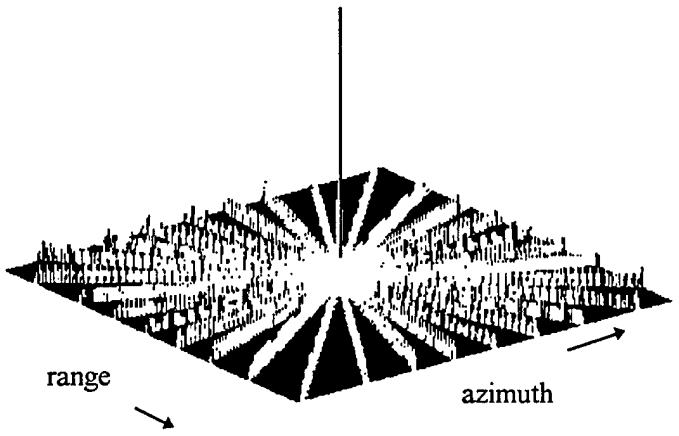


Figure 2. The point spread function (SAS range, cross-range ambiguity function) of a tomographic SAS with 12 degree angle increments over 180 degrees. Peak-to-sidelobe ratio = number of echoes used to construct the image = 180/12 = 15.

The actual reflectivity distribution is convolved with (or smeared by) the function in Figure 2, which is known as the point spread function (PSF). The function in Figure 2 also represents the response of a receiver that makes a hypothesis that a point scatterer is present at the center of the asterisk, when the actual point scatterer position is at various points on the image plane in the Figure. The function in Figure 2 is thus the range, cross-range ambiguity function (RCRAF) of the imaging system, as well as the point spread function [17,18]. This ambiguity function has a peak-to-sidelobe ratio of M, where M is the number of different sensor positions or elements in the synthetic array. *When M is small (as in a monaural or binaural system with no SAS capability), a strongly reflecting point at one location can severely affect the image of a weakly reflecting point at a different location.* PSF sidelobes produce artifacts when a scattering point is much larger than its neighbors for at least one aspect angle. The artifacts appear as lines that pass through the strong scattering point.

If the peak-to-sidelobe ratio (P/S) of the range, cross-range ambiguity function (RCRAF) is small, different points on a distributed target can interfere with one another, leading to a self-clutter effect. A target that is surrounded by other scatterers also will be difficult to detect. *A psychometric procedure that measures angular accuracy or resolution between closely spaced points may depend only on the sharpness of the central peak of the binaural RCRAF, and can be misleading with respect to detection in clutter and classification of distributed targets.*

For a binaural system, the RCRAF has P/S=2 and is scissor-shaped. As a binaural sonar approaches a target, the angle between the scissor blades increases. If range resolution is sufficiently fine (if the width of the scissor blades is small) and if echo samples corresponding to each point on the target are summed as the sonar approaches the target, then P/S becomes larger and the image becomes less ambiguous as the target is approached. Binaural processing thus can be used with forward-looking SAS to create an acoustic image. Biomimetic nonlinear processing can be used to reduce the effects of large point spread function sidelobes.

5. Comparison of biologically inspired SAS with conventional SAS

Biologically inspired SAS has at least two properties that make it different from conventional Doppler tolerant (tomographic) SAS:

1. Biomimetic nonlinear suppression of the sidelobes of the point spread function (PSF) of SAS images;
2. Noncoherent (phase tolerant) summation over different aspect angles, which decreases sensitivity of the SAS image to aspect-dependent phase shifts that depend on target shape.

Different phase shifts are expected to occur when various target structures are viewed from different aspects. When a flat metal plate is tilted so that it does not reflect energy directly back toward a receiver, its impulse response (as approximated by physical optics) changes from a single positive impulse to a positive impulse at the leading edge and a negative impulse at the trailing edge. The impulse response of a spherical scatterer is approximated by an impulse followed by a rectangular function. The rectangular function is synthesized via an integrator with positive weight followed by another, delayed integrator with negative weight. In general, the echo from a complex target can be represented as a weighted sum of delayed versions of the transmitted signal, integrated versions of the signal, and differentiated versions of the signal. The weights and delays in this sum are aspect dependent; the weights can change sign (e.g., from positive to negative) depending on aspect. The integration and differentiation operations induce 90 degree phase shifts.

An example of an aspect dependent sign change is shown in Figure 3. A tilted rectangular plate is composed of material with higher acoustic velocity than water (lower acoustic impedance). The physical optics approximation to the target impulse response is a positive impulse from the near edge of the plate followed by a negative impulse from the far edge. The impulse response of an edge that is initially closest to the sonar changes from positive to negative as the plate rotates.

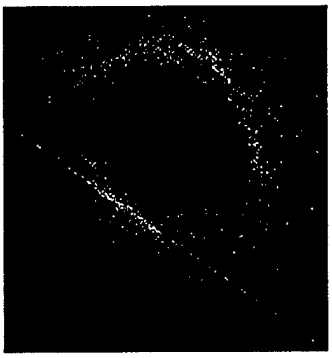
Conventional SAR/SAS systems implicitly assume that targets are composed of independent point scatterers with aspect independent phase shifts. This assumption has not been deleterious because conventional systems generally do not view targets over observation angles that exceed 120 degrees. Insensitivity to aspect-dependent phase shifts is obtained via noncoherent summation of matched filtered echoes over different aspect angles. Such noncoherent summation is feasible with minimum loss of resolution when signals are very broadband, as in biological sonar systems.

PSF (point spread function) sidelobe suppression and tolerance of aspect-dependent phase shifts tend to make BioSAS comparatively tolerant to sparse aspect sampling over large angular observation intervals. Sparse sampling in aspect decreases the peak-to-sidelobe ratio of the PSF, and a method that suppresses the effect of such sidelobes leads to better images when aspect samples are far apart. Large aspect changes may result in different phase shifts from the same scattering point as in Figure 3, and tolerance of such phase shifts tends to increase image quality.

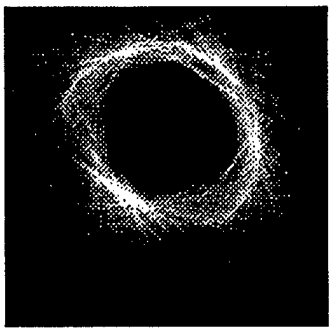
BioSAS can be compared with conventional techniques by ascertaining the separate effects of each BioSAS property, i.e., (1) PSF sidelobe suppression and (2) insensitivity to aspect-dependent phase shifts. This comparison is illustrated in Figure 4. The images in Fig. 4 are generated from echoes obtained from the Manta mine in the SPAWAR/ARL data set (SPAWAR Code 351, San Diego, CA). The Manta is viewed over 360 degrees at four degree intervals. An image generated by a conventional SAS processor is shown in Figure 4a. In Figure 4b, phase tolerance is included, but sidelobe suppression is lacking. In Figure 4c, sidelobe suppression is used without phase tolerance. Finally, in Figure 4d, both sidelobe suppression and phase tolerance are used, thus converting the conventional SAS into BioSAS. The images in Figure 4 indicate that a majority of target points have echo phase shifts that are sensitive to aspect changes. These target points appear on the phase tolerant images but are suppressed on the phase sensitive images.

To assess the effect of sparse aspect sampling, the four processors in Figure 4 can be compared for a large aspect sampling interval of 20 degrees (a total of eighteen "looks" at the target over a 360 degree interval). The results are shown in Figure 5.

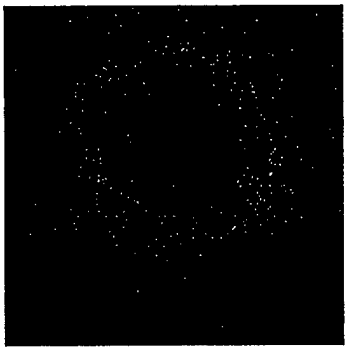
Figure 3. A tilted metal plate in water has a back-scatter impulse response composed of a positive impulse followed by a negative one, even when the plate rotates 180 degrees. Cancellation of edge images can occur with phase-sensitive processing over 180 degrees or more.



a. No PSF sidelobe suppression, no phase tolerance (conventional SAS).



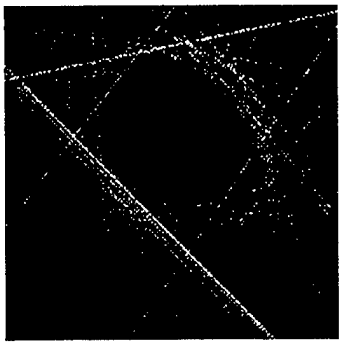
b. No PSF sidelobe suppression, with phase tolerance.



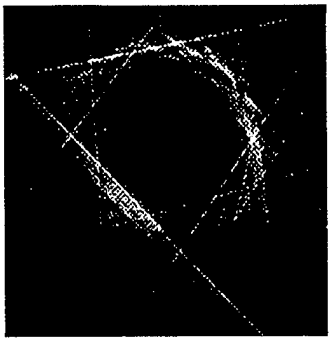
c. With PSF sidelobe suppression, no phase tolerance.

d. With PSF sidelobe suppression, with phase tolerance (BioSAS).

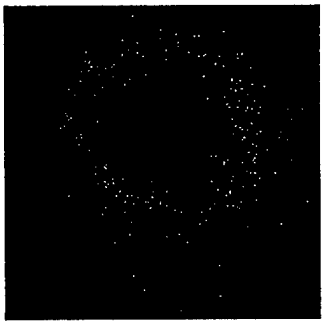
Figure 4. Comparison of SAS images generated with four different processors using a signal with 100 kHz bandwidth and 100 kHz center frequency, with aspect samples that are 4 degrees apart, and with an angular observation interval of 360 degrees. (a) Fully coherent processing. (b) Semicoherent processing (noncoherent summation over aspect for tolerance of aspect-dependent phase shifts). (c) Full coherence combined with nonlinear processing to reduce sidelobes of the SAS point spread function (the range, cross-range ambiguity function). (d). Semicoherent processing for tolerance of aspect-sensitive phase shifts combined with nonlinear PSF sidelobe reduction; biologically inspired SAS processing.



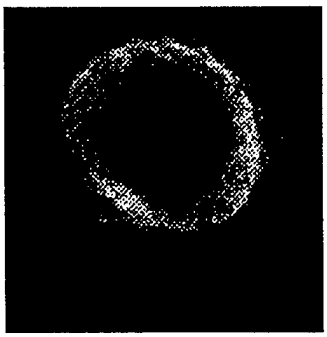
a. No PSF sidelobe suppression, no phase tolerance (conventional SAS).



b. No PSF sidelobe suppression, with phase tolerance.



c. With PSF sidelobe suppression, no phase tolerance.



d. With PSF sidelobe suppression, with phase tolerance (BioSAS).

Figure 5. Comparison of SAS images generated with four different processors using a signal with 100 kHz bandwidth and 100 kHz center frequency, with aspect samples that are 20 degrees apart, and with an angular observation interval of 360 degrees. (a) Fully coherent processing. (b) Semicoherent processing (noncoherent summation over aspect for tolerance of aspect-dependent phase shifts). (c) Full coherence combined with nonlinear processing to reduce sidelobes of the SAS point spread function (the range, cross-range ambiguity function). (d). Semicoherent processing for tolerance of aspect-sensitive phase shifts combined with nonlinear PSF sidelobe reduction; biologically inspired SAS processing.

6. The algebraic reconstruction technique (ART) and top-down, bottom-up processing

A range sample of a pulse-compressed echo amplitude vs. range plot (A-scan) corresponds to the projection (the sum of the reflectivities) of all the pixels in a constant-range surface. If the sonar moves, the measured sample value represents one of many simultaneous linear equations that theoretically can be solved to obtain the reflectivity distribution. The corresponding matrix equation can be solved by a gradient descent optimization technique. Echo samples that are generated from a model of the reflectivity distribution are compared with actual echo data samples, and the difference is used iteratively to improve the model. This method is called the algebraic reconstruction technique (ART) [10]. ART is a special case of an iterative sharpening or deconvolution algorithm. Solution of the matrix equation is equivalent to applying an inverse matrix to the observations, where the matrix that is inverted includes the point spread function. If ART processing can invert the point spread function (i.e., convert the PSF into an impulse), then the ART image should be superior to back projection.

Top-down, bottom-up processing is a cognitive model that describes the interaction between an internal (top-down) representation and sensory input (bottom-up) data [21,60]. A comparison between predicted and observed data results in a correction or modification of the internal representation. The comparison and correction can be implemented as a gradient descent algorithm in which a high resolution, internal model is used to synthesize the input data that would be observed at a neuronal processing center. The comparison can occur in a relatively low resolution representation, e.g., in a cochlear time-frequency representation or in an image that is represented by the superior colliculus [22].

The mean-squared error between the echo and model representations is

$$MSE(\underline{\underline{A}}) = (2\pi M R_{\max})^{-1} E\{\sum_{m=1}^{M} \int_0^{2\pi}\int_0^{R\max} [echo_m(r,\theta) - mdl_m(\underline{\underline{A}};r,\theta)]^2 \, dr d\theta\} \qquad (1)$$

where the parameter matrix $\underline{\underline{A}}$ contains sample values a_{ij} of the high-resolution representation and $E\{\bullet\}$ denotes an ensemble average over various realizations of the echo and the model. The echo varies because it is corrupted by noise, and the model may vary because it can include stochastic neuronal responses. The sum over m represents observations from M different aspect angles, as in Figure 1. The simplest echo model is a smeared version of the high resolution representation:

$$mdl_m(\underline{\underline{A}};r,\theta) = \sum_{i,j} a_{ij} \, smear_m(r - r_i, \theta - \theta_j). \qquad (2)$$

The smearing function represents the loss of angular resolution that is associated with a wide physical beam width. If the signal is a short duration pulse, the m^{th} echo of a given target point is corrupted by a line-like smearing function that is orthogonal to the propagation direction, as in Figure 1.

The gradient descent technique iteratively solves for the high resolution samples a_{ij} via the recursion

$$a_{kl}(n+1) = a_{kl}(n) - \mu \frac{\partial MSE(\underline{\underline{A}})}{\partial a_{kl}} \quad \text{for all values of } k \text{ and } l \tag{3}$$

where n is the number of times the iteration has been repeated. The LMS stochastic gradient algorithm uses only the squared error rather than the mean-squared error, yielding an update equation

$$a_{kl}(n+1) = a_{kl}(n) - \mu \frac{\partial SE(\underline{\underline{A}}; r, \theta, m)}{\partial a_{kl}} \quad \text{for all values of } k \text{ and } l. \tag{4}$$

In the LMS algorithm, the iteration is computed at successive values of the variables (r, θ) and observations m=1,...,M. The iteration is sufficiently repetitive to approximate an ensemble average over multiple observations. The LMS (or LMS stochastic gradient) algorithm is very simple to implement and has many engineering applications [23-25].

The gradient in (4) depends only upon the error at a particular (r, θ) value and the partial derivative of the model with respect to a_{kl}:

$$\frac{\partial SE(\underline{\underline{A}}; r, \theta, m)}{\partial a_{kl}} = -2[echo_m(r, \theta) - mdl_m(\underline{\underline{A}}; r, \theta)][\partial mdl_m(\underline{\underline{A}}; r, \theta) / \partial a_{kl}]. \tag{5}$$

In (5), *the error between the (bottom-up) echo and the (top-down) model is measured in a low resolution or smeared representation*. For the simple convolution or smearing model in (2),

$$\partial mdl_m(\underline{\underline{A}}; r, \theta, m) / \partial a_{kl}] = smear_m(r - r_k, \theta - \theta_l). \tag{6}$$

In the LMS algorithm, the gradient of the squared error equals [the error at (r, θ), as measured in a low resolution representation] × [the smearing function for a relevant area of the high resolution image]. If the smearing function is broad in bearing and narrow in range as in Figure 1, the update equation uses the error at a given range value to update a swath of high-resolution pixels in the internal model. This update operation is the same as ART. *The high resolution representation is updated by using the error in the low resolution representation along with the known smearing function.*

Images that are obtained from back projection BioSAS and from ART-BioSAS are shown in Figure 6 for a Manta mine that is observed over a ninety degree interval with ten degree increments between observations (nine observation points at 1 deg, 11 deg, ..., 89 deg). As expected, ART-SAS yields a better image at the expense of the increased processing time that is associated with multiple iterations.

Conventional image sharpening techniques such as high pass filtering via a Laplacian operator and/or lateral inhibition are prevalent in biological systems [61-63], but seem to be unnecessary for top-down, bottom-up gradient descent.

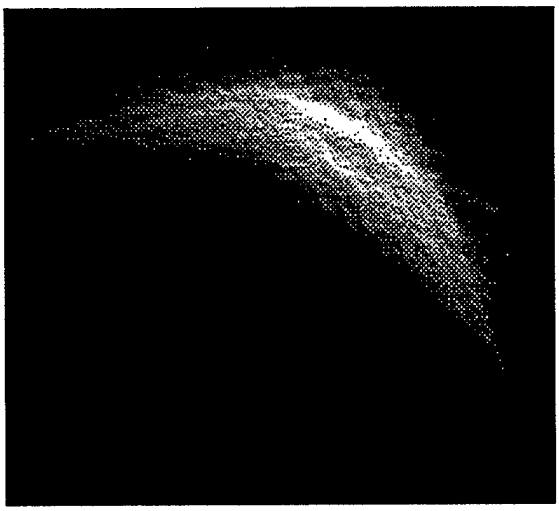


Figure 6. Manta with actuator, high frequency signal, observations at ten degree intervals over 90 degrees (nine observation points), no shadow compensation.
Top: Bionic SAS back projection image. Bottom: Bionic SAS ART image.

7. Biologically inspired techniques for accelerating a gradient descent algorithm and finding a global minimum - associative gradient descent

The iterative operations in (4)-(6) can be accelerated via parallel processing. All of the elements of the A-matrix, for example, can be updated simultaneously by a parallel implementation of (4) for all (k,l) values, as in an ideal gradient computation. A global minimum can be pursued in spite of local minima by using the simplex method [59] or a genetic algorithm [26], which is similar to the simplex technique. For a given local minimum, an associative memory can suggest other solutions (other A-matrices or interpretations of the same data) that may correspond to a smaller, global minimum.

Perceptual alternation (e.g., the Necker cube) and visual illusions suggest that the brain uses prior knowledge and other information to speed up the iteration process and to organize sparse image data into a picture that is commensurate with the viewer's experience. A relatively small number of observations is used to hypothesize the final, high-resolution image, and this image hypothesis is introduced into the top-down data representation [27]. Multiple hypotheses can be introduced simultaneously via parallel processing. An image hypothesis can be generated from prior expectations and an associative memory that is activated by a smeared (low resolution) image, resonance phenomena, echo time-frequency distributions, data from nonacoustic sensors, and other data that are not directly associated with imaging. A model for an associative memory is a k-nearest neighbor classifier in feature space [28], or a neural network that can be trained to make the correct association for various versions of an incomplete image. All kinds of relevant information can be inserted into an ART-like tomographic SAS imaging algorithm in order to accelerate the iterative process and to obtain a global minimum. This process might be called "associative gradient descent."

If the predicted echo incorporates hypotheses about multiple propagation paths, then the resulting comparator is part of a RAKE or matched field receiver. A RAKE receiver correlates input data with the expected version of the data (e.g., with a predicted target echo that is passed through a multipath channel) [29]. A "matched field" receiver performs the same operation at multiple receiving sites (the locations of physical or synthetic array elements). The correlation process can be implemented as part of a mean-square error computation. For energy normalized echoes and models, a receiver that is equivalent to a correlator can be obtained by squaring the difference between received and predicted echo data $echo_m(r,\theta) - mdl_m(\underline{\underline{A}};r,\theta)$ and integrating the squared error over range and cross-range coordinates.

8. Top-down, bottom-up gradient descent and inverse filtering

Top-down, bottom-up gradient descent can be applied in the time-frequency plane to yield a process that is equivalent to inverse filtering (deconvolution of echoes with respect to the transmitted signal). In this case, the A-matrix, which represents the high-resolution, top-down internal model, is a sampled version of the hypothesized target impulse response. The top-down cochlear representation is formed by convolving the hypothesized target impulse response with the transmitted signal and passing the resulting echo through a cochlear model. This top-down version of the cochlear output is compared with the cochlear response to the actual echo. The comparison is used to improve the high resolution internal model of the target impulse response via the LMS algorithm.

The cochlear time-frequency representation involves nonlinear processing, and a major advantage of the LMS algorithm is that it easily handles nonlinearities. The only inconvenience with respect to a nonlinearity is that the partial derivative $\partial mdl(\underline{\underline{A}};r,\theta)/\partial a_{kl}$ in (6) may be difficult to evaluate analytically, and may require empirical evaluation via the difference equation

$$\partial mdl(\underline{\underline{A}};r,\theta)/\partial a_{kl} \approx [mdl(\underline{\underline{A}}^{+}_{kl};r,\theta) - mdl(\underline{\underline{A}};r,\theta)]/\varepsilon \tag{7}$$

where $\underline{\underline{A}}^{+}_{kl}$ is the same as $\underline{\underline{A}}$ except that a small increment ε has been added to element kl. For energy normalized cochlear representations, calculating the mean-square error is equivalent to correlating the cochlear representation of the echo with a reference function that is based on a hypothesized target model (spectrogram correlation).

The relation between matched and inverse filtering can be understood by considering a filter that forms a minimum mean-square error estimate of the target impulse response from an echo time series. If the prior estimate of the target transfer function (the Fourier transform of the unknown target impulse response) is $H_{est}(f)$, the estimating filter has transfer function [30,31]

$$V(f) = \frac{E\{|H_{est}(f)|^2\}U^*(f)}{\Delta N(f) + E\{|H_{est}(f)|^2\}|U(f)|^2} \quad . \tag{8}$$

where $U(f)$ is the Fourier transform of the transmitted signal, Δ is the expected duration of the target impulse response, and $N(f)$ is the noise power spectral density. Two approximations to the right-hand side of (8) are

$$V(f) \approx [U(f)]^{-1} \quad \text{if signal-to-noise ratio (SNR) is large} \tag{9}$$

and

$$V(f) \approx \left[\frac{E\{|H_{est}(f)|^2\}}{\Delta N(f)}\right] U^*(f) \quad \text{if SNR is small.} \tag{10}$$

If SNR is large, then the target impulse response (which has been modeled as a projection of the reflectivity distribution onto the range axis) is estimated with an inverse filter, i.e., a filter that deconvolves the transmitted signal from the echo. If SNR is small, if the noise is white, and if there is no prior information about the target transfer function, then the target impulse response is estimated by a filter that is matched to the transmitted signal, i.e., a filter with transfer function proportional to $U^*(f)$. The filter in (8) performs pulse compression regardless of SNR.

9. Top-down, bottom-up gradient descent and pulse compression with spectrograms

Top-down, bottom-up gradient descent processing can be used to solve a controversial modeling problem in animal echolocation: How can animals implement a pulse compression filter with ~1cm range resolution and an interaural processor with time difference acuity of ~7μs when the traditional model of the auditory system uses a bank of bandpass filters followed by envelope detectors with relatively long integration times? The envelope detectors square-and-smooth (energy detect) filter outputs over ~260 μs for dolphins [32,33], corresponding to a range interval of 20 cm in water (Figure 7). For bats, the integration interval is approximately ~400 μs [34], corresponding to a range interval of 7 cm in air.

To solve the modeling problem, it is necessary to show that pulse compression can occur via operations on time-frequency (spectrogram) representations such as the one in Figure 7. This possibility can be demonstrated by considering the following relation between spectrograms and magnitude-squared cross-ambiguity functions [15]:

$$\iint S_{echo,filter}(t,f)S_{signal,filter}(t+\tau,f)dtdf = \iint |\chi_{signal,echo}(t,f)|^2 \, |\chi_{filter,filter}(t+\tau,f)|^2 \, dtdf \tag{11}$$

where $S_{echo,filter}(t,f)$ is the observed echo spectrogram in Figure 7 and "filter" refers to a baseband version of the bandpass filters that are shown in the figure. $S_{signal,filter}(t,f)$ is the observed spectrogram of the transmitted signal. The desired echo representation for a semicoherent, Doppler tolerant SAS processor is the envelope detected signal-echo cross-correlation function

$$|R_{signal,echo}(t)| = |\chi_{signal,echo}(t,0)| \tag{12}$$

which is known if $|\chi_{signal,echo}(t,f)|^2$ is known. The magnitude-squared filter auto-ambiguity function $|\chi_{filter,filter}(t,f)|^2$ is a smooth function of time and frequency that presumably is known to the animal.

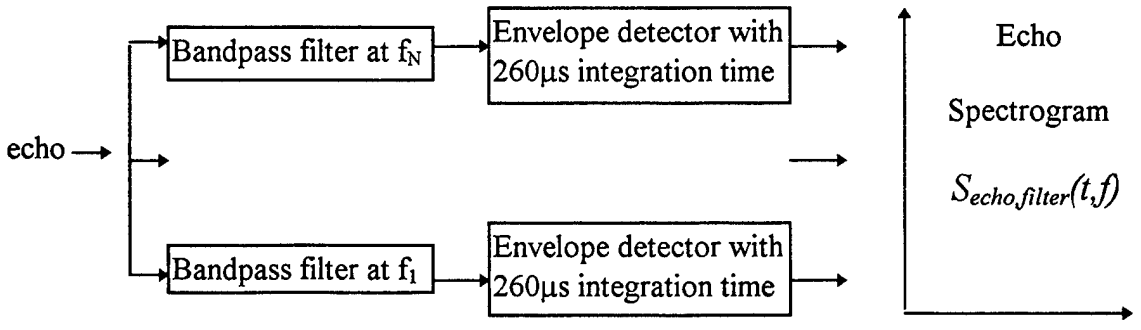


Figure 7. Traditional model of the peripheral auditory system at frequencies above five kHz, using integration times corresponding to the critical interval in dolphins.

Pulse compression can thus occur if (11) can be solved for $|\chi_{signal,echo}(t,f)|^2$, where all the other functions in the integral equation are assumed to be observable or known a priori.

The solution $|\chi_{signal,echo}(t,f)|^2$ can be obtained from (11) by using gradient descent to find the signal-echo magnitude-squared cross-ambiguity function that minimizes the mean-squared error

$$MSE = \int \left[\iint S_{echo,filter}(t,f) S_{signal,filter}(t+\tau,f) dt df - \iint |\hat{\chi}_{signal,echo}(t,f)|^2 |\chi_{filter,filter}(t+\tau,f)|^2 \, dt df \right]^2 d\tau$$

$$= \int error^2(\tau) d\tau. \qquad (13)$$

The LMS update equation is

$$\left[|\hat{\chi}_{signal,echo}(t,f)|^2 \right]_{n+1} = \left[|\hat{\chi}_{signal,echo}(t,f)|^2 \right]_n - \mu \frac{\partial error^2(\tau)}{\partial |\hat{\chi}_{signal,echo}(t,f)|^2}$$

$$= \left[|\hat{\chi}_{signal,echo}(t,f)|^2 \right]_n + 2\mu [error(\tau)] |\chi_{filter,filter}(t+\tau,f)|^2, \qquad (14)$$

iterated over all t,f, and τ values. This equation can be implemented with a top-down, bottom-up process as illustrated in Figure 8.

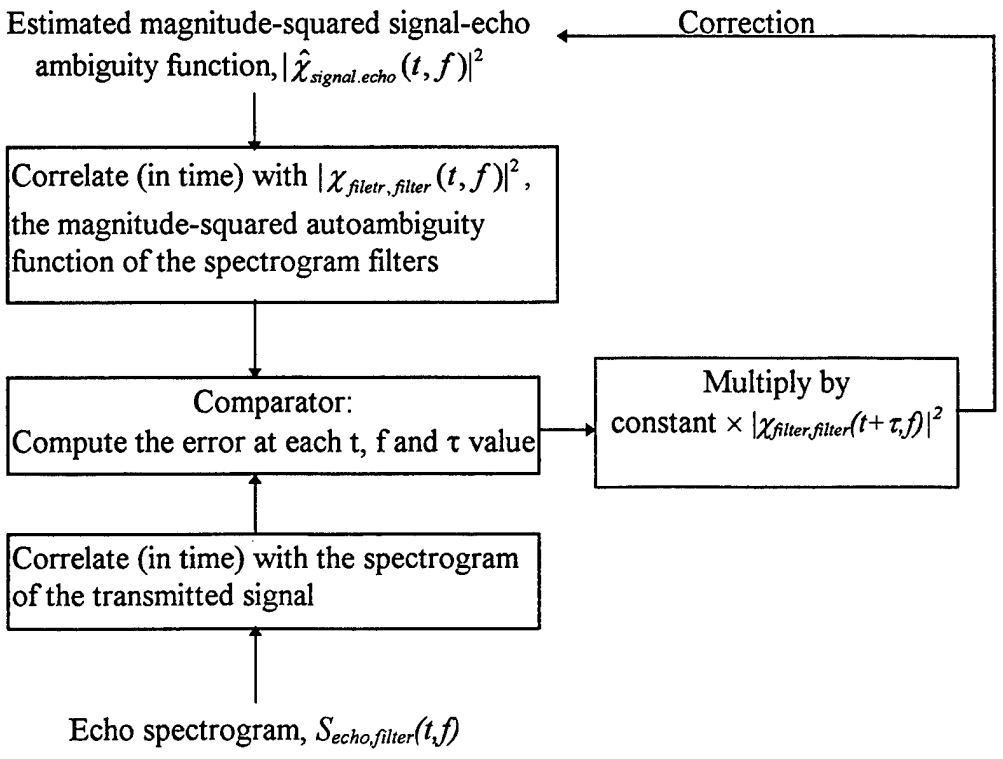


Figure 8. Pulse compression via top-down, bottom-up processing of the echo spectrogram.

A binaural version of the processor in Figure 8 estimates interaural delay by substituting the echo at the second ear for the signal, as illustrated in Figure 9. The processor correlates the spectrogram from ear 1 with the spectrogram from ear 2 in order to estimate the cross correlation function magnitude corresponding to the *input* signals at the two ears, *before* the signals are transformed into spectrograms. Spectrogram correlation implies cross correlation of pairs of envelope detected bandpass filter outputs from the two ears at each frequency channel. This cross correlation operation can be approximated by coincidence detectors that operate on the neuronally coded signals in two delay lines, one from each ear. This operation corresponds to the Jeffress-Licklider model for binaural localization [35], which has been neurologically verified in the barn owl by Konishi [36].

Figure 9 suggests that the Jeffress model may be incomplete at high frequencies; it should perhaps be augmented by a bottom-up, top-down process that sharpens the estimate of interaural delay by converting the spectrogram cross-correlation function into the magnitude-squared cross-ambiguity function of the two input signals. Such augmentation applies to frequencies that are too high for auditory neurons to carry phase information. Since bats are sensitive to echo phase shifts [7,8], phase information in bat auditory neurons may deteriorate only at frequencies above 20 kHz.

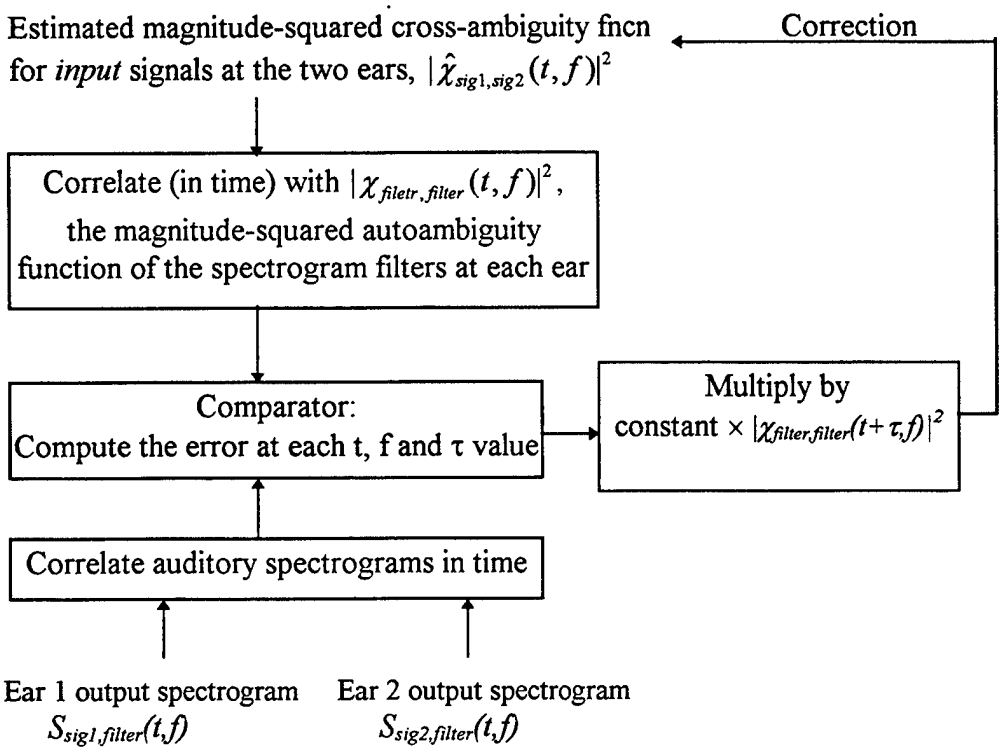


Figure 9. Interaural delay estimation via top-down, bottom-up processing of echo spectrograms. Interaural delay is obtained by cross-correlating the *input* signals at the two ears. Cross correlation of spectrograms, which are the output signal representations at the two ears, is part of the processor, but top-down, bottom-up iteration is also used.

At frequencies that are sufficiently low for transmission of phase information along the auditory nerve, the spectrogram model in Figure 7 can be used without the envelope detectors. The spectrogram without envelope detection is

$$s_{signal,filter}(t,f) = \int signal(x)filter(t-x)\exp[j2\pi f(t-x)]dx \qquad (15)$$

where filter(x) is the impulse response of the baseband filter function that is used to form the spectrogram. Cross-correlation of phase-sensitive spectrograms from the two ears yields

$$\iint s_{signal1,filter}(t,f)s^{*}_{signal2,filter}(t+\tau,f)dtdf = E_{filter}R_{signal1,signal2}(\tau) \qquad (16)$$

where E_{filetr} is the energy of the filter impulse response. When the cross correlation in (16) is synthesized by first time-correlating the outputs of each filter pair at a given frequency and then summing the results over filter pairs at different frequencies, the processor can be implemented with a Jeffress model.

Equations (11-16) pertain to spectrogram analysis with constant-bandwidth filters. The corresponding equations for proportional-bandwidth filters (which are more biomimetic) are given in Appendix B.

10. Top-down, bottom-up ART-SAS processing and beam deconvolution

Dolphin head scanning behavior suggests that SAS may be augmented by beam deconvolution. This deconvolution process can be included in an ART-type gradient descent algorithm. A high resolution internal model is convolved or smeared with the known beam patterns and is then compared with incoming multi-beam data to generate corrections to the model. A set of overlapping beam patterns can be generated by a binaural system that implements multiple direction-of-arrival hypotheses in parallel. Head scanning generates extra independent observations by changing the cross-range distribution of transmitted power. Binaural representations can be incorporated into an ART process by predicting the data at each ear.

11. Tracking and motion compensation with top-down, bottom-up ART-SAS processing

Hypotheses about translational and rotational motion of the sonar and the target can be incorporated into an ART processor. The best motion hypothesis corresponds to the least error between the echo data and the prediction. The best motion hypothesis can be used to track the target, to predict its location and orientation at the next echo, and to characterize body motion in a fish or wing beats in a bat. This process is a form of image-based tracking, which will be discussed in Section 18.

12. Top-down, bottom-up gradient descent with low resolution neuronal maps

In Figures 8 and 9, the data representation at the comparator is a smeared (low resolution) version of the echo spectrogram, not a sharpened one. Similarly, the inferior colliculus is used for integration of acoustic, visual, and somatosensory data via spatially registered, low resolution maps [37]. Top-down, bottom-up processing can use such low resolution maps to create a high resolution image at a higher level of the brain.

A basic question in neurophysiology is how the locations of point stimuli (e.g., in the retina) can be inferred from the collective discharge of a neuronal population [38]. The use of a low resolution map to improve a high resolution internal model via top-down, bottom-up gradient descent provides an answer to this question. The collective discharge is a consequence of divergent connections from various sensor elements, and it is analogous to a low resolution map. In this case, the locations of the point stimuli can be estimated with top-down, bottom-up stochastic gradient descent, using the LMS algorithm. To implement this process, the estimator must have a sufficiently accurate ensemble-average model of the divergent process (the smearing function) that maps a single point into a collective discharge. A payoff for such divergent stimulus coding is that neighboring sensors with overlapping excitation curves contribute to the representation of a point stimulus. The gradient descent deconvolution process can use these extra observations to obtain a more accurate stimulus representation than can be obtained from a single sensor.

13. Top-down, bottom-up gradient descent and hyperacuity

The gradient descent deconvolution process provides a mechanism for hyperacuity [39]. Hyperacuity involves sensitivity to a small difference between a reference stimulus and another stimulus, e.g., two parallel line segments on a vernier scale, which may be colinear or slightly displaced (—|— vs. —| —). For sensitivity to small differences, the mean-squared error in (1) can be changed to mean absolute error:

$$MAE(\underline{\underline{A}}) = (2\pi MR_{\max})^{-1} E\{\sum_{m=1}^{M} \int_0^{2\pi} \int_0^{R\max} |echo_m(r,\theta) - mdl_m(\underline{\underline{A}};r,\theta)| dr d\theta\}. \qquad (17)$$

In this case, the LMS update equation becomes

$$a_{kl}(n+1) = a_{kl}(n) - \mu \frac{\partial AE(\underline{\underline{A}};r,\theta,m)}{\partial a_{kl}} \quad \text{for all } k,l \text{ values} \qquad (18)$$

where

$$\frac{\partial AE(\underline{\underline{A}};r,\theta,m)}{\partial a_{kl}} = -\text{sgn}[echo_m(r,\theta) - mdl_m(\underline{\underline{A}};r,\theta)][\partial mdl_m(\underline{\underline{A}};r,\theta) / \partial a_{kl}]. \qquad (19)$$

In (19), sgn(error) equals one if error>0 and minus one if error<0. The gradient changes its sign but not its magnitude when the error becomes positive rather than negative, even for extremely small absolute error values. This behavior is analogous to computing the difference between the responses of two tuned neurons with slightly displaced and extremely steep tuning curves, where the actual stimulus value is midway between the best stimulus values for the two neurons. Even more sensitivity to small errors can be introduced by using the p^{th} root (p=2,3,...) of the absolute error value in (17). In this case, the algorithm must be designed to cope with unbounded values of the partial derivatives in (18) when the error approaches zero.

14. Feature images

A simplified, conservative model for biological SAS forms an image by noncoherent summation of all echo samples corresponding to each scattering point as the sonar moves. This simplification allows the formation of images that represent features other than reflectivity as a function of location. Volume clutter feature images are insensitive to reflectivity, and represent a measure of the number of small, Rayleigh scatterers in a pixel-sized volume element. Rough/smooth feature images are sensitive to the variation of reflectivity as the aspect changes.

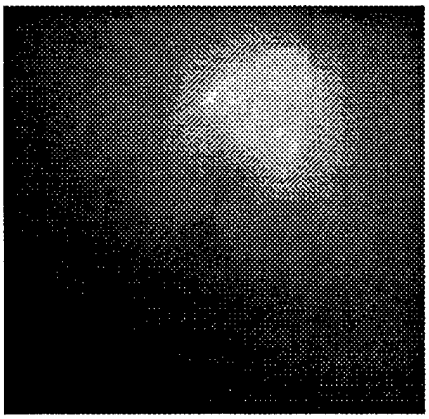
15. The volume clutter feature image

A volume clutter feature image is obtained by detecting Rayleigh scatterers in an amplitude-normalized version of the pulse compressed echo [40]. Amplitude normalization eliminates conventional reflectivity as a target feature. The volume clutter feature image is bright when a pixel contains many Rayleigh scatterers, e.g., bubbles or particles that are small relative to a wavelength. If the small scatterers are displaced by a comparatively large target, the volume clutter feature image becomes dark. A comparatively large target with very small backscatter cross section shows up as a target-shaped hole in the volume clutter feature image.

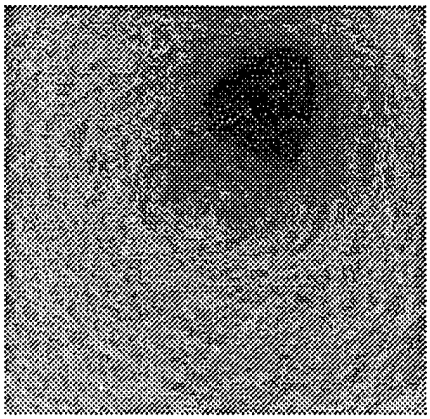
Figure 10 (top, left) shows a SAS reconstruction of a Rockan mine, without the adaptive thresholding that is usually applied to suppress clutter in the final image. The Rockan is a low-cross section target (for back-scattered sound) that resembles a large, trapezoidally shaped clam shell. The target is surrounded by volume clutter consisting of small bubbles and particulate matter in the lake water where the measurements were made. The volume clutter appears as a hazy, fog-like image surrounding the target. The top, right part of Figure 10 shows a volume clutter feature image constructed from the same echo data as in the top, left image. In this image, the target is suppressed and the volume clutter is accentuated. Since the image has high resolution, the target appears as a hole or cavity with a distinctive shape. The bottom, left part of Figure 10 shows an enhanced target image that is obtained by multiplying the clutter feature image by a constant and subtracting it from the reflectivity image. The bottom, right part of Figure 10 shows an enhanced clutter image that is obtained by multiplying the reflectivity image by a constant and subtracting it from the clutter feature image.

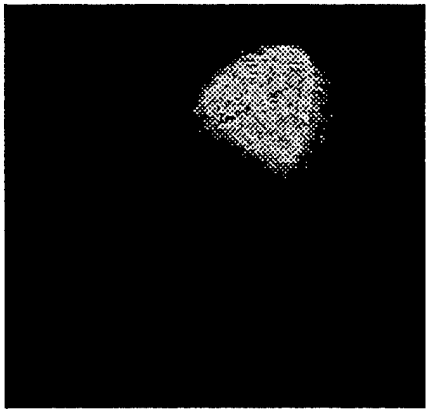
If the volume clutter were to become more reflective relative to the target, the image at the top, left in Figure 10 would become nearly uniform, and detection/classification of the target with a reflectivity image would become very difficult. The clutter feature image on the top, right of Figure 10, however, would become even better at revealing the presence and shape of the target. *Target detection with a clutter feature image is not predicted by the sonar equation,* although this type of detection is familiar to radiologists who work with medical ultrasound. (In medical ultrasound, organs such as the gall bladder and other structures are sometimes identified as dark shapes that are defined by the surrounding clutter echoes or "speckle.") The sonar equation is a logarithmic version of the signal-to-interference ratio at the receiver output, where the "signal" is associated with target reflectivity and the "interference" is associated with clutter echoes and noise [41]. Detection with a reflectivity image is predicted by the sonar equation, but *detection with a volume clutter feature image is predicted by the inverse of the sonar equation* (the clutter-to-signal ratio). Detection and classification with a volume clutter feature image as in Figure 10 may explain how dolphins can find buried fish [42].

SAS Reflectivity Image

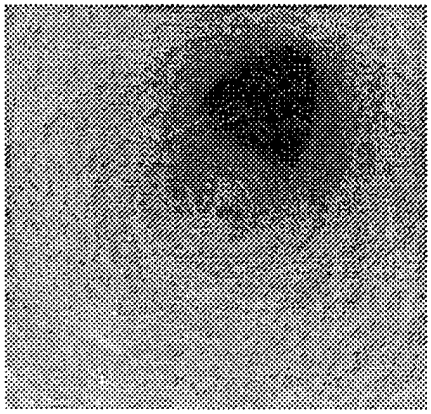


Volume Clutter Feature Image





Reflectivity image -
(weight × feature image)



Feature image -
(weight × reflectivity image)

Figure 10. Top: A SAS reflectivity image and a volume clutter feature image of the same low-reflectivity target. Bottom: Images constructed from weighted sums of the reflectivity and feature images in order to enhance or suppress the target relative to the volume clutter. The images shown here were constructed from echoes obtained over 360 degrees of rotation with 3.7 degree increments between echoes.

16. The rough/smooth feature image

Another feature image can represent smooth or rough reflectors. For smooth and rough feature images, the sum in the delay-and-sum SAS beam former is replaced by a quantity that depends upon the aspect-dependent variation of the echo samples that contribute to the sum. Pixels with large aspect-dependent amplitude variation are associated with smooth surfaces that have large back-scatter reflections at a few aspect angles and weak reflections at other angles. Pixels with small aspect-dependent variation are associated with small, isotropic scattering points that are found on rough surfaces.

Figure 11 shows composite rough/smooth bioSAS feature images for four different mines: Rockan, Manta, VEMS, and M0-8. The smooth-target feature image and the rough-target feature image compete for representation of each pixel in Figure 11. Pixel values in the composite image represent the larger of the two feature images at each sampling point. Smooth surface pixels are colored blue and rough surface pixels are red. Because the Rockan is designed to suppress back-scattered specular reflections, the rough-surface feature image dominates the Rockan image, making it almost totally red. The Manta resembles a truncated upright cone with a rough plate (the actuator) on the top planar surface. A SAS reflectivity image shows the outer shell with no indication of the rough plate; the interior appears to be hollow. A smooth-target feature image also shows only the outer shell. A rough-target feature image, however, shows only the rough plate. The ability to "see" the actuator and to use it for target recognition is greatly enhanced by using the rough-smooth composite color feature image.

BioSAS feature images, like more conventional bioSAS reflectivity images, degrade gracefully when the aspect sampling interval becomes large. Large aspect sampling intervals correspond to high area coverage rate. Figure 12 shows a composite rough/smooth feature image of a Manta mine, obtained with an aspect sampling interval of twenty degrees over an aspect observation interval of 360 degrees. The image is thus constructed with only eighteen echoes.

17. Motion-based feature images

Other feature images can be sensitive to motion. Semi-coherent processing of HFM/LPM (hyperbolic frequency modulated, linear period modulated) bat-like signals, for example, can be used to estimate acceleration from frequency shifts [14], and these acceleration estimates can be represented by an image. Motion-based images may be valuable for motion compensation to obtain a more accurate SAS image, for separating a moving target from surrounding clutter, or for discriminating moving clutter from stationary targets (e.g., seaweed that shifts with surge in shallow water).

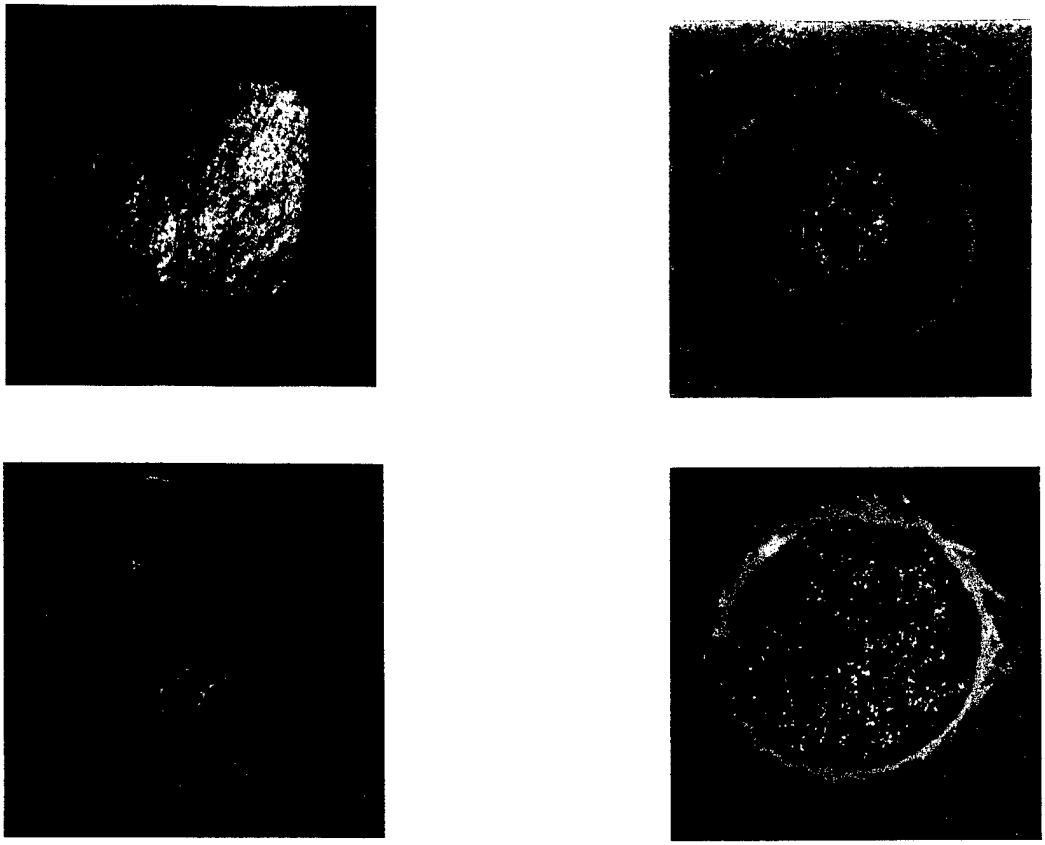


Figure 11. Composite color images of four mines with smooth scatterer image in blue and rough scatterer image in red (3.7 deg aspect sampling intervals over 360 deg).
Top left: Rockan. Top right: Manta. Lower left: VEMS. Lower right: M0-8.

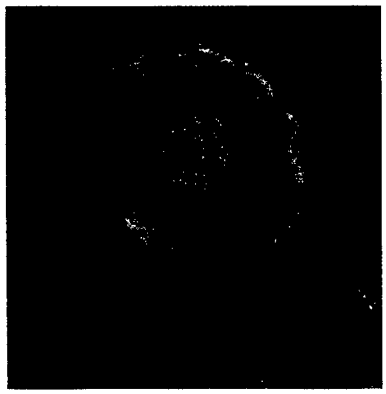


Figure 12. A composite smooth/rough BioSAS feature image obtained from echoes measured at 20 degree intervals over a 360 degree observation interval (18 different aspect angles). Smooth surfaces are blue; rough surfaces are red. The images are constructed with echoes from a Manta mine, using a dolphin-like signal. The echoes are from the SPAWAR/ARL data set.

18. Image-based tracking

A realistic acoustic imaging model for a biological sonar system must allow for freedom of sensor and target (prey) motion. One way to achieve this goal is to use the images themselves to compensate for deviations between actual motion and predicted motion [43]. A different motion compensation method relies on cross-correlation between overlapping echoes.

A theoretical counter-example to cross-correlation for image registration occurs for convex targets with a single specular reflection at each aspect angle. For the target in Figure 1, different area elements on the target surface have maximum reflectivity as they become orthogonal to the propagation direction. The echo cross-correlator response is maximized by moving the effective center of rotation so that the single, large echo at each aspect always corresponds to the same image pixel, even though the echoes are from different parts of the target surface. A similar phenomenon has been encountered at the Naval Coastal Systems Station (CSS) in Panama City, FLA. CSS has found that a SAS echo cross-correlation motion compensation system breaks down when there are an insufficient number of small scatterers to yield an accurate echo cross-correlation measure.

Another counter-example for cross-correlation processing occurs when there is a large aspect difference between "looks." In this case, the echoes from a random array of point scatterers will become decorrelated because of the aspect change, and a motion compensation method that relies on echo cross correlation will again deteriorate. Large aspect differences are not expected for conventional SAS because of area coverage rate constraints (avoidance of spatial undersampling), but wide-band bioSAS can form images from observations with large aspect differences.

Image-based tracking has been introduced in the context of ART processing (Section 11), but it also can be used for delay-and-sum (back projection) SAS imaging. A delay-and-sum SAS processor can construct an image sequentially, by adding the appropriate sample from the latest echo to a sum of sample values from previous echoes (one sample from each echo). Each of these echo samples presumably corresponds to the range of a specific point scatterer, as seen from different sensor locations. Hypothesized motion is represented by delay corrections that are inserted into the delay-and-sum SAS beam former. For the most recent echo, different motion hypotheses result in different "test images." The best test image corresponds to the best motion hypothesis.

A criterion for choosing the best test image is obtained from the SAS point spread function or range, cross-range ambiguity function (RCRAF) in Section 4 (Figure 2). The back projection SAS image is the convolution of the SAS RCRAF with the actual reflector distribution. In the frequency domain, the 2D or 3D Fourier transform of the RCRAF is multiplied by the 2D or 3D Fourier transform of the actual reflector distribution. The best image is obtained when the Fourier transform of the SAS RCRAF is constant, i.e., when the SAS RCRAF most resembles an impulse.

If the most recent delay in the SAS delay-and-sum process is incorrect, a line that should go through the center of the asterisk in Figure 1 is displaced. The resulting SAS RCRAF is less impulse-like, and its mean-square bandwidth is reduced. Since the Fourier transform of the SAS image is the product of the actual reflectivity distribution with the Fourier transform of the RCRAF, the mean-square bandwidth of the SAS image is reduced when the image is formed with an incorrect version of the most recent delay. These observations imply that the best test image has the largest mean-square bandwidth. The mean-square image bandwidth can be computed from a

test image, without using a multidimensional Fourier transform. If u(x,y) is a test image, then its mean-square bandwidth is

$$B_u^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[|\partial u(x,y)/\partial x|^2 + |\partial u(x,y)/\partial y|^2\right]dxdy \Big/ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|u(x,y)|^2\,dxdy \ . \qquad (20)$$

The bandwidth is a measure of variation or image sharpness.

A block diagram of an image-based tracker is shown in Figure 13. Delay corrections are utilized by a dynamic model that predicts the next delay hypotheses. This model can include sensor motion and translation/rotation of various parts of a moving target [43]. The image-based tracker has been demonstrated by using it to compensate for accidental range deviations during acquisition of echoes from a target that was suspended from a rotating platform by thin lines. The range deviations are caused by wobbling of the suspended target. Image-based tracking has been tested further by deliberately introducing a spurious step discontinuity in range halfway through the imaging process. The results of these demonstrations are shown in Figures 14 and 15.

19. Motion compensation, dynamic programming, and dynamic models

Motion variability of a single point scatterer can be modeled as distortion of the predicted range vs time curve of the scatterer. If the hypothesized delay history of echoes from the point scatterer is adaptively modified to match the distortion, the resulting motion compensation is equivalent to adaptive focusing with a multi-pulse matched filter.

A similar distortion problem arises in speech recognition. A phoneme in speech data is often a time warped version of a reference function corresponding to the same phoneme. The problem of classifying the distorted phoneme has been solved by dynamic programming [44]. This method is an efficient technique for sequential implementation of a likelihood function and is similar to the Viterbi algorithm for decoding communication signals [45,46]. The need for such efficiency can be appreciated by considering the number of different trajectories that are implied by a simple three alternative model; the target delay (or the delay of a phoneme sample) at a given sampling time is the same as the predicted delay or is slightly larger or smaller than the predicted value. Figure 16 illustrates that there are three admissible trajectories (or phoneme time distortions) at time t_1. Nine different admissible trajectories or time warps terminate on the five nodes at t_2, twenty-seven such paths exist at t_3, and 3^n admissible trajectories reach the nodes at time t_n. After n echoes or time samples, correlation of the delay history corresponding to each admissible trajectory with echo data and computation of the maximum correlator output will result in a receiver that is focused on an isolated point scatterer, despite the unpredictable delay perturbations. Unfortunately, 3^n different reference functions are needed to form testable hypotheses for unpredictable perturbations (time warps) after n echo time samples.

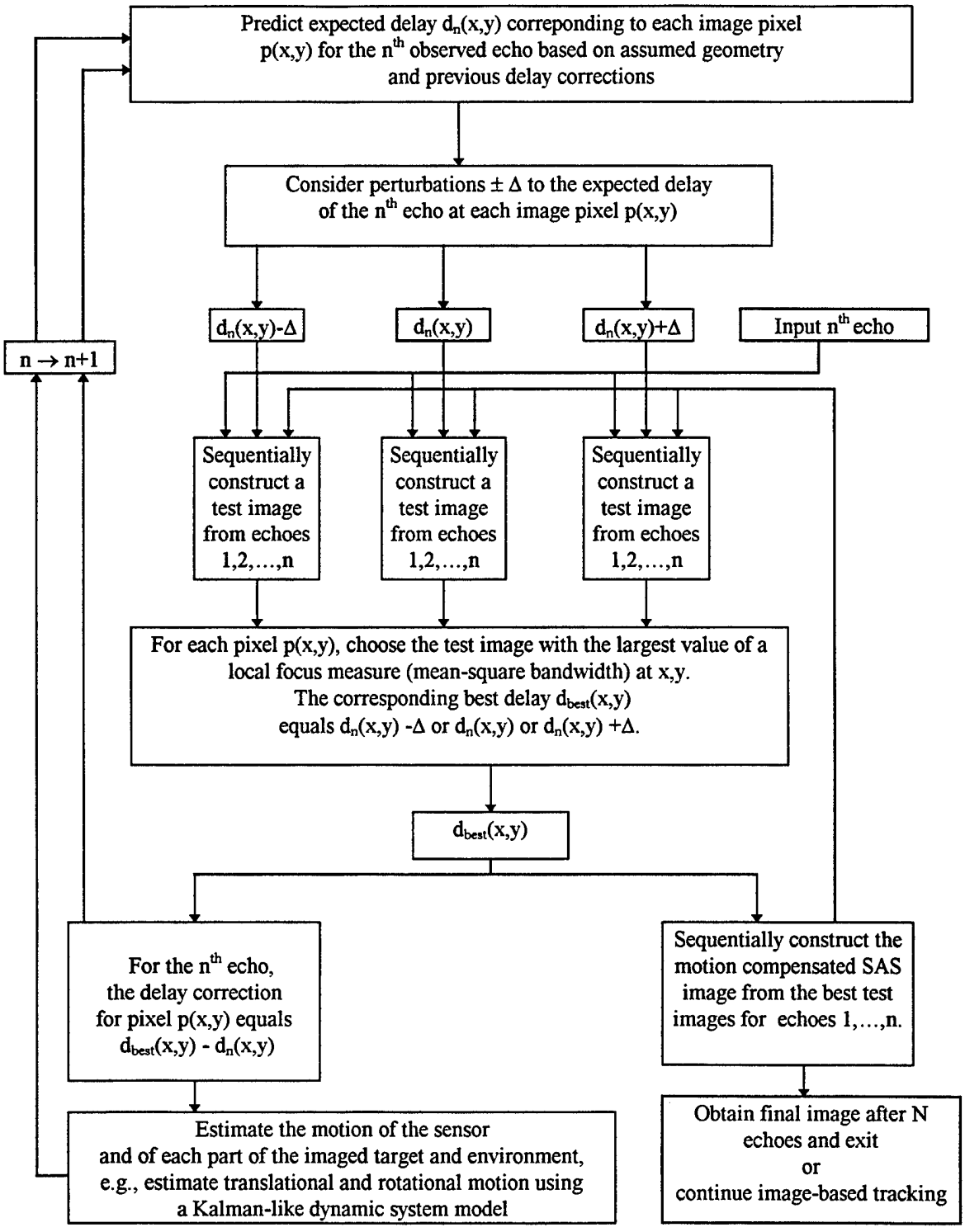
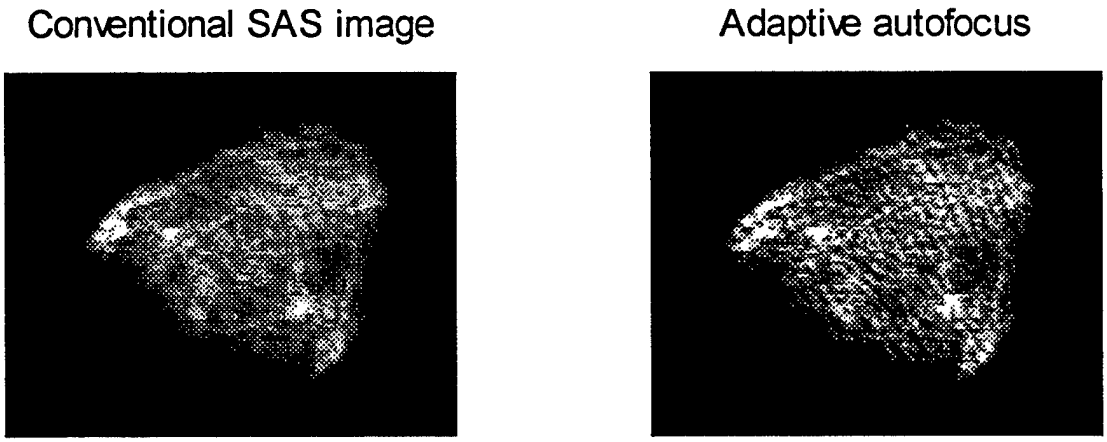


Figure 13. Sequentially formed test images indicate how well delays have been predicted and what corrections should be applied. This flow diagram shows adaptive focusing operations for target/sensor motion compensation, image-based tracking, and SAR/SAS imaging of maneuvering targets.

Figure 14. Sequential adaptive focusing of original echo data using test images. Wobble of the suspended target introduces a small range deviations.

(a) Images obtained by applying conventional and adaptive focusing



(b) Adaptive range corrections applied to original echo data for best focus of target centroid. Different range corrections may be applied to other parts of the target.

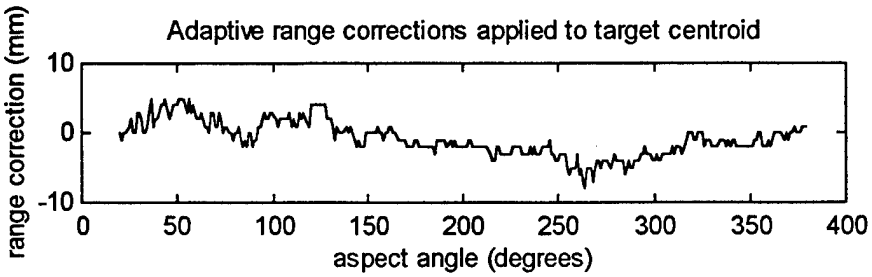
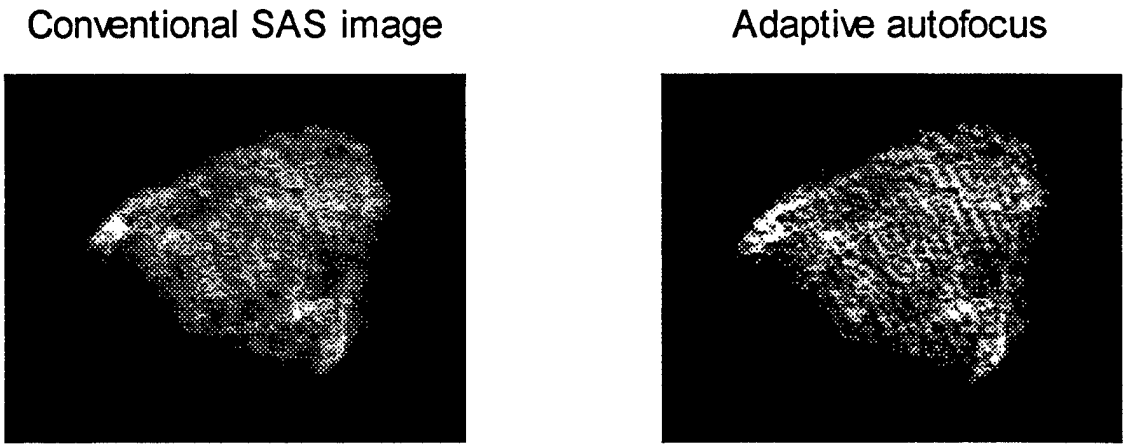
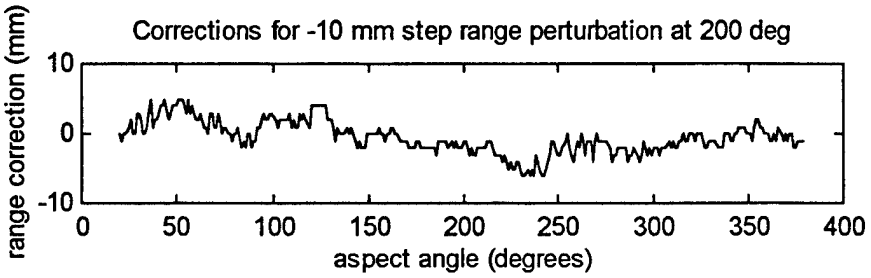


Figure 15. Sequential adaptive focusing of perturbed echo data using test images. Unpredicted sensor or target motion is simulated by artificially reducing range by 10 mm after the target is rotated by 200 degrees.

(a) Images obtained by applying conventional and adaptive focusing.



(b) Adaptive range corrections applied to echo data for best focus of the target centroid when range is artificially reduced by 10 mm after the target is rotated by 200 degrees. Different range corrections may apply to other parts of the target.



Computational complexity is greatly reduced by realizing that only the cumulative maximum correlator response is required at any given time. Suppose that each node in Figure 16 at time t_n is labeled with the maximum cumulative correlator response for all paths connecting to the node. For a given node, this maximum is the incremental increase in the correlator response for the range perturbation represented by the node, plus the largest cumulative correlator response for all nodes at the preceding sample time t_{n-1} that are connected to the node. For an isolated point scatterer, the maximum cumulative correlator response at time t_n denotes the best model for the perturbed trajectory up to that time. Since the perturbations are admissible distortions in the hypothesized range vs time function for a moving point target, the maximum correlator response is proportional to the reflectivity of the point target.

The trajectory of a given target point is of interest in order to determine what kind of maneuver was inferred by the adaptive focusing algorithm and to focus on reflecting points at intevals that are shorter than the overall processing time. This history can be estimated by tracing back from the node with largest cumulative correlator response at time t_n to the connected node with largest cumulative correlator response at time t_{n-1}. The trace-back is continued by finding the connected node with largest cumulative correlator response at time t_{n-2}, etc.

For targets with multiple scattering points, the dynamic programming algorithm should not attempt to maximize cumulative correlator response, since a strongly reflecting point may be ephemeral as the target rotates. A more reliable performance measure is the peak-to-sidelobe ratio of the SAS range, cross-range ambiguity function or point spread function, which increases with the mean-square image bandwidth in (20). The image formation process is sequential, with each new transmitted pulse adding another increment to each pixel. The image after n pulses is the target reflectivity distribution convolved with an asterisk-shaped point spread function (RCRAF) with peak-to-sidelobe level equal to n. As this asterisk becomes more impulse-like with higher P/S, the image becomes less smeared, and the mean square bandwidth increases. The sharpness measure in (20) is thus expected to increase monotonically with n. This monotonic increase of the performance measure is required for dynamic programming [47,48]. A dynamic programming solution to the delay perturbation problem can use a mean-square bandwidth measure to represent the efficacy of delay compensation, and trajectories of target points can be chosen to maximize this image sharpness measure.

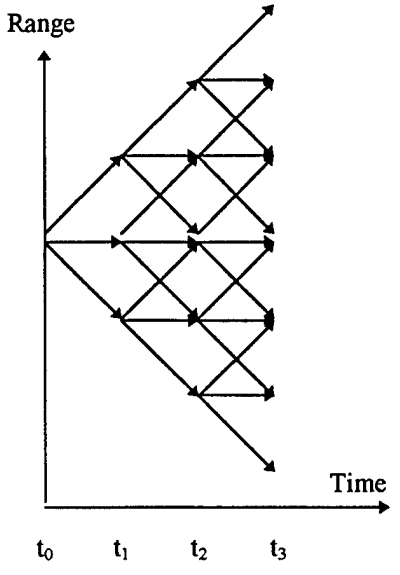


Figure 16. Three admissible perturbations of the predicted range at each echo time-of-arrival yield a large set of admissible range vs. time functions.

The size of a test image should be as large as possible to provide a reliable measure of mean-square bandwidth. For maximum stability of the algorithm, the mean-square bandwidth focus measure should be computed over the whole target. If the target is rigid (or behaves as though it were rigid over the observation interval) the delay perturbation at any target point can be computed from the perturbations of pitch, roll, and yaw, i.e., from unpredictable rotations about the three axes that pass through the target center. Instead of three test images corresponding to no change in predicted delay, an admissible delay increment, and an admissible delay decrement, the algorithm must now compute twenty seven test images, corresponding to all possible combinations of no change, an admissible increment, and an admissible decrement of the predicted pitch, roll, and yaw angles. In each image, different target points have different delay corrections. If the target is contained within a cube of volume elements (voxels), then the hypothesized delay perturbation of each target point or voxel can be computed from the perturbations of pitch, roll, and yaw angles. Corrections to predicted pitch, roll, and yaw angles can be inferred from the test image with the largest mean-square bandwidth.

A motion compensation decision at each node in Figure 16 can be obtained by computing different test images corresponding to proposed local delay corrections or equivalent roll, pitch, and yaw corrections. The test image with the largest mean-square bandwidth corresponds to the best local delay corrections, or the best proposed pitch, roll, and yaw correction. *This criterion is based on the assumption that the 2D target distribution is the same for all test images. The only difference between the test images is the shape of the point spread function that is used to estimate them.* The image after n echoes is sequentially constructed from the best test images for echoes 1, ..., n.

The interpretation of delay perturbations in terms of pitch, roll, and yaw is convenient because these angles can be used as state variables in a dynamic model that describes target behavior. One version of the state equations is given by (21) for the x, y, and z components of the range to the target centroid and by (22) for 3-D target rotations. Corrections to the predicted state vector in (21)-(22) are obtained from evaluation of the mean-square bandwidth of test images, and the corrected state vector is used to predict the next observation as in Kalman filtering [49,50]. The dynamic model improves the prediction of individual pixel delays by computing these delays in the context of overall target motion.

$$\begin{bmatrix} R_x \\ R_y \\ R_z \\ (d/dt)R_x \\ (d/dt)R_y \\ (d/dt)R_z \end{bmatrix}_n = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} R_x \\ R_y \\ R_z \\ (d/dt)R_x \\ (d/dt)R_y \\ (d/dt)R_z \end{bmatrix}_{n-1} \tag{21}$$

$$\begin{bmatrix} pitch \\ roll \\ yaw \\ (d/dt)pitch \\ (d/dt)roll \\ (d/dt)yaw \end{bmatrix}_n = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} pitch \\ roll \\ yaw \\ (d/dt)pitch \\ (d/dt)roll \\ (d/dt)yaw \end{bmatrix}_{n-1} . \quad (22)$$

20. Image models for motion compensation

When dynamic programming is applied to speech recognition, admissible time warps correspond to constraints on the discrete delay choices in Figure 16. These admissible time warps are chosen in order to maximize the correlation of time-warped data with each phoneme in a set of reference phonemes. A classifier picks the reference phoneme that has maximum correlation with admissibly time-warped data. This process can be adapted to imaging by combining focusing with classification. Instead of using mean-square bandwidth to evaluate test images that correspond to different proposed range and rotation corrections, the test images are correlated with reference templates. The updated range and rotation estimates correspond to the energy normalized test image that has the largest correlation with a given reference image. A different sequence of admissible range and rotation corrections may be obtained for different reference templates. Correlation of each final, focused version of the image with the corresponding reference template is used to classify the target and to identify the range and rotation corrections that best describe the motion perturbations that occured during the imaging process. This modeling approach is similar to associative gradient descent in ART-SAS, where comparisons between predicted and observed echo data are used for tracking as well as for image synthesis.

21. Image-based tracking and the 2D target distribution invariance assumption

The utilization of test image bandwidth to make a decision at each node in Figure 16 and to choose the best test image at each iteration in Figure 13 assumes that the 2D target distribution in Figures 1 is the same for each test image at a given aspect angle. This assumption is circumvented when the SAS imaging algorithm is given the freedom to make local adjustments to the predicted delay of each image point. For a 3D target, parts of the target will lie above and/or below the image plane that is determined by target rotation relative to the sonar in Figure 1. The freedom to adjust the predicted delay for motion compensation can be interpreted as the freedom to move outside the image plane in Figure 1 by including elevation (z) changes. A sharpness measure such as mean-square bandwidth could be increased by moving outside the designated image plane as well as by finding the best delays to maximize the peak-to-sidelobe ratio of the point spread function. The delay-corrected image will have large mean-square bandwidth and will appear to be well focused, but *it may not be identical to the 2D image that is formed without delay corrections, even when no corrections are needed.*

The monotonic increase of mean-square bandwidth in (20) with the number of pulses is not affected by focusing on points that are outside the original image plane, since such focusing occurs when the mean-square bandwidth is larger than in the original image plane. On a new 2D image surface, the algorithm can still place all the line segments in Figure 1 so as to intersect at a point,

yielding a sharp point spread function as in Figure 2. Defocusing via smearing of the point spread function may be tolerated, however, in order to exploit an image on a different constant-z plane that has particularly large bandwidth when observed from a new aspect angle. If this situation occurs, then the estimated delay of each point, considered by itself, becomes an unreliable indicator of target motion at the point. The relation between the delay perturbation of a point and target motion at the point is known provided that delay perturbations are contained in the plane of target rotation. The relation becomes inaccurate when the estimated delay has an unknown out-of-plane z-component. This problem can be mitigated by using a 3D motion model as in (21)-(22) to hypothesize the delay of each point and to form the corresponding test images.

Although very little motion compensation is needed in Figure 14, application of the image-based tracking algorithm has a large effect on the SAS image. This effect seems to be associated with the fact that the target is not planar; different parts of the target lie at different elevations. The algorithm automatically chooses a delay correction that corresponds to the elevation which maximizes the local focus criterion. Adaptive focusing extracts a maximally focused two-dimensional image from three dimensional target data. Conventional SAS is restricted to a single range-azimuth plane without the ability to vary elevation for best focus.

In Figure 15, at an aspect angle of 200 degrees, the target suddenly "lurches" toward the sensor via a simulated delay purturbation, causing an unpredicted range decrease of 1 cm for all succeeding echoes. Comparison of the conventional SAS images in Figs. 14 and 15 indicates that the conventional image is smeared by the uncompensated motion, although the target is still recognizable. As expected, the adaptively focused images in Figs. 14 and 15 are nearly identical with and without the simulated motion perturbation of the sensor. The delay corrections in Figure 15b, however, should represent the ten mm delay step that was used to deliberately degrade the data at 200 degrees, if delay corrections are assumed to occur in the original image plane. Comparison of Figures 14b and 15b, however, indicates that only part of the delay step is accounted for by delay compensation in Figure 15. This inconsistency is caused by the limited step size of the admissible delay correction at each observation, and by the fact that 3D delay corrections do not necessarily correspond to delays in the original 2D image plane.

22. Summary of adaptive focusing and motion compensation via image-based tracking

Mean-square image bandwidth is a measure of the peak-to-sidelobe ratio of the asterisk-shaped point spread function (the range, cross-range ambiguity function) of a SAS imaging system. This measure can be combined with a dynamic programming approach that evaluates test images in order to sequentially correct for delay perturbations. The corresponding algorithm has been tested with wideband sonar data. Improvements and modifications of the basic algorithm include (i) a dynamic model for prediction of position-dependent delays and for maneuver estimation from test images and (ii) a SAS analogue to a time-warped speech classifier, where the SAS processor evaluates delay corrections in the context of various reference templates. Echo cross-correlation as in traditional motion compensation methods can be included in the focusing criterion, although counter-examples indicate that cross-correlation should not be used exclusively. The performance of the algorithm in multipath and other adverse conditions has not yet been tested, but further improvements are possible to cope with such conditions. For example, an animal can change depth so as to minimize multipath, and it can roll on its side so as to use binaural processing as an adaptive null-steering mechanism for multipath removal.

Other applications of the method include

(i) A SAR/SAS processor that can focus on moving targets;
(ii) Tracker design to avoid confusion caused by targets with scintillating multiple highlights or scattering centers; (An image-based tracker exploits these highlights to form a target image that is consistent with apparent changes in highlight locations.)
(iii) For medical ultrasound, large-array adaptive focusing in nonhomogeneous media and motion compensation for tracking and imaging blood flow and heart valves;
(iv) Large-array adaptive focusing in nonhomogeneous media for seismic geophysical prospecting and imaging of land mines.

23. Biological feasibility of simplified SAS

A Doppler tolerant, semicoherent, tomographic SAS model is comparatively simple to implement and thus appears to be feasible for animal sonar. For dolphins, the model can be completely noncoherent; *echoes from a given scattering point are noncoherently summed as the animal moves.* Massively parallel processing can be used to obtain similar noncoherent sums from many neighboring points, thus forming an image. For bats, a pulse compression operation is necessary, e.g., an operation as in (8) that may be equivalent to matched filtering or inverse filtering. A deconvolution process which is equivalent to inverse filtering can be implemented by applying top-down, bottom-up gradient descent to cochlear echo representations, as shown in Section 9. For semicoherent SAS, pulse compression is followed by envelope detection. After pulse compression, the bat model is the same as for dolphins.

Evidence already exists for the capability of echolocating animals to form a noncoherent sum of echoes from a given scattering point, although more experiments are needed. Echo summation or integration capability can be inferred from dolphin target recognition experiments [51]. Summation also can be inferred from neurophysiological experiments on bats [52] in which the excitation threshold of a neuron is decreased by repeated stimulation of the neuron. If the neuron is modeled as a sequential likelihood ratio test for a particular stimulus in Gaussian noise [53], then the prior probability of the stimulus increases monotonically with the sum of preceding stimuli. For a Bayesian hypothesis test, an increase in the prior probability of the stimulus is associated with a decrease of the excitation threshold [54]. Echo summation is thus encoded as a decrease in neuronal threshold. The ability to concentrate on the same point in space as the animal moves is implied by range tracking neurons in bats [55,56].

As the animal moves, the ability to average the echo from each point seems to imply that a separate neuron is assigned to each image pixel, as in a topographic neuronal map like the ones found in the superior colliculus [37]. The relatively poor resolution of the superior colliculus map can be improved by a top-down, bottom-up gradient descent sharpening process, as discussed in Section 6. *This improvement is not manifested in the map itself, but in an interpretation of the map by a higher processing center.*

An important counter-argument for the existence of biological SAS is that high resolution topographic neuronal maps of reflectivity as a function of range and direction have yet to be discovered at higher processing centers in echolocating animals. There are several possible explanations for this missing observation:

1. Insufficient dimensionality may exist because of amplitude maps. Amplitopic representations have been found in the bat auditory cortex. A neuron's best amplitude varies monotonically with its physical position relative to other amplitude-specific neurons [55]. If an amplitopic representation is used for reflectivity in an acoustic image, then a map of reflectivity as a function of range, azimuth, and elevation requires four dimensions. Such a map cannot be constructed topographically with a three dimensional neuronal array. The brain is forced to use either a non-topographic representation or multiple maps that are different projections of a higher dimensional representation. Range, azimuth, and elevation, for example, may be coded non-topographically as additional constraints to neuronal excitations within a population of amplitopically organized neurons.
2. An advantage of a topographic map is that stimulus representations can be sharpened by localized lateral inhibition. Range/angle sharpening, however, also can be implemented via top-down, bottom-up gradient descent. Topographic maps are thus convenient but not necessary for obtaining better resolution.
3. Images may be dynamically coded, such that a neuronal topographic map represents changes in an acoustic image rather than the image itself. Pulse-to-pulse jitter in the range and cross-range coordinates of a point scatterer, for example, may be represented topographically, while a point that does not move or scintillate between transmissions may not be represented. Range-rate and rate of angular change may be represented by ordered neuronal maps.
4. A relevant map may be associated with (or projected onto) a different sensory modality such as vision or somatosensation. This type of projection or association is suggested by facial sensations that are experienced by blind people [57]. In this case, a visual or somatosensory map with comparatively high resolution is the top-down part of a top-down, bottom-up gradient descent sharpening process. Comparisons between predictions and data are made in a low resolution representation with registered spatial maps from different sensors, such as the superior colliculus.

24. Coherent SAS

At present, there is insufficient evidence to conclude that bats or dolphins can perform coherent pulse-to-pulse integration, which is necessary for conventional, coherent SAS. Nevertheless, there is some motivation to consider coherent SAS. Bats are sensitive to a constant, frequency-independent echo phase shift [7,8], which is necessary but not sufficient for coherent SAS. Coherence implies that the SAS range, cross-range ambiguity function (RCRAF) is a coherent sum of rotated versions of the physical RCRAF of the sonar. In this case, the sonar RCRAF can be designed such that the SAS RCRAF has very high peak-to-sidelobe ratio, even with large angle increments (a small number of aspect angles and echoes). The required physical RCRAF corresponds to a particular set of signals that must be measured at various positions around the sonar transmitter. These signals resemble waveforms that have been measured around an echolocating dolphin [18,58].

25. Summary and Conclusion

A biological version of synthetic aperture imaging appears to be feasible, and it is advantageous for detection and classification of prey (e.g., finding buried fish). The simplifications and generalizations that are used to obtain a biological model lead to new insights and capabilities for man-made SAS systems. These capabilities include generalized trackers for range-extended targets, new kinds of acoustic images that represent various target features, the capability to obtain acoustic images with observations from relatively few aspects, and an associative gradient descent model that utilizes prior information or hypotheses for fast convergence to a global minimum.

26. References

[1] Altes, R.A. 1971. Optimum waveforms for sonar velocity discrimination. Proceedings of the IEEE, 39:1615-1617.

[2] Altes, R.A. and D.P. Skinner. 1977. Sonar velocity estimation with a linear-period modulated pulse. Journal of the Acoustical Society of America, 61: 1019-1030.

[3] Brown, W.M. and L.J. Porcello, 1969. An introduction to synthetic aperture radar. IEEE Spectrum, 6(Sept):52-62.

[4] Cutrona, L.J. 1975. Comparison of sonar performance achievable using synthetic aperture techniques with the performance achievable by more conventional means. Journal of the Acoustical Society of America, 58:336-348.

[5] Wehner, D.R. 1987. High resolution radar. Artech House, Norwood, MA.

[6] Menne, D. and Hackbarth,H. 1986. Accuracy of distance measurement in the bat *Eptesicus fuscus*: Theoretical aspects and computer simulations. J.Acoust. Soc. Amer. 79:386-397.

[7] Menne, D., I.Kaipf, I.Wagner, J.Ostwald, and H.U. Schnitzler. 1989. Range estimation by echolocation in the bat *Eptesicus fuscus*: Trading of phase versus time cues. Journal of the Acoustical Society of America, 85:2642-2650.

[8] Simmons, J.A., M. Ferragamo, C.F.Moss, S.B.Stevenson, and R.A.Altes. 1990. Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: The shape of target images in echolocation. Journal of Comparative Physiology A 167:589-616.

[9] Munson, D.C. Jr., J.D. O'Brien, and W.K. Jenkins. 1983. A tomographic formulation of spotlight-mode synthetic aperture radar. Proceedings of the IEEE, 71:917-925.

[10] Macovski, A. 1983. Medical imaging systems. Prentice-Hall, Englewood Cliffs, NJ.

[11] Altes, R.A., P.W. Moore, and D.A. Helweg. 1998. Tomographic image reconstruction of MCM targets using synthetic dolphin signals. Technical Document 2993. SPAWAR Systems Center, San Diego, CA and Defense Technical Information Center, Fort Belvoir, VA.

[12] Altes, R.A., and E.L. Titlebaum. 1970. Bat signals and optimally Doppler tolerant waveforms. Journal of the Acoustical Society of America, 48:1014-1020.

[13] Altes, R.A. 1973. Some invariance properties of the wide-band ambiguity function. Journal of the Acoustical Society of America, 53:1154-1160.

[14] Altes, R.A. 1990. Radar/sonar acceleration estimation with linear-period modulated, hyperbolic frequency modulated (LPM/HFM) waveforms. IEEE Transactions on Aerospace and Electronic Systems, 26:914-923.

[15] Altes, R.A. 1980. Detection, estimation, and classification with spectrograms. Journal of the Acoustical Society of America, 67:1232-1246.

[16] Saillant, P.A., J.A. Simmons, S.P. Dear, and T.A. McMullin. 1993. A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver. J. Acoust. Soc. Amer. 94:2691-2712.

[17] Altes, R.A. 1979. Target position estimation in radar and sonar, and generalized ambiguity analysis for maximum likelihood parameter estimation. Proceedings of the IEEE, 67: 920-930.
[18] Altes, R.A. 1995. Signal processing for target recognition in biosonar. Neural Networks, 8:1275-1295.
[19] Freedman, A. 1962. A mechanism of acoustic echo formation. Acustica, 12:10-21.
[20] Altes, R.A. 1976. Sonar for generalized target description and its similarity to animal echolocation systems. Journal of the Acoustical Society of America, 59:97-105.
[21] Anderson, J.R. 1980. Cognitive psychology and its implications. W.H. Freeman and Company, San Francisco.
[22] Capuano, U. and J.T. McIlwain. 1981. Reciprocity of receptive field images and point images in the superior colliculus of the cat. The Journal of Comparative Neurology, 196:13-23.
[23] Widrow, B. and S.D. Stearns. 1985. Adaptive signal processing. Prentice-Hall, Englewood Cliffs, NJ.
[24] Honig, M.L. and D.G. Messerschmitt. 1984. Adaptive filters. Kluwer Academic Publishers, Boston.
[25] Rumelhart, D.E., G.E. Hinton, and R.J. Williams. 1986. Learning internal representations by error propagation. Pp. 318-362, *in* Parallel distributed processing, vol. 1 (D.E. Rumelhart and J.L.McClelland, eds.). The MIT Press, Cambridge, MA, 547 pp.
[26] Fogel, D.B. 1995. Evolutionary computation. IEEE Press, New York, 272 pp.
[27] Gregory, R.L. 1973. Eye and brain. McGraw-Hill, New York, 251 pp.
[28] Cover, T.M. and P.E. Hart. 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13:21-27.
[29] Proakis, J.G. 1989. Digital communications, 2nd ed. McGraw-Hill, New York.
[30] Turin, G.L. 1957. On the estimation in the presence of noise of the impulse response of a random, linear filter. IRE Transactions on Information Theory, 3:5-10.
[31] Altes, R.A. 1977. Estimation of sonar target transfer functions in the presence of clutter and noise. Journal of the Acoustical Society of America, 61:1371-1374.
[32] Vel'min, V.A. and Dubrovskiy, N.A. 1976. The critical interval of active hearing in dolphins. Sov. Phys. Acoust. 2:351-352.
[33] Au, W.W.L. 1993. The sonar of dolphins. Springer-Verlag, New York.
[34] Simmons, J.A., Freedman, E.G., Stevenson, S.B., Chen, L., and Wohlgenant, T. 1989. Clutter interference and the integration time of echoes in the echolocating bat, *Eptesicus fuscus*. Journal of the Acoustical Society of America 86:1318-1332.
[35] Jeffress, L.A. 1948. A place theory of sound localization. Journal of Comparative Physiology and Psychology 41:35-39.
[36] Konishi, M. 1993. Listening with two ears. Scientific American 268:66-73.
[37] Drager, U.C. and D.H. Hubel. 1975. Responses to visual stimulation and relationship between visual, auditory, and somatosensory inputs in mouse superior colliculus. Journal of Neurophysiology 38:690-713.
[38] McIlwain, J.T. 1986. Point images in the visual system: New interest in an old idea. Trends in Neuroscience, 9:354-358.
[39] Altes, R.A. 1989. Ubiquity of hyperacuity. J. Acoust. Soc. Amer. 85:934-942.
[40] Altes, R.A. 1998. Feature imaging and adaptive focusing for synthetic aperture processor. Patent application. US Patent Office Serial No. 09 222 635.
[41] Urick, R.J. 1975. Principles of underwater sound. McGraw-Hill, New York.
[42] Herzing, D. 1999. To appear in Biological Sonar Systems. J. Thomas, C. Moss, and M. Vater, eds. Proceedings of the Biological Sonar Conference, May-June 1998, Carvoeiro, Portugal.

[43] Altes, R.A. 1998. Adaptive ISAR focusing of distributed time-varying targets. Pp. 261-272 *in* Radar processing, technology, and applications III, Volume 3462 (W.J. Miceli, ed.). Society of Photo-Optical Instrumentation Engineers, Bellingham, WA.
[44] H. Sakoe and S. Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust., Speech, and Signal Process. ASSP-26:43-49. Reprinted in Automatic Speech and Speaker Recognition (N. R. Dixon and T.B. Martin, eds.) IEEE Press, New York, 1979.
[45] Viterbi, A.J. 1967. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE Trans. on Information Theory IT-13:260-269.
[46] Gibson, J.D. 1993. Principles of Digital and Analog Communications, 2nd Ed., Sections 10.7 and 12.4. Macmillan, New York.
[47] Dreyfus, S.E. 1965. Dynamic Programming and the Calculus of Variations, Chapter 1. Academic Press, New York.
[48] Larson, R.E. and Casti, J.L. 1978 & 1982. Principles of Dynamic Programming, Parts I and II. Marcel Dekker, New York.
[49] Gelb, A. (ed.) 1974. Applied Optimal Estimation. MIT Press, Cambridge, MA.
[50] Sorenson, H.W. 1980. Parameter Estimation. Marcel Dekker, New York.
[51] Moore, P.W.B., H.L. Roitblat, R.H. Penner, and P.E. Nachtigall. 1991. Recognizing successive dolphin echoes with an integrator gateway network. Neural Networks 4:701-709.
[52] Grinnel, A.D. 1963. The neurophysiology of audition in bats: Resistance to interference. Journal of Physiology, 167:114-127.
[53] Altes, R.A. 1989. An interpretation of cortical maps in echolocating bats. Journal of the Acoustical Society of America, 85:943-952
[54] Van Trees, H.L. 1968. Detection, estimation, and modulation theory, part I. John Wiley & Sons, New York.
[55] Suga, N. and W.E. O'Neill. 1980. Auditory processing of echoes: Representation of acoustic information from the environment in the bat cerebral cortex. Pp. 589-611 *in* Animal Sonar Systems (R.-G. Busnel and J.F. Fish, eds.). Plenum Press, New York, 1135 pp.
[56] Wong, D., M. Maekawa, and H. Tanaka. 1992. The effect of pulse repetition rate on the delay sensitivity of neurons in the auditory cortex of the FM bat, *Myotis lucifugus*. Journal of Comparative Physiology A. Sensory, Neural, and Behavioral Physiology, 170:393-402.
[57] Rice, C.E. 1966. The human sonar system. Pp. 719-755 *in* Animal Sonar Systems, Biology and Bionics, vol. II (R.-G. Busnel, ed.). Laboratoire de Physiologie Acoustique, Jouy-en-Josas, France, 1233 pp.
[58] Altes, R.A. 1992. The line segment transform and sequential hypothesis testing in dolphin echolocation. Pp. 317-355 in Marine Mammal Sensory Systems (J.A. Thomas, R.A. Kastelein, and A. Ya. Supin, eds.). Plenum Press, New York, 773 pp.
[59] Nash, J.C. 1979. Compact numerical methods for computers. Adam Hilger Ltd., Bristol, UK.
[60] Pribram, K.H. 1971. Languages of the brain. Brooks/Cole, Monterey, CA.
[61] Marr, D. 1982. Vision. W.H. Freeman, San Francisco, CA.
[62] Kandel, E.R. 1985. Central representation of touch (Chapter 25) *in* Principles of Neural Science, 2nd ed. (E.R. Kandel and J.H. Schwartz, eds.) Elsevier, New York.
[63] Ratliff, F. 1965. Mach Bands. Holden-Day, San Francisco, CA.
[64] Vetterli, M. and Kovacevic, J. 1995. Wavelets and subband coding. Prentice-Hall, Englewood Cliffs, NJ.
[65] Altes, R.A. 1990. Wide-band, proportional bandwidth Wigner-Ville analysis. IEEE Transactions on Acoustics, Speech, and Signal Processing 38:1005-1012.

APPENDIX A: Back Projection and Synthetic Aperture Processing

In order to compare back projection with synthetic aperture imaging, it is helpful to review some properties of two-dimensional Fourier transforms. The first property is the expression for the 2D Fourier transform in cylindrical coordinates. In Cartesian coordinates, the 2D Fourier transform is

$$f(x,y) = (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\omega_x,\omega_y) \exp[j(\omega_x x + \omega_y y)] d\omega_x d\omega_y. \tag{A1}$$

In cylindrical (r,θ) frequency domain coordinates, $\omega_x = r\cos\theta$, $\omega_y = r\sin\theta$, and

$$F(\omega_x,\omega_y) = F(r\cos\theta, r\sin\theta) \equiv F_{cyl}(r,\theta). \tag{A2}$$

Cylindrical (ρ,ϕ) coordinates in the spatial domain are such that $x = \rho\cos\phi$, $y = \rho\sin\phi$, and

$$f(x,y) = f(\rho\cos\phi, \rho\sin\phi) \equiv f_{cyl}(\rho,\phi). \tag{A3}$$

Substituting (A2) and (A3) into (A1), changing variables , and noting that the Jacobian of the joint change of variables $\omega_x = r\cos\theta$, $\omega_y = r\sin\theta$ is

$$\begin{vmatrix} \partial\omega_x/\partial r & \partial\omega_y/\partial r \\ \partial\omega_x/\partial\theta & \partial\omega_y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = |r| \tag{A4}$$

yields

$$f_{cyl}(\rho,\phi) = (2\pi)^{-2} \int_{-\pi/2}^{\pi/2} \int_{-\infty}^{\infty} F_{cyl}(r,\theta) \exp[jr\rho(\cos\theta\cos\varphi + \sin\theta\sin\phi)] |r| dr d\theta. \tag{A5}$$

The desired expression for the 2D Fourier transform in cylindrical coordinates is obtained by using the identity

$$\cos\theta\cos\phi + \sin\theta\sin\phi = \cos(\theta - \phi) \tag{A6}$$

in (A5), which results in

$$f_{cyl}(\rho,\phi) = (2\pi)^{-2} \int_{-\pi/2}^{\pi/2} \int_{-\infty}^{\infty} F_{cyl}(r,\theta) \exp[jr\rho\cos(\theta - \phi)] |r| dr d\theta. \tag{A7}$$

Rotation of an image in the x,y plane corresponds to a similar rotation of the 2D Fourier transform of the image in the ω_x,ω_y plane. This property follows easily from (A7). Rotation in the x,y plane by γ radians transforms $f_{cyl}(\rho,\phi)$ to $f_{cyl}(\rho,\phi+\gamma)$. Replacing ϕ by ϕ+γ on the right hand side

of (A7) and changing variables by letting $\theta'=\theta-\gamma$, the right side becomes the 2D Fourier transform of $F_{cyl}(r,\theta+\gamma)$. It follows that

$$f_{cyl}(\rho,\phi+\gamma) \leftrightarrow F_{cyl}(r,\theta+\gamma) \tag{A8}$$

or

$$f(x\cos\gamma-y\sin\gamma, y\cos\gamma+x\sin\gamma) \leftrightarrow F(\omega_x\cos\gamma-\omega_y\sin\gamma, \omega_y\cos\gamma+\omega_x\sin\gamma) \tag{A9}$$

where the double arrow indicates a 2D Fourier transform pair.

Another property of 2D Fourier transforms is that the projection of f(x,y) onto the x-axis is the 1D inverse Fourier transform of $F(\omega_x,0)$. The projection of f(x,y) onto the x-axis is

$$P_0(x) \equiv \int_{-\infty}^{\infty} f(x,y)dy. \tag{A10}$$

Integrating

$$f(x,y) = (2\pi)^{-2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\omega_x,\omega_y)\exp[j(\omega_x x+\omega_y y)]d\omega_x d\omega_y$$

with respect to y yields

$$P_0(x) \equiv \int_{-\infty}^{\infty} f(x,y)dy = (2\pi)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\omega_x,\omega_y)\exp(j\omega_x x)\delta(\omega_y)d\omega_x d\omega_y$$

$$= (2\pi)^{-1}\int_{-\infty}^{\infty} F(\omega_x,0)\exp(j\omega_x x)d\omega_x \tag{A11}$$

which is the 1D inverse Fourier transform of $F(\omega_x,0)$. It follows that

$$\int_{-\infty}^{\infty} P_0(x)\exp(-j\omega_x x)dx = F(\omega_x,0). \tag{A12}$$

The above projection property can be generalized to rotated versions of f(x,y). Rotating f(x,y) by θ radians yields

$$f_\theta(x,y) \equiv f(x\cos\theta - y\sin\theta, y\cos\theta + x\sin\theta). \tag{A13}$$

It follows from (A9) that the 2D Fourier transform of $f_\theta(x,y)$ is a similarly rotated version of $F(\omega_x,\omega_y)$:

$$f_\theta(x,y) \leftrightarrow F_\theta(\omega_x,\omega_y) \equiv F(\omega_x \cos\theta - \omega_y \sin\theta, \omega_y \cos\theta + \omega_x \sin\theta). \tag{A14}$$

Letting $P_\theta(x)$ denote the projection of $f_\theta(x,y)$ onto the x-axis, i.e.,

$$P_\theta(x) \equiv \int_{-\infty}^{\infty} f_\theta(x,y)dy, \tag{A15}$$

(A10)-(A12) imply that the 1D Fourier transform of $P_\theta(x)$ is $F_\theta(\omega_x,0)$. Using r instead of ω_x,

$$\int_{-\infty}^{\infty} P_\theta(x)\exp(-jrx)dx = F_\theta(r,0) = F(r\cos\theta, r\sin\theta) \equiv F_{cyl}(r,\theta). \tag{A16}$$

The projection of a rotated version of the image f(x,y) can be Fourier transformed in one dimension to obtain the 2D Fourier transform of the image in cylindrical coordinates, evaluated along a constant-θ slice in the frequency domain. This result is known as the projection-slice theorem [9]. It implies that a sequence of projections of incrementally rotated images can be used to obtain a sequence of constant-θ slices of the 2D Fourier transform of the image in cylindrical coordinates. The image can be reconstructed from its projections by computing an inverse 2D Fourier transform in cylindrical coordinates, as in (A7). This form of reconstruction is known as back projection.

To obtain a more explicit expression for the reconstructed image in terms of its projections, (A16) is solved for $P_\theta(x)$ by taking the inverse 1D Fourier transform of both sides of the equation,

$$P_\theta(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} F_{cyl}(r,\theta)\exp(jrx)dr. \tag{A17}$$

A gradual high pass filter (similar to differentiation without the corresponding phase shift) can be applied to the projection, yielding

$$P_{\theta,HP}(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} F_{cyl}(r,\theta)\exp(jrx)|r|dr. \tag{A18}$$

The integral on the right hand side of (A18) is contained in the 2D inverse Fourier transform expression (A7) in the form

$$(2\pi)^{-1} \int_{-\infty}^{\infty} F_{cyl}(r,\theta)\exp[jr\rho\cos(\theta-\phi)]|r|dr = P_{\theta,HP}[\rho\cos(\theta-\phi)]. \tag{A19}$$

Substituting (A19) into (A7) yields

$$f_{cyl}(\rho,\phi) = (2\pi)^{-1} \int_{-\pi/2}^{\pi/2} P_{\theta,HP}[\rho\cos(\theta-\phi)]d\theta. \quad \text{(A20)}$$

The original image $f_{cyl}(\rho, \phi)$ can be reconstructed by summing high-pass filtered projections. To obtain the image f(x,y) in Cartesian coordinates, recall from (A3) that

$$f_{cyl}(\rho,\phi) \equiv f(\rho\cos\phi, \rho\sin\phi) = f(x,y)$$

and from (A6) that

$$\cos(\theta-\phi) = \cos\theta\cos\phi + \sin\theta\sin\phi..$$

It follows that (A20) can be written

$$f(\rho\cos\phi, \rho\sin\phi) = (2\pi)^{-1} \int_{-\pi/2}^{\pi/2} P_{\theta,HP}[(\rho\cos\phi)\cos\theta + (\rho\sin\phi)\sin\theta]d\theta \quad \text{(A21)}$$

or

$$f(x,y) = (2\pi)^{-1} \int_{-\pi/2}^{\pi/2} P_{\theta,HP}(x\cos\theta + y\sin\theta)d\theta\,. \quad \text{(A22)}$$

Back projection algorithms typically utilize the projection-slice theorem to obtain the 2D Fourier transform of the image in cylindrical coordinates. The image is then reconstructed via a 2D inverse Fourier transform operation. The equivalent expression in (A22), however, is useful for illustrating the similarity between back projection and synthetic aperture processing.

For radar/sonar/ultrasound processing, suppose that a transducer is placed on the negative x-axis. A target is rotated about the origin of the coordinate system, and the range is defined to be zero at the center of rotation. Projections of the target are obtained by rotating the target clockwise and recording reflectivity vs range (A-scan) data at each rotation after filtering to obtain an estimate of the target impulse response, e.g., matched filtering. The integration surfaces for each projection correspond to points with constant delay, e.g., spherical shells. The thickness of the integration surfaces or shells are determined by the range resolution cell of the system , i.e., by the system bandwidth.

To track a point on the target with initial position (x,y), the matched filtered echo (A-scan) from the target is evaluated at range $x\cos\theta+y\sin\theta$ as the target is rotated. Equation (A22) describes a sum of high pass, matched filtered echoes from the point on the target at initial position (x,y) as the target is rotated.

The same A-scan data can be obtained by moving the transducer in a circle around the target, or by using a large array of transducers that are arranged in a circle with the target at the center. The second alternative is an actual array, while the first is a synthetic array. The array is focused on a target point by delay-and-sum beam forming. Consider a transducer that is located on a circle, θ radians counterclockwise relative to the negative x-axis. A signal is transmitted toward the target from this transducer, and the resulting echo is received by the same transducer and matched filtered

or otherwise processed to estimate target impulse response. The contribution of this filtered transducer output to the beam former image of the target point at x,y is a sample of the matched filtered echo. This sample is chosen to correspond to the range of the target point, i.e., to a range of $x\cos\theta+y\sin\theta$ when range zero is at the center of the circle . The delay-and-sum beam former for the real or synthetic array approximates the integral in (A22) by a finite sum over a sequence of aspect (θ) values. Such a finite sum approximation is also used in back projection. If the matched filtered echoes are high pass filtered by using a filter with transfer function $|\omega|$, delay-and-sum synthetic aperture imaging and back projection are equivalent processes.

One way to exploit the equivalence of SAS and BP is to form a 3D image when the transducer is above the plane of rotation. If the transducer is located above the negative x-axis such that the line between the transducer and the origin forms an angle α relative to the negative x-axis, and if the target is in the far field of the transducer, then (A22) becomes

$$f(x,y,z) = (2\pi)^{-1} \int_{-\pi/2}^{\pi/2} P_{\theta,HP}[(x\cos\theta + y\sin\theta)\cos\alpha - z\sin\alpha]d\theta . \qquad \text{(A23)}$$

Appendix B: Cross correlation of proportional bandwidth spectrograms for wide-band signal processing

Proportional bandwidth spectrogram analysis is closely related to wavelet analysis. Two signals $u_1(t)$ and $u_2(t)$ may represent a received signal and a reference function or the input signals at the two ears. In either case, the first goal is to cross correlate the two signals when the signals are represented by phase-sensitive outputs of two proportional bandwidth spectrogram analyzers.

As in the narrow-band, constant-bandwidth case, the signal cross-correlation function can be obtained by cross-correlating the outputs of corresponding spectrogram filters in time and then summing the resulting cross-correlation functions over all the filters. The phase-sensitive spectrograms that are generated by proportional bandwidth filters are

$$s_n(t,\alpha) = \alpha^{-1/2} \int_{-\infty}^{\infty} U_n(\omega) V(\omega/\alpha) e^{j\omega t} d\omega \quad , \quad \text{n=1,2.} \tag{B1}$$

Cross correlating the two spectrograms in time at each α-value (cross correlating corresponding filter outputs) and summing over all the filters yields

$$\int_0^{\infty} d\alpha \int_{-\infty}^{\infty} dt[s_1(t,\alpha) s_2^*(t+\tau,\alpha)] = \int_{-\infty}^{\infty} U_1(\omega) U_2^*(\omega) [\int_0^{\infty} \alpha^{-1} |V(\omega/\alpha)|^2 \, d\alpha] e^{-j\omega\tau} d\omega \tag{B2}$$

where a change of variables in the α-integral yields

$$\int_0^{\infty} \alpha^{-1} |V(\omega/\alpha)|^2 \, d\alpha = \int_0^{\infty} \frac{|V(x)|^2}{x} dx \equiv C_v \tag{B3}$$

as in wavelet analysis [64]. It follows that

$$\int_0^{\infty} d\alpha \int_{-\infty}^{\infty} dt[s_1(t,\alpha) s_2^*(t+\tau,\alpha)] = C_v R_{u_1 u_2}(\tau) \tag{B4}$$

where

$$R_{u_1 u_2}(\tau) = \int_{-\infty}^{\infty} u_1(t) u_2^*(t-\tau) dt \tag{B5}$$

is the desired phase-sensitive cross-correlation function between the two input signals.

Another version of proportional bandwidth spectrogram analysis utilizes spectral distortion defined by

$$U_E(\omega) = U(e^{\omega}). \tag{B6}$$

Since *exp(ω)* varies between zero and infinity, it is assumed that *U(ω)* is analytic, and thus has support only for nonnegative frequencies. The mapping in (B6) automatically performs pulse compression on signals that have a logarithmic frequency domain phase function *exp[-jklog(ω)]*. The corresponding time signals before pulse compression have linear period modulation and hyperbolic frequency modulation, and they resemble the echolocation signals of many bats.

The phase-sensitive spectrogram for signal and filter functions that are spectrally distorted as in (B6) is [65]

$$s_{U_E V_E}(t, \log\alpha) = \int_{-\infty}^{\infty} U_E(\log\omega) V_E(\log\omega - \log\alpha) e^{jt\log\omega} d(\log\omega) \tag{B7}$$

where

$$U_E(\log\omega) = U(\omega) \tag{B8}$$

and

$$V_E(\log\omega - \log\alpha) = V(\omega/\alpha). \tag{B9}$$

Cross correlation of two such time-frequency representations yields

$$\int_{-\infty}^{\infty} d\log\alpha \int_{-\infty}^{\infty} dt [s_{U_{E1}V_E}(t, \log\alpha) s^*_{U_{E2}V_E}(t+\tau, \log\alpha)]$$

$$= \int_{-\infty}^{\infty} U_{E1}(\log\omega) U^*_{E2}(\log\omega) [\int_{-\infty}^{\infty} |V_E(\log\omega - \log\alpha)|^2 \, d\log\alpha] e^{-j\tau\log\omega} d\log\omega$$

$$= E_V R_{U_{E1}U_{E2}}(\tau) \tag{B10}$$

where E_V is the filter energy and $R_{U_{E1}U_{E2}}(\tau)$ is the cross correlation function of the spectrally distorted input signals.

The phase-insensitive, squared-envelope spectrogram is obtained by envelope detecting the outputs of the filter bank;

$$S_{U_E V_E}(t, \log\alpha) = |\int_{-\infty}^{\infty} U_E(\log\omega) V_E(\log\omega - \log\alpha) e^{jt\log\omega} d(\log\omega)|^2 . \tag{B11}$$

From (11), cross correlation of two magnitude-squared spectrograms yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{U_{E1}V_{E1}}(t,\omega)S_{U_{E2}V_{E2}}(t+\tau,\omega)dtd\omega = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\chi_{U_{E1}U_{E2}}(t,\omega)|^2 |\chi_{V_{E2}V_{E2}}(t+\tau,\omega)|^2 \, dtd\omega \tag{B12}$$

where $|\chi_{U_{E1}U_{E2}}(t,\omega)|^2$ is the magnitude-squared version of the narrowband ambiguity function

$$\chi_{U_{E1}U_{E2}}(t,\omega) = \int_{-\infty}^{\infty} U_{E1}(\omega')U_{E2}^{*}(\omega'-\omega)\exp(-j\omega' t)d\omega' . \tag{B13}$$

Solving (B12) for $|\chi_{U_{E1}U_{E2}}(t,\omega)|^2$ via the iterative procedure in Figure 8 yields

$$|\chi_{U_{E1}U_{E2}}(\tau,0)|^2 = |R_{U_{E1}U_{E2}}(\tau)|^2 , \tag{B14}$$

the squared envelope of the cross-correlation function of the spectrally distorted input signals. As in the constant-bandwidth case, cross correlation of proportional bandwidth spectrograms can be used for pulse compression.