OCLC FirstSearch: Display

Your requested information from your library GALLAUDET UNIV LIBR

Return

SHIPPED - Lender

134755

27800133

General Record Information

Request Identifier: 27800133 **Status:** SHIPPED
Request Date: 20070214 **Source:** ILLiad
OCLC Number: 1226792
Borrower: UAF **Need Before:** 20070331
Receive Date: **Renewal Request:**
Due Date: N/A **New Due Date:**
Lenders: *GQG, PLF, ABC, WFN, WLU
Request Type: Copy

Bibliographic Information

Call Number:
Title: The Journal of the Acoustical Society of America.
ISSN: 0001-4966
Edition: any
Imprint: [New York, etc.] American Institute of Physics for the Acoustical Society of America.
Article: Erbe, Kin, Yedlin, Farmer: Computer models for masked hearing experiments with beluga whales (Delphinapterus leucas).
Volume: 105
Number: 5
Date: May 1999
Pages: 2967-78
Verified: <TN:134755><ODYSSEY:206.107.42.144/ILL> OCLC

My Library's Holdings Information

LHR Summary: v.1-(1929-)
Lending Policies: Unknown / Unknown
Location: GQGV
Format: unspecified

Borrowing Information

Patron: Rosa, Cheryl
Ship To: Univ of Alaska Fairbanks/Rasmuson Library - ILL Rm. 320/P.O. BOX 756807/Fairbanks, Alaska 99775-6807

Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*)

Christine Erbe
Institute of Ocean Sciences, Acoustical Oceanography, 9860 W Saanich Rd., Sidney, British Columbia V8L 4B2, Canada

Andrew R. King
School of Earth Sciences, Macquarie University, Sydney, New South Wales 2109, Australia

Matthew Yedlin
University of British Columbia, Earth & Ocean Sciences, 2219 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada

David M. Farmer
Institute of Ocean Sciences, Acoustical Oceanography, 9860 W Saanich Rd., Sidney, British Columbia V8L 4B2, Canada

(Received 24 July 1998; accepted for publication 28 January 1999)

Environmental assessments of manmade noise and its effects on marine mammals need to address the question of how noise interferes with animal vocalizations. Seeking the answer with animal experiments is very time consuming, costly, and often infeasible. This article examines the possibility of estimating results with software models. A matched filter, spectrogram cross-correlation, critical band cross-correlation, and a back-propagation neural network detected a beluga vocalization in three types of ocean noise. Performance was compared to masked hearing experiments with a beluga whale [C. Erbe and D. M. Farmer, Deep-Sea Res. II **45**, 1373–1388 (1998)]. The artificial neural network simulated the animal data most closely and raised confidence in its ability to predict the interference of a variety of noise sources with a variety of vocalizations.
 [S0001-4966(99)00905-4]

PACS numbers: 43.80.Lb, 43.80.Nd, 43.60.Lq [WA]

INTRODUCTION

Over the past couple of years, public awareness of human impact on nature has steadily increased. Living in the era of decreasing animal diversity, public concern about protecting species from becoming extinct has led to a rapid increase in environmental assessments of various human activities. History has shown that we often act too late, i.e., when a dying species cannot be saved anymore. Understanding, foresight, and preventative action is therefore of utmost importance.

Long considered as vast, hence invulnerable, our world's oceans have experienced extensive human abuse posing threats to all marine life. In particular, the protection of whales and dolphins has recently earned unprecedented public interest. Threats to marine mammals include accidental or intended takings (killings); entanglement in debris or fishing gear; habitat destruction; water contamination due to industrial pollution, oil spills, toxic chemicals, waste and sewage; changes in water temperature and salinity; physical alteration of habitat during offshore construction; overfishing of prey; and underwater noise exposure. Since the beginning of the industrial revolution, the world's oceans have become increasingly noisy. Ship traffic, oil and mineral exploration, and offshore construction all contribute to the noise pollution of marine-mammal habitat.

Noise can have a variety of effects on marine mammals such as behavioral disturbance, physiological damage to their auditory system as well as other organs and tissues, and masking. Marine mammals rely primarily on acoustics for communication and orientation. Manmade noise, however, has the potential of interfering with animal communication signals, odontocete (toothed whale) echolocation signals, environmental sounds animals might listen to for orientation, and predator and prey sounds. An understanding of the extent of masking is crucial for the writing of regulations for industrial underwater noise emission.

A few studies have examined masking with cetaceans (whales and dolphins) both psychophysiologically[1–7] and electrophysiologically.[8,9] Most of these looked at high-frequency signal discrimination; only three[1,4,7] provided masked hearing data at frequencies below 10 kHz, where most of the industrial noise spectra prevail. Masking has generally been studied with the signal being a pure tone and the masker being either a pure-tone or broad-band (white) and temporally consistent noise. These studies provided valuable information on the basic characteristics of the animal auditory filter, such as the relationship between the amplitude, frequency, and frequency bandwidth of signal and masker.

The study by Erbe and Farmer[7] was different in that it presented a "holistic" approach to masking where the masked signal was a complex animal vocalization and the masker was structured noise. Figure 1 shows power density spectrograms of the sounds used. The signal was a typical 2-s beluga vocalization consisting of six pulses with frequency components between 700 Hz and 8 kHz. Bubbler

 0001-4966/99/105(5)/2967/12/$15.00

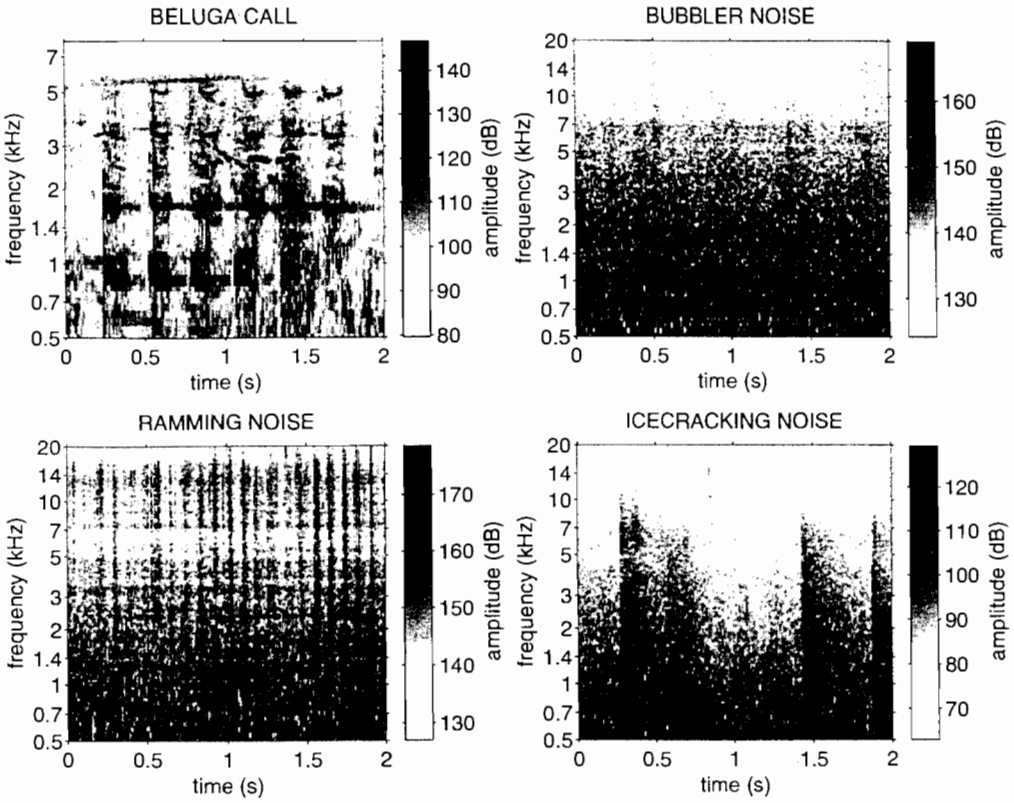

FIG. 1. Power density spectrogram of the beluga vocalization, an icebreaker's bubbler and propeller (ramming) noise, and natural ice-cracking noise in dB *re* 1 (μPa^2/Hz)@1 m. A source level of 160 dB *re* 1 μPa@1 m was assumed for the beluga call. The source levels of the noises were, respectively, 194, 203, and 147 dB *re* 1 μPa@1 m.

noise emitted by the bubbler system of an icebreaker was temporally continuous and broadband, with most of the energy below 2 kHz. Propeller cavitation noise consisted of sharp broadband pulses occurring 11 times per s. It was also called ramming noise, because this sample was taken during the ice-ramming action of the icebreaker. Naturally occurring ice-cracking noise (ambient Arctic noise) exhibited broadband pulses at irregular intervals. The vocalization was mixed with 2-s samples of each of the three noises in various signal-to-noise ratios. Mixed sounds were played back to a beluga whale named Aurora, who indicated signal discrimination in a behavioral go/no-go paradigm. Aurora's response is replotted in Fig. 2. Bubbler noise exhibited the strongest interference with a detection threshold at a "critical" noise-to-signal ratio (nsr) of 15.4 dB (signal-to-noise ratio −15.4 dB). Propeller cavitation noise was second strongest in masking with a critical nsr of 18.0 dB. Natural ice-cracking noise was least masking, with an nsr of 29.0 dB.

As a means of assessing the degree of masking of a variety of industrial noises, animal experiments are inefficient because of the amount of time and cost involved. It would be preferable to have a fast, ground-truthed model simulating masked hearing experiments and thus predicting masking effects in cases where direct experiments with animals are infeasible. This article applies a variety of software tools, some of which are standard signal-processing methods, to the problem of detecting animal vocalizations in noise. Mellinger,[10] and Mellinger and Clark[11] used a matched filter, a hidden Markov model, and spectrogram image convolution to detect bowhead whale vocalizations in noise. The methods were compared with respect to their false alarm and miss rates. Spectrogram image convolution performed the best, having the smallest combined error rate. The hidden Markov model followed; matched filtering had the highest error rate. Potter *et al.*[12] designed an artificial neural network trained with back-propagation and tested it on the same bowhead data set. The neural net performed even better than spectrogram image convolution. Mellinger and Clark[13] tested

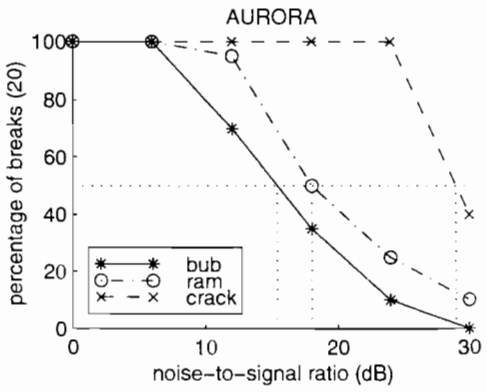

FIG. 2. Masked-hearing thresholds of the beluga whale, Aurora, in bubbler, ramming, and ice-cracking noise. The *x*-axis denotes the pressure noise-to-signal ratio in dB. Every mixture of the call with the three noises in the 6 nsr's shown was played exactly 20 times. The *y*-axis indicates how often Aurora heard the call in the noise, i.e., how often she broke away from the target. Defining the hearing threshold at 50% yields the following critical nsr's: 15.4 dB for bubbler system noise, 18.0 dB for ramming noise, and 29.0 dB for natural ice-cracking noise. [Reprinted from Ref. 7, Deep Sea Res. II **45**, 1383 ©1998, with permission from Elsevier Science.]

matched filtering and spectrogram image convolution on a few mysticete sounds from blue, fin, and minke whales. They found that matched filtering worked well if the signals to be detected were buried in white background noise. Spectrogram image convolution excelled if the interfering noise was structured, e.g., harmonic. In the case that the animal vocalization was a highly repetitive sequence of sounds occurring at regular intervals (a pulse train), a summed autocorrelation method proved useful. These previous studies compared the signal detection methods under the criterion of yielding the highest hit rate. In our study, the signal detectors are compared under the criterion "How closely does the model's performance resemble that of the whale?" Performance is judged in terms of the order and level of the maskers as determined by our earlier study.[7]

I. METHODS AND RESULTS

A. Matched filtering

In this section, we hypothesize that matched filtering can successfully model beluga masked-hearing experiments. For linear, time-invariant systems, a filter performs the convolution of an incoming time series $x[t]$ with its impulse response $h[t]$ to yield

$$y[t]=\sum_{k=-\infty}^{\infty} h[k]\cdot x[t-k]. \tag{1}$$

In problems of signal detection in noise, one wants to design a filter which—while convolving along the time series of input data—produces maximum output when there is complete overlap between the signal buried in noise and the desired pure signal. It can be shown that the optimal impulse response of such a filter has to be as long as the signal to be detected.[14] Furthermore, the filter coefficients have to be equal to the product of the inverted autocorrelation matrix of the noise and the time-reversed pure signal. For white noise, the autocorrelation matrix turns into the identity matrix, and the filter response equals the time inverse of the pure signal. As the filter coefficients are matched to the signal time series, one calls this filter a matched filter. The convolution of a time-reversed signal is equivalent to the cross correlation of the signal without time reversal. Therefore, matched filtering is equivalent to cross correlation of the input time series with the desired signal. We applied the matched filtering technique to "near-white" noises, as Fig. 3 indicates. The autocorrelation coefficients of the three noises (bubbler, ramming, and ice-cracking) and artificially created, Gaussian-distributed white noise are shown for comparison. In each case, a 2-s noise sample was correlated with a 3-s noise sample of the same type, and lags between 0 and 1 s were plotted. For zero lag, the coefficients were equal to 1. For all other lags, they oscillated around 0. In other words, the autocorrelation matrices of the noises were zero except for zero lag, i.e., the matrices were diagonal. The noises were, hence, white in the sense that they were uncorrelated with themselves.

In the discussion of our previous paper,[7] we put forward the idea that natural ice-cracking noise masked the least, because of its irregular temporal structure. The whale managed

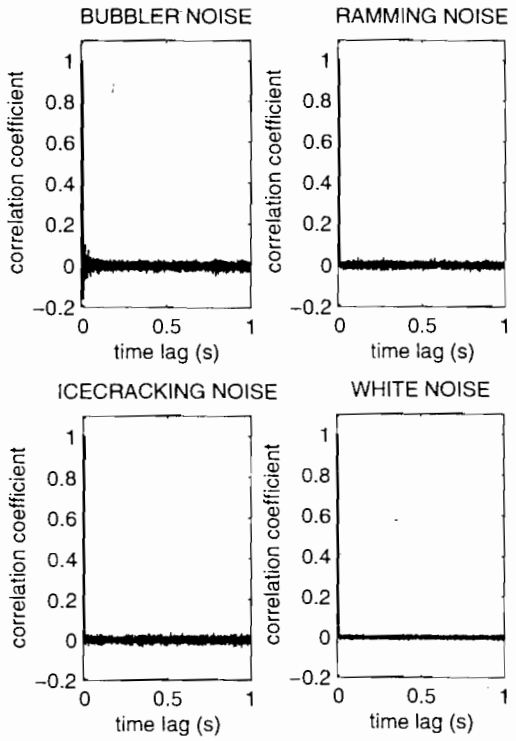

FIG. 3. Autocorrelation of the noise time series showing "near-white" behavior for the three Arctic noises.

to identify the vocalization in noise from short fractions which emerged through quieter gaps in the noise field. For the current analysis, we therefore assume that the whale can identify the vocalization if one of the six pulses in the call is audible. In order to calculate the impulse response $h[t]$ of the matched filter (simulating the whale's auditory filter), we cut each of the six 200-ms-long pulses, which we consider "phonemes," out of the time series of the 2-s vocalization. We then computed the covariance matrix of the six phonemes and searched for its eigenvalues and eigenvectors. The eigenvector corresponding to the largest eigenvalue of the covariance matrix was the one correlating the best with each of the six phonemes. This eigenvector in a sense represented a "mean" phoneme and was therefore selected as the impulse response $h[t]$ of our matched filter. During the matched filtering process, the filter $h[t]$ then "correlated along" the time series of incoming sounds. These incoming sounds were mixed sounds $x[t]$, i.e., the point-by-point summation of the pressure time series of the vocalization $s[t]$ and each of the three noises $n[t]$

$$x[t]=s[t]+\alpha\cdot n[t], \tag{2}$$

where α determined the noise-to-signal ratio (nsr). The time series above had the same length $N=90\,112$ samples (sampling frequency 44 kHz, 2 s of sound). We did not vary the time offset between the signal and the noise; i.e., the signal always happened at the same time in the noise. This was the same during the animal experiments.[7] The cross-correlation coefficient at each time lag β as a function of the nsr could then be computed as Ref. 15

$$R_\beta(\alpha)=\frac{\sum_{t=1}^{M}h[t]x[t+\beta]}{\sqrt{\sum_{t=1}^{M}h^2[t]\cdot\sum_{t=1}^{M}x^2[t+\beta]}}. \tag{3}$$

The total number of lags β is $N-M+1$, with N denoting the length of the time series $x[t]$ and M being the number of samples in $h[t]$.

The normalization in Eq. (3) was chosen such that the filter's output would be 1 for perfect detection and 0 for no correlation. The behavior of the cross-correlation coefficients R_β as a function of α can be understood by looking at the limits $\alpha=0$ and $\alpha\to\infty$. With $x=s+\alpha\cdot n$,

$$R_\beta(\alpha)=\frac{\Sigma h[t]x[t+\beta]}{\sqrt{\Sigma h^2[t]\cdot\Sigma x^2[t+\beta]}}=\frac{\Sigma h[t](s[t+\beta]+\alpha n[t+\beta])}{\sqrt{\Sigma h^2[t]\cdot\Sigma(s^2[t+\beta]+\alpha^2n^2[t+\beta]+2\alpha s[t+\beta]n[t+\beta])}}$$

$$=\frac{\Sigma h[t]s[t-\beta]+\alpha\Sigma h[t]n[t+\beta]}{\sqrt{\Sigma h^2[t]\cdot\Sigma s^2[t+\beta]+\alpha^2\Sigma h^2[t]\cdot\Sigma n^2[t+\beta]+2\alpha\Sigma h^2[t]\cdot\Sigma s[t+\beta]n[t+\beta]}}. \tag{4}$$

For $\alpha=0$, the cross-correlation coefficients become

$$R_\beta(0)=\frac{\Sigma h[t]s[t+\beta]}{\sqrt{\Sigma h^2[t]\cdot\Sigma s^2[t+\beta]}}. \tag{5}$$

The value of the cross-correlation coefficient depends on how well the filter $h[t]$ matches the desired signal $s[t]$ over the length of the filter. At the lag of best overlap, the cross-correlation coefficient will be equal to 1, if the impulse response of the filter equals the time series of the desired signal: $h[t]=s[t]$. If the filter is not equal but only similar to the desired signal, as is the case in our model, the correlation coefficients will be slightly less than 1.

For $\alpha\to\infty$, we divide the numerator and denominator by α and let all the terms with α in the denominator go towards 0.

$$R_\beta(\alpha)=\frac{(1/\alpha)\Sigma h[t]s[t+\beta]+\Sigma h[t]n[t+\beta]}{\sqrt{(1/\alpha^2)\Sigma h^2[t]\cdot\Sigma s^2[t+\beta]+\Sigma h^2[t]\cdot\Sigma n^2[t+\beta]+(2/\alpha)\Sigma h^2[t]\cdot\Sigma s[t+\beta]n[t+\beta]}}, \tag{6}$$

$$\lim_{\alpha\to\infty}R_\beta(\alpha)=\frac{\Sigma h[t]n[t+\beta]}{\sqrt{\Sigma h^2[t]\cdot\Sigma n^2[t+\beta]}}. \tag{7}$$

Independent of the type of noise, all the plots of $R_\beta(\alpha)$ will always converge to the product of the filter's impulse response with the pure noise. In other words, if there is no signal in the noise, the output of the filter will not necessarily be 0 but slightly greater or smaller (negative) than 0. Thinking of the time series as vectors, $R_\beta(\alpha)$ converges to the cosine of the angle between the filter and the noise. Therefore, the more "similar" the filter and the noise are, the smaller the angle, the greater the cosine.

In matched filtering, as soon as the filter's output surpasses a preset threshold, the signal detection is deemed successful. Picking a threshold, however, is tricky. We first examined the fluctuation of the filter, when presented with pure noise. We therefore computed the cross correlation of the mean phoneme with each 2-s noise time series over all time lags β. This yielded a series of correlation coefficients R_β for each of the three noises. The standard deviations of the correlation coefficients were 0.033, 0.045, and 0.023 for bubbler, ramming, and ice-cracking noise, respectively. Given that we hypothesized the whale would detect the call in the noise if just one of the phonemes were audible, we looked for the maximum R over all lags β. A detection threshold should therefore be greater than the maximum correlation coefficient when pure noise is presented. For a comparison across methods, we picked a detection threshold at 0.5 in analogy to the whale experiment. The matched filter was normalized such that it would give an output close to 1 when the signal without added noise was presented, in analogy to a 100% signal-detection probability from Aurora when no or low noise was added, see Fig. 2. The matched filter was further chosen to give an output close to 0 in the case of pure noise, in analogy to a near-zero response probability of Aurora and because the animal was trained not to respond upon hearing pure noise.

Figure 4(a) shows the cross correlation of the mean phoneme with the call $s[t]$. The impulse response of the filter correlates well with the first four pulses in the call, and poorly with the last two. Figure 4(b) shows the results of the matched filtering. Taking thresholds at 0.5 yields the following critical nsr's: 1.6 dB for bubbler noise, 3.2 dB for ramming noise, and 10.0 dB for ice-cracking noise. The order of the noises is the same as with Aurora (Fig. 2); the relative distances of the thresholds are also similar. If one was to add an offset of 15.0 dB to the estimated thresholds of the matched filter, Aurora's results would be reproduced with a

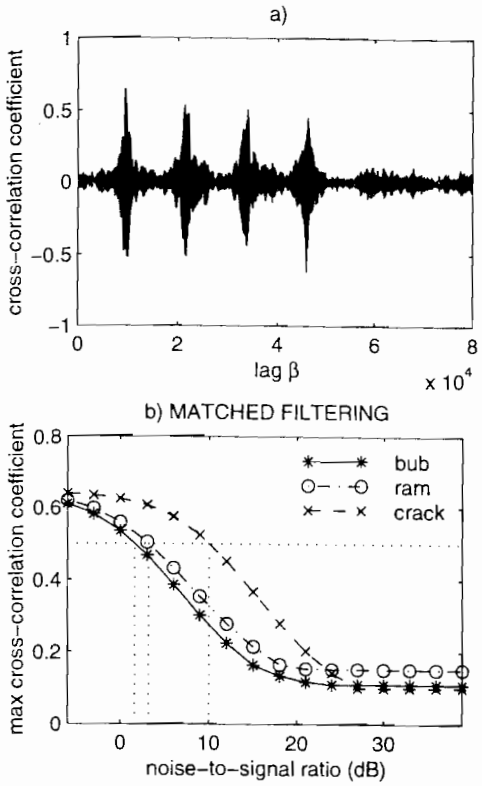

FIG. 4. (a) Cross correlation of the matched filter's impulse response with the time series of the beluga vocalization. The first four phonemes in the vocalization are detected well; the last two phonemes correlate poorly. (b) Matched filtering of the beluga vocalization buried in bubbler, ramming, and ice-cracking noise. The maximum cross-correlation coefficient over all lags β was plotted. Taking a threshold at 0.5, as in the whale experiment, classifies the noises in the same order as the whale did and with similar relative thresholds. The critical nsr's are 1.6 dB for bubbler noise, 3.2 dB for ramming noise, and 10.0 dB for ice-cracking noise.

maximum error of 13%, which we consider a good similarity. This offset of 15.0 dB was calculated such that the sum-squared-error between Aurora's critical nsr's and the modeled nsr's was minimized.

There are many different possible implementations of matched filters. Erbe[16] chose the full call $s[t]$ as the impulse response of the matched filter $h[t]$, rather than picking a mean phoneme. The order of the noises, however, was such that bubbler system noise exhibited the strongest interference, followed by ice-cracking noise, then ramming noise. In other words, the degree of masking for ice-cracking and ramming noise was reversed. Thresholds were also very close to each other, at 4.8 dB, varying only by 0.1 dB. Given that the mean phoneme impulse response presented in this paper worked better, a possible explanation could include that the animal indicated the call detection as soon as one phoneme was audible.

We also tried to take the exact time series of one 200-ms phoneme as $h[t]$, rather than calculating a mean phoneme. Each of the six phonemes in the beluga vocalization was taken as the impulse response of the filter one at a time. The particular phoneme chosen was cut out of the call $s[t]$ and correlated with mixtures of the shortened call and the three noises. The phoneme chosen as the filter was taken out of the time series of the call in order to avoid large autocorrelation values in the desired cross correlation. Also, in the wild, an animal could never have the exact time series of an incoming signal stored in its memory due to propagation effects and the inherent variability of time series of vocalizations. The results of this matched filtering method were that the order of the noises in the correlation coefficient plots varied depending on which phoneme was chosen as the filter. This was because each phoneme did not correlate equally well with the others.

Altogether, the matched filter presented in this article successfully modeled the relative degrees of masking of bubbler, ramming, and ice-cracking noise as measured with the trained beluga whale, Aurora. However, a generalization of these results to other vocalizations and noises needs to be treated with caution. This is for two reasons: First, the theory of a matched filter is based on white noise. A matched filter can therefore be expected to perform poorly in nonwhite, structured noise, as shown by Mellinger and Clark.[13] Spectrogram image convolution can be expected to work better for structured noise. Second, in the case of complex vocalizations (those having multiple frequencies which are not harmonics), two calls of the same type generally correlate poorly, which makes it difficult, if not impossible, to design the impulse response of a matched filter that will correlate well with the signals to be detected. From our experience, only constant frequency tones or single frequency whistles (with consistent time structure) correlate well in the time domain. We hypothesize that multifrequency (nonharmonic) vocalizations correlate poorly, because from utterance to utterance, only amplitude and frequency structure will be consistent, but not the phase. Sound propagation effects in the ocean further change the time series of animal vocalizations. Signal-detection methods based on spectrograms rather than time series might thus be more promising, because they no longer contain phase information.

B. Spectrogram cross correlation

In this section, we hypothesize that spectrogram cross correlation will model beluga masked hearing experiments better than matched filtering. In Mellinger and Clark's study[11] on automatic bowhead call detection in noise, spectrogram cross correlation (which they called spectrogram image convolution) had a higher call-detection rate than matched filtering and a hidden Markov model. Mellinger and Clark[13] showed that spectrogram cross correlation outperformed matched filtering, particularly if the interfering noise was nonwhite.

We want to compare spectrogram cross correlation to matched filtering under the criterion of how closely the thresholds and the order of noises resemble those of a beluga whale. The motivation for trying spectrogram cross correlation was twofold. First, the time series of two vocalizations or phonemes often correlate poorly because of varying phase. Phase information is lost in spectrograms. Spectrograms contain frequency, amplitude and time information which is consistent amongst calls of the same type. Therefore, spectrograms of calls or phonemes often correlate better than time series. Second, spectrogram representation is biologically more justified than time series representation: In the mammalian ear (of terrestrial as well as marine mammals),

the inner ear and auditory nerve serve simplistically as a series of tuned resonators or filters providing spectral information in real time to the brain.[17,18]

For each of the 2-s-long sounds (beluga vocalization, bubbler noise, ramming noise, ice-cracking noise, and mixtures of the call with the three noises), we computed a spectrogram by calculating the fast Fourier transform in blocks of 1024 data points, using Hamming windows with 50% overlap. We took the magnitude of the complex Fourier components. Negative frequencies and frequencies outside the call spectrum (>6 kHz) were discarded. With a sampling frequency of 44 kHz, the resulting spectrogram matrices were of size 139×175.

We constructed our kernel in a different way than Mellinger and Clark.[11] They actively placed positive numbers where there was energy in the spectrogram of the call and negative numbers on either side, and normalized the kernel such that the sum over all kernel values was 0. In order to construct our kernel, we selected the six call phonemes from the call spectrogram. Each phoneme spectrogram was of the size 139×11. Similar to our matched filter design, we computed the covariance matrix of the spectra of the six phonemes in the call and looked for the largest eigenvalue. The corresponding eigenvector was chosen as the kernel for spectrogram cross correlation after subtracting its mean.

Spectrogram cross correlations were then calculated according to

$$R_\beta(\alpha)=\frac{\sum_{t=1}^{11}\sum_{f=1}^{139}h[t,f]x[t+\beta,f]}{\sqrt{\sum_{t=1}^{11}\sum_{f=1}^{139}h^2[t,f]\cdot\sum_{t=1}^{11}\sum_{f=1}^{139}x^2[t+\beta,f]}},\tag{8}$$

where $h[t,f]$ is the kernel and $x[t,f]$ is the spectrogram of the mixed sound. The time lag β had the size of one time sample in the spectrogram. Figure 5(a) shows the correlation of the kernel with the spectrogram of the vocalization. The first four phonemes are detected clearly; the last two phonemes correlate more poorly with the kernel.

Again, we hypothesize that the call can be detected if the correlation of the kernel with one phoneme in the call lies above a detection threshold. While correlating the kernel with the spectrograms of the mixed sounds, we therefore looked for the maximum $R_\beta(\alpha)$ over all lags β. The correlation curves $\max[R_\beta(\alpha)]$ for the three noises are shown in Fig. 5(b). The order of the noises was again the same as with the whale: bubbler noise was identified as the strongest masker, followed by ramming noise, then ice-cracking noise. For comparison, the detection threshold was taken at 0.5 yielding the following critical nsr's: 11.4 dB for bubbler noise, 13.8 dB for ramming noise, and 17.9 dB for ice-cracking noise. If the critical nsr's estimated by spectrogram cross correlation are shifted towards higher nsr's by an offset of 6.4 dB (taken to minimize the sum-squared-error), the critical nsr's of Aurora are reproduced with a maximum deviation of 16%.

Critical nsr's computed with spectrogram cross correlation depend on the number of Fourier components used during the fast Fourier transform. We tried spectrogram cross correlation with 512 instead of 1024 Fourier components. This gave the same correct order of noises, though with a maximum mismatch of 19% after shifting the critical nsr's such as to minimize the sum-squared-error between the modeled nsr's and Aurora's. In the case of 256 Fourier components the maximum mismatch was 27%; in the case of 2048, the maximum mismatch was 20%. The order of the noises was correct in all four cases. The best modeling results were achieved with 1024 Fourier components.

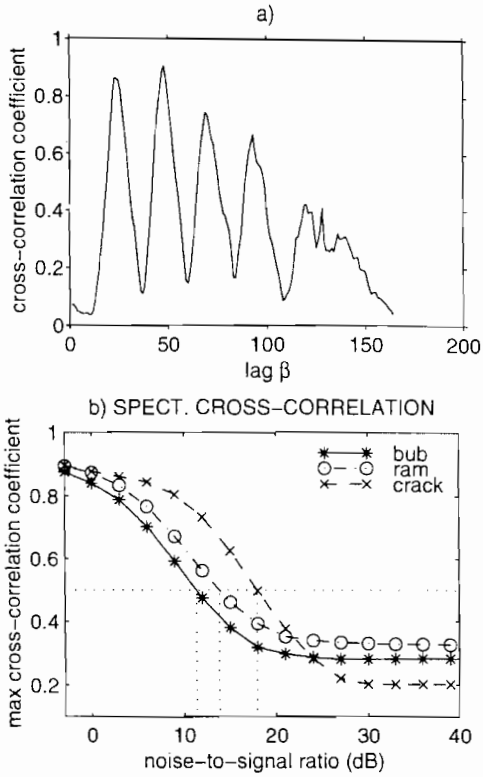

FIG. 5. (a) Cross correlation of the spectrogram kernel with the spectrogram of the beluga vocalization. Similar to the matched filter, the first four phonemes in the vocalization are detected well; the last two phonemes correlate poorly. (b) Spectrogram cross correlation of the beluga vocalization buried in bubbler, ramming, and natural ice-cracking noise. Plotted is the maximum cross-correlation coefficient over all lags β. With a threshold taken at 0.5, the critical nsr's are 11.4 dB for bubbler, 13.8 dB for ramming, and 17.9 dB for ice-cracking noise. The order of noises is correct.

Both matched filtering and spectrogram cross correlation managed to classify the three noises from strongest masking to least masking in the same order as the whale. Comparing critical nsr's, their relative distance is slightly closer to that of the whale in the case of matched filtering.

Erbe[16] chose the spectrogram of the entire vocalization as the kernel, which resulted in ice-cracking noise being identified as the strongest masker, followed by ramming noise, then bubbler noise. The order of the noises was exactly opposite to that of the whale, Aurora. This could mean that the detection of one phoneme is enough to recognize the vocalization and hence the correct modeling approach.

C. Critical band cross correlation

A quantity often measured in masked-hearing experiments is the critical ratio.[19] In the case that a pure tone is masked by a broadband white noise, the critical ratio is defined as the ratio of the intensity of the tone I_t divided by the

noise spectral density SI_n (intensity per Hz) at the level when the tone is just audible through the noise. The critical ratio is often expressed in decibels

$$\mathrm{CR} = 10 \cdot \log_{10} \frac{I_t}{SI_n}. \tag{9}$$

The mammalian auditory system can generally be represented as a series of overlapping bandpass filters.[18,19] The idea is that a listener trying to detect a signal in noise will "choose" an auditory filter with a center frequency close to that of the signal. Then, only the amount of noise coming through this filter will have an effect on masking the signal; noise at frequencies far away from the signal frequency will have no effect. Fletcher[20] hypothesized that at detection threshold, the intensity of the tone equaled the total intensity of the noise in the corresponding auditory filter (equal-power assumption)

$$I_t = SI_n \cdot \Delta f. \tag{10}$$

Substituting into CR yields

$$\mathrm{CR} = 10 \cdot \log_{10} \Delta f. \tag{11}$$

The critical ratio can therefore also be expressed in Hz as the approximated width of the auditory filter around the test tone

$$\Delta f = 10^{\mathrm{CR}/10}. \tag{12}$$

In 1940, Fletcher[20] did a masked-hearing experiment with human subjects, in which the signal was a pure tone and the masker was narrow-band white noise. Fletcher measured the intensity of the signal at detection threshold as a function of the noise bandwidth. For narrow-band noise (narrower than the bandwidth of the auditory filter), the signal-detection threshold increased with increasing noise bandwidth, i.e., increasing total noise power in the filter. Once the bandwidth of the noise reached the bandwidth of the auditory filter, the signal-detection threshold remained constant during further widening of the noise band. The noise bandwidth above which masking could not be increased was called the critical bandwidth. The critical bandwidth is thus the width of the auditory filter measured via Fletcher's band-widening technique; whereas the critical ratio is the filter width calculated from masked-hearing thresholds in broadband white noise.

Johnson[2] measured critical ratios for a bottlenose dolphin (*Tursiops truncatus*) for signal frequencies between 5 and 100 kHz and continuous broadband noise. In a double-logarithmic plot, a straight line could be fitted through the data points of CR as a function of frequency. This indicated that the dolphin auditory system may be modeled as a bank of constant Q filters. The quality factor Q of a bandpass filter is defined as the ratio of center frequency to filter width, measured at half-peak power. Constant Q implies an increase in filter width with center frequency. Au and Moore[5] also measured critical ratios of a bottlenose dolphin and obtained similar results to Johnson.[2] Johnson *et al.*[4] measured critical ratios between 40 Hz and 115 kHz in a beluga whale (*Delphinapterus leucas*). For frequencies less than 1 kHz, the critical ratio no longer increased with frequency, but appeared to become constant, indicating that a constant Q model only holds for frequencies above 1 kHz. A comparison across species further showed that the critical ratios of the beluga whale were about 3 dB smaller than those of the bottlenose dolphin. Au and Moore[5] also measured critical bandwidths for a bottlenose dolphin using Fletcher's band-widening technique. On average, the critical bands were 7.5 dB greater than the critical ratios, indicating that Fletcher's equal-power assumption might hold less well in the case of bottlenose dolphins. Larger critical bands mean that the signal power is less than the noise power at threshold.

Critical band analysis is based on simultaneous masking, i.e., when the signal and the masker happen at the same time. However, forward masking and backward masking exist, too, in which a time-limited masker precedes or follows the signal, respectively. For masking to occur, signal and masker must be less than a critical time interval apart. Vel'min and Dubrovskii,[21] Moore *et al.*,[22] and Au *et al.*[23] measured this interval with bottlenose dolphins for high-frequency echolocation signals. It was, on average, 300 μs. There are no data for critical time intervals at lower communication frequencies and no data for beluga whales. Vel'min and Dubrovskii[24] pointed out that dolphins might be equipped with two functionally (and possibly morphologically) independent auditory subsystems, one for communication and one for echolocation, in which case these critical intervals could not be applied to communication signals.

There are no data on critical bandwidths in beluga whales. For the current analysis, we therefore calculated the widths of the beluga auditory filters from Johnson's critical ratio data[4] using Eq. (12). For example, at 700 Hz (the lower limit of the call spectrum), the critical ratio is about 18 dB according to Johnson's experiment, yielding a filter bandwidth of 63 Hz. At 6 kHz (the upper limit of the call spectrum), the critical ratio is about 24 dB, yielding a bandwidth of 251 Hz. On average, for low frequencies, the filter width was about 6% of the center frequency. Or, in other words, the critical ratios were about 1/12th of an octave wide. Picking a center frequency f_0, the lower limit of the filter can be calculated as $f_l = 2^{-1/24} \cdot f_0$, the upper limit is $f_u = 2^{1/24} \cdot f_0$. Table I lists the center frequencies of the adjacent, i.e., nonoverlapping 12th-octave bands used in our study.

The computation of spectrogram cross correlation in the previous section involved linear frequency distribution and linear amplitudes. If cross correlation takes place in the mammalian brain, hypothetical spectrograms created by our ears are more likely logarithmic in frequency and amplitude. Therefore, we first computed power-density spectrograms of the beluga vocalization, the three noises, and all mixed sounds, using 1024 Fourier components and Hamming windows with 50% overlap. Power is the area underneath a power-density spectrum. Therefore, the total power passing through one frequency band could be calculated as the integral over the power density spectrum from f_l to f_u at each time in the power density spectrogram. We took $10 \cdot \log_{10}$ of the amplitudes to yield band levels in dB *re* 1 μPa. Furthermore, we adjusted the amplitudes relative to the beluga audiogram, which relates the amplitude of a pure tone at detection threshold (in the absence of noise) to the tone's frequency. According to the beluga audiogram, high frequen-

TABLE I. Center frequencies (Hz) of adjacent 12th-octave bands.

40	42	45	48	50	53	57	60	63	67
71	76	80	85	90	95	101	107	113	120
127	135	143	151	160	170	180	190	202	214
226	240	254	269	285	302	320	339	359	381
403	427	453	479	508	538	570	604	640	678
718	761	806	854	905	959	1 016	1 076	1 140	1 208
1 280	1 356	1 437	1 522	1 613	1 709	1 810	1 918	2 032	2 153
2 281	2 416	2 560	2 712	2 874	3 044	3 225	3 417	3 620	3 836
4 064	4 305	4 561	4 833	5 120	5 424	5 747	6 089	6 451	6 834
7 241	7 671	8 127	8 611	9 123	9 665	10 240	10 849	11 494	12 177
12 902	13 669	14 482	15 343	16 255	17 222	18 246	19 331	20 480	

cies were amplified. This audiogram was the mean of six published beluga audiograms[25,26,4] and the four frequencies measured by Erbe and Farmer.[7] Only the frequency bands between 700 Hz and 6 kHz, which occupy energy of the call, were chosen for this analysis. The kernel was again selected as the eigenvector of the largest eigenvalue of the covariance matrix of the call phonemes. We subtracted its mean.

Figure 6(a) shows the correlation of the kernel with the 12th-octave band spectrogram of the call. Compared to Fig. 5(a), the correlation peaks at the location of the six phonemes in the call are less sharp. The correlation maxima vary less from phoneme to phoneme than with the spectrogram cross-correlation method. Figure 6(b) shows the maximum cross-correlation coefficients of the kernel correlated with the band-averaged spectrograms of the mixed sounds. Although the curves change order with the nsr, nowhere does the order resemble that of the whale. The results are curious. The time resolution of the spectrograms used in the previous section and the band-averaged spectrograms used here was the same. The difference in correlation behavior therefore lies in the audiogram normalization, logarithmic amplitude, and band averaging. The failure of the critical band cross correlation compared to the spectrogram cross correlation (under the criterion of similarity to the whale's results) could indicate that a beluga whale's critical bands are narrower than the ones we calculated from Johnson's critical ratio data under the equal-power assumption: The spectrograms in the previous section contained 70 frequency samples, compared to 38 frequency bands used here.

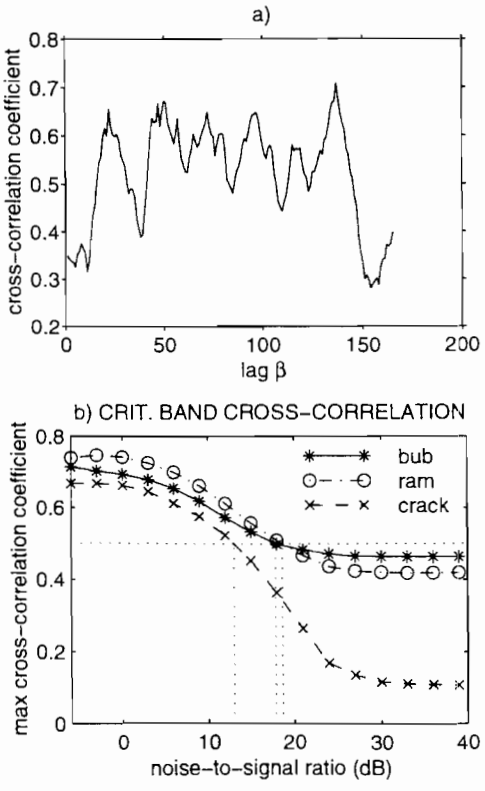

FIG. 6. (a) Critical band (CB) cross correlation of the kernel with the CB-averaged spectrogram of the beluga vocalization. The six phonemes are detected, though less sharply than in Fig. 5(a). (b) Critical band cross correlation of the beluga vocalization in bubbler system noise, ramming noise, and natural ice-cracking noise using a time resolution of 23 ms. Taking a threshold at 0.5, this method could not reproduce the correct order of noises.

A different interpretation of Fig. 6(b) would involve picking a different threshold. The animal might pick a different threshold for each noise depending on how large the cross-correlation coefficients of the correlation of the pure noises with the kernel are. The cross correlations for infinite nsr in Fig. 6(b) are fairly large, particularly for bubbler and ramming noise. The difference between the maximum cross-correlation coefficient at low noise levels (when the call is easily detectable) and at high noise levels (when the call is undetectable) is largest for ice-cracking noise, then ramming, then bubbler noise. One possible interpretation of this is that it would be "easier" for the animal to detect the call in the ice-cracking noise than in the other two noises, in the sense that there is a larger range over which the animal can set a threshold in order to indicate call detection with confidence. This interpretation would put the three noises back into the correct order.

It is obvious that spectrogram cross-correlation methods depend on the frequency and time resolution of the spectrograms, i.e., the number of Fourier components used in the Fourier transform. Limited by the fixed sampling rate, there is a tradeoff between the frequency and time resolution in spectrograms. The sampling frequency was $f_s = 44$ kHz for our recordings. With $n_{\text{fft}} = 1024$ Fourier components, the spectrograms used in the previous section had a frequency resolution of $\Delta f = f_s / n_{\text{fft}} = 43$ Hz. Frequency samples were averaged into 38 12th-octave bands in the current section. The time resolution was $\Delta t = 1/\Delta f = 23$ ms in both sections. This was considerably wider than the critical time interval for dolphin echolocation. In order to test the effect of decreasing the time interval, critical band cross correlation was tried with $n_{\text{fft}} = 512$ and 256. This decreased the frequency

resolution to 86 and 172 Hz, respectively. The time resolution doubled to $\Delta t = 12$ and 6 ms. A further increase in time resolution would result in fewer frequency samples than 38 bands. Taking thresholds at 50% still classified the noises in the wrong order to the whale. However, using the difference between starting and end point for each curve showed that ice-cracking noise correlations had the largest variation, followed by ramming, then bubbler, noise. This could be interpreted in terms of the whale's facility to set a threshold with confidence, yielding a "better" call detection in ice-cracking noise. This would give the same order of noises as with the whale.

D. Artificial neural network

Artificial neural networks (ANN) have successfully been applied to automatic detection of human speech signals.[27] Potter *et al.*[12] used an artificial neural network to detect calls of the bowhead whale in noisy Arctic recordings. Their neural network outperformed a matched filter, a hidden Markov model, and spectrogram cross correlation. We hypothesize that a neural network can also outperform our previous methods when judged under the criterion of closeness to the whale's response. Our hypothesis is based on two facts. First, the previously employed correlation techniques were linear filters. The mammalian auditory system, however, is highly nonlinear.[18,19] The neural network designed in this section is nonlinear. Second, the impulse responses (kernels) of the filters in the previous sections were chosen as eigenvectors of the covariance matrix of call phonemes. This might not be the best approach to the whale's filter. In neural network analysis, a "kernel" does not need to be chosen by the operator *a priori*. Rather, the network learns to recognize features from a set of training data. This process is biologically more justified than picking a present kernel.

An ANN can be regarded as a very simple model of the biological neural system, the brain. Both networks are complex, massively parallel information-processing systems which learn from experience and store their knowledge. The major types of ANNs, their history, and applications are described by Haykin.[28] For our ANN design, we chose a fully connected two-layer network trained with back propagation.[29] During the training phase, the ANN was repeatedly presented with a training matrix of noisy beluga calls and pure noise. The network's weights and biases were adjusted from iteration to iteration such that the sum-squared-error (between the desired and actual output) of the network was minimized. We found two modifications to the back-propagation learning rule useful. The inclusion of momentum[30] avoided getting stuck in a high local minimum of the ANNs error performance surface; an adaptive learning rate helped to speed convergence. During the generalization phase, weights and biases were kept constant, and the ANN was presented with the mixed sounds played to the beluga whale. Its call detection performance could then be compared to the methods of the previous sections.

We tested different numbers of neurons in the hidden layer. From our experience, the fewer neurons, the faster the computation of one epoch; the more neurons, the faster the convergence of the neural network (in fewer epochs) to its preset minimum error. Rather than picking the number of hidden-layer neurons corresponding to the minimal computation time, we chose the minimum number of neurons required for the network to converge to its desired minimum error, at the expense of a larger computation time. The reason was that the performance of the neural network during generalization seemed more stable with fewer neurons, i.e., less dependent on the initial conditions. The desired sum-squared-error was set to 0.001. Our ANN had three neurons in the hidden layer and one output neuron. Initial weights and biases were chosen randomly.

We tried both sigmoid and hyperbolic tangent transfer functions and did not find major performance differences. We eventually settled on sigmoid functions, because they have the convenient advantage that their output lies between 0 and 1, where one can interpret 0 as "no recognition" and 1 as "full recognition" of a particular input pattern. The second argument for choosing sigmoid transfer functions was their biological motivation. They have been said to simulate the refractory phase of real neurons.[31]

The idea was to train our ANN to detect beluga call features in spectrograms of mixed sounds. For the training matrix, we needed a large number of noisy versions of the beluga vocalization. In order to create the training matrix, we first computed the spectrogram of the beluga vocalization using 512 Fourier components and Hamming windows with 50% overlap. This resulted in 89 856 amplitude data, which would require the same amount of neurons in the input layer. ANNs with too many neurons generalize poorly, and computation time increases rapidly with the number of neurons. Finally, ANN analysis which is based on matrix multiplications easily exceeded the computer's virtual memory. Therefore, dramatic data reduction and compression prior to neural network computation was essential. For data reduction, we discarded the first and last 200 ms of the spectrogram before call onset and behind the last phoneme. Furthermore, we limited the frequency band to between 700 Hz and 6 kHz. This way, only the exact time and frequency range occupied by the call was selected. For data compression, we averaged the call spectrogram into a square grid of 20 time steps and 20 frequencies. Modeling the beluga auditory system as a bank of constant Q filters, Table I gave about 38 bandwidths between 700 Hz and 6 kHz. We thus chose a filter array half as fine. Along the time axis, each of the 20 boxes was 82-ms long, which was very much coarser than the critical interval of 300 μs reported for bottlenose dolphins in the case of echolocation.

A training matrix for the neural network was then created by adding noise to the call, not in the time series but in the averaged spectrogram (subscript s for spectrogram domain). We deliberately did not include any of the three test noises in the training matrix in order to prevent the ANN from deriving a cue from the particular noise characteristics rather than call features. In order to evenly span the input space, we added noise matrices that were orthogonal to each other. In particular, we chose doubly periodic background noise, where the amplitude varied both with frequency and time. Altogether, 800 noise matrices of the form

$$n_s[t,f]=\sin\left(\frac{2\pi\omega_t t}{T}+\phi_t\right)\cdot\sin\left(\frac{2\pi\omega_f f}{F}+\phi_f\right)+0.1\cdot\mathrm{rand}(t,f), \tag{13}$$

were computed with $T=20$ and $F=20$. This equation describes a two-dimensional sine wave overlapped by random values. The angular frequencies ω_t and ω_f were taken from the set [0,1/2,1,...,6] with equal numbers of each combination. The frequencies were less than half the grid size, because some networks had difficulty converging if higher frequencies were included. This was in analogy to visual discrimination of training matrices. When n_s matrices were plotted with colors representing amplitude, the human eye could easily detect the call in low-frequency, but not high-frequency, noise. The phases ϕ_t and ϕ_f were chosen randomly between 0 and π. Random white noise was added with values between 0 and 0.1. We created another 50 noise matrices of entirely random values.

The root-mean-square (rms) amplitudes of the call and the 850 noises were calculated in the spectrogram domain. The spectrograms were divided by their rms amplitude in order to normalize the sounds. Sounds were then mixed according to

$$x_s[t,f]=s_s[t,f]+\alpha_s\cdot n_s[t,f], \tag{14}$$

with α_s varying between 0 and 1. The same recording of the vocalization was chosen every time. All the 20×20 matrices of mixed sounds were reshaped into 400-element-long column vectors by placing the 2nd column of 20 frequency values underneath the 1st column, following with the 3rd column and so forth. We put the 850 column vectors of the mixed spectrograms and the 850 column vectors of the pure-noise matrices together into a training matrix of the size 400×1700. The desired output was set to 1 for the first 850 training vectors representing noisy versions of the call spectrogram. The desired output was set to 0 for the last 850 training vectors representing pure noise. Given that the nsr of the training vectors was low, an output of 1 would thus indicate successful call detection, while an output of 0 would indicate call absence.

With the training matrix and desired output thus chosen, we trained the ANN with back propagation until the sum-squared-error was less than 0.001. This required about 1000 iterations, i.e., repeated presentations of the training matrix. After completed convergence, we held the weights and biases constant. In order to check for a proper generalization of the ANN, we presented the network with the averaged and reshaped spectrogram matrices of the beluga vocalization and bubbler, ramming, and ice-cracking noise. The output was 1 for the call and 0 for the three noises, as desired.

The network was then presented with the averaged and reshaped spectrogram matrices of the original mixtures of the call with the three noises in (time series) nsr's of 0, 3,...,30 dB. The ANN was trained ten times with varying (random) initial conditions (weights and biases). Figure 7 shows the network's mean output at each nsr. Taking the thresholds at 50% yielded the following critical nsr's: 1.6 dB for bubbler noise, 4.7 dB for ramming noise, and 14.0 dB for ice-cracking noise. The standard deviations were 0.9 dB for bubbler and ramming noise, and 1.4 dB for ice-cracking noise. Therefore, not only the order of the noises from strongest to weakest masking was the same as for the whale, but also the relative degrees of masking were similar. Subtracting the net's thresholds from Aurora's yielded an offset of 13.8 dB for bubbler, 13.2 dB for ramming, and 15 dB for ice-cracking noise. Thus, shifting the net's thresholds an average of 14.0 dB to higher nsr's gave Aurora's results with a maximum error of 6%.

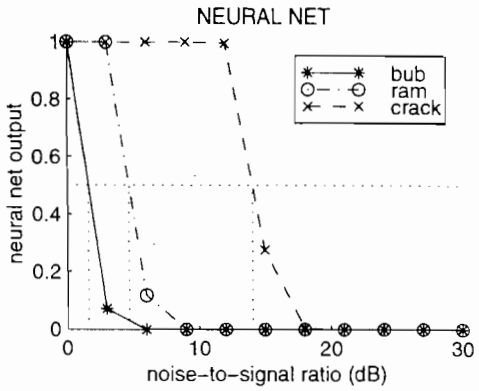

FIG. 7. Mean masked-detection thresholds of the neural network in bubbler, ramming, and ice-cracking noise after running the ANN ten times with different initial conditions. Taking the threshold at 0.5 yielded the following critical nsr's. 1.6 dB for bubbler, 4.7 dB for ramming, and 14.0 dB for natural ice-cracking noise. With an offset of 14.0 dB, the network modeled Aurora's thresholds within 6%.

This "calibration" of the neural network to Aurora's performance was necessary because the neural net did not get any species or individual-specific input data. We tried a normalization of training and generalization sounds with respect to the beluga audiogram. However, as the vocalization used was limited to between 700 Hz and 6 kHz, where the audiogram was basically linear, this normalization did not affect the neural network's performance. For more broadband sounds, though, audiogram normalization might become important and a good means of including biological background data in the computer modeling.

II. DISCUSSION

The purpose of this study was to develop an automatic and objective detector for beluga vocalizations in noise. Rather than designing the detector such that it would find the quietest calls with the smallest possible false alarm and miss rates, the detector was supposed to detect beluga calls with the same probability as a beluga whale, in particular the beluga whale trained and used for masked-hearing experiments by Erbe and Farmer.[7] A detector that models beluga masked-hearing experiments could then be used efficiently for environmental assessments of manmade noise with respect to masking of animal vocalization signals. A reliable model is particularly desirable in cases where direct experiments with trained animals are impractical.

We presented a variety of software algorithms which are commonly used for signal detection in noise. Their performance was compared to the masked-hearing experiments with the beluga whale Aurora.[7] In the first section, a matched

filter correlated the time series of a mean call phoneme with the time series of mixed sounds containing both call and noise. Signal detection was deemed successful when the filter's output surpassed a threshold of 0.5. The matched filter classified the three noises in the same order as the beluga whale: bubbler system noise was identified as the strongest masker, followed by propeller cavitation, then natural ice-cracking noise. The relative distances between the critical noise-to-signal ratios (nsr), at which the call was just detectable in the three noises, were similar to those of the whale with a maximum deviation of 13%. We considered this a good similarity and hence accepted our hypothesis that a matched filter could successfully model the whale's results.

In the second section, a mean phoneme spectrogram was correlated with spectrograms of mixed sounds containing both call and noise. Taking a detection threshold at 0.5, the noises were again classified in the correct order. Relative distances between critical nsr's, however, had a maximum deviation of 16%, slightly higher than for the matched filter. Mellinger and Clark[11] found that spectrogram cross correlation outperformed matched filtering, if compared under the criterion of the smallest miss and false-alarm rates. In our analysis, matched filtering slightly outperformed spectrogram cross correlation under the criterion of closeness to the whale's response. We therefore had to reject our hypothesis that spectrogram correlation could outperform matched filtering.

In the third section, we modified the spectrogram correlator to include some basic bioacoustic information about the beluga auditory system, such as the width of the animal's auditory filter and the animal's audiogram. Taking the threshold at 0.5 again led to the wrong order of noises. We suggested the possible explanation that the critical bands of the beluga whale were narrower than the ones we calculated from critical ratio data. If so, Fletcher's equal-power assumption,[20] which works reasonably well in humans,[19] would not be valid for beluga whales. We further offered a different interpretation of the critical band correlation results involving different thresholds for the three noises. This would put the noises into the correct order again. If this same argument were applied to the spectrogram cross correlation and matched filtering section, however, these two methods would not be able to classify the noises in the correct order. It would, at this stage, clearly be useful to collect more masked-hearing data from the animal to corroborate one or the other method of picking a detection threshold.

In the fourth section, we hypothesized that an artificial neural network trained with back propagation to detect call features in noise would outperform the previous methods. Potter *et al.*[12] found that this type of neural network had the smallest error rate when compared to spectrogram cross correlation and matched filtering. The neural network we designed outperformed our spectrogram correlator and matched filter under the criterion of closeness to the whale's response, leading us to accept our hypothesis. The noises were classified in the correct order with a maximum deviation of the relative distances between critical nsr's of 6%. The threshold was taken at 0.5 again. The alternative interpretation offered for the critical band cross correlation cannot be applied to the neural network, because the signal-only and noise-only outputs were the same for the three noises.

In summary, the neural network modeled the whale's performance the best. It thus raised confidence in its ability to predict the degree of masking of other types of noise. At this stage, it would be advisable to let the ANN perform on further noises and subsequently return to the aquarium to test the same sounds on the beluga whale. We also suggest inclusion of further vocalizations of different time and frequency structure. This would allow conclusions about the interference of manmade noise with beluga communication sounds in general. Furthermore, it would be desirable to train beluga whales of different age and sex in order to average out individual differences. Ultimately, we would like to apply an integrated tool of animal experiments and subsequent computer modeling to a variety of marine mammal species.

ACKNOWLEDGMENTS

We wish to express our thanks to Kurt Fristrup and two anonymous reviewers for many constructive suggestions. This work was supported by the Canadian Coast Guard, Central and Arctic Region, under supervision of Patrice St-Pierre.

[1] V. I. Burdin, V. I. Markov, A. M. Reznik, V. M. Skornyakov, and A. G. Chupakov, "Ability of *Tursiops truncatus* Ponticus Barabasch to distinguish a useful signal against a noise background," in *Morphology and Ecology of Marine Mammals*, edited by K. K. Chapskii and V. E. Sokolov (Wiley, New York, 1973), pp. 162–168.

[2] C. S. Johnson, "Masked tonal thresholds in the bottlenosed porpoise," J. Acoust. Soc. Am. **44**, 965–967 (1968).

[3] C. S. Johnson, "Auditory masking of one pure tone by another in the bottlenose porpoise," J. Acoust. Soc. Am. **49**, 1317–1318 (1971).

[4] C. S. Johnson, M. W. McManus, and D. Skaar, "Masked tonal hearing thresholds in the beluga whale," J. Acoust. Soc. Am. **85**, 2651–2654 (1989).

[5] W. W. L. Au and P. W. B. Moore, "Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **88**, 1635–1638 (1990).

[6] J. A. Thomas, J. L. Pawloski, and W. W. L. Au, "Masked hearing abilities in a false killer whale (*Pseudorca crassidens*)," in *Sensory Abilities of Cetaceans—Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York, 1990), pp. 395–404.

[7] C. Erbe and D. M. Farmer, "Masked hearing thresholds of a beluga whale (*Delphinapterus leucas*) in icebreaker noise," Deep Sea Res. II **45**, 1373–1388 (1998).

[8] A. Supin and V. Popov, "Frequency-selectivity of the auditory system in the bottlenose dolphin, *Tursiops truncatus*," in *Sensory Abilities of Cetaceans—Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York, 1990), pp. 385–393.

[9] A. Y. Supin, V. V. Popov, and V. O. Klishin, "ABR frequency tuning curves in dolphins," J. Comp. Physiol. A **173**(5), 749–656 (1993).

[10] D. K. Mellinger, "Handling time variability in bioacoustic transient detection," in *Oceans '93, Engineering in Harmony with the Ocean, Proceedings IEEE*, Vol. III, pp. 116–121 (1993).

[11] D. K. Mellinger and C. W. Clark, "A method for filtering bioacoustic transients by spectrogram image convolution," in *Oceans '93, Engineering in Harmony with the Ocean, Proceedings IEEE*, Vol. III, pp. 122–127 (1993).

[12] J. R. Potter, D. K. Mellinger, and C. W. Clark, "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262 (1994).

[13] D. K. Mellinger and C. W. Clark, "Methods for automatic detection of mysticete sounds," Mar. Fresh. Behav. Physiol. **29**, 163–181 (1997).

[14] J. H. Karl, *An Introduction to Digital Signal Processing* (Academic, San Diego, 1989).

[15] E. C. Ifeachor and B. W. Jervis, *Digital Signal Processing: A Practical*

Approach (Addison-Wesley, Wokingham, England, 1993).
[16] C. Erbe, "The masking of beluga whale (*Delphinapterus leucas*) vocalizations by icebreaker noise," Ph.D. thesis, University of British Columbia, Canada (1997).
[17] G. von Békésy, *Experiments in Hearing* (McGraw-Hill, New York, 1960).
[18] J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic, San Diego, 1988).
[19] B. C. J. Moore, *An Introduction to the Psychology of Hearing* (Academic, San Diego, 1997), 4th ed.
[20] H. Fletcher, "Auditory patterns," Rev. Mod. Phys. **12**(1), 47–65 (1940).
[21] V. A. Vel'min and N. A. Dubrovskii, "The critical interval of active hearing in dolphins," Sov. Phys. Acoust. **22**(4), 351–352 (1976).
[22] P. W. B. Moore, R. W. Hall, W. A. Friedl, and P. E. Nachtigall, "The critical interval in dolphin echolocation: what is it?" J. Acoust. Soc. Am. **76**, 314–317 (1984).
[23] W. W. L. Au, P. W. B. Moore, and D. A. Pawloski, "Detection of complex echoes in noise by an echolocating dolphin," J. Acoust. Soc. Am. **83**, 662–668 (1988).
[24] V. A. Vel'min and N. A. Dubrovskii, "Auditory analysis of pulse tones by dolphins," Dokl. Akad. Nauk SSSR **225**, 470–473 (1975).
[25] M. J. White, Jr., J. Norris, D. Ljungblad, K. Baron, and G. di Sciara, "Auditory thresholds of two beluga whales (*Delphinapterus leucas*)," Report by Hubbs/Sea World Research Institute for Naval Ocean System Center, Report 78–109 (San Diego, CA, 1978).
[26] F. T. Awbrey, J. A. Thomas, and R. A. Kastelein, "Low-frequency underwater hearing sensitivity in belugas (*Delphinapterus leucas*)," J. Acoust. Soc. Am. **84**, 2273–2275 (1988).
[27] D. P. Morgan and C. L. Scofield, *Neural Networks and Speech Processing* (Kluwer Academic, Boston, 1991).
[28] S. Haykin, *Neural Networks: A Comprehensive Foundation* (Macmillan College Publishing, New York, 1994).
[29] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, edited by D. E. Rumelhart and J. L. McClelland (MIT Press, Cambridge, MA, 1986), pp. 318–362.
[30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature (London) **323**, 533–536 (1986).
[31] F. J. Pineda, "Generalization of backpropagation to recurrent and higher order neural networks," in *Neural Information Processing Systems*, edited by D. Z. Anderson (American Institute of Physics, New York, 1988), pp. 602–611.