

Detection of whale calls in noise: Performance comparison between a beluga whale, human listeners, and a neural network

Christine Erbe^{a)}

Institute of Ocean Sciences, Acoustical Oceanography, 9860 West Saanich Road, Sidney, British Columbia V8L 4B2, Canada

(Received 11 May 1999; accepted for publication 24 March 2000)

This article examines the masking by anthropogenic noise of beluga whale calls. Results from human masking experiments and a software backpropagation neural network are compared to the performance of a trained beluga whale. The goal was to find an accurate, reliable, and fast model to replace lengthy and expensive animal experiments. A beluga call was masked by three types of noise, an icebreaker's bubbler system and propeller noise, and ambient arctic ice-cracking noise. Both the human experiment and the neural network successfully modeled the beluga data in the sense that they classified the noises in the same order from strongest to weakest masking as the whale and with similar call-detection thresholds. The neural network slightly outperformed the humans. Both models were then used to predict the masking of a fourth type of noise, Gaussian white noise. Their prediction ability was judged by returning to the aquarium to measure masked-hearing thresholds of a beluga in white noise. Both models and the whale identified bubbler noise as the strongest masker, followed by ramming, then white noise. Natural ice-cracking noise masked the least. However, the humans and the neural network slightly overpredicted the amount of masking for white noise. This is neglecting individual variation in belugas, because only one animal could be trained. Comparing the human model to the neural network model, the latter has the advantage of objectivity, reproducibility of results, and efficiency, particularly if the interference of a large number of signals and noise is to be examined. © 2000 Acoustical Society of America. [S0001-4966(00)01007-9]

PACS numbers: 43.66.Dc, 43.66.Gf, 43.80.Lb [WA]

INTRODUCTION

Acoustic interference of noise with sound signals is an experience of our daily life. Noise obscures or masks signals, making it more difficult or even impossible to detect signals important to us. Masking is defined as the process or amount by which the threshold of audibility for one sound is raised by the presence of another sound.¹ Defining SL_0 as the detection threshold of a signal in the absence of noise (measured in dB) and SL_n as the detection threshold of the same signal in the presence of masking noise, then the masking M is expressed as the shift in threshold²

$$M = SL_n - SL_0 \text{ [dB]}. \quad (1)$$

There are two types of masking: simultaneous masking (when signal and noise occur at the same time) and nonsimultaneous masking (backward masking and forward masking). The physiological processes responsible for masking are very complex and still not fully understood.^{3,4} Masking is strongest when signal and noise contain the same or very similar frequencies. The frequency selectivity of the auditory system, i.e., the ability to resolve the sinusoidal components of a complex sound, plays an important role. The mammalian auditory system can generally be represented as a series of overlapping bandpass filters. The critical bandwidth is a measure of the width of the auditory filters;⁵ Fletcher hypothesized that at detection threshold, the intensity of the signal

equaled the intensity of the noise in the corresponding auditory filter (equal-power assumption).

Our understanding of masking is based on studies of the human auditory system, which have been carried out for many decades. Research on masking in marine mammals is fairly recent. In a few projects the signal and the noise were simple acoustic sounds, such as pure tones and white noise.⁶⁻¹³ Erbe and Farmer¹⁴ studied masking with underwater sounds as they occur in a marine mammal habitat. With a trained beluga whale, masked hearing data were collected in a go/no-go paradigm according to a titration (staircase) method. The signal was a typical beluga call, frequently used by the population in Maxwell Bay. The noise was icebreaker noise and natural ice-cracking noise. Figure 1 shows power-density spectrograms of the sounds. Bubbler noise was recorded from the bubbler system of the Canadian Coast Guard icebreaker HENRY LARSEN. This system blew high-pressure air into the water in order to push floating ice debris away. Ramming noise was propeller cavitation noise recorded from the same vessel during ice ramming, particularly when an ice ridge withstood the ship's force and stopped the ship despite the propeller still turning at full rpm. Sound of single ice-cracking events caused naturally by local temperature and pressure fluctuations was used for comparison. Results were that bubbler noise masked strongest. The call-detection threshold lay at a critical noise-to-signal ratio, NSR, of 15.4 dB (signal-to-noise ratio (SNR)-15.4 dB). The NSR was chosen over the SNR, because in the original experimental procedure, the noise was the variable and was increased re-

^{a)}Electronic mail: erbec@pac.dfo-mpo.gc.ca

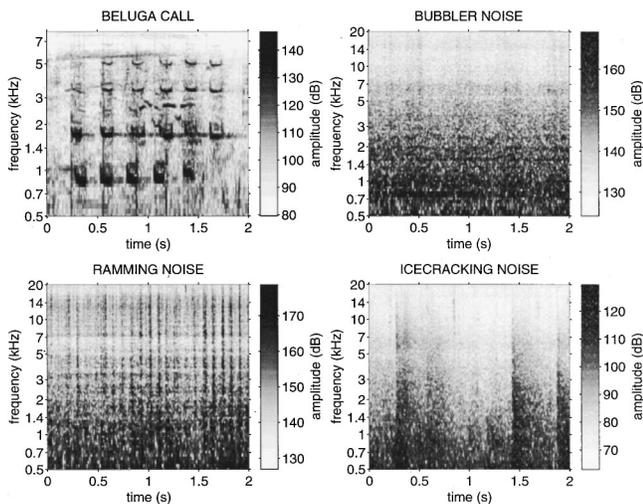


FIG. 1. Power density spectrogram of the beluga call, an icebreaker's bubbler and propeller (ramming) noise, and natural ice-cracking noise in dB re $1 \mu\text{Pa}^2/\text{Hz}$ @ 1 m. A source level of 160 dB re $1 \mu\text{Pa}$ @ 1 m was assumed for the beluga call. The source levels of the noises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

flected in increasing NSR. Ramming noise followed with a critical NSR of 18.0 dB. Ambient ice-cracking noise masked the least with a critical NSR of 29.0 dB.

Controlled experiments with captive marine mammals are both time- and cost expensive. It would be far more efficient if a fast, ground-truthed model for masking experiments existed. The interference of a large variety of anthropogenic noise with signals used by animals could be studied more easily. The literature on the mathematics of signal detection in noise is vast. A few studies have applied standard techniques to the problem of finding whale calls in noise.^{15–18} These studies compared techniques under the aspect of achieving the highest hit rate under the smallest false-alarm rate. Erbe *et al.*¹⁹ compared techniques with respect to the order and level of the maskers as determined by masking experiments with a beluga whale in an earlier study.¹⁴ Various implementations of matched filters, spectrogram cross-correlators, and an artificial neural network detected the same beluga call as the whale in the same three noises, Fig. 1. The neural network, trained with backpropagation,²⁰ modeled the beluga hearing data best, leading to the same order of masking noises and similar critical NSRs.

The fact that beluga calls and much of the noise in their habitat fall into the hearing range of humans poses the interesting question of whether human listening experiments could substitute for beluga experiments efficiently and with sufficient accuracy. In this article, I compare masked-hearing experiments with human subjects to the earlier beluga experiments¹⁴ using the same signal and noise stimuli. The human results and the neural network developed previously¹⁹ are then used to predict the degree of masking that a fourth type of noise, Gaussian white noise, would have on the beluga call. Performance is compared and rated by going back to the aquarium and measuring masked-hearing thresholds of the beluga whale in white noise.

I. METHODS AND RESULTS

A. Experiments with human subjects

The sounds used in this study were recorded digitally with 16-bit resolution and a sampling frequency of 44 kHz.¹⁴ The beluga call chosen was about 2 s long. A sample of equal length was selected from each of the three noises, bubbler, ramming, and ice-cracking noise. In order to study the relative degree of masking, all four sounds were normalized by their root-mean-square (rms) pressure. Sounds were then digitally mixed by adding their time series according to

$$x[t] = s[t] + \alpha n[t]. \quad (2)$$

Here, $x[t]$ represents a 2-s time series of a mixed sound; $s[t]$ is the 2-s time series of the call; $n[t]$ is the time series of one of the three noises. The factor α denotes the noise-to-signal ratio (NSR). NSRs of 0, 6, 12, 18, 24, 30, and 36 dB were computed. The mixed sounds $x[t]$ were subsequently normalized by their rms pressure. Therefore, for increasing NSR, the noise in a mixed sound would get louder and the call quieter at the same time.

The sounds were digitally stored on a notebook computer. They were converted into analog form by an external soundcard (PORTABLE Sound Plus™ from Digispeech) and amplified by a 20-W audio amplifier (Realistic model MPA-30). Each human subject sat in a quiet room wearing headphones connected to the amplifier. First, the volume control was adjusted such that the normalized sound was comfortably loud. I then determined the call-detection threshold in the absence of noise for each subject using a titration (staircase) method²¹ similar to the one described below for the masked-hearing study. This was to ensure that the signal content in mixed sounds with an NSR of 36 dB was at least 6 dB above the pure call-detection threshold. This way, masked-hearing thresholds would not be influenced by the call content dropping below audibility.

Masked-hearing data also were collected in a titration method. The subjects were asked to say “yes” when they detected a signal in the noise and “no” if they perceived only noise. Subjects received feedback after each trial about whether a signal had been present or absent. Starting with a high call content, mixed sounds were played at increasing NSR in steps of 6 dB as in the early beluga experiment.¹⁴ The subjects were given 2 s to respond. If the subject missed a signal, I stepped back to the previous NSR. If the subject successfully detected a signal, the NSR was stepped down. This staircase method was continued until three consecutive lower reversal points had the same NSR or until the total number of trials exceeded 40. One such staircase was run for each of the three noises; all data were collected in one session. To reinforce the no response and to get information about the subject's response bias, one out of four signals was a catch trial (pure noise). According to the titration procedure, the masked-hearing threshold was calculated as the mean of all (upper and lower) reversal points²¹ plus/minus one standard deviation. As is generally done, this ignores the fact that dB is a logarithmic ratio and thus does not give symmetric confidence intervals about the mean. With T_i

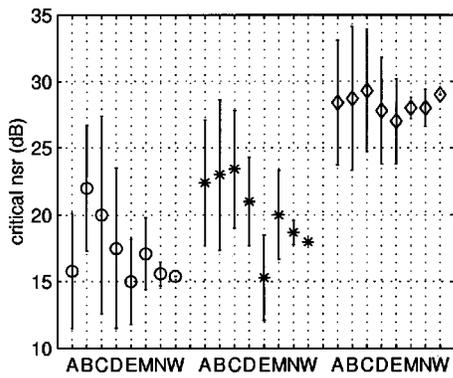


FIG. 2. Masked-hearing thresholds of five human listeners (A, B, C, D, and E) and their mean (M), compared to a neural network¹⁹ (N) and the beluga whale Aurora¹⁴ (W) for a beluga call buried in three types of noise: bubbler (O), ramming (*), and icecracking (◇).

$\pm\sigma_i$ being the threshold of one human subject, the average human threshold \bar{T} was calculated as the weighted mean²²

$$\bar{T} = \frac{\sum_{i=1}^5 \frac{T_i}{\sigma_i^2}}{\sum_{i=1}^5 \frac{1}{\sigma_i^2}}. \quad (3)$$

The standard deviation of the mean threshold was calculated as the average weighted variation of the individual thresholds around the mean²²

$$\sigma_{\bar{T}} = \sqrt{\frac{\sum_{i=1}^5 \frac{1}{\sigma_i^2} (T_i - \bar{T})^2}{\sum_{i=1}^5 \frac{1}{\sigma_i^2}}}. \quad (4)$$

Figure 2 summarizes the data for the human subjects and compares them to the neural network¹⁹ and the beluga whale Aurora.¹⁴ Results show that all five human subjects classified the three noises in the same order as the neural network and the whale: Bubbler noise was identified as the strongest masker, followed by ramming, then natural ice-cracking noise. Maximum deviation amongst the human listeners was 8.1 dB in the case of ramming noise. The probability of false alarms was very small for all human subjects. Subjects A and E gave no false alarms for any of the noises. Subject B had false-alarm rates of 0.25, 0, and 0.13 for bubbler, ramming, and ice-cracking noise, respectively. For Subject C, the corresponding values were 0.25, 0, and 0.14; for Subject D, 0.28, 0, and 0. With such “conservative” bias, calculated thresholds can be expected slightly shifted towards smaller NSRs than if the subjects had adopted a more “liberal” (risky) attitude.

With the exception of Subject B’s bubbler threshold and Subject C’s ramming threshold, the whale’s thresholds lay within one standard deviation of the human thresholds. This indicates that, at least for these types of signal and noise, human experiments could substitute for beluga experiments. The mean plus/minus one standard deviation of the human responses correctly surrounded the beluga thresholds for

bubbler and ramming noise. For ice-cracking noise, however, the beluga threshold fell just outside the interval. In fact, the beluga threshold lay at a higher NSR than the mean human threshold. This could indicate that beluga whales can detect their own calls in this type of ambient arctic noise better than the average human. This might point at a masked-hearing specialization of these animals living in a naturally noisy environment. On the other hand, the standard deviation for the mean human ice-cracking threshold was very small, smaller than the deviations for the other two noises. This occurred because of a smaller individual variation for ice-cracking noise. With only five human subjects, this might be coincidental, and testing of further human subjects could change the mean and deviation. Similarly, one would expect an individual variation amongst beluga whales and testing of further animals would allow more solid conclusions.

The neural network’s thresholds approached the beluga’s thresholds more closely than the human thresholds did. For examining the masking of a large variety of signals and noise, it is thus the more efficient technique. The neural network further poses the advantage of objectivity with its results being reproducible and independent of a potentially changing response bias.

During and after the experiment, all human subjects reported a “frustration” with the experimental procedure. The problem identified was that all sounds played had an equal length of 2 s and the call always happened at the same time in the noise. The initial reason for adding signal and noise this way was to fix a phase between them. With signal and noise exhibiting temporally varying spectral characteristics, masking was expected to depend on the time lag between signal and noise. Obviously, in the case of a pulsed signal and a pulsed noise, masking would be minimal if the pulses of the signal happened in between the pulses of the noise, and maximal if the pulses of signal and noise were in phase. Fixing the phase, however, meant that the subjects knew when to expect the signal. For large NSRs and occasionally for pure noise, subjects reported a frustration about the conflict between whether they thought they heard a signal only because their mind knew what it would sound like and when it would happen or whether they actually heard it. I therefore modified the experiment to include two more beluga calls with very different spectral characteristics.²³ Human subjects were asked not to simply say “yes” if any signal was detected but to identify which of the three signals was detected. The two subjects tested still had a problem with the procedure, because they perceived all three signals simultaneously at large NSRs or pure noise. It was then decided to modify the experiment such that noise would play continuously and the signal would be injected at random times so subjects did not receive any onset cues.

B. Modified human experiments

The experimental setup differed only slightly from the previous one. Two portable computers were used with the same soundcards. One computer stored 2-s sound files of the beluga call at varying sound-pressure levels. The other computer stored 15 min of pure noise. The analog outputs of the soundcards were connected to the two line-ins of the audio

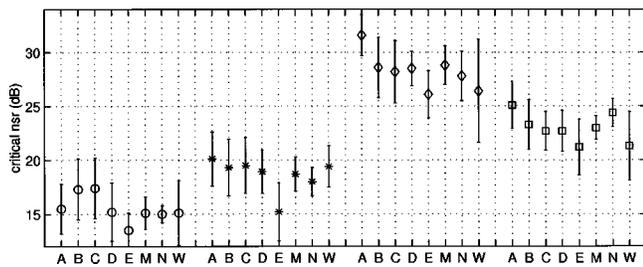


FIG. 3. Masked-hearing thresholds of five human listeners (A, B, C, D, and E) and their mean (M), compared to a neural network (N) and the beluga whale Aurora (W) for a beluga call buried in four types of noise: bubbler (○), ramming (*), icecracking (◇), and Gaussian (□). Modified experimental procedure.

amplifier, which mixed them at a volume ratio of 1:1 and sent a mixed sound to the headphones. Rather than taking a 15-min noise recording, the 15-min noise files were created by repeating the previous 2-s noise file 450 times. This way, no matter when the call happened in the noise, the NSR defined as the ratio of rms voltages was constant over the 2-s duration of the call. NSRs varied in steps of 3 dB, yielding a less coarse analysis than with the previous 6 dB. A fourth type of noise, Gaussian white noise, was created digitally and assigned the same bandwidth of 22 kHz as the other three noises. While the noise played continuously, beluga calls were inserted according to a titration method starting out with a loud volume and decreasing until the human missed a signal, then stepping up again, and so forth. The subjects received feedback after each trial. Every fourth trial was a signal-absent trial. A session (series of trials) ended after the lower reversal point had been constant three times in a row or if the total number of trials exceeded 40. If this took longer than 15 min, the noise playback was restarted after a brief 2-s break.

Results are summarized in Fig. 3. All five human subjects classified bubbler noise as the strongest masker, followed by ramming noise, then Gaussian white noise, then natural ice-cracking noise. Thresholds were slightly different from before. For all five subjects, the critical NSRs for bubbler and ramming noise decreased. For three out of five subjects, the critical NSR for ice-cracking noise decreased, while it increased for the other two. The mean ice-cracking threshold increased. These shifts, however, were small. A test of statistical significance was done on the difference of means, using 98 confidence intervals and student-*t* probabilities [Eq. (13.14a)²⁴]. With the exception of Subject B's bubbler thresholds, all shifts in threshold were not statistically significant. Individual variation amongst the five humans was maximal with 5.5 dB in the case of ice-cracking noise and hence smaller than previously. An analysis of catch trials revealed a strongly conservative attitude in all subjects with a probability of false alarms close to zero throughout the experiment. In both human experiments, Subject E consistently provided thresholds at lower NSRs than other human subjects and also had the lowest false-alarm rate. Based on the conclusion from the previous experiment, that human listeners can successfully model beluga masking experiments at least with the signal chosen and bubbler, ramming, and

ice-cracking noise, the human data predict a beluga masked hearing threshold of 23.0 ± 1.1 dB for white noise.

C. Modified beluga experiments

The experimental setup was identical to the previous beluga experiment.¹⁴ At the Vancouver Aquarium, the beluga whale Aurora had been trained for a yes/no response according to a behavioral go/no-go paradigm. The instrumentation was the same as that used for the human experiments described in this article with the exception of an Argotec J9 underwater sound projector (with a bandwidth of 40 Hz–20 kHz) replacing the headphones. Noise played continuously for 15 min in the experiment pool. Aurora was asked to station against a stationing bar 1 m in front of the J9. At random times within 30 s after stationing, the beluga call was injected into the noise at a certain NSR according to a titration method with a step size of 3 dB. Aurora would break away from the stationing bar upon call detection. Otherwise, she would hold station and be recalled at the end of the 30-s period. She was given 2 s to respond on the signal-present trials; her reaction time was usually 1 s or less. Catch trials made up one-fourth of all trials. The animal received feedback after each trial, being rewarded for correct rejections and correct detections. Sessions were kept short, limited to 20 trials. About four sessions were run per noise.

Results are included in Fig. 3. Aurora's thresholds were similar to the ones from the previous experiment with the old critical NSRs falling into one standard deviation of the new NSRs. False-alarm rates were less than 0.1. Aurora classified the four noises in the same order, from strongest to weakest masking, as did the humans. Aurora's thresholds fell into one standard deviation of most of the individual human responses. For bubbler and ramming noise, Aurora's thresholds fell within one standard deviation of the average human threshold. For ice-cracking noise, however, the mean human threshold was previously smaller than and now larger than Aurora's threshold. As far as the predictability of the degree of masking of white noise is concerned, results of the human experiment slightly overestimate the beluga threshold. The discrepancy, on the other hand, can be considered small if one takes Aurora's standard deviations into account. In this case, the thresholds of Aurora and the humans overlap for all four noises.

Figure 4 illustrates masking as a shift in threshold. The *x* axis denotes frequency on a logarithmic scale; the *y* axis gives sound-pressure levels in dB *re* 1 μ Pa @ 1 m. The gray shaded area is the area underneath the beluga audiogram. Sound falling into this area is considered not audible to beluga whales. This audiogram is the mean of seven published beluga audiograms.^{9,14,25,26} The top solid line is the spectrum of the noise as it was played continuously for 15 min in the experiment pool. Plotted are 12th-octave band levels corresponding to the width of the beluga critical bands in this frequency range, listed in Table I in Erbe *et al.*¹⁹ All noises had a broadband noise level of $NL = 160$ dB *re* 1 μ Pa @ 1 m. The lower solid line represents 12th-octave band levels of the call at threshold in the absence of noise; the broadband signal level was¹⁴ $SL_0 = 108$ dB *re* 1 μ Pa @ 1 m. The middle line gives band levels of the call at threshold in the presence

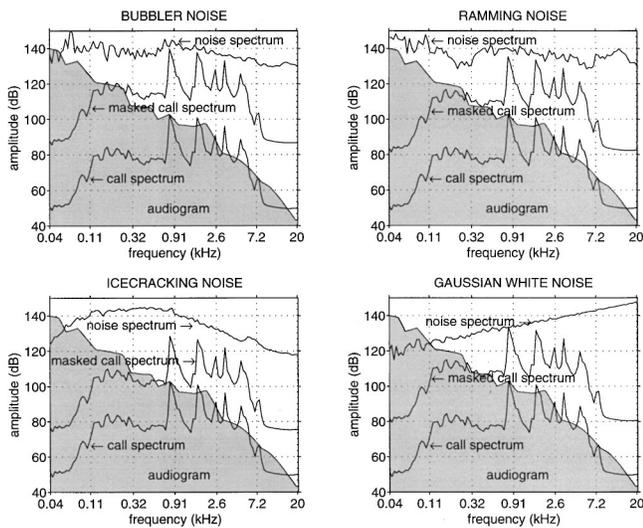


FIG. 4. Shift in threshold of the call from noise absence (call spectrum) to noise presence (masked call spectrum). Values for masking are 36.9 dB for bubbler, 32.6 dB for ramming, 25.6 dB for ice cracking, and 30.7 dB for white noise.

of noise. The amount of masking M can thus be read as the upwards shift of the masked call spectrum compared to the call spectrum. From Eq. (1), it can also be calculated as

$$M = SL_n - SL_0 = NL - NSR_c - SL_0, \quad (5)$$

where NSR_c are the critical NSRs from Fig 3. In dB, the NSR is the noise level minus the signal level. The masking was 36.9 dB for bubbler noise, 32.6 dB for ramming noise, 25.6 dB for ice-cracking noise, and 30.7 dB for white noise.

There are a few important conclusions to be drawn from Fig. 4. First, in the absence of noise, the animal stopped responding “yes” to the call as soon as the low-frequency peaks dropped below audibility. The animal should, however, still have heard the call components at 3 and 5 kHz to much lower signal levels. I therefore like to call the call spectrum at threshold the “call recognition level” rather than the “call detection level.” This has important implications for masking in the wild. I hypothesize that an animal would be able to hear parts of a call over long ranges but recognize the call over shorter ranges.

In the presence of noise, for bubbler and ramming noise, the masked call spectrum was shifted upwards such that the three major peaks of the call just “touched” the noise spectrum. Spectrum levels plotted were means over the 2-s duration of the sounds. As can be seen from Fig. 1, the call consisted of six spectrally similar pulses (phonemes) which together occupied only about half of this time interval. In fact, the mean spectrum of the phonemes is about 6 dB higher than the mean over 2 s. Therefore, at the masked-hearing threshold, the power of the call equaled the power of the noise in the three critical bands surrounding the major peaks in the call. This corroborates Fletcher’s equal-power assumption.⁵ For Gaussian white noise, the two low-frequency peaks of the masked call spectrum just surpassed or equaled the noise power in the corresponding critical bands. This might indicate that the low-frequency components are more important than the high-frequency compo-

nents for signal detection. For ice-cracking noise, the plot is misleading. This noise consisted of loud pulses occurring about twice a second. As concluded in an earlier study,¹⁴ masking for ice-cracking noise was so low because the animal (and humans) could identify the call from the phonemes falling in between the pulses of the noise. It is this important temporal structure of the intermittent ice-cracking noise that is not represented in mean noise levels over 2 s. From Fig. 1, in between the pulses of nearby ice-cracking events, the spectrum levels of ambient ice-cracking noise drop at least 20 dB. From Fig. 4, for a phoneme falling in between two noise pulses, five out of the six call peaks would surpass or equal the noise power in the corresponding critical bands if one lowered the noise level by 20 dB.

D. Modified neural network

A backpropagation neural network was previously proven to successfully model beluga masked-hearing experiments.¹⁹ The same neural net design was chosen to model the modified experiments with continuous noise. This neural network was a fully connected two-layer network with 400 input neurons, three hidden layer neurons, and one output neuron. It was trained to detect features of the beluga spectrogram in Fig. 1, buried in noise. As previously, random and sinusoidal noise was used during the training phase. The network gave an output close to 1 if it easily detected a signal and an output close to 0 if it failed to detect a signal. After training had been completed, the network was presented with spectrograms of the same mixed sounds as the beluga and the humans in this study. Critical NSRs were taken at an output of 0.5 resembling a 50% hit rate. As the time lag between the call and the noise was random in the modified experiment, thresholds of the humans and the whale represented means over all time lags encountered. Critical NSRs for the neural network were therefore averaged over 40 different time lags. This was done in the following way. While creating mixed sounds for the neural network, the time series of the noises were shifted against the time series of the call in steps of 40 ms. Figure 5 shows how the neural network’s critical NSRs varied as a function of time lag between signal and noise. The plot can be wrapped around the x axis, i.e., the NSRs at the 41st time lag equal those at the first time lag. The maximum variation for bubbler and white noise was about 3 dB, which was small due to the fact that the spectral characteristics of these two noises didn’t vary with time. Ramming noise exhibited a strong time dependence with sharp pulses occurring regularly every 90 ms. This time scale, however, was smaller than the time scale of the beluga call, where pulses occurred every 250 ms lasting for about 150 ms. The pulses (phonemes) of the beluga call therefore encountered the same amount of ramming noise relatively independent of the time lag. For ice-cracking noise the situation was different. Its pulses occurred irregularly and lasted a few hundred ms. Masking depended largely on the time lag between the noise pulses and the call phonemes. When they coincided, masking was greater; when they were out of phase, masking was less. Figure 5 shows six major maxima in the critical NSRs for icecracking noise. They are about 250 ms apart, as are the call phonemes.

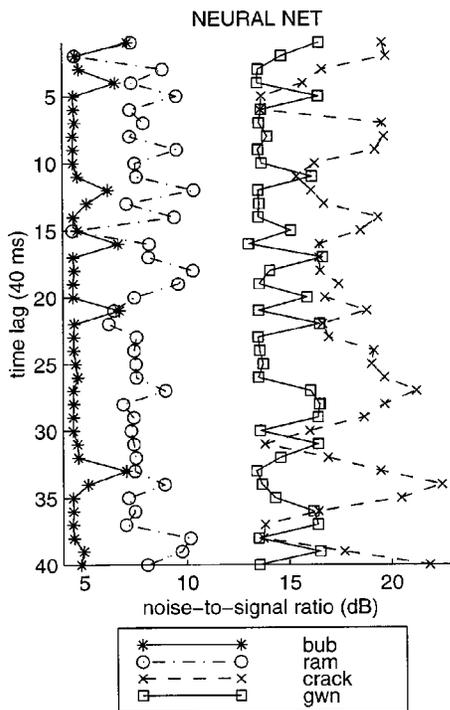


FIG. 5. Critical NSRs of the neural network for bubbler, ramming, ice cracking, and Gaussian white noise. The time lag between the call and the noises was changed in steps of 40 ms, indicating a strong dependence on when the call happened in the noise only for ice-cracking noise.

The mean critical NSRs over all time lags were 4.9, 7.9, 14.6, and 17.7 dB for bubbler, ramming, white, and ice-cracking noise, respectively. The standard deviations were 0.8, 1.3, 1.3, and 2.3 dB, respectively. As explained earlier,¹⁹ the neural network (like other software techniques) required the setting of an offset. Otherwise, thresholds would only have relative meaning. Based on earlier results that this type of neural network can successfully model beluga masking experiments with bubbler, ramming, and ice-cracking noise,¹⁹ I determined the offset using these three noises. Adding a constant 10.1 dB to the network's critical NSRs yielded Aurora's thresholds with a minimum sum-squared error. The neural network was then asked to predict the degree of masking of white noise. Results are printed in Fig. 3. The neural network slightly overestimated the degree of masking of white noise for beluga whales.

II. SUMMARY AND CONCLUSION

The purpose of this study was to find a model for beluga masked-hearing experiments which could accurately, reliably, and efficiently predict the degree of interference of anthropogenic noise with beluga communication signals. Masked-hearing thresholds of a beluga call in two types of icebreaker noise (bubbler system and propeller cavitation/ramming noise) and ambient arctic ice-cracking noise had previously been measured with a trained beluga whale.¹⁴ A variety of software models had subsequently been developed identifying a backpropagation neural network as the most suitable model for beluga masking experiments.¹⁹ Given that beluga calls and most of the noise in their habitat fall into the hearing range of humans, this article hypothesized that hu-

man listening experiments could estimate results of beluga experiments yielding data much faster without lengthy training. Five human subjects participated in this study. Their performance was compared to that of the whale¹⁴ and the neural network.¹⁹

The humans, the whale, and the neural network classified the three noises in the same order, identifying bubbler noise as the strongest masker, followed by ramming, then natural ice-cracking noise. The mean plus/minus one standard deviation of the human responses correctly encompassed the beluga thresholds for bubbler and ramming noise, and were slightly lower than the beluga's threshold for ice-cracking noise. The beluga threshold fell within one standard deviation of the neural network results in the case of all three noises. The neural network therefore modeled the beluga data better than did the humans. The human subjects identified a psychoacoustic problem with the experiment procedure. All sounds played had an equal length of 2 s and the call always happened at the same time in the noise. Subjects therefore knew when to expect the call and had difficulty deciding whether they only imagined it at high NSRs.

I therefore changed the experiment procedure such that noise played continuously for 15 min and the call was injected at random times. Thresholds were not statistically significantly different compared to the previous experiment and were in fact within one standard deviation of the previous thresholds. Humans were then asked to detect the beluga call in a fourth type of noise, Gaussian white noise. The mean threshold was used to predict the masking of white noise in beluga whales. The human experiment predicted that masking of white noise would be stronger than that of ice-cracking noise but less than that of ramming noise.

I returned to the aquarium to measure beluga masked-hearing thresholds with the modified experimental procedure (continuous noise). The human experiment correctly predicted masking to lie in between that of ramming and ice-cracking noise. A comparison of spectrum levels of the masked call at threshold and the noise corroborated Fletcher's equal-power assumption⁵ in that the call was just masked when the power in the critical bands surrounding the major call frequencies equaled the power of the noise. The animal attended strongly to the lower-frequency peaks of the call and stopped responding as soon as these were masked by noise or fell below the audiogram in the absence of noise. Higher-frequency peaks of the call were likely still audible. This fits in with the idea of call recognition rather than call detection.

The backpropagation neural network from an earlier study¹⁹ was used to model the modified masking experiment with continuous noise. The network nicely illustrated how masking depended on the time lag between signal and noise. For ice-cracking noise, which had a temporal structure similar to the pulsed beluga call, masking was maximum when the call phonemes coincided with the noise pulses and minimum if the call phonemes fell in between. The neural network ordered the noises the same way as the human subjects and the whale. As with the human experiment, the neural network slightly overpredicted the amount of masking for white noise in beluga whales.

Unfortunately, due to time constraints at the Vancouver Aquarium, only one animal could be trained for the experiment. A variation between individuals in masked-hearing thresholds of a few dB can be expected. The testing of further animals would thus allow a better judgment of the ability of human and software models to estimate masking in beluga whales. Comparing the neural network model to the human model, the neural network has the advantage of being independent of subjectivity and of giving easily reproducible results.

ACKNOWLEDGMENTS

First of all, I wish to thank the Vancouver Aquarium for supporting this research by providing animals and trainers. Without the help, patience, good will, cheerful attitude, and commitment of Michelle Brown, Indy Canagaratnam, Dennis Christen, Joanne Cottrell, Jeremy Fitz-Gibbon, Christine Fritzsche, Kyle Jenkins, Sascha Melnechuck, Jane Osen, Christopher Porter, Alysoun Seacat, Todd Shannon, Brian Sheehan, Gwyneth Sheppard, Clint Wright, and Anne Young, the beluga experiments would simply not have taken place. The marine mammal staff miraculously squeezed my experiments into their already busy daily schedule. Thanks to Andrew, Kuan-Neng, Frank, Roblyn, and Chad for lending me their ears. Dr. Sue Cosens and Dr. Larry Dueck kindly gave me their recordings of beluga whale calls. Dr. Yunbo Xie recorded the icebreaker noise and ambient arctic noise. Last but not least, thanks to Patrice St-Pierre and the Canadian Coast Guard, Central and Arctic Region, for continued financial support.

¹American Standards Association "Acoustical Terminology S1, 1-1960" (American Standards Association, New York, 1960).

²N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90–119 (1947).

³J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic, San Diego, 1988).

⁴B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego, 1997).

⁵H. Fletcher, "Auditory patterns," *Rev. Mod. Phys.* **12**(1), 47–65 (1940).

⁶C. S. Johnson, "Masked tonal thresholds in the bottlenosed porpoise," *J. Acoust. Soc. Am.* **44**, 965–967 (1968).

⁷C. S. Johnson, "Auditory masking of one pure tone by another in the bottlenose porpoise," *J. Acoust. Soc. Am.* **49**, 1317–1318 (1971).

⁸V. I. Burdin, V. I. Markov, A. M. Reznik, V. M. Skornyakov, and A. G. Chupakov, "Ability of *Tursiops truncatus* Ponticus Barabasch to distin-

guish a useful signal against a noise background," in *Morphology and Ecology of Marine Mammals*, edited by K. K. Chapkii and V. E. Sokolov (Wiley, New York, 1973), pp. 162–168.

⁹C. S. Johnson, M. W. McManus, and D. Skaar, "Masked tonal hearing thresholds in the beluga whale," *J. Acoust. Soc. Am.* **85**, 2651–2654 (1989).

¹⁰W. 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).

¹¹J. 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.

¹²A. 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* (Ref. 11), pp. 385–393.

¹³A. Y. Supin, V. V. Popov, and V. O. Klishin, "ABR frequency tuning curves in dolphins," *J. Comp. Physiol. A* **173**(5), 749–656 (1993).

¹⁴C. Erbe and D. M. Farmer, "Masked hearing thresholds of a beluga whale (*Delphinapterus leucas*) in icebreaker noise," *Deep-Sea Res., Part II* **45**, 1373–1388 (1998).

¹⁵D. K. Mellinger, "Handling time variability in bioacoustic transient detection," *Oceans '93, Engineering in Harmony with the Ocean, Proceedings IEEE Vol. III*, pp. 116–121 (1993).

¹⁶D. K. Mellinger and C. W. Clark, "A method for filtering bioacoustic transients by spectrogram image convolution," *Oceans '93, Engineering in Harmony with the Ocean, Proceedings IEEE Vol. III*, pp. 122–127 (1993).

¹⁷J. 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).

¹⁸D. K. Mellinger and C. W. Clark, "Methods for automatic detection of mysticete sounds," *Mar. Freshwater Behav. Physiol.* **29**, 163–181 (1997).

¹⁹C. Erbe, A. R. King, M. Yedlin, and D. M. Farmer, "Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*)," *J. Acoust. Soc. Am.* **105**, 2967–2978 (1999).

²⁰D. 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.

²¹W. W. L. Au, *The Sonar of Dolphins* (Springer, New York, 1993).

²²S. B. Richmond, *Statistical Analysis*, 2nd ed. (Ronald, New York, 1964).

²³C. Erbe, "The masking of beluga whale (*Delphinapterus leucas*) calls by icebreaker noise," Ph.D. thesis, Earth & Ocean Sciences, University of British Columbia, Canada (1997).

²⁴R. L. Mason, R. F. Gunst, and J. L. Hess, *Statistical Design and Analysis of Experiments* (Wiley, New York, 1989).

²⁵M. 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, 1978).

²⁶F. 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).