THE MASKING OF BELUGA WHALE *(DELPHINAPTERUS LEUCAS)* VOCALIZATIONS BY ICEBREAKER NOISE

By

Christine Erbe
M. Sc. (Physics) University of Dortmund, Germany

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
EARTH AND OCEAN SCIENCES

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
November 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Earth & Ocean Sciences

The University of British Columbia
Vancouver, Canada

Date Dec. 22, 1997

Abstract

This thesis examines the masking effect of underwater noise on beluga whale communication. As ocean water is greatly opaque for light but well conducting for sound, marine mammals rely primarily on their hearing for orientation and communication. Man-made underwater noise has the potential of interfering with sounds used by marine mammals. Masking to the point of incomprehensibility can have fatal results–for the individual, but ultimately for the entire species. As part of our understanding of whether marine mammals can cope with human impact on nature, this thesis is the first to study the interference of real ocean noises with complex animal vocalizations.

At the Vancouver Aquarium, a beluga whale was trained for acoustic experiments, during which masked hearing thresholds were measured. Focus lay on noise created by icebreaking ships in the Arctic.

As experiments with trained animals are time and cost expensive, various techniques were examined for their ability to model the whale's response. These were human hearing tests, visual spectrogram discrimination, matched filtering, spectrogram cross-correlation, critical band cross-correlation, adaptive filtering and various types of artificial neural networks. The most efficient method with respect to similarity to the whale's data and speed, was a backpropagation neural net.

Masked hearing thresholds would be of little use if they could not be related to accessible quantities in the wild. An ocean sound propagation model was applied to determine critical distances between a noise source, a calling whale and a listening whale. Colour

diagrams, called *maskograms*, were invented to illustrate zones of masking in the wild. Results are that bubbler system noise with a source level of 194 dB re 1 μPa at 1 m has a maximum radius of masking of 15 km in a 3-dimensional ocean. Propeller noise with a source level of 203 dB re 1 μPa at 1 m has a maximum radius of masking of 22 km. A naturally occurring icecracking event with a source level of 147 dB re 1 μPa at 1 m only masks if the listening whale is within 8 m of the event. Therefore, in the wild, propeller cavitation noise masks furthest, followed by bubbler system noise, then icecracking noise.

Contents

List of Figures

List of Tables

Acknowledgments

The biggest hug goes to Dr. David M. Farmer, my supervisor, who–to me–had this lovely father-like character. He was extremely supportive, in every sense, and–last but not least–always found the right words of motivation when things didn't quite work out as expected. Thanks also to Dr. Matthew Yedlin, who supported me here at geophysics and to Dr. John K.B. Ford for doing the same at the aquarium. I would like to thank Dr. Richard A. Pawlowicz for jumping onto my supervisory committee two weeks before the departmental defense; and Dr. Tad J. Ulrych for working miracles in getting my public defense to run smoothly.

I am highly indebted to the Marine Mammal Staff of the Vancouver Aquarium, in particular Michelle Brown, Indrajit 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 for their patience during long, daily, sometimes tedious and often cold and wet data collection, for sacrificing their backs for science by holding the 17 kg transducer and stationing bar for 15 to 20 minutes at a time and last but not least for making time available for training and experiments in their already very busy schedule. Without the commitment of the above mentioned trainers, this study would not have been possible.

I would also like to thank Patrice St-Pierre and the Canadian Coast Guard for supporting an environmental study like this. Mankind usually learns too late about its impacts on nature and I hope that the Coast Guard can set an example for early concern about possible damage.

My thanks go out to the support staff at Earth and Ocean Sciences at UBC, and at the Acoustical Oceanography Research Group at the Institute of Ocean Sciences in Sidney, BC. I put Grace Kamitakahara-King and Richard Outerbridge through regular software license tragedies. Philippa Sumsion never (at least not obviously) lost her patience during endless bureaucratic and financial hassles. Big hugs! Although I spent way, way too little time with Dave Farmer's group on Vancouver Island, everybody there always instantly made me feel so much at home!

Thank you to my long-term, childhood friend Miriam Kroeske for drawing the two sketches for me after jokes told to me on conferences (after my presentations–if that is of any significance). Also, to all my other good old German friends who have maintained contact over this long distance and who reassured and believed in me when I initially doubted whether I shouldn't pack my suitcase again.

Love and kisses to my dear parents, Ingrid and Heinz Erbe, and my sweet parents-in-law, Joan and Mike King. When I left for Canada four years ago, I said they must not take it personally, but if time and money permitted, I would discover America rather than spend my vacations back home. Well, I still haven't seen anything of North America except

for a compass drawn circle of 200 km radius around Vancouver and the odd conference hotel. Instead, I went home every single year, and believe me, I didn't regret a minute of it. I've got the nicest two families a child can dream of and miss all of you very much.

And last but not least, all my love and thanks (the rest that's left) to the best hubby of all, Andrew. He had to be cook and cleaning man during hectical times, skater, skier, dancer and recipient of excess energy during happy times, comforter, masseur and recipient of excess energy during less happy times and whale during equipment testing times. If I don't stop, his supervisor will wonder how he wrote his own thesis at the same time...

To all living creatures

that once were, are currently, or will one day be

on the endangered species list.

May this thesis arouse awareness

of human impact on nature and ultimately

help to set up control regulations.

Chapter 1

Introduction

In the end we will conserve only what we love;

We will love only what we understand;

We will understand only what we have been taught.

–Baba Dioum

This thesis presents a novel approach to studying the interference of man-made underwater noise with marine mammal communication. The focus lies on beluga whales merely for reasons of convenience. Beluga whales were available for acoustic experiments at the Vancouver Aquarium. The methods designed and presented in this thesis can readily be transferred to other marine mammal species.

There are three major parts to this thesis. First, acoustic experiments with beluga whales were conducted to measure their hearing ability in noisy environments. Second, computer software was developed to model the whales' auditory process. This software

can be used to estimate the interference of underwater noise with beluga vocalizations. This method yields results much faster than costly training and experiments with animals. The third part comprises the application of ocean sound propagation models in order to relate the measured and computed data to the wild. Zones of masking around a noise source are plotted as a function of distances between a calling whale, a listening whale and the noise source.

In detail, Chapter 1 gives the motivation for this thesis and introduces the biological background. What are the effects of underwater noise on marine mammals? How do whales produce sound and how do they hear? What is masking? Chapter 2 presents the physical and acoustical background. How is sound measured? How does it propagate through the ocean? The animal vocalizations and underwater noises used in this thesis are introduced. Chapter 3 describes the whale training and the subsequent acoustic experiments. It also characterizes the acoustic properties of the aquarium pool. In Chapter 4, human subjects were asked to do the same task as the whale. Results led to important psychophysical modifications of the experiment procedure. Chapter 5 presents computer software developed to model the whale's hearing abilities. This continues through Chapter 6 which focuses on artificial neural network models. Chapter 7 picks up on the idea for experiment modification presented in Chapter 4. Animal and human subjects were "retrained", new data were collected, computer models were adjusted. Chapter 8 comprises ocean acoustic models to convert the measured hearing data to accessible quantities in the wild. At what distance does a noise source, e.g. an approaching ship, prevent two beluga

whales from communicating with each other? A graphic technique, called "maskogram", is developed to illustrate zones of masking around a noise source. Chapter 9 presents conclusions and ideas for future research.

This thesis is a greatly interdisciplinary work combining marine mammal biology with computer science and ocean acoustics. I've tried to give relevant background wherever appropriate for readers from any of these three fields. Different people may thus find different sections too long or maybe even superfluous. Finally, I would like to express that I do not see the value of my thesis in new advancements in neural network theory or in ocean sound propagation modeling, both of which are fields which still receive great attention and active research. In fact, the corresponding chapters are fairly basic employing techniques which have been around for some years. Rather, I see the value of my thesis in the unique combination of techniques from zoology, computer science and oceanography. The study of masking of complex marine mammal vocalizations by real underwater noise is novel to each of the three sciences.

The outcome of this work is the development of a systematic approach to assessing industrial noise interference on marine mammals. Previously, there has been no such systematic approach to masking available.

1.1 Motivation

In the era of decreasing richness of life due to disappearing animal species, knowledge of the manner in which mankind impacts on nature is indispensable. 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 impact posing threats to all marine life. Waste disposal as well as accidental outpourings or toxic chemical spills lead to water contamination and changes in water properties such as temperature and salinity. Offshore construction alters the physical characteristics of the local environment. Apart from such obvious chemical and physical effects, noise pollution plays an ever more serious role. Ships ranging from small pleasure boats to large transport vessels, oil production, mineral mining, construction, seismic exploration, ocean-acoustics research, all these sources emit sounds which have the potential of interfering with marine life.

1.1.1 Effects of Noise on Marine Mammals

Focusing on aquatic mammals, three distinct effects of noise can be identified. First, loud and sudden bursts of noise as from underwater explosions as well as prolonged or repeated exposure to high levels of non impulsive noise have the potential of causing permanent or temporary hearing loss. Unfortunately, no concrete data exist on noise

levels that cause *temporary threshold shift (TTS)* or *permanent threshold shift (PTS)* in marine mammals. Conclusions are often drawn from human hearing [Kryter 1985] or experiments with terrestrial mammals held under water [Yelverton *et al.* 1973]. Ears of dead whales which were probably killed due to blast injury have been dissected, though received sound levels could not be reconstructed [Ketten *et al.* 1993]. Only estimates of blast injury zones around underwater explosions exist [Ketten 1995].

Second, noise can disturb animals, i.e. change their current behavior, whether they were hunting, feeding, resting, vocalizing, mating etc. at the time and scare them away from their current location. Many people have observed *disturbance* reactions of marine mammals to various noise sources [Richardson *et al.* 1995, Ch. 9]. In cases without overt avoidance reactions, indicators for stress such as changes in mean durations of surfacings and dives, or the number of and the interval between respiratory blows have been measured [Richardson *et al.* 1985, 1986, 1990]. Nothing is known about any long-term consequences these disturbances may have on a population of animals. Repeated exposure to noise sources can lead to either *sensitization* or *habituation.* In the first case, animals react more and more strongly to repeated exposure to human activities [Richardson *et al.* 1995, Ch. 9.14.3]. In the latter case, animals eventually tolerate a noise source, although it is unknown whether the animals adapt to the noisy environment or whether they simply have nowhere else to go and remain stressed or otherwise adversely affected. An example for habituation might be the different degrees of tolerance of Arctic beluga populations and the St. Lawrence population. Arctic belugas show extremely large-scale avoid-

ance reactions to icebreakers at distances of about 50 km [LGL and Greeneridge 1986, Cosens and Dueck 1998, Cosens and Dueck 1993, Finley *et al.* 1990]. St. Lawrence belugas, however, which are much more exposed to heavy ship traffic, have been reported rather tolerant of large vessels [Pippard 1985, Sergeant 1986].

The third effect of noise on marine mammals is the *masking* of their communication signals. We know that most marine mammal species possess a complex acoustical communication system. Various delphinid sounds have been related to excitement, alarm, fright, threat, foraging, traveling, resting and copulation [Herman and Tavolga 1980]. Some baleen whales, such as the humpback and bowhead whale, sing complex songs during mating season which probably play a role in courtship [Tyack 1981, Ljungblad *et al.* 1982, Helweg *et al.* 1992]. In particular for dolphins, researchers have proposed the existence of signature whistles which broadcast individual identity and maintain contact between individuals [Caldwell and Caldwell 1965, Sayigh *et al.* 1990]. Masking of communication signals to the point of incomprehensibility can have fatal results. This is obvious in the case of emitted warning signals, but also for instance in the case that male and female can't "find" each other during mating season, because the oceans are too noisy. Depending on species, *communication* signals usually range from a few Hz to about 20 kHz [Richardson *et al.* 1995, Ch. 7], though higher frequency signals have recently been recorded [Lammers and Au 1996]. This coincides with the frequency range of most underwater noises [Wenz 1962, Richardson *et al.* 1995, Ch. 5,6]. Masking has mainly been studied for high frequency *echolocation* signals and white noise [Au 1993]. *Odonto-*

cetes (toothed whales and dolphins) emit high frequency sonar signals of up to 200 kHz [Richardson *et al.* 1995, Ch. 7]. In the wild, these signals are less affected by most underwater noises than are communication signals, because sonar occupies a much higher frequency band. Furthermore, excellent directional hearing abilities of dolphins for high frequencies [Renaud and Popper 1975, Zaitseva *et al.* 1980, Au and Moore 1984] can increase the signal-to-noise ratio considerably.

The only study on low frequency masking in whales I found addressed the interference of white noise with pure tones in a beluga whale [Johnson *et al.* 1989]. There are no data on low frequency directional hearing for marine mammals. It is assumed that the improvement of the signal-to-noise ratio due to directional hearing is poor for communication signals [Zaitseva *et al.* 1980]. From a pure physical point of view, directional hearing must be limited for low frequencies because of increasing wavelength. As mammalian hearing is highly nonlinear and depends on both frequency and temporal structure of signal and noise, results from pure tone experiments cannot be superposed to predict the masking of complex communication signals by structured noise. A data gap was identified showing the need to study low frequency masking with signals and noise "more similar" to those occurring in the ocean [Richardson *et al.* 1995]. This is the topic of my thesis.

For completeness it shall be mentioned that noise will also mask prey sounds or environmental sounds animals might listen to, though it is unknown to what extent they process such cues.

1.2 Biology Background

Marine Mammals (Class *Mammalia*) fall into three groups. The Order *Cetacea* comprises whales and dolphins; the Order *Carnivora* includes seals, sea lions, walrus, otters and polar bears. The Order *Sirenia* is formed by manatees and dugongs (sea cows).

Beluga whales (Species *Delphinapterus leucas*), together with narwhals, constitute the Family *Monodontidae*, which belongs to the Superfamily *Delphinoidea*. This in turn belongs to the Suborder *Odontoceti* (toothed whales). Together with *Mysticeti* (baleen whales), these two Suborders build the Order *Cetacea*.

1.2.1 Delphinoid Sound Production and Reception

Differing theories on the mechanisms of dolphin sound production and reception exist. A discussion would be beyond the framework of this thesis. A recent study [Aroyan 1996] applied geophysical techniques to a numerical simulation of biosonar sound emission and reception, and provides a list of further references. Currently, most researchers seem to agree that the site for sound production is not the larynx as in most mammals, but the nasal sac area. Aroyan fixed the locus of sound production near the monkey lip / dorsal bursae complex in the nasal region. The bony structure of the skull and the arrangement of air sacs in the dolphin head focus emitted high frequency echolocation signals forward. The fatty melon functions primarily as an impedance match for sound entering the water, but also as a focusing acoustic lens. In the case of sound reception, it is widely agreed that

dolphins hear not through their auditory meatus as terrestrial mammals do but through their lower jaw. Sound enters through a thinning in the jaw bone and is conducted to the auditory bullae via a fatty channel. The inner ear is similar in structure to our ears. The basilar membrane vibrates at different loci for different frequencies. The organ of Corti sits on the basilar membrane. Its hair cells convert mechanical energy into chemical energy, which in turn is converted to electrical energy. This electrical information is then transported to the brain via nerve fibres.

1.2.2 Beluga Audiogram

As indicated above, not much is known about beluga hearing which could be useful in a study of masking. An *audiogram* is a diagram showing hearing sensitivity as a function of frequency. I found six published beluga audiograms [White *et al.* 1978, Awbrey *et al.* 1988, Johnson *et al.* 1989]. Fig. 1.1 is a plot of the mean thresholds. The x-axis denotes the frequencies of pure tone sine waves on a logarithmic scale. The y-axis denotes the sound pressure level of the tones when they are just audible. It can be seen that the hearing of beluga whales is most sensitive in the frequency range between 20 kHz and 80 kHz. For comparison, an underwater audiogram of a killer whale [Bain 1992] and a bottlenose dolphin [Johnson 1967], and an average human audiogram in air [Sivian and White 1933] are plotted as well.

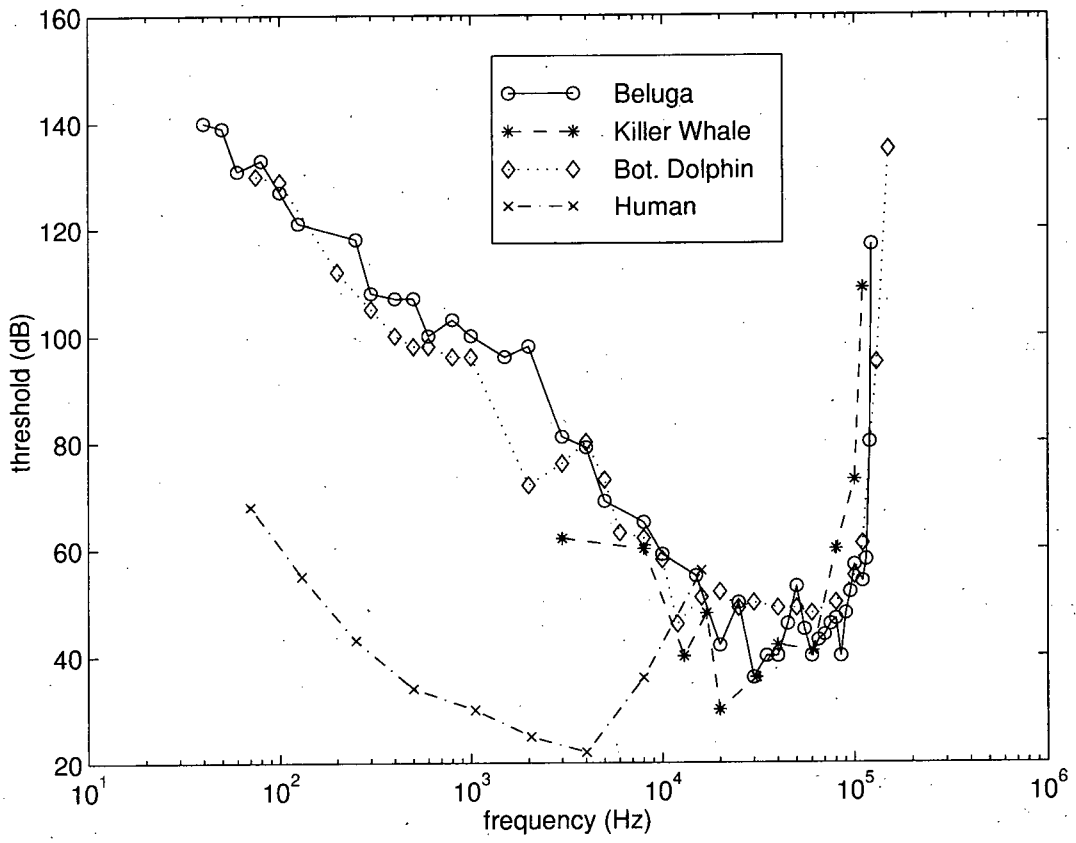


Figure 1.1: Underwater audiograms for three odontocete species in comparison to a human in-air audiogram. Thresholds are sound pressure levels in dB re 1 μPa.

1.2.3 Directivity Index

The *directivity index* in acoustics is a measure of the directionality of a receiving instrument. In psychoacoustics, it measures the directional hearing ability of an individual. The directivity index is defined as the logarithmic ratio of the power received by an omnidirectional receiver and the power received by a directional receiver:

$$\mathrm{DI} = 10 \cdot \log_{10} \frac{L_o}{L_d} \quad . \tag{1.1}$$

As an omnidirectional receiver listens in all directions, it generally receives more power than a selective, directional receiver. Therefore, the directivity index is a positive quantity.

To the best of my knowledge, directional hearing has not been measured for beluga whales. In my thesis, I assume that directivity indices of beluga whales would be similar to those published for bottlenose dolphins [Au and Moore 1984]. These were measured for 30 kHz, 60 kHz and 120 kHz. Extrapolating the data to lower frequencies yields a directivity index of about 0, i.e. no directivity, for frequencies of less than 10 kHz.

1.3 Masking

In human audiology, the *sensation level* of a sound is the attenuation needed to reduce the "loudness" to the threshold of hearing in the absence of noise. The *masking* is the shift in threshold due to the presence of noise. Let β_0 denote the level of a tone at threshold in the absence of noise, and β_n the level of the same tone when it is just audible in the presence of noise, then the masking M is

$$M = \beta_n - \beta_0 \quad , \tag{1.2}$$

when the quantities are measured in dB. The plot of M as a function of frequency is called the *masking spectrum* of the noise. From such masking spectra we know that in humans, low frequencies mask high frequencies more than the other way round. The closer the frequency of the test tone is to the frequency of the masker, the greater the masking. In tone-tone experiments where the noise is limited to one frequency, the masking exhibits a

local minimum when the frequency of the test tone is equal to the frequency of the noise. This is due to the tone and the masker producing beats [Wegel and Lane 1924]. This dip does not show up when the masker is a narrow-band white noise [Egan and Hake 1950]. In dolphins, interband masking (low frequencies masking high frequencies and vice versa) is slightly more symmetrical than in humans [Johnson 1971].

Masking can generally be attributed to three different components: 1) intraband masking due to noise in the same critical band as the desired signal, 2) interband masking due to noise in adjacent bands and 3) temporal masking due to components directly preceding the wanted signal [French and Steinberg 1947].

1.3.1 Critical Bands

Mammalian ears can generally be represented as a bank of overlapping, constant Q filters. The quality factor Q of a *bandpass filter* is defined as the ratio of center frequency and filter width, measured at half peak power (compare with Equation 3.6). Constant Q implies an increase in filter width with center frequency. In the case that broadband white noise masks a pure tone, critical bandwidth theory says that only the part of the noise falling into the filter around the pure tone is effective at masking. The only data available on filter characteristics of the beluga ear, are experimentally determined critical ratios [Johnson *et al.* 1989]. The *critical ratio* is defined as the ratio of the intensity of a pure tone I_t divided by the spectral intensity SI_n (intensity per Hz) of a broadband white noise at the level when the tone is just audible through the noise. It is generally

expressed in decibels:

$$\mathrm{CR} = 10 \cdot \log_{10} \frac{I_t}{SI_n} \quad . \tag{1.3}$$

For example, at 700 Hz, the critical ratio is about 18 dB according to Johnsons's experiment. At 6 kHz (these two frequencies will be important during neural network analysis, Chapter 6), the critical ratio is about 24 dB. It increases more rapidly for even higher frequencies. *Critical bandwidths* can be calculated from critical ratios under the equal-power assumption [Fletcher 1940]. The idea is that at detection threshold, the intensity of the tone equals the total intensity of the noise in the corresponding critical band:

$$I_t = SI_n \cdot \Delta f \quad .$$

Substituting into CR, yields

$$\mathrm{CR} = 10 \cdot \log_{10} \Delta f \quad ,$$

which can be solved for the critical bandwidth

$$\Delta f = 10^{\frac{\mathrm{CR}}{10}} \quad . \tag{1.4}$$

One thus obtains a critical bandwidth of 63 Hz for the center frequency of 700 Hz, and a width of 251 Hz at 6 kHz. On average, for low frequencies, the critical bandwidth is about 6 % of the center frequency. Or, in other words, the critical bands are about $\frac{1}{12}$th of an octave wide. Picking a center frequency f, the lower limit of the critical band can be calculated as $2^{-\frac{1}{24}} \cdot f$, the upper limit is $2^{\frac{1}{24}} \cdot f$. Table 1.1 lists the center frequencies of the adjacent, i.e. non-overlapping, 12th octave bands, used throughout this thesis.

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	1016	1076	1140	1208
1280	1356	1437	1522	1613	1709	1810	1918	2032	2153
2281	2416	2560	2712	2874	3044	3225	3417	3620	3836
4064	4305	4561	4833	5120	5424	5747	6089	6451	6834
7241	7671	8127	8611	9123	9665	10240	10849	11494	12177
12902	13669	14482	15343	16255	17222	18246	19331	20480	

Table 1.1: Center Frequencies (Hz) of adjacent 12th Octave Bands.

In analogy to critical bandwidths which define a range in the frequency domain which can successfully mask a tone, critical time intervals have been described for the time domain. The idea is that a time-limited noise can only mask a time-limited signal if the time in between them is less than the critical interval. For bottlenose dolphins, this interval was reported to be about 200-300 μs long [Altes 1979].

As mentioned above, little is known about the beluga auditory system which could be used to estimate or predict the masking of complex animal vocalizations by real underwater noises. Data available are very basic, such as pure tone audiograms or critical

bandwidths which were derived from masking experiments involving pure tones and artificial, broadband white noise. Due to the non-linear nature of mammalian hearing, results from pure tone experiments cannot simply be added to predict the masking of complex signals. My thesis therefore presents a different approach to the analysis of masking of animal sounds by man-made noise. Audiograms and critical bandwidths will aid in data interpretation throughout the chapters.

Chapter 2

Underwater Acoustics of Sounds Studied

The longest period of time for which a modern painting has hung upside down in a public gallery unnoticed is 47 days. This occurred to "Le Bateau" by Matisse in the Museum of Modern Art, New York City. In this time 116,000 people had passed through the gallery.

–The Guinness Book of Records

Sadly I had to discover that there is much confusion in the literature with respect to units of sound in particular after a Fourier transform. One finds sound spectra plotted in units of pressure, intensity, power, energy, or any of these per Hertz. It is often impossible to compare published data, because of lacking signal processing information. I therefore want to explain step-by-step how I calculated calibrated sound spectra.

2.1 Acoustics Background

For plane waves, sound pressure is proportional to particle velocity. Sound intensity is equal to pressure times particle velocity, hence proportional to pressure squared. Power is intensity integrated over area. Energy is power integrated over time.

Underwater sound is recorded with a hydrophone which converts received sound pressure to an output voltage. The voltage is generally amplified, digitized and stored as a discretely sampled time series. Hydrophone sensitivity, amplifier gain and the response of the digitizing unit allow a straight-forward conversion to received sound pressure units.

2.1.1 Fourier Transforms

To examine sound spectra, the time series have to be Fourier transformed. The *Fourier transform* of a time series $x(t)$ is defined as

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{2\pi i f t} dt \quad . \tag{2.1}$$

The inverse transform from the frequency domain into the time domain is

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{-2\pi i f t} df \quad . \tag{2.2}$$

It is apparent that if the time series is given in pressure units such as Pascal, the Fourier transform has units of Pascal times second, or Pascal per Hertz. For a discrete time series, the integrals become summations; the differentials dt and df become sampling intervals Δt and Δf. The Fourier transform can be implemented as the *Fast Fourier Transform*

(FFT) algorithm [Press *et al.* 1992], which computes the Fourier coefficients without the multiplication by $\Delta t = \frac{1}{f_s}$, where f_s is the *sampling frequency.*

The "calibration" of the Fourier coefficients has to be done via *Parseval's Theorem* which states that the total energy of a time series is the same in the time domain and in the frequency domain:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \quad . \tag{2.3}$$

I read the time series of the sound files in blocks of $T = 1$ s length, used a Hamming window and 50 % overlap between subsequent blocks. For each block, I computed the Fast Fourier Transform and divided the coefficients by the sampling frequency. For a time series of real values, the Fourier coefficients of the negative frequencies are equal to the complex conjugate of the positive frequencies. The negative frequencies can therefore be discarded by introducing a factor 2 in Parseval's Theorem. I squared the Fourier coefficients, selected the positive frequencies and multiplied by 2. At this stage, I had computed $|X(f)|^2$, an "energy density". The energy of a time series depends on its total length; a power density spectrum is more useful. Therefore, I divided by the length T, i.e. multiplied by $\Delta f = \frac{f_s}{\mathrm{NFFT}}$, to yield the *power density spectrum* in units of Pascal squared per Hertz. The sampling interval Δf as the ratio of sampling frequency and number of Fourier coefficients NFFT is equal to 1 Hz in the case that data are read in blocks of $T = 1$ s length. The power density spectra of the individual blocks were added and the mean was plotted giving frequency on a linear scale along the x-axis, and *power density spectrum levels* in dB re (relative to) 1 $\frac{\mu\mathrm{Pa}^2}{\mathrm{Hz}}$ along the y-axis (Figs. 2.3, 2.4, 2.5

and 2.6).

In order to illustrate the temporal features of the sounds, I also calculated colour *power density spectrograms.* I chose a 2 s long time series for each of my sounds and computed the Fast Fourier Transform on blocks of NFFT = 1024 data points. With a sampling frequency of 44 kHz, this corresponds to a time resolution of $T = 23$ ms. (To avoid confusion, the time resolution in the time domain is $\Delta t = \frac{1}{f_s}$. In the spectrogram domain, however, the time resolution is $T = \frac{\text{NFFT}}{f_s}$.) The frequency resolution of the spectrograms is $\Delta f = \frac{1}{T} = 43$ Hz. Again, choosing Hamming windows and 50 % overlap, I calculated the Fast Fourier Transform on a total of 172 blocks. The individual spectra were not averaged, but plotted next to each other to yield a 3-dimensional spectrogram showing the variation of power spectrum with time. For the spectrograms, I chose a logarithmic frequency scale to correspond to the logarithmic nature of mammalian hearing. A quick proof of non-linear frequency hearing is that we can tell the difference between 1,000 Hz and 1,005 Hz. But we can't tell 10,000 Hz and 10,005 Hz apart. Last but not least, we couldn't enjoy music if it wasn't based on our ear anatomy. The "concert A" has a frequency of 440 Hz, the "A" an octave below it has a frequency of 220 Hz, the "A" an octave above it has a frequency of 880 Hz. Frequency doubles from octave to octave which defines a logarithmic relationship. In order to avoid colour pixels becoming denser and denser for increasing frequency, I laid a square 172x172 grid over the logarithmic spectrogram and calculated the mean power density in each of the squares. The result is still a power density spectrogram representing amplitudes as colours in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$

(Figs. 2.11, 2.12, 2.13 and 2.14).

I always chose a logarithmic amplitude scale to account for logarithmic volume discrimination of mammalian ears. We can tell if one or two persons are talking at the same time, but we can't tell the difference between 100 and 101 speakers. In analogy, Gustav Mahler's 8th symphony for orchestra and choir ("Symphony of the Thousand") might actually be more enjoyable as the "Symphony of the 1024"!

To illustrate how a beluga whale might perceive the sounds studied, I calculated band levels in adjacent 12th octave bands simulating the beluga auditory filter. Taking the initially computed power density spectrum levels, the *12th octave band levels* are the area underneath the spectrum density curve in the corresponding bands. In other words, for each of the critical bands whose center frequencies were listed in Table 1.1, I summed up the spectrum density levels (on a linear scale, not in decibels) and multiplied by the frequency resolution, i.e. the "width" of each sample, 1 Hz. The resulting pressure levels were then plotted in dB re 1 μPa versus a logarithmic frequency scale giving center frequencies in kHz (Figs. 2.7, 2.8, 2.9 and 2.10). Comparing 12th octave band levels to power density spectrum levels, it is obvious that 12th octave band levels increase with frequency because of increasing bandwidth, i.e. more and more samples of the linear plot are summed to yield a band level. White noise would be a horizontal line in a power density spectrum, but a diagonal from the bottom left to the upper right corner in an octave band level plot.

2.1.2 The Decibel

The dimensionless "unit" *decibel* defines a logarithmic amplitude scale. It is used whenever one has to deal with a large range of amplitudes. This is particularly the case in bioacoustics. For instance, the "volume" of a beluga whale echolocation call at 1 m can be as strong 100,000 Pa. However, it's hearing sensitivity at 20 kHz is as low as 0.0001 Pa. To plot such vast ranges together, the decibel scale is helpful. It was named in honor of Alexander Graham Bell. Quantities which are measured in decibels are the *sound pressure level* and the *sound intensity level*:

$$SPL \quad [\mathrm{dB}] = 20 \cdot \log_{10} \frac{P}{P_{ref}} \tag{2.4}$$

$$SIL \quad [\mathrm{dB}] = 10 \cdot \log_{10} \frac{I}{I_{ref}} \tag{2.5}$$

Standard reference intensities and pressures are for air:

$$P_{ref} = 20\mu\mathrm{Pa}, \quad I_{ref} = \frac{P_{ref}^2}{\rho_a c_a} = 1 \times 10^{-12} \frac{\mathrm{W}}{\mathrm{m}^2} \quad ,$$

with $\rho_a = 1.2 \frac{\mathrm{kg}}{\mathrm{m}^3}$ for air density and $c_a = 333 \frac{\mathrm{m}}{\mathrm{s}}$ for the speed of sound. In water:

$$P_{ref} = 1\mu\mathrm{Pa}, \quad I_{ref} = \frac{P_{ref}^2}{\rho_w c_w} = 6.7 \times 10^{-19} \frac{\mathrm{W}}{\mathrm{m}^2} \quad ,$$

with $\rho_w = 1000 \frac{\mathrm{kg}}{\mathrm{m}^3}$ for water density and $c_w = 1500 \frac{\mathrm{m}}{\mathrm{s}}$ for the speed of sound. A doubling of sound pressure is equivalent to an addition of 6 dB. A doubling of sound intensity results in an increase by 3 dB. The sound pressure level of the echolocation call thus becomes 220 dB, and the beluga hearing threshold at 20 kHz becomes 40 dB, quantities which can easily be plotted together on a linear axis.

In human audiology and hence also in animal bioacoustics, decibels seem to be the prevalent measure of sound amplitude. In underwater and engineering acoustics, however, Watts appear more common. In particular, very recently, underwater acoustics projects have received hostile criticism and rejection due to a confusion of dB reference levels in air and in water. Therefore, underwater acousticians try to avoid decibels and rather use Watts. Expecting that people from various fields might read this thesis, I will give source levels both in dB (pressure or intensity measure) and W (power measure). No matter whether sound pressure levels or sound intensity levels are measured, the absolute value in dB is the same.

The world of bioacoustics seems to split further into two halves, researchers preferring pressure units and researchers preferring energy units. Amusingly enough, I have met people from both parts defending their viewpoint with the same argument: "Because it's what mammalian ears hear." Looking at ear mechanics, the pressure approach seems "straight-forward". In humans, the tympanic membrane between the outer and middle ear receives a pressure wave; the three ossicles of the middle ear then amplify the pressure by a factor 20 and transmit it to the inner ear through the oval window. Here, the basilar membrane basically does a Fourier transform. This is what first-year biology books have taught us for decades [Purves *et al.* 1992]. However, there are good indications that mammalian ears are energy detectors rather than pressure detectors ([Green and Swets 1966] for humans, [Au 1990] for dolphins). This is most likely related to electrophysiological processes in the inner ear.

I myself adopt the pressure approach in my thesis, because I study the interference of signal and noise, the underlying physical process of which is an addition of sound pressure waves.

2.1.3 Sound Propagation in the Ocean

Books which give an excellent introduction to this topic include [Clay and Medwin 1977], [Brekhovskikh and Lysanov 1982] and [Urick 1975]. Detailed descriptions of computational sound propagation models can be found in [Jensen *et al.* 1994] and [Etter 1996]. For low-frequency transmission loss in the Arctic ice cover, the reader is referred to [LePage and Schmidt 1994]; and for geoacoustic models of the sea floor see [Hamilton 1980].

When a sound wave propagates in the ocean, its path is subject to reflection, refraction and diffraction. All three processes are based on *Huygens' Principle*:

> Each point on an advancing wave front can be considered as a source of secondary waves, which move forward as spherical wavelets in an isotropic medium. The outer surface that envelops all these wavelets constitutes the new wave front. (Christian Huygens 1629-1695)

If Huygens' Principle is applied to spherical wavefronts hitting a planar surface, new wavefronts arise which appear to have originated in the *image space* at an *image source* on the opposite side of the planar interface. *Ray vectors* which point in the direction of wave

propagation and are hence perpendicular to the wavefronts, change direction at the planar interface such that the incident angle equals the reflected angle. In general, *reflection* occurs when a wave hits an object whose dimensions are greater than the wavelength.

In contrast to reflection, which is a reflection of sound energy back into the space where the sound originated, *diffraction* is the bending of sound waves around an object leading to sound energy in the space behind the object. If the planar surface of the previous example was semi-infinite, i.e. had an edge, an end, to it, each point along this edge would act as a Huygens point source radiating spherical wavelets. These wavelets would extend from the source space around the edge into the space behind the interface. The advancing wave as a whole would thus bend around the edge.

Refraction occurs when an advancing wavefront hits an interface separating two sound conducting media with different speeds of sound. Part of the energy of the incident wave penetrates into the second medium. If the sound speed in the second medium is greater than that in the first medium as is the case for a wave traveling from air into water and from water into most sediment, the refracted ray bends away from the normal. In the case that the sound speed in the first medium is greater than that in the second medium, the refracted ray bends towards the normal of incidence. The incident and the refracted angles are directly related to the two sound speeds via *Snell's Law.*

Altogether, if a wavefront hits an interface which is large compared to the wavelength of the incident wave, part of the energy will be reflected and the other part refracted. The ratio of the reflected energy (or often pressure) and the incident energy (pressure)

is termed the *reflection coefficient*; the ratio of refracted to incident energy (pressure) is termed the *transmission coefficient*. An interesting special case exists when sound travels from a medium of low sound speed into a medium of high sound speed. At angles greater than a critical incident angle, *total reflection* occurs. There is no refracted ray. At the critical angle, the refracted ray travels parallel to the interface in the second medium and continuously leaks energy back into the first medium, giving rise to the *geophysical head wave*.

The *speed of sound* in the ocean is on average about 1,500 $\frac{\mathrm{m}}{\mathrm{s}}$. It increases with temperature, salinity and pressure. Therefore, the exact speed of sound depends on the exact location in a three-dimensional ocean. It changes with season, time of day, depth, geographical position and proximity to rivers and melting ice. If one was to do CTD-casts measuring conductivity, temperature and depth at various places in the world's oceans, one would find that isohalines, surfaces of constant salinity (conductivity), are mostly horizontal. The same is true for isotherms and isobars. This fact is referred to as the *stratified ocean*. As a result, the speed of sound is also horizontally stratified.

In mid-latitudes, the water at the surface of the oceans is warm on the top and cools with depth. The speed of sound decreases with temperature. At a certain depth, however, the pressure dependence of the sound speed outweighs the temperature dependence and the speed of sound increases with increasing depth. The resulting minimum in the speed of sound lies at a depth corresponding to the channel axis of the so-called *SOFAR (sound fixing and ranging) channel* or *deep sound channel*. The stratified ocean acts like an

acoustic lens bending sound rays towards the channel axis. This is because in every "layer" to either side of the layer of minimal sound speed, the rays refract and bend away from the normal.

For a sound source near the channel axis, rays emitted nearly horizontally will get trapped in the sound channel. They do not interact with the sea surface or bottom, which otherwise leads to energy loss in shallow water. In deep water, sound can thus travel over vast distances, all across the major oceans.

In order to predict the received sound intensity at any location in the ocean, a method called *ray theory* is useful in the case of deep water. Individual rays are thereby traced from the projector to the receiver by computing Snell's Law at constant increments in space. For a sound source near the channel axis, rays emitted almost horizontally have the smallest excursion from the axis, hence the shortest travel path. However, due to the minimum in the speed of sound, they travel slowest. Near-axis rays contain most of the energy, because due to their short travel path, absorption by the sea water is minimal. Rays that leave the sound source at higher angles have greater excursions from the axis. They travel fastest but carry less energy than near-axis rays due to increased absorption.

Absorption in ocean water is based on the conversion of acoustic energy into heat. A minor component is the shear viscosity of ocean water, i.e. frictional forces during relative motion between adjacent layers of water. The major contribution comes from the bulk viscosity of ocean water based on molecular rearrangements. At frequencies less than 10 kHz, energy is lost due to rearrangements in boric acid. Between 10 kHz and 1 MHz,

magnesium sulfate absorbs most of the energy. For even higher frequencies, the freshwater component dominates.

The increase of absorption loss with travel path length and the difference in travel times for different sound rays results in the reception of multiple echos of varying intensity at the location of the receiver.

For a sound source far away from the channel axis, rays fill the ocean sound channel "less evenly" and converge in so-called *convergence zones.* For a continuously transmitting sound source, ray theory approximates the received sound intensity via the density of sound rays. It is thus elevated in convergence zones and near zero in so-called "shadow zones". These zones are only illuminated by surface or bottom reflected rays which lose energy with each reflection.

In mid-latitudes, the depth of the deep sound channel axis is about 1 km. At high latitudes, i.e. in the polar regions, the surface water is not warmer than deeper water. The speed of sound increases monotonically with increasing depth. Therefore, the channel axis rises towards the polar regions. The minimum of the speed of sound lies at the sea surface in the Arctic and Antarctic. Sound channeling in these regions is therefore often called the *Arctic surface duct.* Sound rays are all bent upwards where they reflect on the sea ice. For low frequencies (10 - 100 Hz), this reflection results in a significant energy loss due to the conversion of acoustic energy to elastic energy and the creation of flexural waves of the ice sheet. For higher frequencies, energy loss depends largely on the roughness of the ice surface with sound propagation in the surface duct being more efficient for smooth

than for rough ice.

Surface channelling also occurs at lower latitudes during the night or during the winter when the surface water layer cools down. In addition, turbulence can create well-mixed surface layers in which the sound speed increases with depth due to the steady pressure increase. At these times in mid-latitudes, two channels exist, the surface channel and the deep sound channel. They correspond to the two local minima in sound speed at the seasonal *thermocline* and the deeper, main thermocline. Near-horizontal rays leaving a sound source near the surface will get trapped in the surface channel. The paths are determined by the alternating repetition of upward refraction and downward reflection. Rays leaving the source at greater angles will "miss" the surface channel and travel into the deep sound channel. In particular scattering at a rough sea surface will lead to a scattering of rays from the surface channel into the deep sound channel. This is termed *leaking*.

In shallow water or in cases of extreme and extended heating of upper layers by solar radiation, sound propagation can turn into *antiwaveguide propagation*. The speed of sound monotonically decreases with depth from the surface to the bottom of the sea. Therefore, sound rays are refracted downwards and reflected upwards with major energy losses due to sound transmission into the sea floor.

Sound propagation in shallow water is often modeled as a superposition of *normal modes* in a waveguide. These normal modes are traveling waves in horizontal directions and standing waves in the vertical direction. They are solutions of the Helmholtz equation

(time-independent, frequency-domain wave equation). As any waveguide, the shallow water channel has a lower cut-off frequency which corresponds to a wavelength of four times the water depth. In other words, only waves with a wavelength small enough such that a quarter wave fits in between the surface and the bottom can be transported. Lower frequencies of sound will decay exponentially. The higher the frequency, the more modes are excited. Often, the sea floor consists of water-saturated sediment which is well-conducting for low-frequency sound. If the acoustic properties of the bottom layer are known, they are included in normal mode modeling.

In cases where the channel characteristics vary with range, eg. if the sea floor is not horizontal but sloped or if the speed of sound varies in the horizontal plane, a technique called the *parabolic equation* is useful. Normal modes are calculated based on the Helmholtz equation. The parabolic equation includes an extra term corresponding to the first derivative in radius. For example in the case of downward sloping sea floor, the volume available for sound propagation/expansion from a surface source increases, leading to greater spreading loss than for a constant volume surface duct. However, incidence angles during reflection off the sea floor decrease with slope leading to a significant reduction of transmission loss. Therefore, a downward slope usually results in received amplitudes greater than expected. This is why ships over continental shelf are audible over long ranges. On the other hand, in case of upward sloping sea floor, the number of bottom reflections and the incidence angle increase leading to enhanced transmission losses.

Sound propagation will further be influenced by currents, eddies, internal waves, tur-

bulence and the presence of air bubbles (free or enclosed in organisms such as the swim bladders of fish, for instance). Which type of sound propagation model fits best for a particular location in the ocean should ideally be determined by a transmission loss experiment prior to any acoustic study. Often, these are infeasible. If no detailed information on the location (water and sea floor characteristics) is available, a simple approach to sound propagation modeling is based on ideas of spherical and cylindrical spreading via the SONAR equation.

2.1.4 The SONAR Equation

The abbreviation SONAR stands for "sound navigation and ranging". The SONAR equation relates the *sound pressure level SPL* or the *sound intensity level SIL* at a particular distance to the *source level SL* of a sound source measured at or related to a distance $R_0 = 1$ m.

I will briefly derive the sonar equation I used for sound calibration. The ship noises and the icecracking noise I studied originate all at the sea surface. Considering the sound source a point source, sound will spread spherically until it hits the sea floor. Thereafter, spreading becomes cylindrical. Assuming the ice layer on the sea surface is an ideal reflector, the total power of a point source integrated over a half sphere pointing into the water is

$$L_0 = \int I dA = \frac{1}{2} 4\pi R_0^2 P_0^2 \frac{1}{\rho c} \quad .$$

The subscript 0 denotes source levels at 1 m. Some distance away from source, the same total power radiates through a larger sphere with radius R. Equating $L = L_0$ yields the condition for *spherical spreading*:

$$P_0^2 R_0^2 = P^2 R^2 \quad . \tag{2.6}$$

The transmission loss is defined as the ratio of the sound pressure 1 m away from the source and the sound pressure at a larger distance R. A division becomes a subtraction on a logarithmic scale. Therefore, in decibels, the transmission loss can be expressed as the positive quantity:

$$TL_{sph} = 20\log_{10}\frac{P_0}{P_{ref}} - 20\log_{10}\frac{P}{P_{ref}} = 20\log_{10}\frac{P_0}{P} = 20\log_{10}\frac{R}{R_0} \quad , \tag{2.7}$$

where the condition for spherical spreading was substituted in the last step.

When R becomes equal to the water depth H, spreading gradually changes from spherical to cylindrical. The total power radiated through a cylinder at reference distance R_0 is

$$L_0 = 2\pi R_0 H P_0^2 \frac{1}{\rho c} \quad .$$

Equating this to the power radiated through a larger cylinder, yields the condition for *cylindrical spreading*:

$$P_0^2 R_0 = P^2 R \quad . \tag{2.8}$$

The transmission loss becomes:

$$TL_{cyl} = 20\log_{10}\frac{P_0}{P} = 10\log_{10}\frac{P_0^2}{P^2} = 10\log_{10}\frac{R}{R_0} \quad . \tag{2.9}$$

The sound files available to me were recorded at a distance R greater than the local water depth H, therefore a combination of spherical and cylindrical spreading applied:

$$TL_{spread} = 20\log_{10}\frac{H}{R_0} + 10\log_{10}\frac{R}{H} = 10\log_{10}\frac{H^2R}{R_0^2H} = 10\log_{10}\frac{R}{R_0} + 10\log_{10}\frac{H}{R_0} \quad . \tag{2.10}$$

Apart from spreading loss, sound energy dissipates along its travel path due to absorption by the surface ice, the sea floor and the sea water itself. Absorption leads to exponential decay:

$$P = P_0 \cdot e^{-\kappa_e R} \quad , \tag{2.11}$$

with an absorption coefficient κ_e. The transmission loss based on absorption becomes:

$$TL_{abs} = 20\log_{10}\frac{P_0}{P} = \kappa_e R \cdot 20\log_{10} e \approx 8.686\kappa_e R \quad . \tag{2.12}$$

The absorption coefficient κ usually includes the factor 8.686, such that the transmission loss is simply:

$$TL_{abs} = \kappa R \quad . \tag{2.13}$$

A review study on sound attenuation [Thiele *et al.* 1990] produced the following equation for the absorption coefficient in an ice-covered ocean as a function of frequency:

$$\kappa = \frac{0.235f^3}{0.0023 + f^3} + \frac{0.11f^2}{1 + f^2} + \frac{43.7f^2}{4,100 + f^2} \quad . \tag{2.14}$$

Putting it all together, a SONAR equation giving the sound pressure level SPL at any distance R from a sound source, is (including only spreading loss and absorption loss):

$$\begin{aligned} SPL &= SL - TL_{spread} - TL_{abs} \\ &= SL - 10\log_{10}\frac{R}{R_0} - 10\log_{10}\frac{H}{R_0} - \kappa R \quad , \end{aligned} \tag{2.15}$$

where $R_0 = 1$ m, H is the water depth, κ is the frequency dependent absorption coefficient and SL is the source level at 1 m. All the terms in this equation are in dB re 1 μPa. The SONAR equation for the sound intensity level SIL looks identical and the terms have the same algebraic value, though decibels would be relative to 1 $\frac{\mathrm{W}}{\mathrm{m}^2}$.

2.2 Icebreaker Noises

Recordings of the Canadian Coast Guard ship CCGS *Henry Larsen* (Fig. 2.1) were obtained from the Institute of Ocean Sciences in Sidney, BC. The *Henry Larsen* is a medium gulf/river icebreaker of 100 m length, 20 m breadth, with a total power of 12 MW. During a cruise in the Beaufort Sea in August 1991, a scientist was flown to a site about 5 km away from the ship. The water depth in this area was about 300 m. Holes were drilled through the 4 m thick ice. Hydrophones were lowered into the water to depths between 2 and 10 m underneath the ice. The hydrophones were MetOcean models of the type NH4123 though modified for low frequency recording. Their bandwidth was 25 kHz, with the low-frequency cut-off of -6 dB V at 32 Hz. Their radiation pattern was omnidirectional; the sensitivity -187 dB re 1 $\frac{\mathrm{V}}{\mu\mathrm{Pa}}$. Sound was prewhitened to enhance the sensitivity for high frequencies, and then recorded digitally onto video tape using a SONY PCM-F1. The resolution was 16 bit, the sampling frequency 44 kHz. Two major noise sources were identified.

2.2.1 Bubbler System Noise

Some icebreakers are equipped with a so-called bubbler system along the sides of the ship. This system blows high-pressure air into the sea in order to push floating ice debris away and leave a clean passage for the ship. The bubbler system is generally used when the ocean is covered with floating ice floes or after the icebreaker has broken an ice ridge. The bubbler system will prevent ice pieces from jamming into the sides of the ship, which could cause serious drag. Once the high pressure air strikes the water, a large number of air bubbles is formed in the surface layer and the surface water is whitened. The striking of pressured air onto the ocean surface and the oscillations of bubbles under water make the ocean extremely noisy. Bubbler system noise is a fairly continuous signal with most of its energy around a few hundred Hz. The amplitude is slightly modulated with 2 Hz, which was the frequency with which the bubbler system motor on the *Henry Larsen* pushed the air into the water. After computing the Fourier transform of the recorded time series, I removed the prewhitening effect and corrected for transmission loss due to 1) spherical spreading up to a distance of 300 m, 2) cylindrical spreading between 300 m and 5 km and 3) frequency dependent absorption under ice. The calculated power density spectrum at 1 m from the source is plotted in Fig. 2.3. The amplitude is given in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$. The total, broadband source level can be regarded as the area underneath the curve. Integration gives a sound pressure level of 194 dB re 1 μPa at 1 m (source power: 211 W). A 2 s spectrogram is shown in Fig. 2.11. Again, the amplitude is in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ at 1 m.

2.2.2 Propeller Cavitation Noise

Cavitation is a common phenomenon in underwater acoustics. It is defined as the rupture of a liquid caused by a reduction of local static pressure. It is thus different from boiling which is induced by an increase in temperature. Local pressure can be decreased in different ways. For example, a sound wave of 220 dB re 1 μPa has a peak pressure of more than 1 atm. In its negative cycle, the wave therefore reduces the local pressure to less than the ambient static pressure (at the surface) and rupture occurs. This limits transducer output. A reduction of local pressure below the static value can also arise during fluid flow around bodies. At the face of the obstacle, fluid motion speeds up. By Bernoulli's principle, this results in a pressure reduction proportional to the density of the fluid and the square of the flow speed. A detailed theoretical and experimental analysis of cavitation with focus on propeller cavitation is given by [Ross 1976]. Any propeller driven ship (except for a submarine operating at great depth) creates propeller cavitation noise. This type of noise is loudest when the thrust produced is small or negative. In the case of the *Henry Larsen* icebreaker, propeller cavitation noise was strongest when the vessel had built up momentum, rammed an iceridge, but failed and was stopped by the ice with the propeller still turning at full speed. The power density spectrum is plotted in Fig. 2.4. The broadband source level is 203 dB re 1 μPa at 1 m (source power: 1680 W). A 2 s spectrogram is shown in Fig. 2.12. The sound consists of short broadband pulses, occurring at a repetition frequency of 11 Hz which is related to the rotation frequency of the propeller. As it is this particular action of the icebreaker which creates the loudest

cavitation noise, I also refer to the sound as ramming noise.

For comparison, the source level of the icebreaker MS *Voima* with 10.2 MW power, has been reported to reach maximum values of 190 dB re 1 μPa at 1 m (84 W) [Thiele 1981]. The maximum source level of the MV *Arctic* during icebreaking was 191 dB re 1 μPa at 1 m (106 W) [LGL and Greeneridge 1986]. This vessel's power was 11 MW. The *Robert Lemeur* with maximum power of 7.2 MW exhibited source levels of 197 dB re 1 μPa at 1 m (422 W) [LGL and Greeneridge 1995].

2.3 Natural Icecracking Noise

For means of comparison, naturally occurring icecracking noise was studied as well. Ambient noise in the Arctic is caused by wind and waves, thermal icecracking and pressure cracking, and biologics (sounds made by marine organisms). Recordings of natural icecracking sound were obtained from earlier studies on acoustical radiation from stressed ice [Farmer and Xie 1989, Xie and Farmer 1991]. These studies not only provided experimentally obtained sound spectra but also theoretical models. The source level of a single icecracking event can be about 147 dB re 1 μPa at 1 m (source power: 4 mW). A typical power density spectrum is shown in Fig. 2.5. Ambient Arctic noise at the time of these experiments was a superposition of individual icecracking events. The broadband sound pressure level was measured to be 87 dB re 1 μPa. It does not make sense to specify a distance for ambient noise. A typical 2 s power density spectrogram is shown in Fig. 2.13.

The sound is highly irregular; one identifies sudden broadband pulses. Other studies reported ambient Arctic noise at 90-95 dB re 1 μPa [LGL and Greeneridge 1995].

2.4 Beluga Whale Vocalizations

The distribution of beluga whales (scientific name: *Delphinapterus leucas*), is circumpolar. They are found in Arctic and Subarctic waters along the northern coasts of Canada, Alaska, the Russian Federation, Norway and Greenland. The total population is estimated at 60,000-70,000, [DFO 1991, Hoyt 1990, Carwardine 1995]. A photo of two beluga whales is shown in Fig. 2.2. They got their name from their colour (beluga meaning white in Russian). Another name for them is sea canaries, which they got due to their large, "bird-like" vocal repertoire. For categorizations of beluga vocalizations see [Sjare and Smith 1986, Angiel 1997]. I kindly received high quality (digitally recorded, 16 bit resolution, 44 kHz sampling frequency) sound of beluga whales in Maxwell Bay from Sue Cosens of the Freshwater Institute in Winnipeg, Manitoba. Unfortunately, these sounds have not yet been calibrated. Comparing with reported source levels of other odontocete species [Richardson *et al.* 1995, Ch. 7.2], I assume a broadband source level of 160 dB re 1 μPa at 1 m (source power: 84 mW) for the following plots. In my later masking analysis I study sounds of differing volume (Chapter 8). From eight hours of recordings, I extracted one vocalization which the whales seemed to use fairly frequently and which was recorded very clearly with hardly any background noise. The power density spectrum is plotted in Fig. 2.6. A 2 s spectrogram is shown in Fig. 2.14.

This vocalization consists of 6 pulses between 800 Hz and 8 kHz. It is about 2 s long.

2.5 Digital Mixing of Signal with Noise

In the ocean, the masking of whale communication by ship noise depends on the distance of the listening whale from both the vocalizing whale and the noise source. The characteristics of the call and noise at the location of the listener depend on the respective travel paths. Amplitudes change with frequency due to increasing absorption with frequency. Furthermore, temporal structures can change due to frequency dispersion, in particular for multipath arrivals resulting from sea floor and surface (ice) reflection. In addition, directional hearing abilities of the listener might facilitate signal detection in noise if the two sounds come from different directions. In order to make my masking study as basic as possible, i.e. independent from all these–often unknown–factors, I normalized all the sounds to an equal *root-mean-square (rms)* amplitude and then mixed the signal with the noise digitally.

For a time-series $x[t]$ with total length T, the rms amplitude is calculated according to

$$\mathrm{rms}(x) = \sqrt{\frac{1}{T}\sum_{t=1}^{T} x^2[t]} \quad . \tag{2.16}$$

The recorded time-series of the two-second-long beluga vocalization and two-second-long samples of each of the three noises were divided by their rms amplitude, yielding normalized versions with rms amplitudes of 1. Sound pressures are additive. Therefore, mixtures

$x[t]$ of the call $s[t]$ with the three noises $n[t]$ were computed simply by adding the signals in the time domain,

$$x[t] = s[t] + \alpha \cdot n[t] \quad . \tag{2.17}$$

The factor α denotes the noise-to-signal ratio (nsr). I chose the noise-to-signal ratio rather than the common signal-to-noise ratio, because–as will be seen–the signal can still be detected in the noise when the noise is louder than the call. Again, to account for the logarithmic nature of mammalian hearing, α was varied logarithmically between 2^0 and 2^5. In detail, mixtures were computed with the noise pressure amplitude being 1, 2, 4, 8, 16 and 32 times that of the call. In decibels, the noise-to-signal ratios were 0, 6, 12, 18, 24 and 30 dB.

Figure 2.1: Photo of the Canadian Coast Guard Icebreaker CCGS *Henry Larsen.*

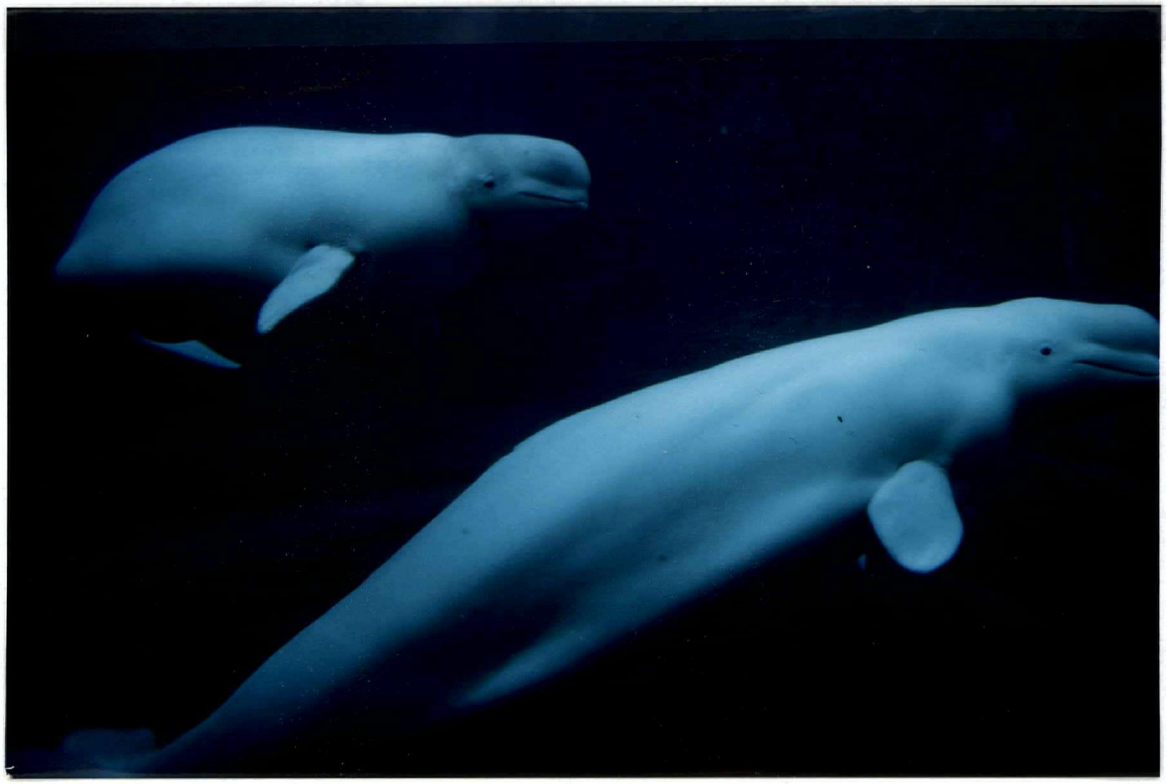


Figure 2.2: Photo of two Beluga Whales, *Delphinapterus leucas.*

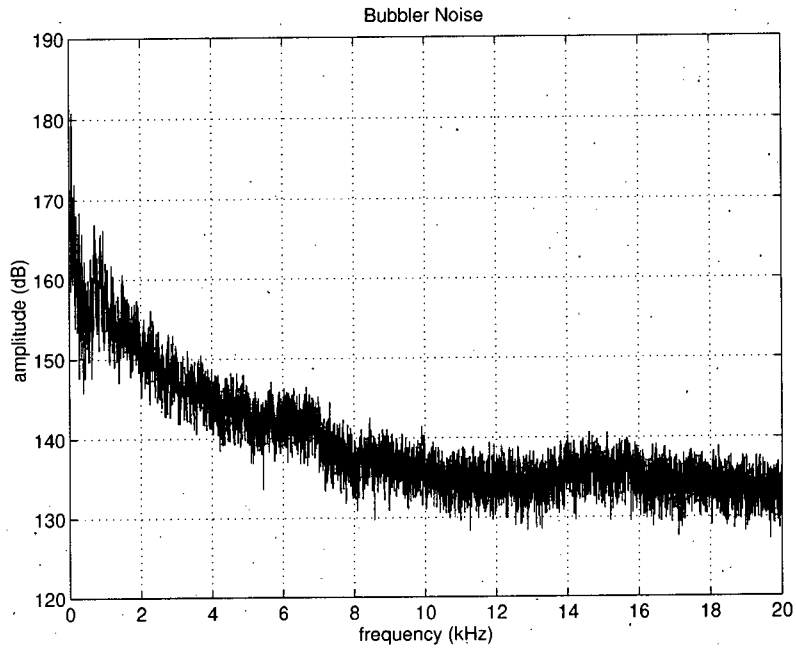


Figure 2.3: Power Density Spectrum of Bubbler System Noise in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

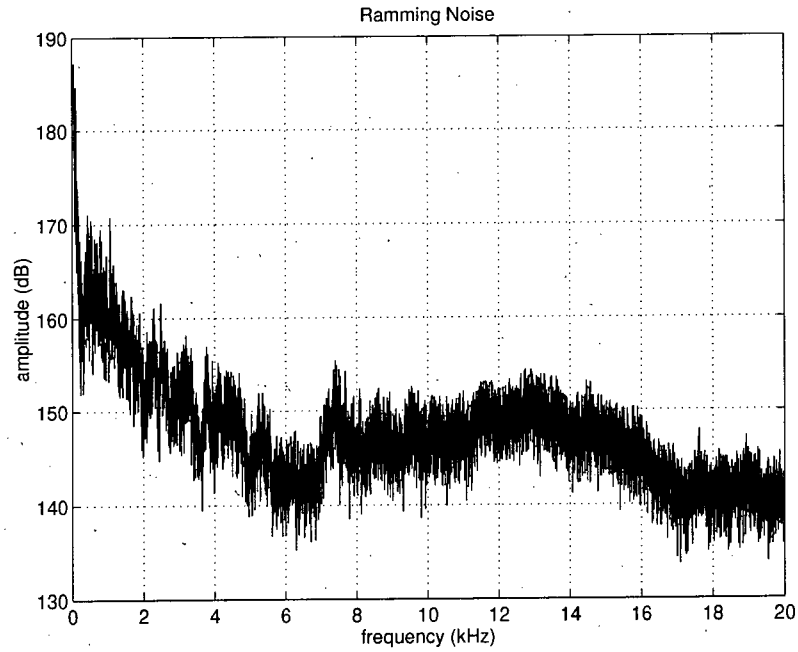


Figure 2.4: Power Density Spectrum of Ramming Noise in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

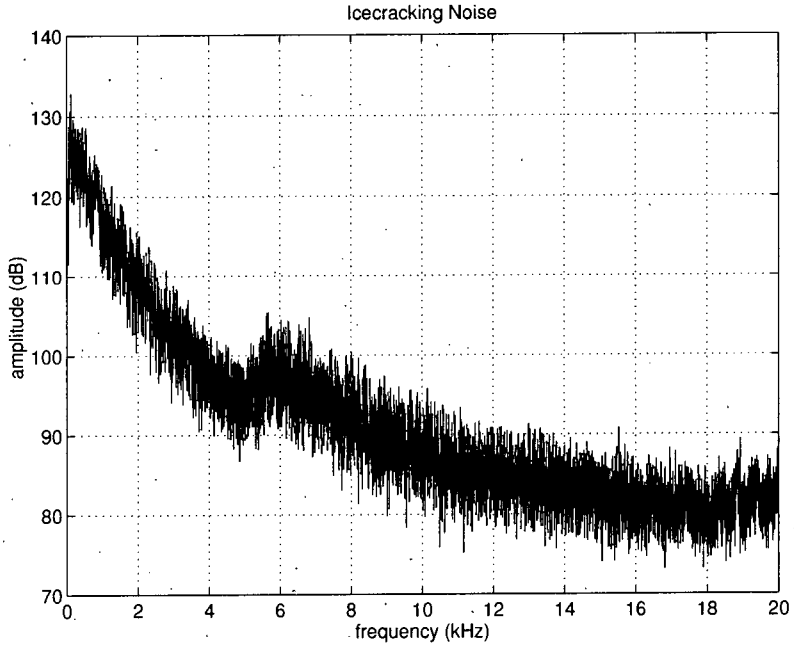


Figure 2.5: Power Density Spectrum of a Single Icecracking Event in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

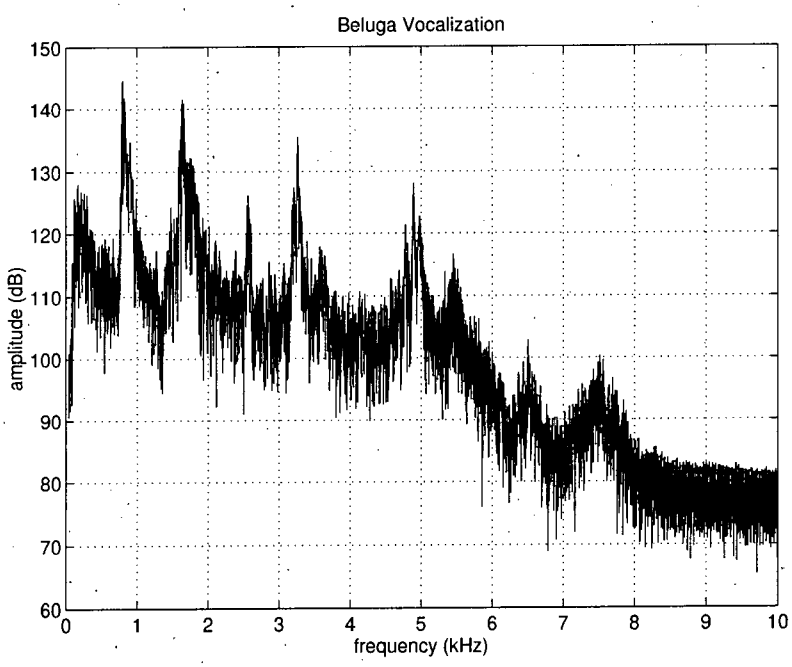


Figure 2.6: Power Density Spectrum of the Beluga Vocalization in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

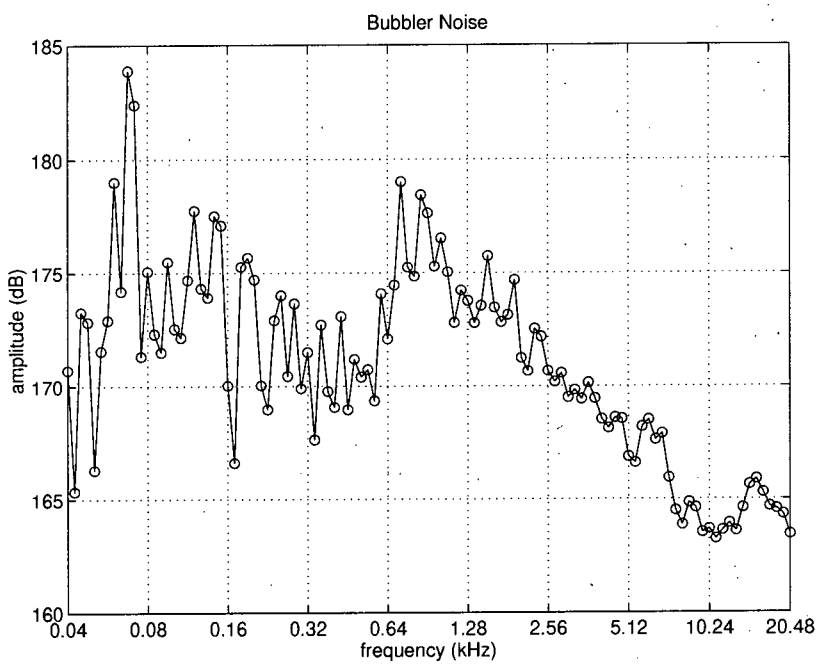


Figure 2.7: 12th Octave Band Levels of Bubbler System Noise in dB re 1 μPa @ 1 m.

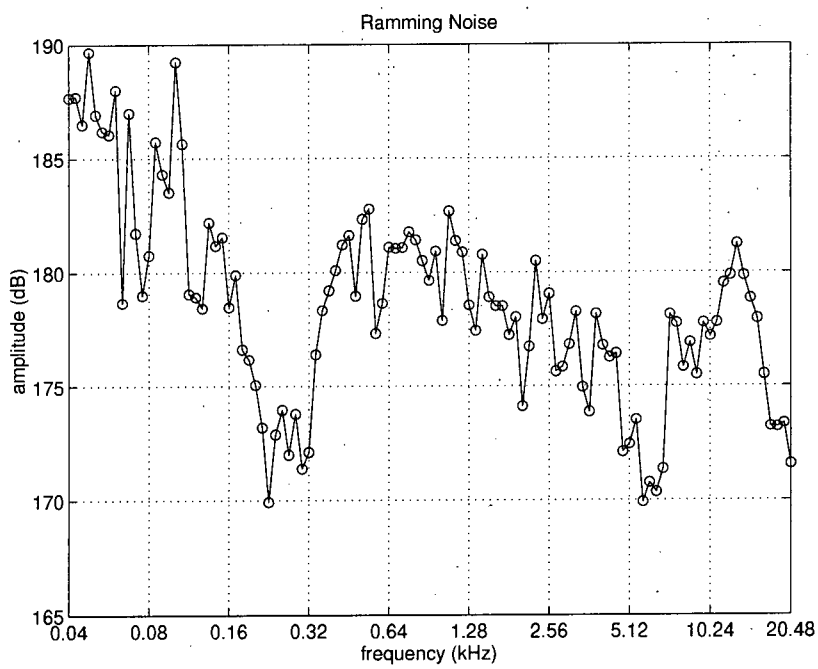


Figure 2.8: 12th Octave Band Levels of Ramming Noise in dB re 1 μPa @ 1 m.

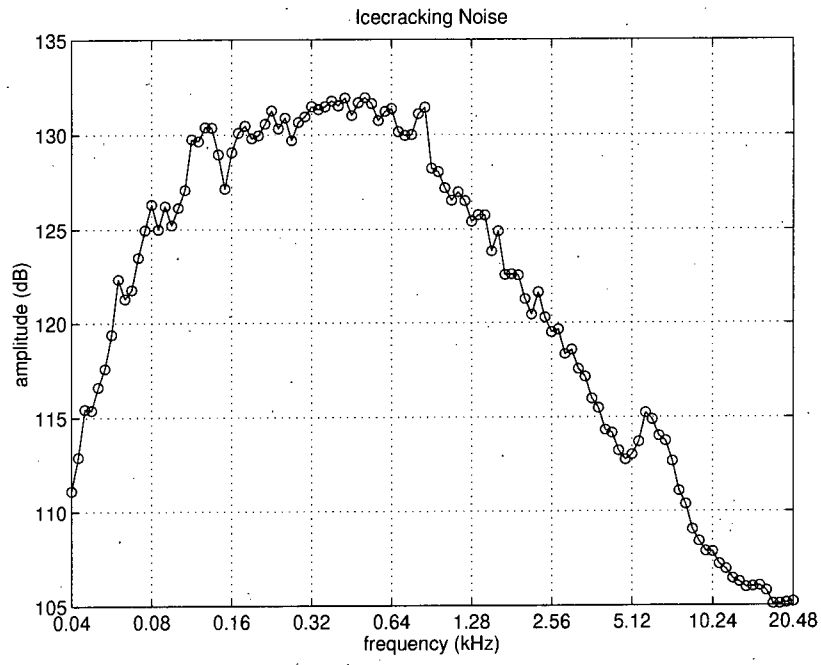


Figure 2.9: 12th Octave Band Levels of a Single Icecracking Event in dB re 1 μPa @ 1 m.

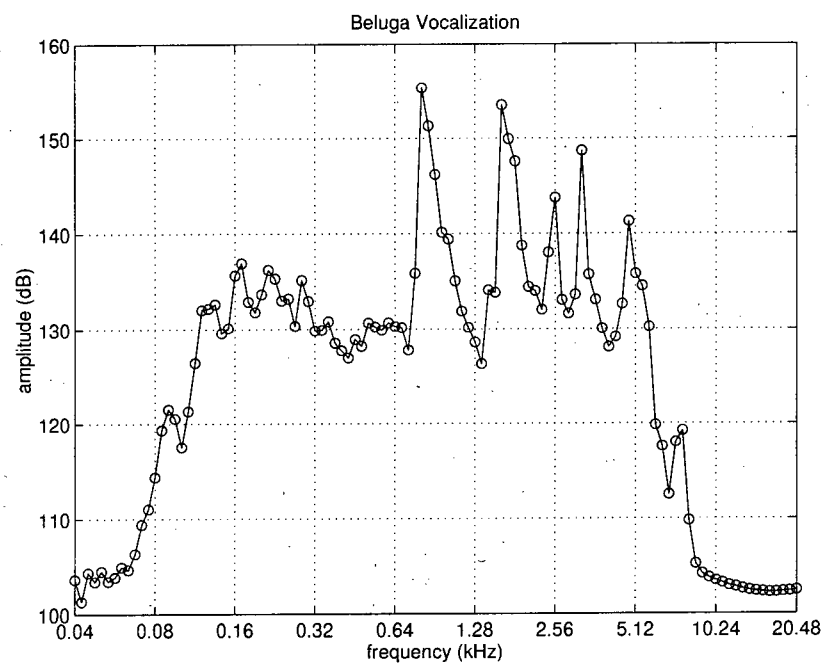


Figure 2.10: 12th Octave Band Levels of the Beluga Vocalization in dB re 1 μPa @ 1 m.

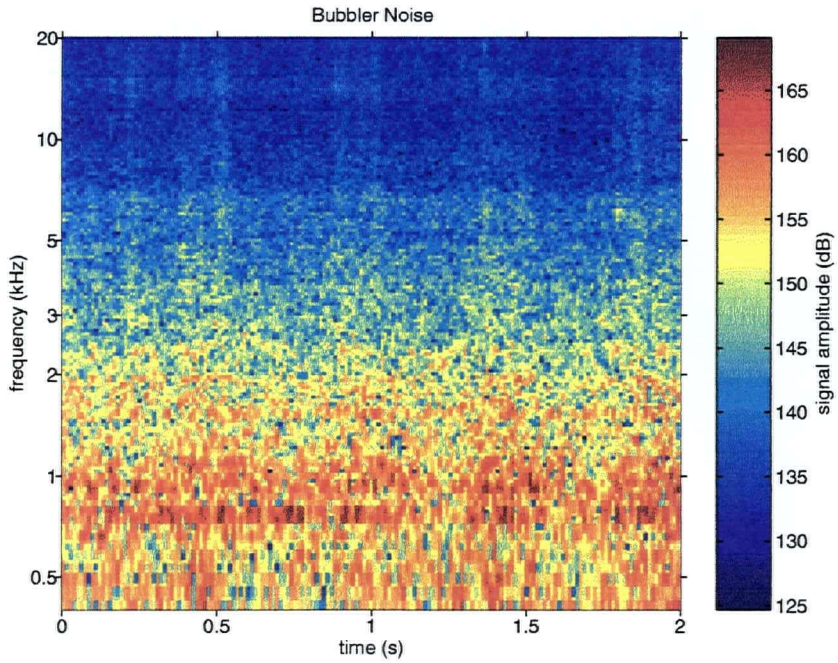


Figure 2.11: Power Density Spectrogram of Bubbler System Noise in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

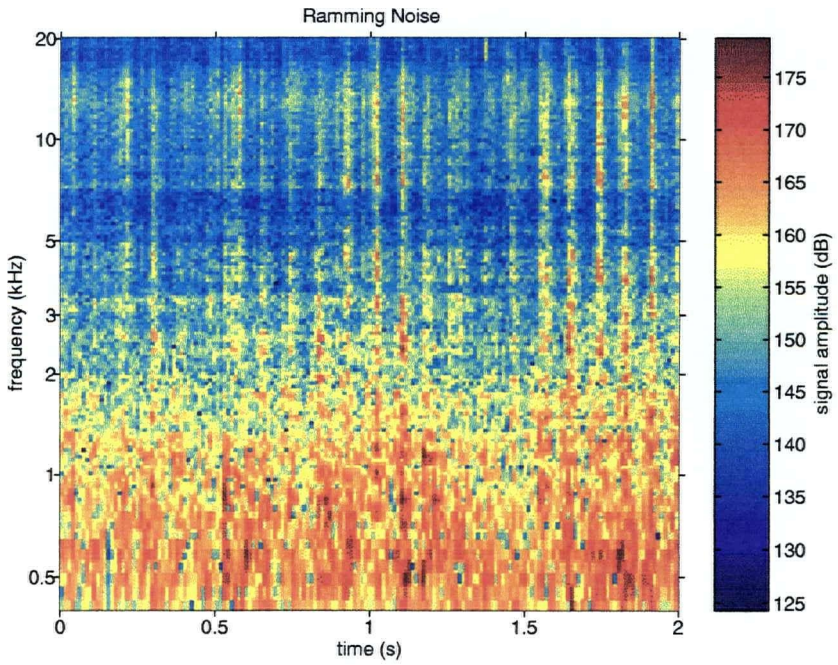


Figure 2.12: Power Density Spectrogram of Ramming Noise in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

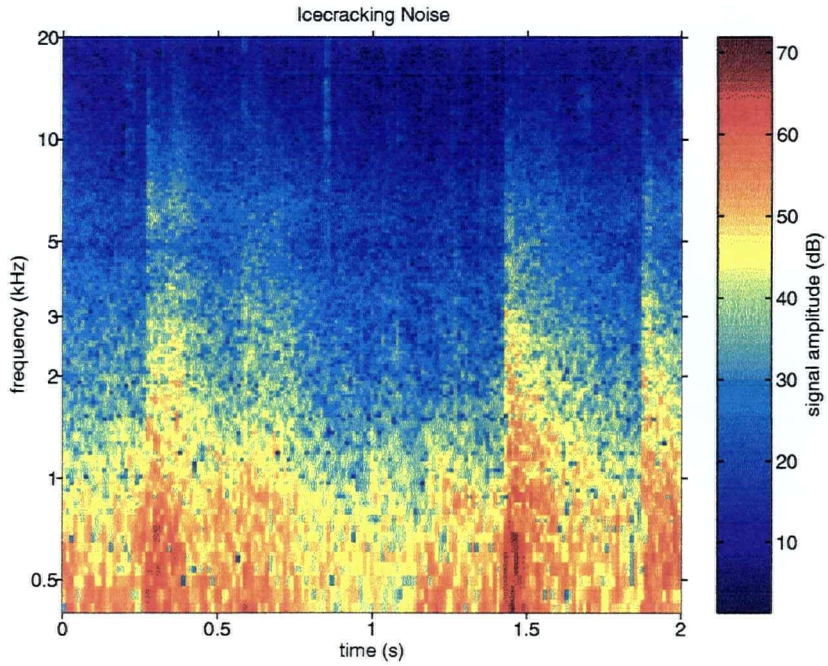


Figure 2.13: Power Density Spectrogram of Ambient Icecracking Noise in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$.

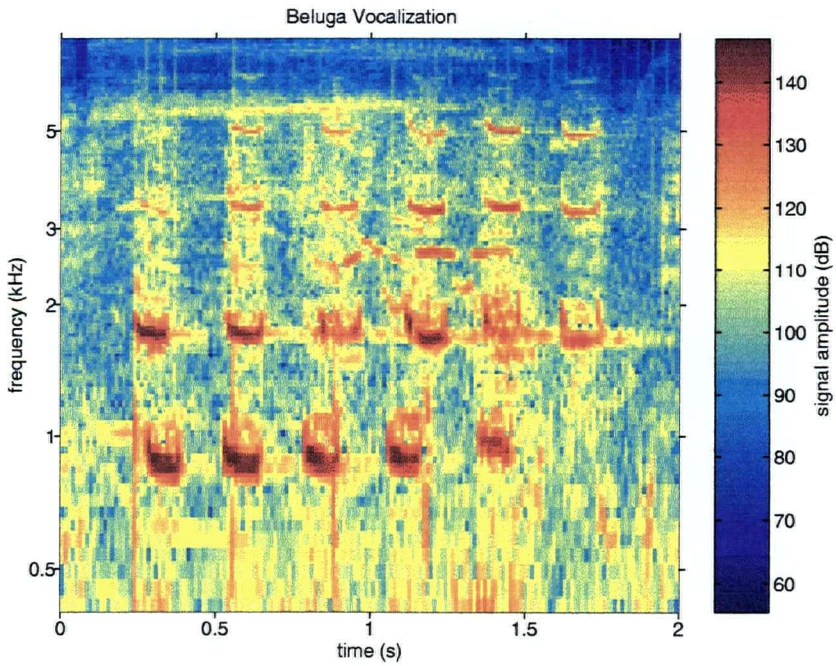
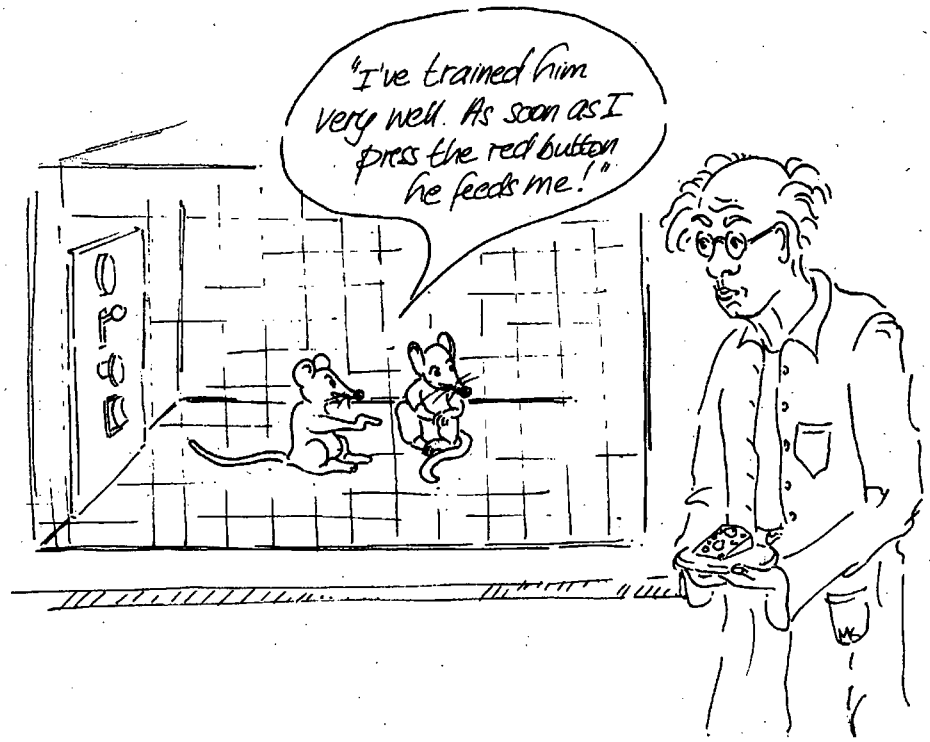


Figure 2.14: Power Density Spectrogram of the Beluga Vocalization in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

Chapter 3

Acoustic Experiments with a Beluga Whale

The original idea was to train three beluga whales of different age and sex for acoustic masking experiments in which the animals would indicate when a noise source was just too loud for the animals to detect a buried vocalization. The Vancouver Aquarium houses six beluga whales, five adults and one calf which was born in captivity in July 1995. The five adults were all caught in Hudson Bay. Kavna, the oldest female, was born in 1970 and caught in 1976. Allua, the second oldest female, was born in 1982 and caught in 1985. Nanuq, the older male, is as old as Allua and was caught in 1990 together with Aurora, the youngest female, born in 1986, and Imaq, the younger male, born in 1986. The marine mammal staff and I started training Nanuq, Imaq and Aurora for the experiments. Eventually, circumstances, however, did not allow training to be completed for more than



one animal. Therefore, data were only collected from Aurora.

3.1 Resonance Frequencies of the Beluga Pool

3.1.1 Theoretical Evaluation

During acoustic experiments at an aquarium, some characteristic features of the pool have to be taken into account. Like resonant strings or vibrating membranes, rooms have normal modes of vibration and natural resonant frequencies. This fact is very important in architectural acoustics where rooms are designed to maximize the intelligibility of speech, the richness of music and freedom from external noise.

When sound waves are produced in a region completely enclosed by walls, they might soon vanish due to absorption (by the walls and the medium inside the chamber) and deconstructive interference after reflection. But there are special wavelengths for which interference is constructive and the sound waves superpose to form standing waves. In this case, they might last for thousands of oscillations before they are totally damped out. As these wavelengths depend on the size and shape of the room, the standing waves are called "eigenmodes" of the closed room or cavity, which is derived from the German word "eigen" meaning "own". The cavity is said to be resonant at the corresponding eigenfrequencies. Even when the source becomes silent, the eigenvibrations persist. This is termed reverberation.

For my acoustic masking experiments, the resonance frequencies of the beluga pool are of particular importance, because a pool, as a resonant cavity, can act as a filter amplifying only selected frequencies and damping all the others. Thus, acoustic signals may dramatically be distorted. An extreme scenario would be the case of the pool having distinct resonances at the frequencies of the call. The pool would amplify these selected components which would make it "easier" for the whale to detect the call in the noise. Therefore, the pool resonances have to be thoroughly examined leading to a proper measurement setup. Fig. 3.1 shows a sketch of the pool.

For a theoretical analysis, it is necessary to find a solution of the three-dimensional wave equation under specified boundary conditions. Sound waves are pressure vibrations occurring periodically in space and time. Let P denote the pressure, then it satisfies the

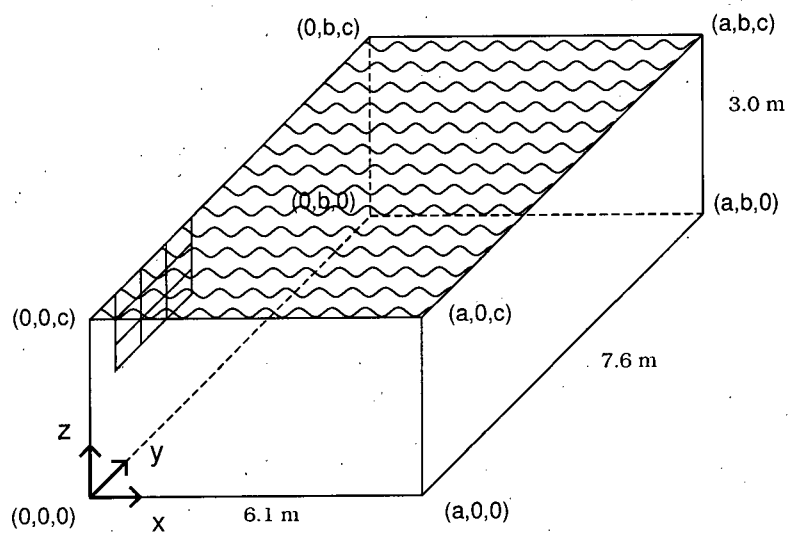


Figure 3.1: Sketch of the Beluga Pool.

following wave equation:

$$\Delta P = \frac{1}{v^2} \cdot \frac{\partial^2 P}{\partial t^2} \quad , \tag{3.1}$$

where Δ is the Laplacian operator. $\Delta = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ in Cartesian coordinates.

The boundary conditions are such that there is a node in the pressure wave wherever the initial wave is reflected at the surface of a less dense material; antinodes occur at dense-denser interfaces. The medium in which the standing waves develop is the ocean water inside the pool. There are five bounding walls which are, of course, denser than water. The top surface is the air-water interface with air less dense than water. Thus, at the top surface one will find pressure nodes and at the other surfaces pressure antinodes (Fig. 3.2).

The problem to solve is an eigenvalue problem. These are analytically solvable only for simple, symmetric shapes. As the pool is almost rectangular, the general solution of

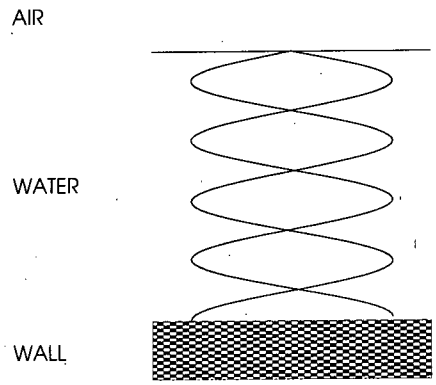


Figure 3.2: Standing Wave between Materials of Different Density.

the above problem is, by separation of variables,

$$P(x,y,z,t) = (A_f e^{ik_x x} + A_b e^{-ik_x x}) \cdot (B_f e^{ik_y y} + B_b e^{-ik_y y}) \cdot (C_f e^{ik_z z} + C_b e^{-ik_z z}) \cdot e^{i\omega t} \quad , \quad (3.2)$$

where ω is the sound wave's frequency, the k's are the wavenumbers in the three directions and A, B, C are the corresponding amplitudes. The indices denote waves traveling forth and back.

Consider a standing wave between the left and the right walls. The boundary condition at the left wall, where $x = 0$, is that $P(0,y,z)$ has a maximum, i.e. the derivative with respect to x has to be zero. Regarding only the x-component, one gets

$$\frac{\partial P}{\partial x} \propto A_f[-k_x \sin(k_x x) + ik_x \cos(k_x x)] + A_b[-k_x \sin(k_x x) - ik_x \cos(k_x x)] \quad .$$

For this to be zero at $x = 0$, it is required that $A_f = A_b$. Let's define a new constant $A = \frac{A_f}{2} = \frac{A_b}{2}$. Doing the same for the walls at $y = 0$ and $z = 0$ and defining new constants without indices gives a pressure distribution of the form

$$P(x,y,z) = A\cos(k_x x) \cdot B\cos(k_y y) \cdot C\cos(k_z z) \quad . \quad (3.3)$$

At the wall $x = a$, again P has to be maximal, i.e. its derivative with respect to x has to be zero:

$$\frac{\partial P}{\partial x} \propto -Ak_x \sin(k_x x) \quad .$$

At $x = a$, this is vanishing only for $k_x a = l \cdot \pi$, where l is an integer. The boundary condition at $y = b$ requires a similar equation: $k_y b = m \cdot \pi$, with an integer number m.

Only the boundary at $z = c$ is different. This is the air-water interface. The pressure has to be zero here requiring that $\cos(k_z c) = 0$. This is satisfied only for distinct k_z: $k_z c = \frac{2n+1}{2}\pi$, where n is integer.

Altogether, the pressure varies as

$$P(x, y, z) = \hat{P} \cdot \cos(k_x x)\cos(k_y y)\cos(k_z z) \cdot e^{-i\omega t} \quad . \tag{3.4}$$

$\hat{P} = A \cdot B \cdot C$ is the maximum amplitude. The wavenumbers in the three directions are

$$k_x = \frac{l\pi}{a}, \qquad k_y = \frac{m\pi}{b}, \qquad k_z = \frac{(2n+1)\pi}{2c} \quad .$$

Substituting this solution into the wave equation leads to the following condition:

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{v^2}$$

or

$$\omega = v\pi\sqrt{\frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{(2n+1)^2}{4c^2}} \quad . \tag{3.5}$$

As l, m, n can only represent the values $0, 1, 2, \ldots$, the eigenvalue problem has a discrete spectrum of eigenvalues. This means that a cavity is resonant at an infinite set of distinct eigenfrequencies. One usually writes these resonance frequencies as ω_{lmn}. For each

eigenfrequency, there is a corresponding eigenmode, i.e. a corresponding geometric configuration of vibrations in the room. The indices l, m, n may be visualized as follows. l counts the pressure nodes in $x-$direction, m counts the pressure nodes in $y-$direction and n counts the pressure nodes in $z-$direction, where the permanent node at $z = c$ is omitted.

If there exists more than one eigenmode for a distinct frequency ω_{lmn}, these modes are said to be degenerate. From a mathematical point of view this means that there are several linearly independent eigenfunctions for one eigenvalue. Ideally, the frequency spectrum is a series of delta peaks. A room that has been excited once should vibrate for an infinite period of time. However, in reality all modes are damped due to absorption by the surrounding walls and the medium inside the cavity. Looking at the resonance spectrum, the peaks are not infinitely sharp but have a finite width $\Delta\omega$ within which the corresponding mode can be excited. One defines the quality factor

$$Q = \frac{\omega_{lmn}}{\Delta\omega} \quad , \tag{3.6}$$

where the width $\Delta\omega$ is measured at the height of half amplitude in the energy spectrum (Fig. 3.3). The higher the quality factor, the sharper the resonance peak and the longer the duration of vibration.

I briefly want to discuss the assumptions underlying the above calculation. Validity of the assumptions will be proven by the agreement between the theoretical results and the experimental ones presented in the following section.

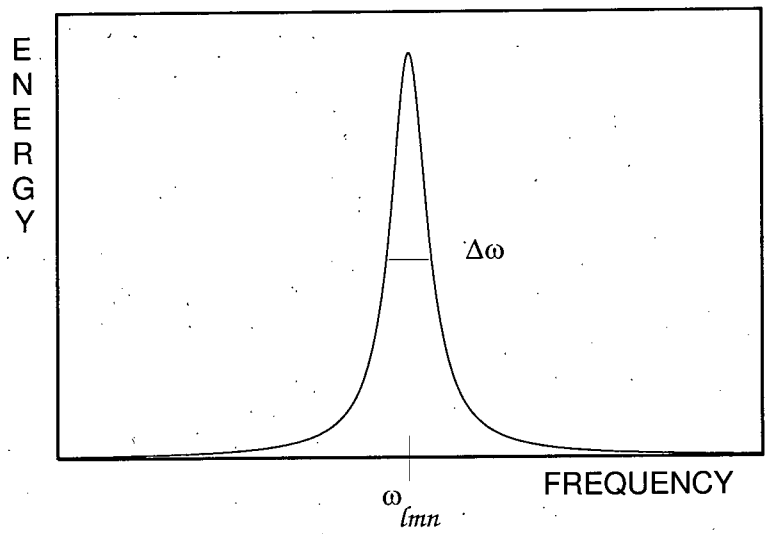


Figure 3.3: Resonance Curve.

In all the above calculations, it was assumed that the surrounding walls be smooth, rigid and ideally reflecting. I didn't allow for the water surface to be rippled, which would result in wave scattering. Friction along the walls was neglected.

For propagating sound waves, each medium has a characteristic impedance which is defined as the ratio of pressure to particle velocity:

$$Z = \frac{P}{u} \quad .$$

Generally, a wave hitting an interface gives rise to a reflected and a transmitted wave. The reflection coefficient R, defined as the ratio of the amplitude of the reflected wave to the amplitude of the incident wave, can be expressed in terms of the impedance of the transporting medium Z_0 and the impedance of the reflecting surface Z_S [Ford 1970, p. 107]:

$$R = \frac{Z_S - Z_0}{Z_S + Z_0} \quad .$$

Usually, all the impedances are complex, leading to a complex reflection coefficient. This means that the incident and reflected waves differ in both amplitude and phase. In the present example of sound waves in a pool, it was assumed that $Z_{wall} = \infty$. Thus, there is no transmitted wave in the walls, i.e. no absorption by the walls. The reflection coefficient is $R = 1$, meaning that the reflected wave has the same amplitude as the incident wave and no phase shift occurs (Fig. 3.2). The characteristic impedance of air was set to be zero, $Z_{air} = 0$. There is no absorption either, but a phase shift of 180 degrees.

Furthermore, attenuation due to the water inside the pool was neglected. A mathematically more elaborate presentation is given in [Morse and Ingard 1968]. During the experiments, the sounds were transmitted into the water by a sound projector. A projector emits soundwaves within a cone, where the opening angle depends on the frequency. Thus, there aren't only normal incidences to the surrounding surfaces but also oblique incidences. Further neglecting the slightly (over a length of approximately 20 cm) rounded corners of the pool and the effect of the gate with a cross section of 1 m^2 through which the whales enter, the low resonance frequencies are calculated.

The experiment pool had the following dimensions: $a = 6.1$ m, $b = 7.6$ m, $c = 3.0$ m. The sound speed of ocean water depends on temperature and salinity. The values for the aquarium pool were $T = 13°C$ and $S = 27.5$ ppm. The value of the corresponding sound speed was obtained from the UNESCO reports [UNESCO 1987]: $v = 1500\ \frac{m}{s}$. Table 3.1 shows some of the lowest resonance frequencies $f = \frac{\omega}{2\pi}$.

For these pool dimensions, there are 15 eigenmodes below 400 Hz. For higher fre-

l	m	n	f/Hz
0	0	0	125
0	1	0	159
1	0	0	175
1	1	0	201
0	2	0	234
1	2	0	264
2	0	0	276
2	1	0	293
0	3	0	321
2	2	0	339
1	3	0	344
0	0	1	375
0	1	1	388

l	m	n	f/Hz
3	0	0	389
1	0	1	395
3	1	0	402
2	3	0	405
1	1	1	407
0	4	0	414
0	2	1	424
1	4	0	432
3	2	0	437
1	2	1	441
2	0	1	448
2	1	1	459
0	3	1	478

l	m	n	f/Hz
2	4	0	482
3	3	0	489
2	2	1	490
1	3	1	493
4	0	0	507
0	5	0	509
4	1	0	517
1	5	0	524
3	0	1	526
3	1	1	535
2	3	1	537
0	4	1	544
4	2	0	544

Table 3.1: Resonance Frequencies of the Aquarium Pool.

quencies, the number of modes within a specified bandwidth increases, i.e. the density of eigenmodes increases with frequency. In fact, it is proportional to the square of frequency, [Ford 1970, p. 64]:

$$\delta N = \frac{4\pi V}{v^3} f^2 \delta f \quad .$$

$V = 139$ m^3 is the volume of the pool. Due to its geometrical derivation, this formula is not accurate for low frequencies, but gives very good results for the higher resonances.

For example, within a bandwidth of ±100 Hz around 10 kHz, there are 10,000 resonances. One may wish to integrate the equation to get the total number of eigenfrequencies below a boundary frequency f_b:

$$N = \frac{4\pi}{3}\frac{V}{v^3}f_b^3 \quad . \tag{3.7}$$

For a boundary frequency of $f_b = 10$ kHz, there are 173,000 resonance frequencies below it! These are obviously too many to measure separately. However, the mathematical evaluation can easily be checked for the low resonances.

3.1.2 Experimental Evaluation and Conclusions

The goal was to verify the previous calculations and to estimate the validity of the assumptions. A block diagram of the measurement setup is shown in Fig. 3.4. The sine wave

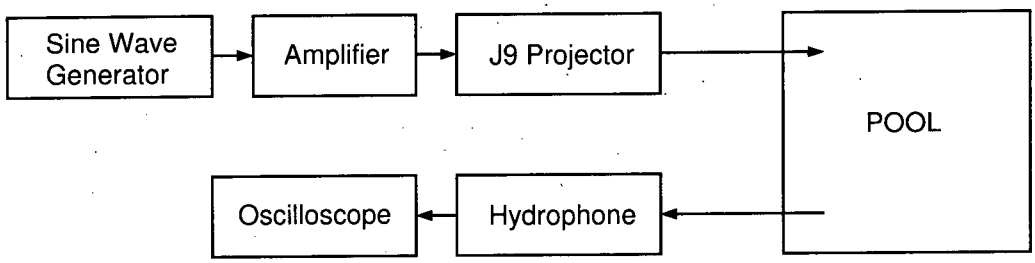


Figure 3.4: Setup for Measuring the Low Resonance Frequencies.

generator provided an analog alternating voltage of determined frequency. This voltage was amplified and converted to an underwater pressure wave by a J9 projector. A special supporting device was manufactured for the projector allowing it to be positioned anywhere inside the pool. This device was mounted on a little cement edge running around

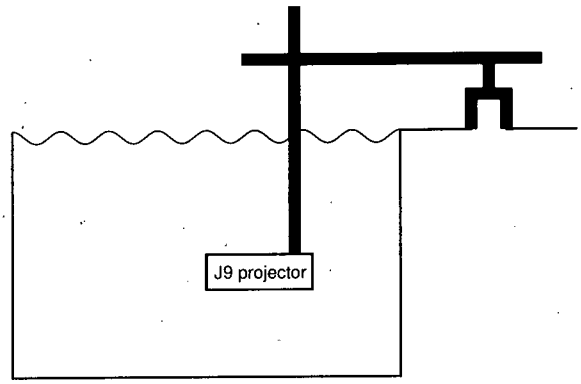


Figure 3.5: Supporting Device for the J9 Projector.

the pool. It mainly consisted of two adjustable metal arms with which the overhang and depth of the projector could be varied. The construction is sketched in Fig. 3.5. A Brüel & Kjær 8101 hydrophone measured the resulting pressure fluctuations in the water and provided an analog voltage at its output. It was simply lowered into the pool on a rod. The voltage was measured by an oscilloscope.

Pure tones within a frequency range of 100-400 Hz were transmitted into the water. It was attempted to detect some of the low resonances given in Table 3.1. In order to excite a special eigenmode, the projector had to be positioned in one of the pressure antinodes. Consider for example the 011-mode. Its pressure distribution due to Equation 3.4 becomes:

$$P(x,y,z) = \hat{P} \cdot \cos(\frac{\pi}{b}y)\cos(\frac{3\pi}{2c}z) \cdot e^{-i\omega_{011}t} \quad .$$

There is no node in $x-$direction. This means, the pressure is uniform in $x-$direction and the projector may be placed anywhere between 0 and a (Fig. 3.1). The mode has one node in $y-$direction at $y = \frac{b}{2}$. Therefore, the pressure antinodes are at $y = 0$ and $y = b$.

For maximum excitation, the projector has to be either near the wall $y = 0$ or the wall $y = b$. In z–direction, there is one node at $z = \frac{c}{3}$. The antinodes are at $z = 0$ and $z = \frac{2}{3}c$. These are the possible positions for the projector in z–direction. Combining the results for all three directions gives the final position in which the projector had to be fixed. A pure tone with a constant frequency of the calculated 388 Hz for the 011-mode was transmitted. The hydrophone was then gently moved through the water while the resultant voltage was observed. Because of the limited output power of the sound generation part and the low sensitivity of the recording hydrophone, it was impossible to measure all pressure nodes and antinodes throughout the pool. Only in close vicinity to the J9 projector, could signals be recorded. Holding the hydrophone fixed in a position of high output voltage, the frequency at the signal generator was adjusted to get an even higher output. This was to account for frequency shifts due to deviations of the pool from an ideal rectangular enclosure. With this method, the lower resonance frequencies of the pool were measured and turned out to match the theoretical ones within 1 %.

For higher frequencies, it is almost impossible to measure single eigenmodes because of their increasing density. As described in the previous section, each resonance can be excited within a frequency band $\Delta\omega$ according to its quality factor. If the resonance frequencies become so dense that the resonance curves overlap, coupling between the modes is possible. This means that if one mode is excited, energy might be transferred into a different mode provided that the mode configurations allow this. Therefore, it becomes difficult to decide which modes were initially evoked by the projector and which

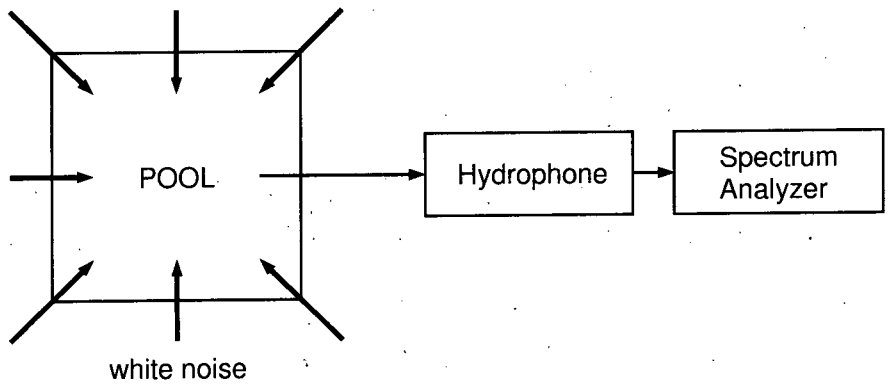


Figure 3.6: Setup for Measuring a Broadband Frequency Spectrum.

modes are measured. For higher frequencies, the measurement was therefore carried out in a different way (Fig. 3.6). The recording hydrophone was lowered into the pool and fixed at a position 1 m below the water surface and 1 m off the walls $x = a$ and $y = b$. No signal was generated inside the pool. It was simply recorded whatever oscillations were established due to random white noise and background noise. Averaging over a long period of time also leads to a characteristic pool spectrum.

Fig. 3.7 shows the spectrum up to 10 kHz, which was recorded overnight for more than 15 hours. There are all sorts of low frequency noise around the pool. Various motors and instruments operate continuously, such as to clean and filter the water and to control water conditions like temperature and oxygen content. Also the water level is raised repeatedly in order to let leaves and filth which collected on the surface flow out. Low frequencies are more persistent and penetrate the walls around the pool more easily than high frequencies. The sound pressure level at very low frequencies is about 100 dB, which is approximately 40 dB below the whale's hearing threshold (Fig. 3.12). I conclude that

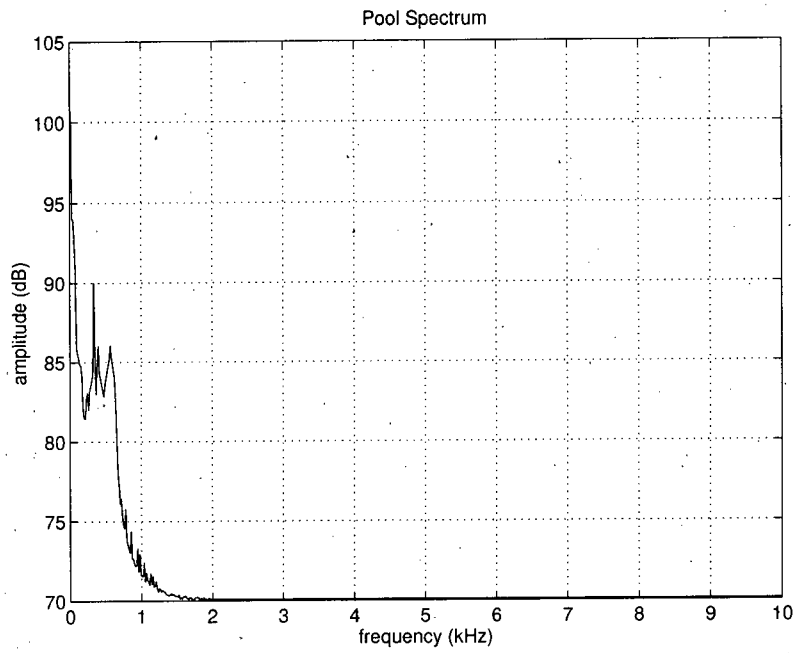


Figure 3.7: Pool Spectrum (dB re 1 μPa) averaged over 15 hours.

they are therefore undetectable and also won't add to the masking by the ship noises.

There is a peak at 350 Hz with an amplitude of 90 dB, 20 dB below the beluga hearing threshold. As the spectrum analyzer had a bandwidth of 20 Hz, this could be the 1-3-0-resonance, which would couple very well to the position of the hydrophone in this measurement. There are two more peaks up to 600 Hz, which are also 20 dB below the beluga hearing threshold. If any of these peaks are resonances, they can be excited by the noises with energy in the low-frequency range. To test this, I played the ship noises into the pool using the J9 projector at the location where it would be during the masking experiments. With the B&K hydrophone at the future location of the whale, I measured the spectra and observed them on the spectrum analyzer. All the spectra did

show a peak at 350 Hz, which was raised about 20 dB above the ship noise spectra. No other distortions to the ship noises could be identified. The 350 Hz component therefore is indeed a beautiful resonance. It could not be seen when the beluga vocalization was played. This is because it cannot be excited by the call whose lowest frequency component lies at 800 Hz. The 350 Hz resonance thus does not "facilitate" call detection in noise. Furthermore, as it lies well below the lowest call component, it is unlikely to add to the masking ability of the noises.

For frequencies greater than 600 Hz, the pool spectrum in Fig. 3.7 converges to 70 dB re 1 μPa. This does not correspond to a broadband white background noise in the pool. It is the sensitivity limit of the spectrum analyzer. With the chosen settings, the spectrum analyzer shows an apparent white noise of 70 dB SPL when no input is connected.

In conclusion, low-frequency noises around the pool lie considerably below the animal's hearing threshold. They are therefore undetectable and also won't add to the masking by the ship noises. The only distinct resonance in the frequency range up to 10 kHz, which is excited by the sounds used and the particular projector/hydrophone configuration chosen for the masking experiment, lies at 350 Hz. It is thus unlikely to have any serious masking effect on the call whose energy lies between 800 Hz and 8 kHz. At least, its contribution to masking would be the same for all the noises, as it shows up with the same amplitude. Altogether, it was not necessary to undertake any alterations in the pool, such as mounting wooden pyramids on the walls in order to convert the pool into an anechoic chamber.

All these calculations and measurements dealt with an empty pool. The presence of

a beluga whale could disturb the geometry of some eigenmodes and shift the resonance frequencies to higher frequencies. The effect can easily be estimated via the beluga volume. Suppose a beluga whale has an average length of 3.20 m and an almost triangular (!) cross section. Let the sides of the equilateral triangle be 60 cm long, then the cross section becomes 0.156 m^2, yielding a beluga volume of $V_b = 0.5$ m^3. The volume of the pool is $V_p = 139$ m^3. The presence of the beluga therefore decreases the volume in which resonance frequencies can develop by 3 per thousand. According to Equation 3.7, also the number of resonance frequencies below 10 kHz decreases by 3 per thousand, which are about 500 resonances less than the previous 173,000, hence negligible. To estimate the shift in resonance frequencies, Equation 3.5 has to be examined. The volume $V = abc$ is not easy to incorporate, therefore a cubic pool is considered instead. For $a = b = c$, ω_{lmn} becomes proportional to $\frac{1}{a}$. If the volume decreases by 3 per thousand, $V = V_0 \cdot 0.997$, the pool length decreases by 1 per thousand, $a = a_0 \cdot \sqrt[3]{0.997} = a_0 \cdot 0.999$. The frequencies would shift by the same amount, 1 per thousand, which is negligible. Furthermore, as the density of a beluga whale (which like human beings is mainly made of water) is so close to the density of the water in the pool, the disturbance of the resonance frequencies would even be much less than assumed.

3.2 Whale Training

The following sections describe the crucial steps of the whale training taken to prepare the animals for the acoustic masking experiments.

3.2.1 Familiarization with Experiment Equipment

As a first step, the animals had to get used to unfamiliar objects in the pool. These were mainly the J9 sound projector and its supporting device (Fig. 3.5). Two different reactions of the belugas were observed. Nervousness, upset or even anxiety on the one hand and curiosity to the extent that the instruments were closely inspected and almost attacked on the other hand. Especially Nanuq quickly developed a strong dislike and animosity and literally threw the instruments out of the pool. In fact, I had always been concerned that the 17 kg heavy J9 construction might fall into the pool and therefore had had the supporting device designed to resist a torque pulling downwards. In order not to waste too much time on long familiarization and to avoid constructing a whole new supporting device for the J9 which would also be secure in case of strong upward pushes, a whale trainer simply held the J9, fixed on a metal bar, into the water by hand. This was accepted by all the whales right away!

3.2.2 Stationing and Waiting in front of the Sound Projector

As a second step, the animals were trained to station against a 20 cm long bar, fixed 1 m in front of the J9 projector, for 30 seconds. In order to train any of the behaviour required in my experiments, the whale trainers made use of a variety of signals well-known by the whales from other trained behaviour. For instance, the animals recognize a long pole with a white ball at its end. A slap onto the water surface with this pole calls a whale from anywhere inside the pool to the trainer. The whales approach the pole with

their head and station their melon against the white ball. A whistle is blown every time a whale does something correctly, and usually, fresh fish or squid follow as a reward. There is no negative reinforcement as such, only denied positive reinforcement. Any behaviour is asked for by the use of handsignals. There are different handsignals for all sorts of behaviour ranging from husbandry behaviour to medical behaviour and show behaviour. Just by combining these previously learned signals and behaviours, it was no problem to send Nanuq, Imaq and Aurora to the J9, to station them against the short bar and to recall them with the trainer's pole.

Making them station still for up to 30 s though turned out to be rather demanding, because it represented unnatural behaviour. Furthermore, the animals were required to station straight not bent upwards, downwards or to the sides. Their body had to form one line with the stationing bar and the J9 projector. Only if the animal directly faced the bar and J9, could the same head position be assured from trial to trial and the same sound pressure level on both sides of the head be assured. The animals were also not allowed to push against the bar which would bend it and result in a changing distance between the animal's head and the sound source from trial to trial. If an animal stationed itself against the bar too strongly, it had to release pressure until the bar would just touch its melon.

3.2.3 Recognition of the Beluga Vocalization

Step three comprised the recognition of the beluga call by the animals. It was originally planned to record vocalizations of the five adult beluga whales at the Vancouver Aquarium, but I was advised otherwise. Whales and dolphins were reported to show extreme disturbance reactions when their own songs were played back to them. Some screamed aggressively, others attacked the projectors. I did not observe any aggression or disturbance reactions from Nanuq, Imaq or Aurora when the beluga call recorded from wild animals in Maxwell Bay was played to them.

At this point, I often get asked whether the belugas in the aquarium might not "know", i.e. "understand", the vocalizations of their wild conspecifics and if that doesn't pose a problem. No it does not, because my study does not deal with the potential meaning of the particular vocalization. I simply want to know if the trained beluga whales can detect the vocalization in noisy backgrounds independent of its actual "meaning". Beluga whales have an extremely large repertoire of vocalizations, and it seems as if different populations use different vocalizations. Therefore, beluga whales from the western Arctic might in fact not be able to "understand" beluga whales from Hudson Bay. For this study it is assumed that the auditory systems of all the beluga whales are the same and that the captive, trained animals are representative for the entire species when it comes to the whales' call detection abilities in noise.

Figure 3.8: Photo of the main pool and the adjacent experiment pool.

Figure 3.9: Michelle stations Aurora. Alysoun holds the J9 projector. I operate the computer.

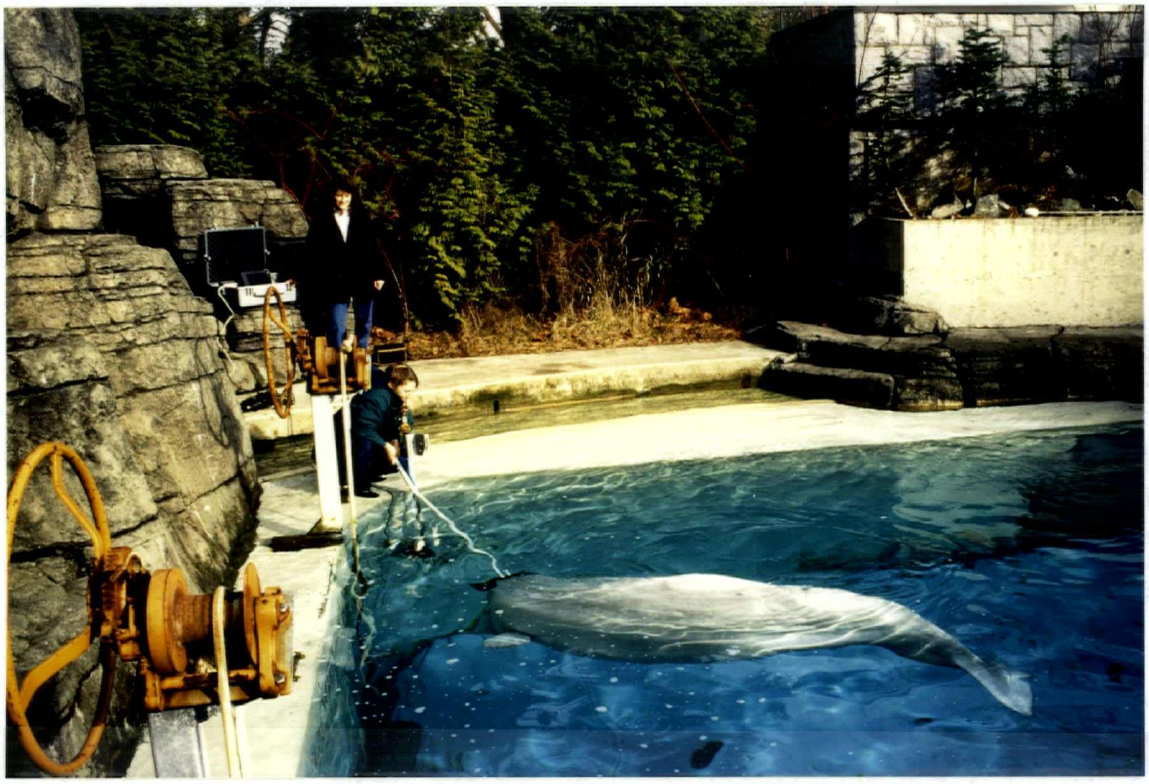


Figure 3.10: Aurora has been sent over to Alysoun and stations in front of the J9 projector.

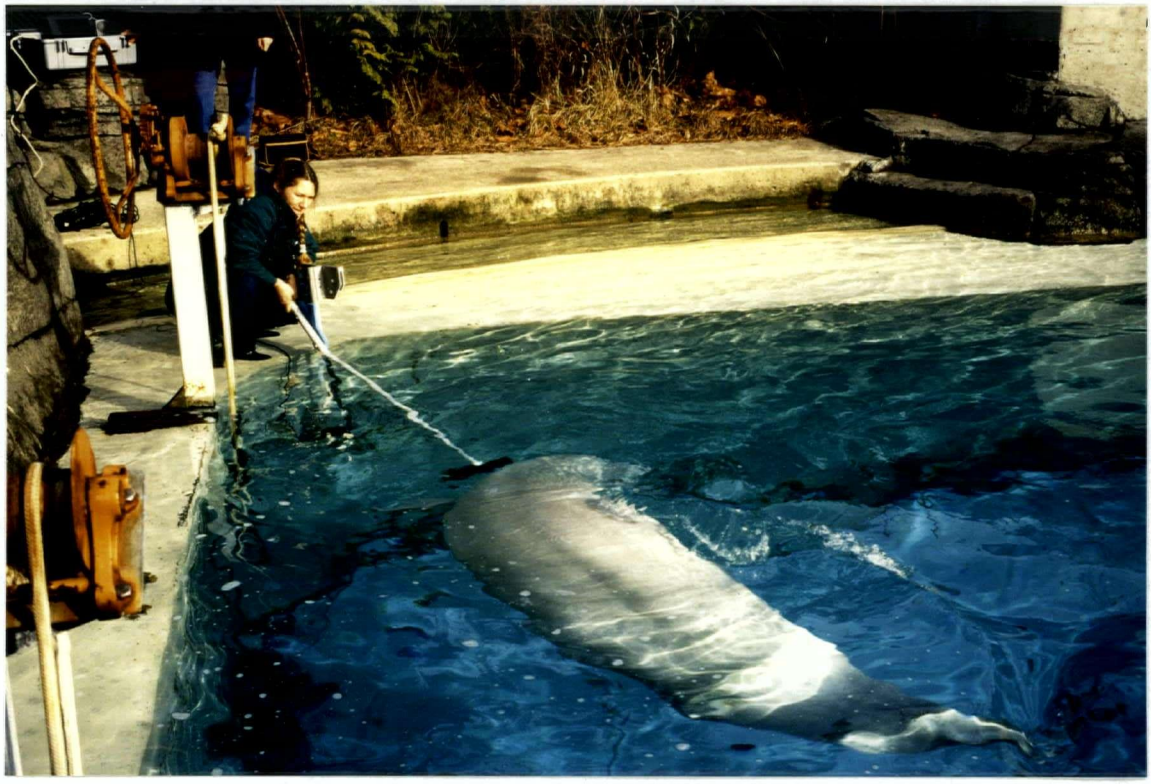


Figure 3.11: Upon transmission of a beluga call, Aurora breaks away from the J9 in less than 1 s.

Fig. 3.8 shows a photo of the beluga pool at the Vancouver Aquarium. The main pool is in the foreground. A smaller pool can just be seen in the back to the right. This is the so-called medical pool in which I did my experiments. The two pools are connected by an underwater gate. During the training and the experiments, there is always only one animal in the small pool.

Nanuq, Imaq and Aurora were initially trained to react to the transmission of the beluga vocalization. Circumstances eventually only permitted training and data collection with Aurora to proceed. In photo 3.9, Michelle stations Aurora on the right side of the pool. Alysoun holds the J9 with the stationing bar into the pool on the opposite side. I operate the sound generation equipment standing on the rocks next to the pool. Michelle then sends Aurora over to the J9 with a long arm pointing. In photo 3.10, Aurora stations against the stationing bar in front of the J9 projector. Within 30 s I transmit the beluga vocalization. Aurora is supposed to break away from the target and swim back to Michelle if she hears it. Otherwise, if she didn't hear the call, she holds station until Michelle recalls her. Photo 3.11 shows Aurora breaking away from the target after a beluga vocalization was transmitted. Her reaction time was less than a second during which she pulled her head off the bar. It then always took another few seconds for the entire "body" to get moving and swim back to Michelle. Training this behaviour was unexpectedly difficult. The marine mammal staff and I tried three different methods.

First, Aurora was sent to the J9 and stationed. When Michelle recalled her by slapping the white pole onto the water, I transmitted the beluga call at the same time. We then

gradually delayed Michelle's slap, i.e. I played the call first and then Michelle slapped the water and recalled Aurora. We thought that Aurora would anticipate the slap and start breaking on the transmission of the call. Unfortunately, this never happened. Beluga whales are some of the few whales whose cervical vertebrae aren't fused, hence allowing Aurora to station with her head but twist her body just enough out of the way to observe Michelle out of the corner of her eye. As soon as Michelle slightly lifted the pole to recall Aurora, Aurora broke. These were the only cues Aurora paid attention to, not the transmitted vocalization. We gave up after a couple of weeks and tried the second method.

Aurora was stationed in front of the J9 projector. I played the beluga vocalization again and again until she finally broke. Michelle's slap followed, a whistle, lots of fish, friendly head patting and praise! By doing this over and over again, Aurora actually started to break earlier and earlier until she eventually reacted to the first transmission of the beluga call. Unfortunately, after a while, Aurora started to anticipate even the first call and broke right after she was stationed with no call transmitted and no pole slap by Michelle. This became serious enough for us to go back to step 2 of the training to increase Aurora's patience and make her station in front of the J9 for up to 30 seconds again.

As a third method, we stationed Aurora in front of the J9; at a random time between 3 and 30 seconds, I transmitted the beluga vocalization; no reaction from Aurora. Then Alysoun pulled the J9 out of the pool which made Aurora slowly swim back to Michelle.

Whistle, fish, patting, laudation! Doing this over and over again to the "pain" of the whale trainer who had to lift the 17 kg heavy J9 construction into and out of the water, and gradually delaying the retrieval of the J9 eventually taught Aurora.

3.2.4 Recognition of Background Noises

Once Aurora reliably broke away from the J9 stationing bar upon transmission of the beluga vocalization, we had to train her to hold station if ship noise or icecracking noise was transmitted. The marine mammal staff and I had expected this to be the most difficult part after the problems with training her to break on a call in step 3. However, step 4 of the training was the shortest of all. Aurora immediately held station upon transmission of a noise signal. It shall be mentioned here again, that all the signals played to her were 2 seconds long and had the same rms pressure amplitude (Section 2.5). In order to keep the difference between call and noise signals, Michelle whistled immediately when Aurora broke on the beluga vocalization and also called her back right away. After the transmission of a noise signal, on the other hand, we let Aurora station for another five seconds and only then did Michelle whistle and slap. Once Aurora reacted reliably to the transmission of the call and the three noises in any order, we introduced mixed signals.

3.2.5 Reaction to Mixtures of Call and Noise

The fifth and last step of the whale training comprised the training of break/station reactions to mixed signals containing both call and noise. Aurora should break on all

mixtures in which she could detect a call and hold station on all others. We used mixtures with equal loudness of call and noise, i.e. noise-to-signal ratios of 1, and trained Aurora to break on those the same way as in step 3 of the training. Once random sequences like pure ramming noise (station), bubbler-call mixture (break), pure call (break), icecracking-call mixture (break), pure bubbler noise (station), pure icecracking (station), pure call (break), ramming-call mixture (break) etc. were handled excellently, we decided to begin the masking experiments.

3.3 Call Detection Experiments in Noise

3.3.1 Pure Tone Audiogram

In order to assure that the animal used for the masked hearing experiments had "normal" hearing, I measured Aurora's pure tone detection threshold at 4 selected frequencies. The main energy of the sounds used in the masking experiments lies between 500 Hz and 10 kHz. Therefore, the test frequencies were chosen to be 500 Hz, 1 kHz, 5 kHz and 10 kHz. Aurora was easily trained to break away from the J9 projector when one of the 4 pure tones was transmitted.

Hearing thresholds were measured in a so-called *titration method.* One starts at a tone volume which the animal can easily detect. The sound pressure level is then reduced gradually. I chose to halve the amplitude from step to step. Once Aurora stopped breaking, the volume was increased again to the last one she had heard. She basically

always broke on this one. If not, the volume would be further increased until she heard the tone again. After that, the tone volume would be decreased until she stopped breaking. Stepping up and down in this fashion, closely approaches the threshold from the quieter and louder side. The threshold is calculated as the average of both the upper and the lower reversal points. I did this titration three times for each frequency and calculated the mean.

Aurora's detection thresholds for 500 Hz, 1 kHz, 5 kHz and 10 kHz were respectively: 107 dB, 90 dB, 82 dB and 64 dB. I found three references in the literature to compare these data with. The earliest reference [White *et al.* 1978] measured an audiogram of a male and a female beluga whale between 1 kHz and 123 kHz. The second publication [Awbrey *et al.* 1988] lists audiograms of three beluga whales between 125 Hz and 8 kHz. The third reference [Johnson *et al.* 1989] gives data for low-frequency hearing in one beluga whale between 40 Hz and 4 kHz. I took all the data available and calculated the mean at frequencies where more than one data point existed. Fig. 1.1 showed the mean of all these published data. I then wrote a short interpolation program which would plot the resulting audiogram at the center frequencies of the beluga auditory filter (see Section 1.3). Results are shown in Fig. 3.12. Aurora's threshold at 500 Hz lies exactly on top of the average published audiogram. Her threshold at 1 kHz is about 10 dB lower. Her threshold at 5 kHz, on the other hand, is 13 dB higher than the threshold of the one comparison animal available for this frequency. The 10 kHz threshold lies 5 dB above the published audiogram. Given that threshold variations from individual to individual

but also from day to day (in particular for behaviorally collected data) can be as large as 20 dB [Awbrey *et al.* 1988], Aurora's audiogram can be regarded as "normal" compared to the six published ones.

3.3.2 Pure Call Hearing Thresholds

In order to calculate the absolute masking, i.e. the threshold shift of the beluga vocalization in the presence of noise (Section 1.3, Equation 1.2), I measured the pure call detection threshold in a "quiet" pool.

To begin with, I played the beluga vocalization with an rms amplitude of 1 and adjusted the hardware settings, i.e. the amplifier gain, such that the transmitted call had a loudness which subjectively resembled the loudness of the beluga vocalizations generated by the animals in the aquarium. This was considered a comfortable loudness for them to listen to. Then I successively decreased the loudness halving the sound pressure from step to step. In a titration fashion, Aurora's pure call detection threshold was measured. It lay at a total sound pressure level of 108 dB re 1 μPa.

In order to see if other methods but the standard titration would give different results, we had one experiment session during which I played the calls in opposite order. I started at the quiet end, which I defined as one 512th of the initial volume, and increased the sound pressure by a factor two from step to step until Aurora broke on one mixture. The sound pressure level was then decreased again and I proceeded stepping around the threshold in the same fashion as during the standard titration. The computed threshold was the

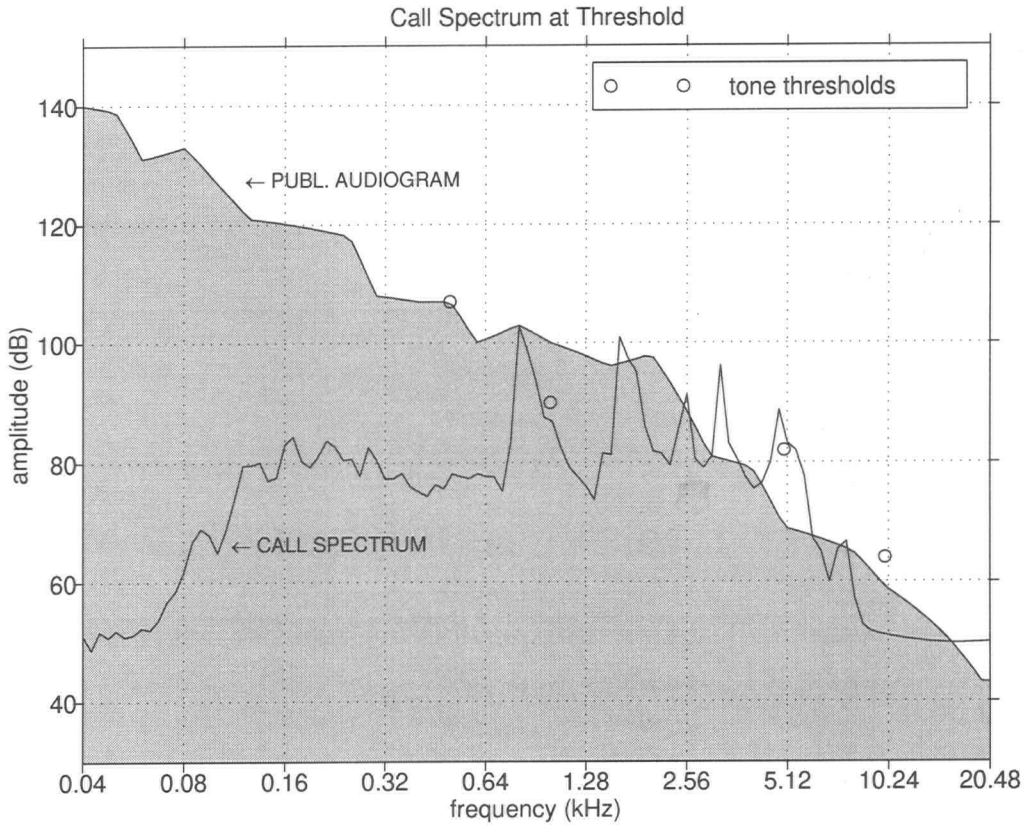


Figure 3.12: The shaded area marks the mean of 6 published audiograms. Aurora's thresholds are indicated as circles. The solid curve shows the 12th octave band levels of the vocalization at detection threshold. Amplitudes are sound pressure levels in dB re 1 μPa.

same. As a third method, I played the vocalizations unordered at random loudnesses and took the threshold at the 50% probability of breaking. For the chosen step size of double pressure, the threshold lay between the same two steps for all the three methods. I did not find biased thresholds for the three different methods, therefore, the above mentioned threshold of 108 dB re 1 μPa is the mean of all the three methods.

Fig. 3.12 shows the 12th octave band levels of the call spectrum at the detection threshold. I find it very interesting that the animal stops reacting to the call when

it can just hear all major frequency components. I had expected Aurora to continue breaking for as long as the 5 kHz component would stand out over the audiogram. By that stage all the other frequencies would have dropped below detectability. From a physical point of view it seems reasonable that Aurora stops breaking as soon as one peak falls below audibility. A pure tone cannot carry much information, and in order to recognize a complex vocalization, information from harmonics is necessary. As most of the call energy lies at the low frequency harmonics, their presence might be of particular importance. It shall be noted that the call spectrum as plotted in Fig. 3.12 represents band levels averaged over the 2 s duration of the call. As the call is of pulsed nature, the sound pressure level of the individual pulses is about 6 dB higher than shown. This slightly raises the call above the audiogram.

3.3.3 Masked Call Hearing Thresholds

A sketch of the experiment setup is drawn in Fig. 3.13. The experiments were carried out in the little medical pool which is adjacent to the big main pool. Aurora, the beluga whale, first stationed against a pole held into the water by a whale trainer. Then she was sent over to the other whale trainer who held the J9 underwater sound projector with a stationing bar fixed 1 m in front of it. The J9 is an Argotec model with a bandwidth of 40 Hz to 20 kHz. For the frequency range I was interested in, the frequency response is flat to within 3 dB between 100 Hz and 10 kHz, and slightly more variable up to 10 dB between 10 kHz and 20 kHz.

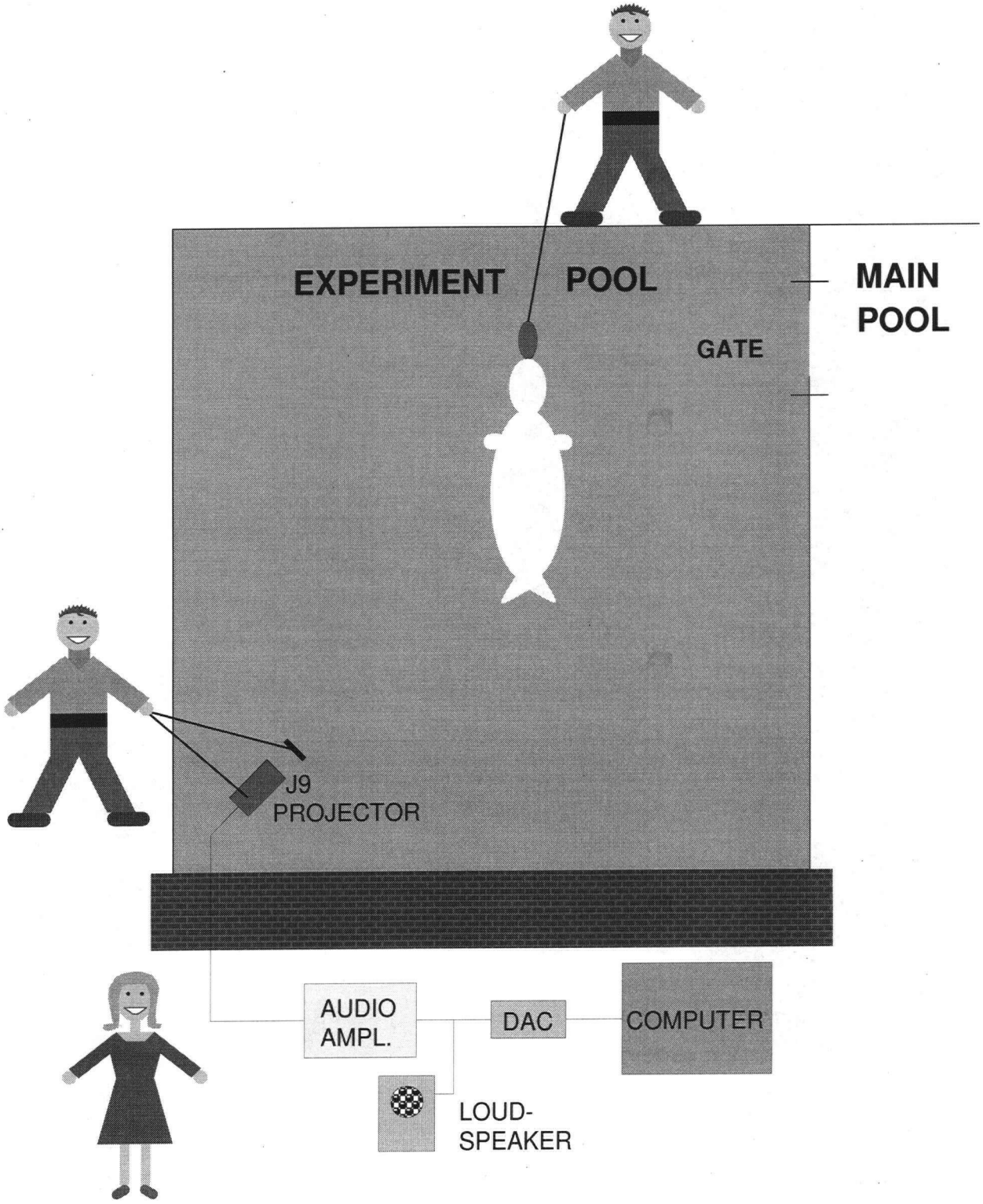


Figure 3.13: Setup for Masked Hearing Experiments.

The experiment conductor, myself, sat behind a wall of rocks out of sight of the animal and the trainers. This was to avoid passing on any cues about when and what kind of signal was transmitted. All the data of the pure beluga vocalization, the three noises and the mixtures were digitally stored on a portable notebook computer. The sampling rate was 44 kHz, the resolution 16 bit. An external soundcard, PORT·ABLE Sound Plus™ from Digispeech, converted the data from digital to analog format. They were amplified by a standard 20 W Radio Shack audio amplifier with 15 kHz bandwidth and were finally transmitted into the pool through the J9 projector. All the hardware settings were kept the same during the experiments, i.e. the volume of the transmitted signals was controlled digitally while the data were still on the computer harddrive.

To illustrate this, Fig. 3.14 shows the menu popping up on the notebook computer during the experiments. By simply pressing the Q-key on the keyboard, a mixture of icecracking noise with the beluga call in a noise-to-signal ratio of 1 was transmitted. The second line of letters referred to bubbler system mixtures and the third line to ice-ramming mixtures. The noise-to-signal ratio increased from left to right. The four number-keys on the top corresponded to pure vocalization and pure noise signals. All the signals were 2 seconds long.

Once Aurora stationed properly against the J9 bar, the whale trainer said "steady" and within 30 seconds thereafter, I played a sound. The time between the stationing and the sound transmission was varied so as to avoid anticipation. With a little loudspeaker, I could check the successful conversion and transmission of the sound. Masked hearing

pure call 1	pure crack 2	pure bub 3	pure ram 4

crack/call nsr=1 Q	crack/call nsr=2 W	crack/call nsr=4 E	crack/call nsr=8 R	crack/call nsr=16 T	crack/call nsr=32 Y
bub/call nsr=1 A	bub/call nsr=2 S	bub/call nsr=4 D	bub/call nsr=8 F	bub/call nsr=16 G	bub/call nsr=32 H
ram/call nsr=1 Z	ram/call nsr=2 X	ram/call nsr=4 C	ram/call nsr=8 V	ram/call nsr=16 B	ram/call nsr=32 N

Figure 3.14: Menu for Transmission of Beluga Vocalization and Noise in Various Noise-to-Signal Ratios.

data were collected in a *go/no-go paradigm*. Aurora held station when she was unable to detect the beluga call in the acoustic signal. To indicate successful discrimination, she released from the pole. Whenever she did so, the whale trainer told me by calling "break". Aurora was given a two-second-reaction time. If the break occurred within this period of time after transmission of a mixed signal, I said "yes". The first trainer whistled and slapped his pole to call Aurora back for reward. If Aurora did not break, she was kept stationed for the rest of the 30 seconds. She was finally released by the first trainer's pole slap and also received a fish under the assumption that she could not hear the beluga vocalization. If Aurora broke later than two seconds after transmission of a mixed signal or even before anything was played, I said "no" and Aurora was stationed again without positive reinforcement.

In order to verify the whale's correct reaction from time to time, pure call and pure

noise signals were played at random. I avoided making any judgments about Aurora's behaviour to intermediate nsr mixtures. Definite mistakes were only breaks on pure noise signals and no breaks on the pure vocalization, the latter of which never happened. The functioning of the experimental concept was based on the fact that it was simply more exciting for Aurora to break away from the J9 as soon as she detected the call in the noise, rather than holding station for the entire 30 seconds. In psychoacoustics, transmission of pure noise signals, which are referred to as *catch trials*, is important, because it indicates the animal's decision bias (Section 3.3.5). In my experiments, one out of four signals was a catch trial.

Data were collected in separate bubbler noise, ramming and icecracking sessions. I tried two ways of mixing the data. One set of signals was created by holding the amplitude of the call constant at double the pressure of the pure call detection threshold, which was measured in the previous section. For increasing noise-to-signal ratio, the noise became louder and louder. The second set of mixtures was created by normalizing every mixture to a root-mean-square amplitude of 1. Thus, for increasing nsr, the noise became louder and the call became quieter at the same time. Furthermore, I tested three different paradigms of data transmission, the standard titration starting with low nsr mixtures, a titration starting at the loud-noise end and the random transmission of mixtures as described in Section 3.3.2.

3.3.4 Results

I found no difference between the three data collection paradigms or the two data mixing methods. Therefore, all the data were analyzed together. Fig. 3.15 is a plot of the number of breaks versus the nsr of the mixed signals. Each mixture was played exactly 20 times. 40 % breaks thus refers to 8 breaks and 12 no-breaks on the corresponding mixture. The call detection threshold was chosen at the 50 % break mark. Results are that naturally occurring, icecracking noise has the weakest masking effect on the beluga vocalization studied. Aurora's detection threshold lies at a noise-to-signal pressure ratio of 29.0 dB, i.e. when the noise amplitude is about 28 times as high as the call amplitude. Bubbler system noise exhibits the strongest interference with a detection threshold of 15.4 dB, in linear units, the noise was 6 times as loud as the call. The masking degree of ramming noise lies in between with a threshold at an nsr of 18.0 dB (linear nsr = 8).

It shall be emphasized here that these are relative degrees of masking based entirely on the frequency and time characteristics of call and noise, and not on their absolute amplitude. In the ocean, masking will depend on a variety of sound propagation parameters (Section 2.1.3). For known local sound propagation characteristics, the noise-to-signal ratios measured here can always be converted to masking distances between a listening whale and a speaker as well as the noise source (Chapter 8).

In the human and dolphin ear, low frequencies are more effective at masking high frequencies than vice versa; masking is maximum if the characteristic frequencies of the masker are similar to those of the probe [Pickles 1988, Ch. 9]. In my study, the main

energy of all the noises lies within the frequency range of the call. One can thus conclude that the determining factor for the degree of masking is the temporal structure of the noises. Natural icecracking noise exhibits the weakest masking effect, because of its high irregularity with respect to time. Most of the energy is concentrated in sharp broadband pulses leaving short quieter gaps in between through which the beluga whale can easily detect pieces of the call. The same argument holds for the ramming noise though its energy is more evenly distributed; the pulses have a repetition frequency of about 11 Hz compared to 2 Hz for my recording of natural icecracking noise (Figs. 2.12 and 2.13). Bubbler system noise is the most continuous one of the three with respect to frequency and time (Fig. 2.11); it therefore masks the beluga vocalization strongest.

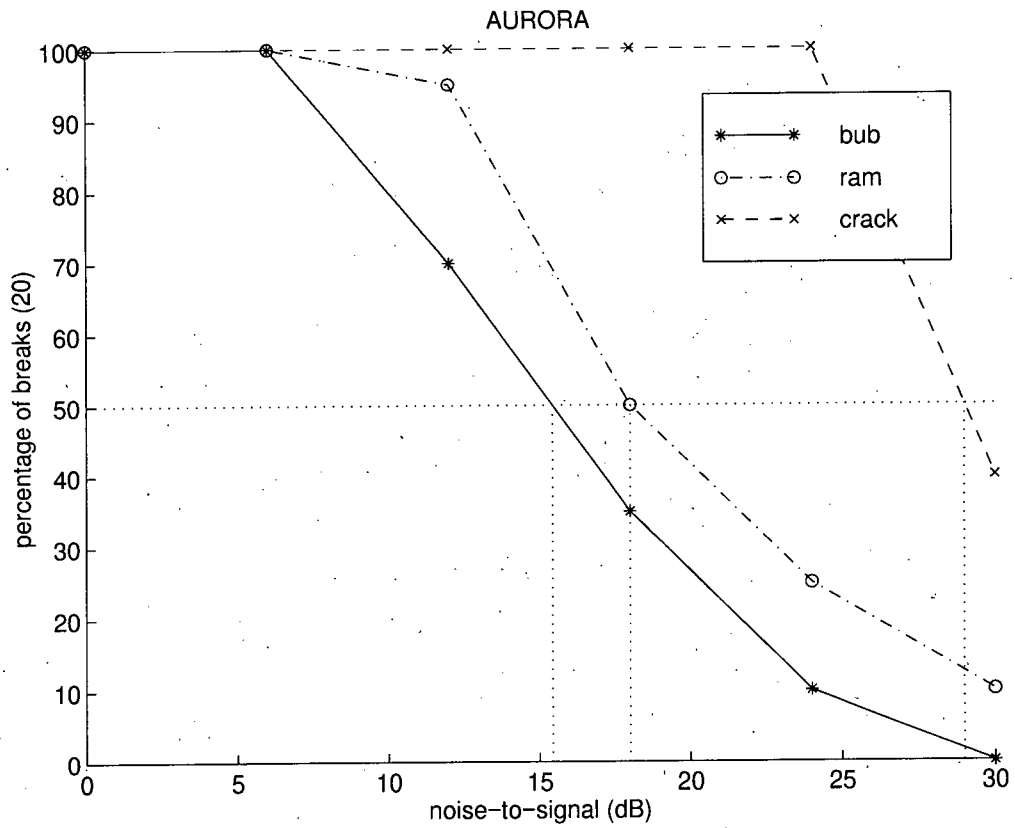


Figure 3.15: Masked hearing thresholds of the beluga whale, Aurora, in bubbler, ramming and icecracking noise. The x-axis denotes the pressure noise-to-signal ratio in dB. Every mixture of the call with the three noises in the 6 nsr's shown was played exactly 20 times. The y-axis indicates how often 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 icecracking noise.

3.3.5 Psychophysical Analysis

Psychophysics is the study of minimal signal levels which a subject can detect. An individual's psychophysical judgment, however, is determined by attitudinal or motivational factors. If correct responses are positively reinforced and incorrect responses are negatively reinforced, a subject will try to maximize the percentage of correct responses during a decision task. If the subject considers the value of the positive reinforcement to be unequal in magnitude to the value of the negative reinforcement, e.g. if one positive reinforcement "makes up for" three negative reinforcements, or if more than one correct response exists with differing reward value, the subject will try to maximize the expected payoff. Under conditions of uncertainty, the subject's decision depends on the risk he/she/it is prepared to take. *Statistical Decision Theory*, which has its foundation in the statistical testing of hypotheses in game theory, has been applied to psychophysical decision-making with the goal of isolating a subject's attitude, the *decision bias*, from the inherent detectability, the subject's *sensitivity*, [Green and Swets 1966].

A common psychophysical procedure is the *yes/no task*, in which the subject has to decide whether the presented *stimulus* was a signal added to background noise or simply the noise alone. Defining $P(Y|sn)$ as the probability that a given stimulus event sn (signal plus noise) will evoke a response Y ("yes", the signal was detected in the noise), the following *stimulus-response matrix* can be set up (Fig. 3.16). If a subject successfully detects a signal, in my case a vocalization, in a call-noise mixture, the response is called a *hit*. If the subject answers "no" to a pure noise signal, the response is called

a *correct rejection.* These are the two possible correct responses which receive positive reinforcement. If the subject's response to a mixed signal is "pure noise", one talks about a *miss.* A *false alarm* is the wrong identification of a call in a pure noise signal. These are the two possible incorrect responses which receive negative reinforcement. The matrix has only two degrees of freedom, not four as its dimension might suggest, because the rows add to 1:

$$P(Y|sn) + P(N|sn) = 1$$

$$P(Y|n) + P(N|n) = 1$$

Stimulus Alternative \ Response Alternative	yes	no
signal+noise	$P(Y\|sn)$ HIT	$P(N\|sn)$ MISS
pure noise	$P(Y\|n)$ FALSE ALARM	$P(N\|n)$ CORRECT REJECTION

Figure 3.16: Stimulus-Response Matrix of the Yes/No Procedure.

With two degrees of freedom, stimulus-response characteristics of a subject can be represented in a two-dimensional graph. With the two independent probabilities $P(Y|n)$ and

$P(Y|sn)$ on the x- and y-axis, such a diagram is called the *receiver-operating-characteristic (ROC) graph.* The first picture in Fig. 3.17 shows a sample ROC graph. Along the major diagonal, the probability of a subject saying "yes" to a signal in noise and to pure noise is equal, $P(Y|sn) = P(Y|n)$. The subject's response is completely random. Therefore, the major diagonal is also called the *chance line.* Above the chance line, the subject successfully tries to detect a signal in the noise, i.e. increases the number of hits over the number of false alarms. Below the chance line, the subject gives more false alarms than hits. Its performance is worse than chance; the subject makes deliberate mistakes. Provided a subject understands what its task is and does not make deliberate mistakes, if the individual is prepared to take great risks, it will increase the hit rate, though at the expense of an increased false alarm rate. Such an individual is characterized as *liberal*; the stimulus-response matrix will be a point in the upper small triangle, to the right of the minor diagonal. If on the other hand, the subject says "yes" only when it is absolutely sure that a call was played, the individual will decrease the false alarm rate but will also have a lower hit rate. The stimulus-response matrix will be a point to the left of the minor diagonal. Such an individual is called *conservative.*

Fig. 3.17 shows the ROC graphs for bubbler, ramming and icecracking noise. Plotted are the stimulus-response matrices for the noise-to-signal ratios of 1, 2, 4, 8, 16 and 32. Common features in all three plots are, that the hit rate is 1 for low nsr and monotonically decreases with increasing nsr. The animal maintained a conservative decision bias with a few exceptions. The point corresponding to an nsr of 4 for bubbler noise lies in the

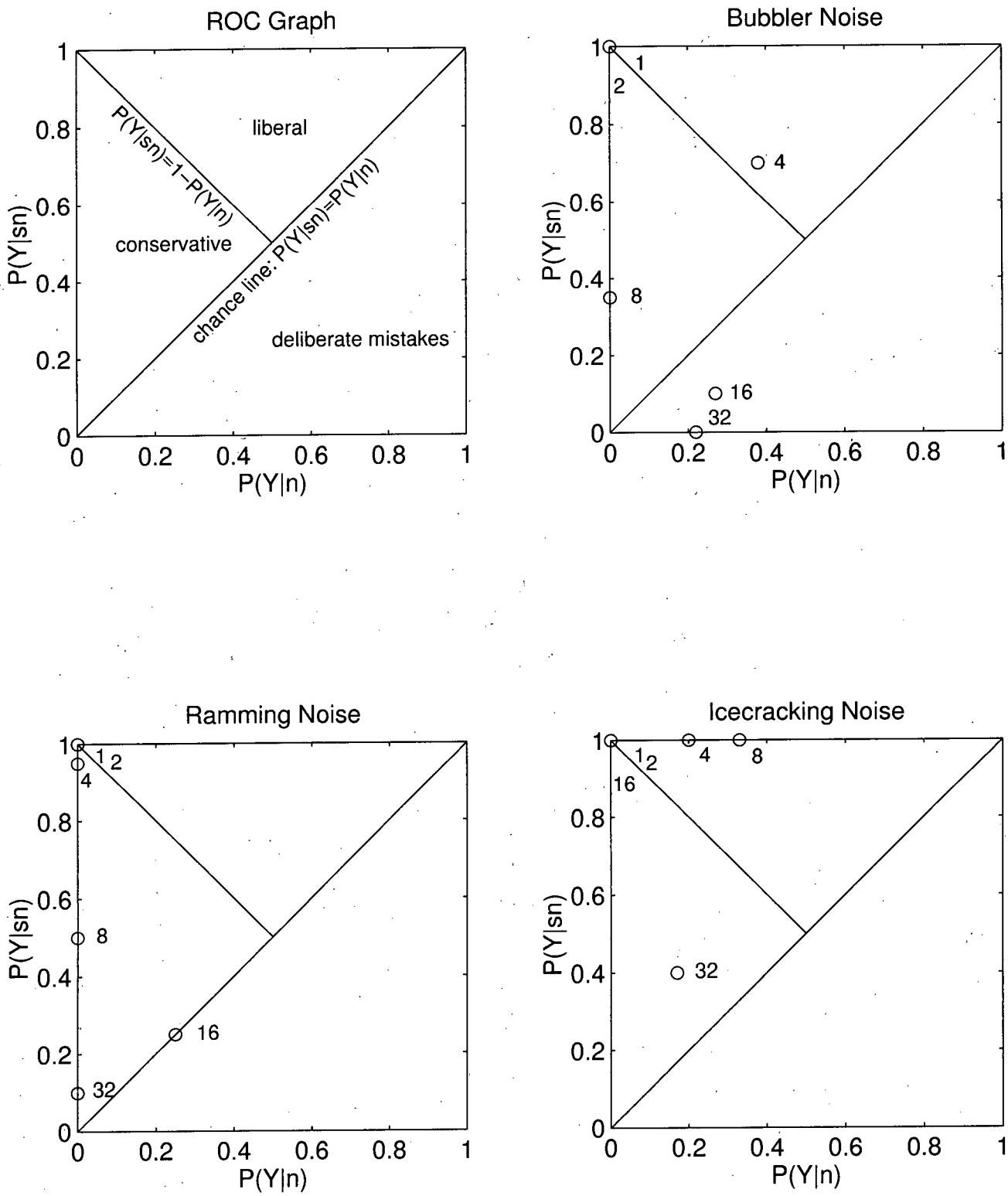


Figure 3.17: ROC graphs for Bubbler, Ramming and Icecracking Noise.

liberal zone; so do the points for an nsr of 4 and 8 for icecracking noise. In the case of bubbler noise with an nsr of 16 and 32, the animal seems to have made deliberate mistakes. One has to keep in mind, that the statistical ensemble is very small; every nsr was played only 20 times and had only 7 catch trials. In a sequence of stimuli of varying nsr, catch trials were assigned to the preceding mixture. A more realistic statistical analysis would require more data points. Furthermore, the *method of constant stimuli* would be preferable. According to this method, only one stimulus, i.e. one particular nsr, is played during a session randomly interspersed with pure noise catch trials. To conclude, despite the limitations of a proper statistical analysis, the ROC graphs indicate that the animal took a conservative (less risky) approach for all the noises.

The actual strength of ROC graphs lies in the measurement of ROC curves. Imagine if we had changed Aurora's *payoff matrix*. Instead of rewarding her with one fish for a hit and a correct rejection, we could have given her three fishes for a hit and only one fish for a correct rejection. Probably Aurora would have started to take a greater risk and more often responded "yes" in cases of uncertainty. This way she would have increased the hit rate but at the expense of the false alarm rate. Her decision bias would have shifted into the liberal zone. Deliberate manipulation of an animal's decision bias during hearing experiments has been done with marine mammals. An animal can be manipulated to favour the "yes" response, i.e. to become more liberal, if catch trials occur less than 50 % of the time [Schusterman *et al.* 1975], if hits are rewarded with more food than correct rejections [Schusterman and Johnson 1975] or if the probability of receiving a reward is

decreased for correct rejections [Schusterman 1976]. If for a fixed nsr, the bias of Aurora was actively changed, the point of her stimulus-response matrix would follow a so-called *ROC curve* in the ROC diagram. The slope of this curve is related to the nsr and the sensitivity of the individual tested [Green and Swets 1966]. ROC curves thus provide the ability to separate an individual's sensitivity from its bias. Unfortunately, training and experiment time at the Vancouver Aquarium was limited; the measurement of ROC curves was beyond the scope of this project.

I would like to give a final note for readers not familiar with psychophysics. When I as a physicist with no training in biology or psychology got the first data from Aurora, I was surprised that her signal detection ability was "smeared out" over many nsr's. I had expected all thresholds to be sharp, i.e. that her percentage of breaks would drop from 100 to 0 in just one step, which in turn would define the critical nsr. Even including some neurophysiology knowledge, the idea that neurons have activation thresholds separating the two possible states "on" and "off", only supported my expecting sharp cut-offs. I was intrigued to learn that stimulation produces a bell-shaped distribution of effects having a mean value (on an appropriate psychological scale) and a variance. This is in accordance with the existence of large numbers of similar receptive and nervous elements. A discussion of various theories on the existence, characteristics and measurability of sensory thresholds can be found in [Green and Swets 1966].

Chapter 4

Acoustic Experiments with Humans

I collected data from three human test persons for two reasons. First, I wanted to compare Aurora's call detection abilities to those of humans, in order to determine if human listening experiments are a good model of beluga masking tests. In general, experiments with animals are very time and cost consuming. If the masking of man-made noise on beluga communication could be simulated by human listeners, results could be achieved very fast. Second, a whale cannot be asked about his impressions from an experiment, but humans can. Not knowing what to expect, I thought that if there were hooks or psychoacoustic traps in my experiment procedure, human listeners might be able to identify and communicate them.

4.1 Data Collection with Human "Guinea-Pigs"

The people sat in a quiet room wearing headphones which were plugged into the audio amplifier. I first adjusted the volume control such that the transmission of a signal with an rms amplitude of 1 was maximally, not painfully, loud. Then I determined the pure call detection thresholds of the test persons. These were of the order rms $= \frac{1}{1024}$. Data were collected only with the second set of mixed signals, in which all the signals had the same rms amplitude. For increasing nsr, the call would become quieter and the noise would become louder. All three paradigms were undertaken and all the data analyzed

together. Test persons were, apart from myself, my husband Andrew, who was quite familiar with the whale experiments and who knew the sounds of the vocalization and noises beforehand, and Kuan-Neng, who had no prior information. Results are plotted in Figs. 4.1, 4.2 and 4.3. I collected data up to higher nsr's than with Aurora.

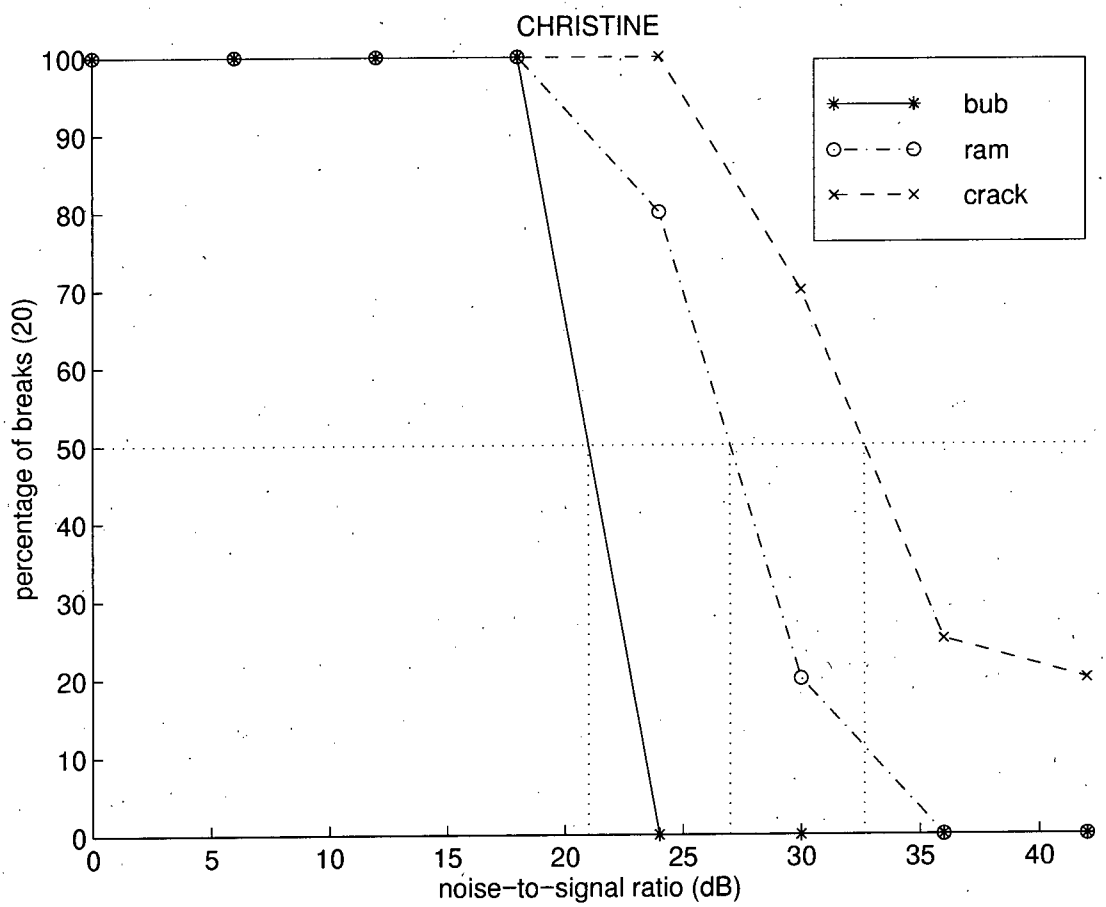


Figure 4.1: Masked hearing thresholds of myself, in bubbler system noise, ramming noise and natural icecracking noise. The same explanations apply as in Fig. 3.15. Defining the hearing threshold at 50 % yields the following critical noise-to-signal ratios: 21.0 dB (linear 11) for bubbler system noise, 27.0 dB (linear 23) for ramming noise and 32.4 dB (linear 44) for natural icecracking noise.

Comparing the plots from Aurora and the three humans, for all four of us, bubbler system noise had worse masking effects than propeller cavitation noise, and natural ice-

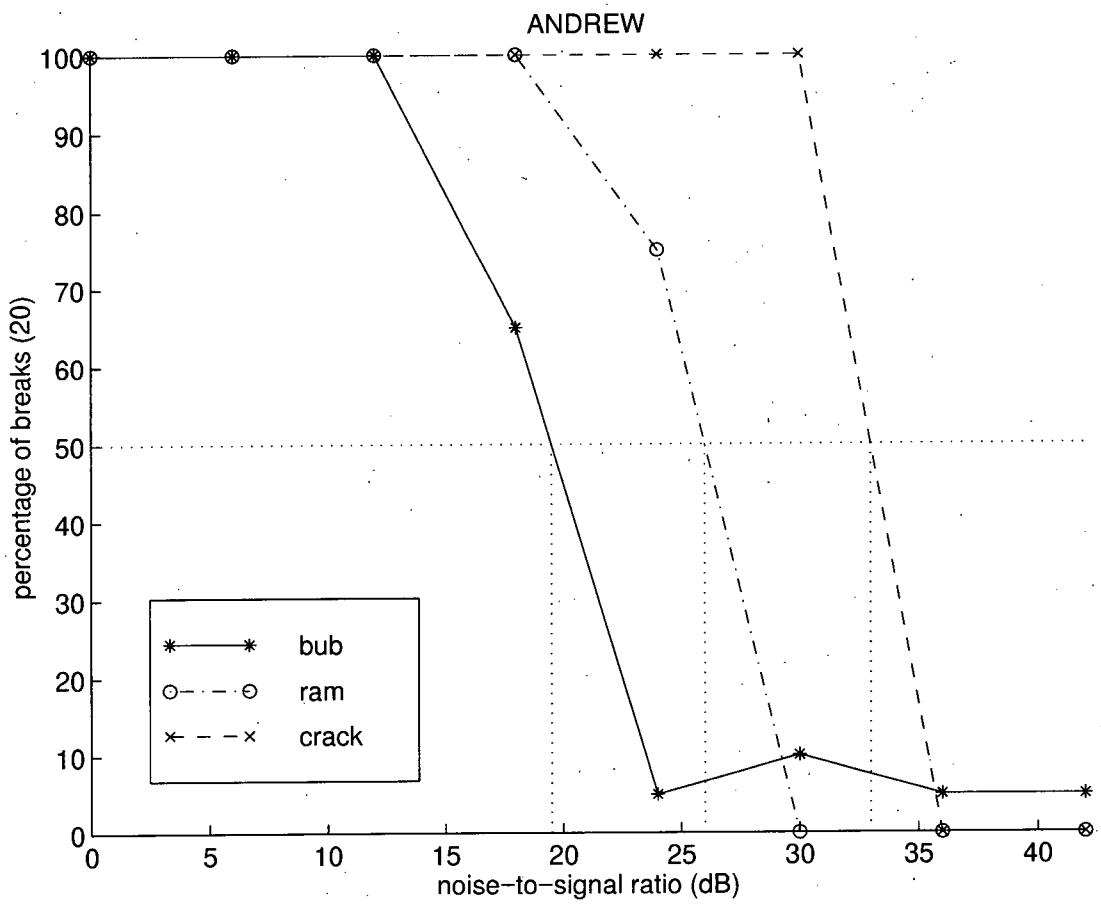


Figure 4.2: Masked hearing thresholds of Andrew, in bubbler system noise, ramming noise and natural icecracking noise. The same explanations apply as in Fig. 3.15. Defining the hearing threshold at 50 % yields the following critical noise-to-signal ratios: 19.8 dB (linear 10) for bubbler system noise, 25.8 dB (linear 20) for ramming noise and 33.0 dB (linear 45) for natural icecracking noise.

cracking noise was the least harmful. The thresholds from all three humans were higher than Aurora's thresholds, i.e. the humans were better than the whale at detecting the beluga call in noise. This leaves room for discussion and speculation. One possible reason is that the beluga call falls into the frequency range of maximum sensitivity for humans, but not for the beluga whale. The maximum sensitivity for beluga whales lies between 20 to 80 kHz, frequencies they use for echolocation (Fig. 1.1). There is also the possibility of

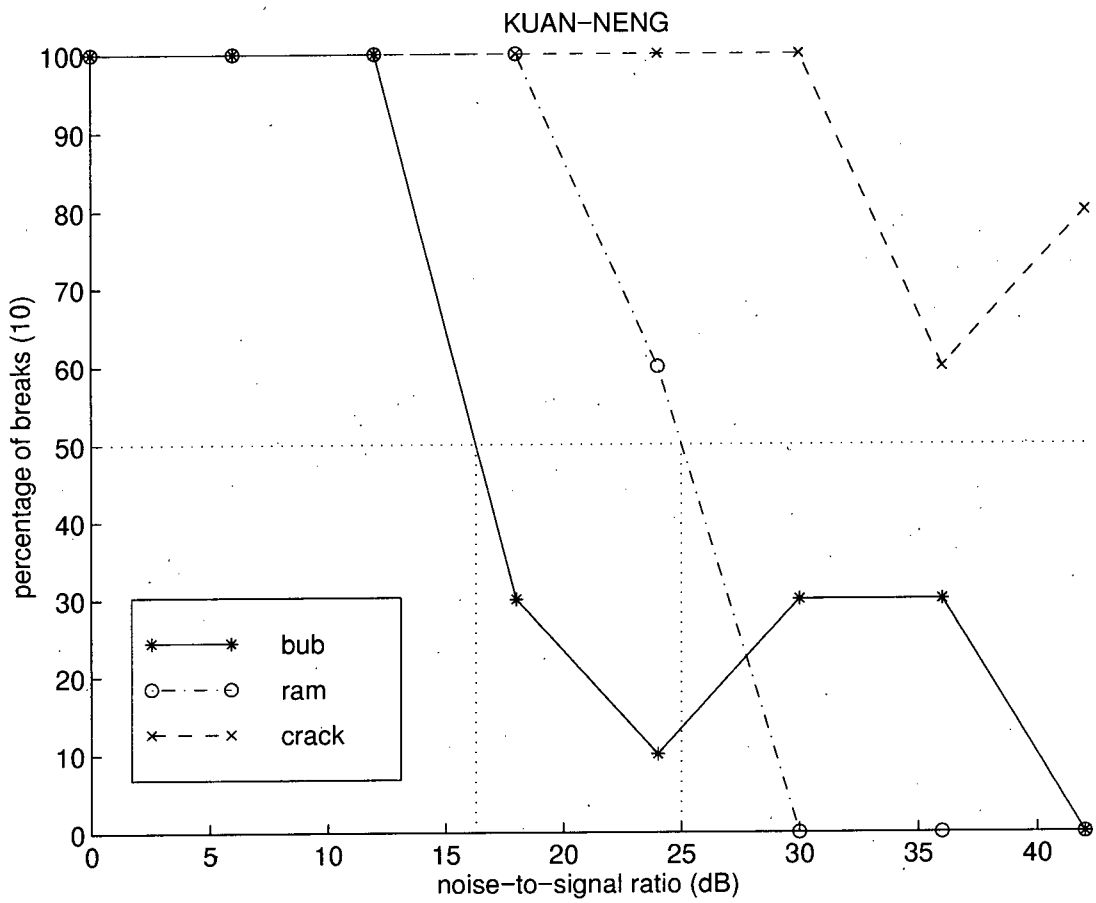


Figure 4.3: Masked hearing thresholds of Kuan-Neng, in bubbler system noise, ramming noise and natural icecracking noise. The same explanations apply as in Fig. 3.15. The critical noise-to-signal ratios are: 16.2 dB (linear 7) for bubbler system noise and 25.2 dB (linear 18) for ramming noise. No 50 % threshold can be defined for icecracking noise.

lack of concentration of the animal, which might explain why Aurora's response curves are much flatter (less sharp) than the humans'. Furthermore, it was shown in Section 3.3.5, that Aurora had a conservative decision bias. She only responded when she was relatively sure that a signal was played. This attitude of less risk leads to apparently lower thresholds. A last difference one notices between the human and the whale responses is that for Aurora, the bubbler and ramming noise curves are close together while the icecracking

noise is set aside at high nsr's. For us humans, the thresholds were separated at almost equal distances.

4.2 Ghost Detection in Noise

There are three "puzzling" features in the human data. Andrew's response to bubbler system noise seems to converge to a chance of 5 % to discriminate call from noise no matter how large the noise-to-signal ratio. Kuan-Neng's responses to bubbler system noise and natural icecracking noise show ups and downs. Fortunately, one can ask humans why they reacted the way they did.

The main problem we (the three test persons) identified was that after listening to the signals only twice or thrice, one develops a very accurate feeling for them. Their frequency distribution as well as their time pattern are stored in one's mind. As all the transmitted signals were of 2 s duration, the call always happened at the same time in the noise. One knows exactly where the pulses of the vocalization are and what they sound like or what they WOULD sound like. Thus, one tends to "identify" the call in mixtures with very high nsr's and even in pure noises when there is no call content at all. It happened to each of us that when the signals were played in a sequence with increasing noise content and decreasing call content, that we thought we could hear the call right through the end of the sequence, even in the last signal which was pure noise.

This problem can quickly be identified by human listeners if there is feedback between

the experiment conductor and the test person. Andrew (Fig. 4.2) was the first subject to be tested; the first noise tested was bubbler system noise. After he had indicated that he had heard the call in a couple of mixtures with very high noise content and even in many pure noise sounds, we talked about his reaction and identified the problem, that the mind is led to think it hears a call although it is not there. When Andrew's data for the ramming noise and the icecracking noise were collected, he was more "careful", trying to respond only when he was "more certain" that a call was played. His attitude probably shifted to more conservative.

Unfortunately, at the time I did not realize the importance of catch trials. I did not conduct equal numbers of catch trials for all the mixed signals. Pure noise sounds were played about a third of the time for high nsr mixtures, about once in ten transmissions for intermediate nsr's and never for low nsr mixtures. The only reason why I conducted catch trials with Aurora was that I wanted to verify if she understood and remembered what she was supposed to do, an issue which does not arise during human experiments.

I was tested after Andrew and therefore knew about the "ghost detection problem". I set myself an artificial threshold higher than the "true" threshold and only responded when I was sure the call was louder than the set threshold. My data (Fig. 4.1) thus do not show a long tail like Andrew's bubbler curve.

When Kuan-Neng's data were collected, he and I sat back to back. There was neither visual nor oral communication, hence no feedback. Kuan-Neng's data (Fig. 4.3) exhibit a "chaotic" up and down fluctuation for high nsr. I assume that this feature is only

apparent in mixtures with an nsr below the "true" threshold. It shall be mentioned that due to time constraints, Kuan-Neng's data were averaged only over 10 trials compared to 20 with Andrew, myself and Aurora.

I like to call this phenomenon, that we tend to think we hear a call in high nsr mixtures and even pure noise, "ghost detection" and set it aside from the possibility of having a decision bias leading to a liberal subject. If humans do not know about this "psychoacoustic trap" and begin the experiment under the pretext of being unbiased, neither too conservative nor too liberal, they cannot avoid falling into this trap. Even if this pitfall begins to become clear when subjects begin to doubt that the experiment conductor would play so many mixtures with detectable call content, even if they start to shift their attitude towards conservative, they can still be led into the trap when data are collected in a standard titration. If the series of transmitted signals starts at a low nsr and gradually increases the noise content, it becomes extremely "difficult" to stop responding at some stage, because our mind keeps detecting the call and the imagined loudness does not decrease noticeably. Basing my conclusions mainly on Kuan-Neng, who was the only subject who did not receive feedback during the experiment, I found that if catch trials are played at random with mixtures of high nsr, the corresponding points in the ROC curve fall onto the diagonal chance line. His response to mixed signals and pure noise could not be told apart and was statistically the same. If the subject runs into the trap and keeps imagining the call, the ROC point lies at (1,1) in the upper right corner. If, however, the subject becomes suspicious, thinking that "certainly by now, the experiment

conductor must play a high nsr mixture or even a pure noise in which I'm not supposed to hear a call", the point in the ROC diagram will slide down the diagonal towards (0.5,0.5) or even lower. If on the other hand, the ghost detection problem is clearly identified, as was the case with Andrew and myself, subjects become extremely conservative.

Having nailed down this psychoacoustic pitfall, I went back to the notes I took during the training of Aurora. I found that she was led into exactly the same trap as the humans. Initially, when mixed signals were played in a sequence of increasing nsr, Aurora did not stop breaking on any of the mixtures and kept breaking on repetitions of pure noise sounds. Furthermore, when mixtures were played at random nsr, her data exhibited a chaotic tail similar to Kuan-Neng's. Having been ignorant about problems with the psychoacoustic procedure at the time, the whale trainers and I interpreted this as Aurora not yet being trained properly, i.e. not yet understanding what she was supposed to do. We strongly reinforced her stationing on pure noise sounds and she received negative reinforcement every time she ran into the trap. This way we unconsciously conditioned a very conservative animal. Data collection only started once Aurora had stopped breaking during a titration.

The first idea I had to avoid this ghost detection was to train the whale to recognize more than one beluga vocalization. I extracted two more vocalizations from the recordings of wild belugas, with the aim of selecting calls as different as possible from each other. Spectrograms of the other two calls are shown in Figs. 4.4 and 4.5. Again, as the recordings have not yet been calibrated, I normalized the amplitudes to a source level of 160 dB.

The first call was a whistle, the second call a pulsed vocalization with a much higher repetition frequency than the previous call. I thought that if we trained the whale to break on all three vocalizations, I could transmit mixtures of any of the three calls with any noise in random order. The whale would not know which call to expect, hence its brain would not know which call to look for and "wrongly identify". In order not to waste time on long whale training, I first tried the idea on my husband and myself. We "trained" ourselves to identify all three beluga vocalizations. I mixed all three calls only with bubbler system noise to start with and I changed the experiment such that we were required not only to indicate whether or not we thought we heard a beluga call but also to say which one. We were both totally amazed that for high nsr's we actually thought we heard all three calls at the same time! The same problem occurred when pure noise was played. The sensation was that all three calls were played at the same time and mixed with a bit of bubbler noise. We would decide for one of the three calls or a pure noise sound mainly by chance. All wrong answers, such as an apparent identification of call 1 in a call 2 / bubbler mixture, were regarded as not heard at all. Thus we actually got rid of most of the long tails as in Andrew's plot of the bubbler system mixtures (Fig. 4.2). But thresholds were still chaotic when we "accidentally guessed" the right call. It was time to drastically revise the experimental approach.

4.3 Need for Experiment Modification

The main problem with the experiment so far was that all data files had exactly the same length and that the call appeared at exactly the same time in all the mixtures. The reason why I initially chose this way of mixing the data was, that I expected the masked hearing threshold to depend on when the call happened in the noise. In particular for pulsed call and noise this is very obvious. The call will easily be detected if it is just out of phase with the noise. On the other hand, if call and noise pulses coincide, masking will be maximal.

If noise was played continuously and the call was injected at random times, the animal or person under study would not know when to expect the call. The experiment would be similar to the pure call detection threshold measurements which were carried out with the whale and the humans prior to the masking experiments. These did not exhibit any psychoacoustic problems. The only difference would be that the pool or room was quiet at the time, whereas now noise would play continuously.

Unfortunately, this experiment idea could not be pursued right away, because Aurora gave birth to her first calf, Qila, at the end of July 1995. Experiments could only be resumed a year later.

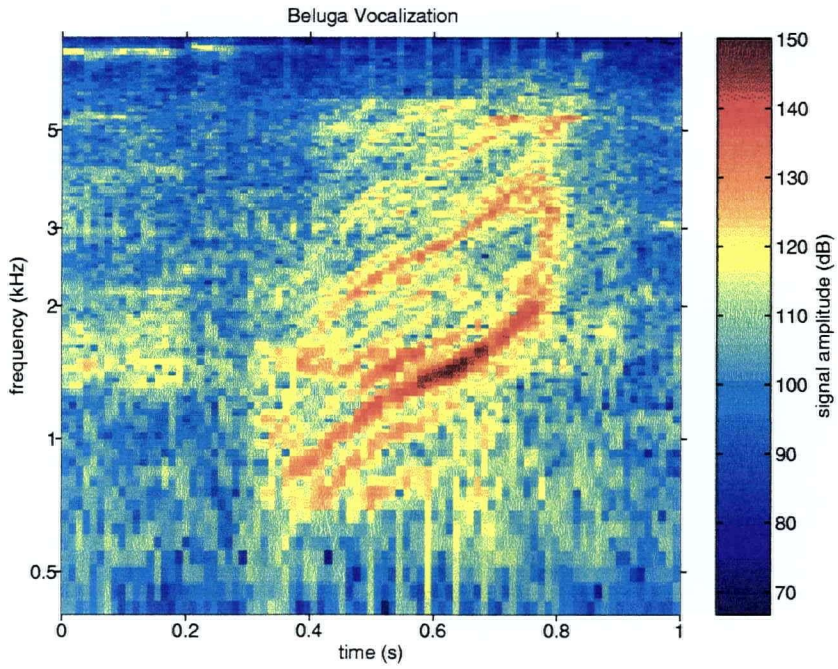


Figure 4.4: Power Density Spectrogram of the 2nd Beluga Vocalization in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

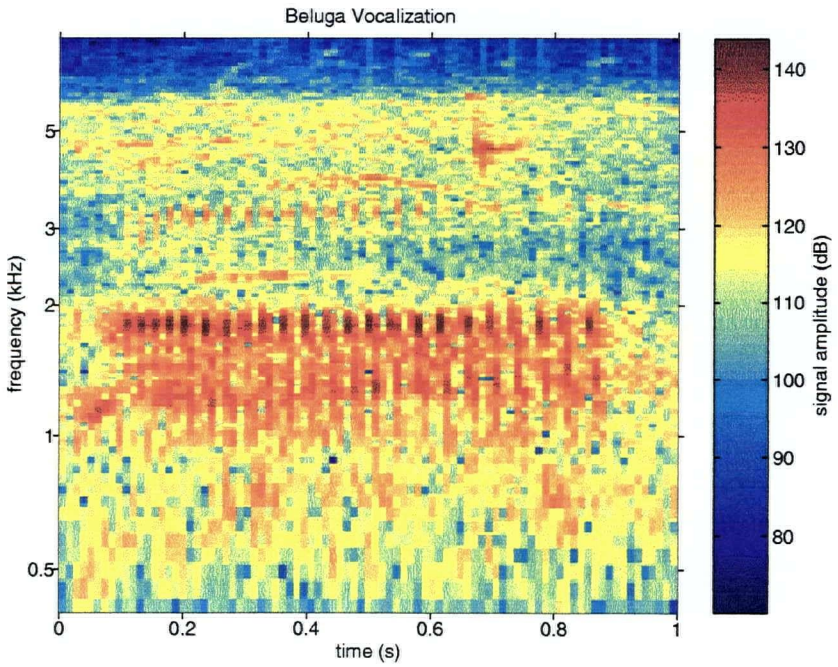


Figure 4.5: Power Density Spectrogram of the 3rd Beluga Vocalization in dB re 1 $\frac{\mu\text{Pa}^2}{\text{Hz}}$ @ 1 m.

Chapter 5

Various Detectors for Animal Calls in Noise

While the husbandry training with Aurora and her new-born calf was going on at the Vancouver Aquarium, I focussed on software models for the whale's call detection process.

5.1 Matched Filtering

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

$$y[t] = \sum_{k=-\infty}^{\infty} h[k] \cdot x[t-k] \quad . \tag{5.1}$$

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 [Karl 1989] that the impulse response of such a filter has to be as long as the signal to be detected. Furthermore, the filter coefficients have to be equal to the product of the inverted auto-correlation matrix of the noise and the time-reversed pure signal. For white noise, the auto-correlation matrix turns into the identity matrix, and the filter response equals the time-reversed 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 equal to the cross-correlation of the signal without time reversal. Therefore, matched filtering is equal to cross-correlating the input time series with the desired signal.

As an approach to matched filtering with non-white noises, I computed the cross-correlation of the beluga vocalization $s[t]$ with each of the mixed signals $x[t]$. It will be seen in Section 6.6 that the three noises are in fact very white in the sense that their auto-correlation matrices are zero except for zero lag, i.e. the matrices are diagonal. Therefore, cross-correlation is a good approximation to matched filtering for the sounds studied. Only the zero-lag cross-correlation coefficient is of importance, because the signals were mixed such that the call always happened at the beginning of the noise. The zero-lag value of the discrete correlation coefficient as a function of the noise-to-signal ratio α is:

$$R_0(\alpha) = \frac{\sum_{t=1}^{T} x[t]s[t]}{\sqrt{\sum_{t=1}^{T} x^2[t] \cdot \sum_{t=1}^{T} s^2[t]}} \quad . \tag{5.2}$$

It is plotted in Fig. 5.1. It can be seen that matched filtering fails to exhibit different degrees of masking for the three noises, because all the three curves basically lie on top of each other. Even if one was to examine the slight splitting for high noise-to-signal ratios,

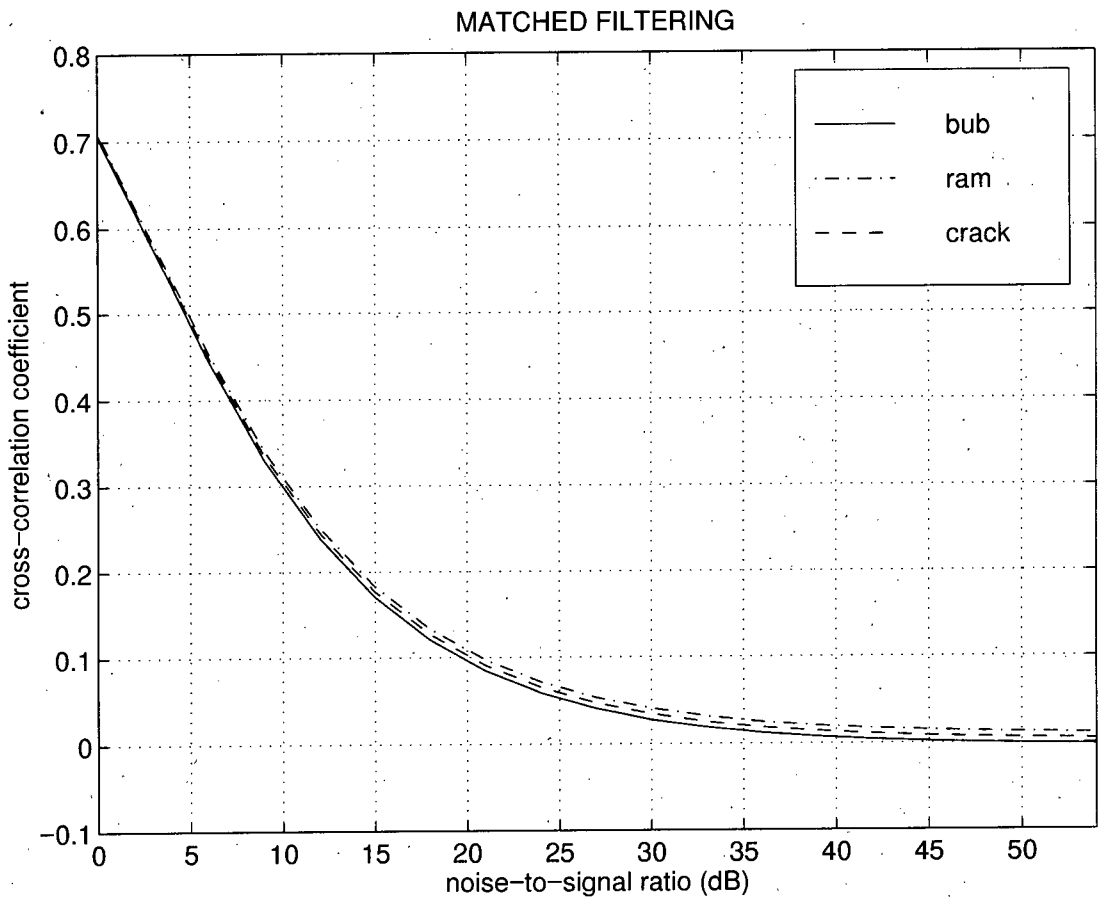


Figure 5.1: Matched filtering of the beluga vocalization buried in bubbler system noise, ramming noise and natural icecracking noise. The zero-lag cross-correlation coefficient of the corresponding time series was computed for noise-to-signal ratios (nsr) increasing logarithmically from 0 to 54 dB. The degree of masking as a function of nsr basically is the same for all the three noises.

the correlation coefficient of ramming noise would be slightly larger than the one for icecracking noise than the one for bubbler system noise. This would mean that ramming noise was least masking, followed by bubbler system noise, then icecracking noise. The behavioral experiments with the beluga whale indicated a completely different order of noises.

The behaviour of the cross-correlation coefficient as a function of α can be understood

by looking at the boundaries $\alpha = 0$ and $\alpha \to \infty$. With $x = s + \alpha \cdot n$,

$$\begin{aligned} R_0(\alpha) &= \frac{\sum xs}{\sqrt{\sum x^2 \cdot \sum s^2}} = \frac{\sum(s^2 + \alpha sn)}{\sqrt{\sum(s^2 + \alpha^2 n^2 + 2\alpha sn) \cdot \sum s^2}} \\ &= \frac{\sum s^2 + \alpha \sum sn}{\sqrt{(\sum s^2)^2 + \alpha^2 \sum n^2 \cdot \sum s^2 + 2\alpha \sum s^2 \cdot \sum sn}} \quad . \end{aligned} \tag{5.3}$$

For $\alpha = 0$, the cross-correlation coefficient is equal to 1. For $\alpha \to \infty$, divide the numerator and denominator by α and let all the terms with α in the denominator go towards 0.

$$R_0(\alpha) = \frac{\frac{1}{\alpha}\sum s^2 + \sum sn}{\sqrt{\frac{1}{\alpha^2}(\sum s^2)^2 + \sum n^2 \cdot \sum s^2 + 2\frac{1}{\alpha}\sum s^2 \cdot \sum sn}} \quad , \tag{5.4}$$

$$\lim_{\alpha \to \infty} R_0(\alpha) = \frac{\sum sn}{\sqrt{\sum s^2 \cdot \sum n^2}} \quad . \tag{5.5}$$

Independent of the type of noise, all the plots of $R_0(\alpha)$ will always start at 1 and converge to the product of the pure signal with the pure noise. Thinking of the time series as vectors, $R_0(\alpha)$ converges to the cosine of the angle between the signal and the noise. Therefore, the more "similar" the signal and the noise are, the smaller the angle, the greater the cosine. This is the only reason for the slight splitting of the correlation curves for large nsr's.

Altogether, matched filtering is not an appropriate tool for modeling the whale's call detection abilities in noise as a function of nsr. Considering that the cochlea of mammalian ears in a sense performs a Fourier transform of received acoustic signals, rather than a pure time series, our brains receive a time series of Fourier spectra. Therefore, if the mammalian brain detects a signal in noise by cross-correlation, a cross-correlation of 3-dimensional spectrograms is more likely than matched filtering in the time domain.

5.2 Spectrogram Cross-Correlation

For each of the mixed signals $x[t]$, a spectrogram was computed by calculating the Fourier transforms in chunks of 256 data points, using Hamming windows with 50 % overlap. I took the magnitudes of the complex Fourier components and kept the amplitude linear. Negative frequencies were discarded. The spectrogram matrices of size 256x351 were then reshaped into 89856 element-long row vectors. This way, the same algorithm could be used as for the time series cross-correlation. Results are plotted in Fig. 5.2. If a detection threshold was defined, e.g. at a cross-correlation coefficient of 0.5, above which the signal can be detected in the noise, the corresponding noise-to-signal ratio would be lowest for the icecracking noise, followed by the ramming noise, and greatest for the bubbler system noise. Therefore, according to spectrogram cross-correlation, the degree of masking should be highest for the icecracking noise and lowest for the bubbler system noise which is exactly opposite to the whale's response.

For $\alpha = 0$, the coefficient for all the three noises is 1. For $\alpha \to \infty$, the curves converge to the cosine of the angle between the spectrogram vector of the pure call and the spectrogram vector of the pure noise. Therefore, the order of the noises in Fig. 5.2 is entirely determined by how much the spectrogram of the call and the noise are alike. This fact exhibits an important ambiguity. In signal processing, the higher the correlation coefficient, the more signal was detected in the noise. Looking at the correlation coefficients for a fixed noise-to-signal ratio, the value is always highest for bubbler noise, followed by ramming noise then icecracking noise. This means that bubbler noise would be the

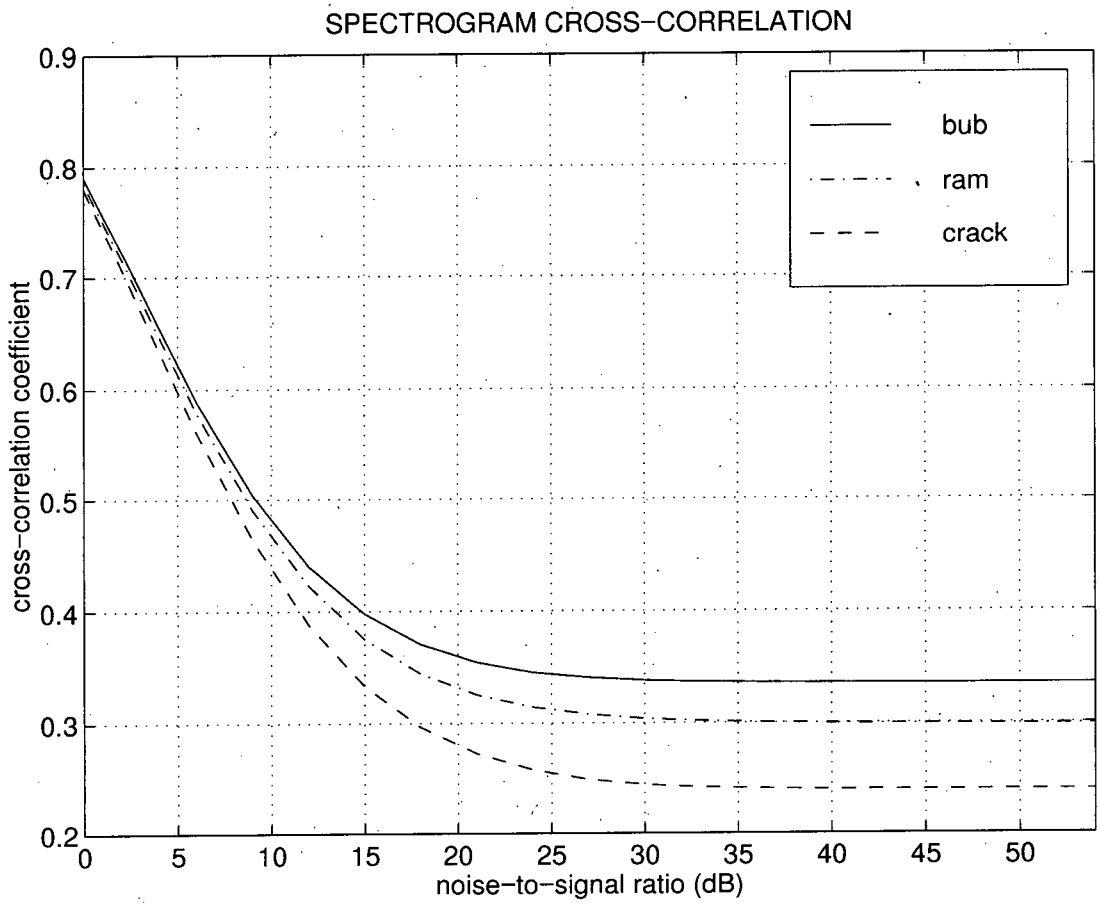


Figure 5.2: Spectrogram cross-correlation of the beluga vocalization and mixtures with bubbler system noise, ramming noise and natural icecracking noise. Fixing an arbitrary correlation coefficient at 0.5, above which the call can be detected in the noise, yields thresholds for the three noises in reversed order compared to the whale's response. According to spectrogram cross-correlation, bubbler system noise should be least masking, followed by ramming noise, then icecracking noise.

least masking and icecracking noise the worst masking. In biology, however, we know from psychoacoustic experiments that the degree of masking is greater, the more signal and noise are alike (Section 1.3). The biological argument is thus exactly opposite to the signal processing argument. It is interesting to note that by looking simply at the values to which the three correlation plots converge for large α and by using the biological argu-

ment, the order of the three noises is the same as for Aurora. Furthermore, the relative distance between the three noises is similar to Aurora's data. The bubbler noise exhibits the strongest degree of masking and is closer to the ramming noise than the ramming noise is to the icecracking noise.

5.3 Critical Band Cross-Correlation

In the previous section, spectrograms with linear frequency distribution and linear amplitude were correlated. If cross-correlation takes place in the mammalian brain, the spectrograms created by our ears will be logarithmic in frequency and amplitude. The spectrograms of the previous section were averaged into 12th octave bands resembling the beluga auditory filter. Amplitudes were converted to decibels and adjusted relative to the beluga audiogram (Fig. 3.12), i.e. high frequencies were amplified. The zero-lag cross-correlation coefficients behave as plotted in Fig. 5.3. The most striking feature is that the curves cross. For nsr's smaller than 8 dB, the coefficient for icecracking noise is greater than the one for ramming noise, followed by the one for bubbler noise. For larger nsr's, the order is exactly opposite. The response curves of Aurora and the humans did not exhibit this behaviour. Therefore, I conclude that cross-correlation in this form does not occur in the mammalian brain.

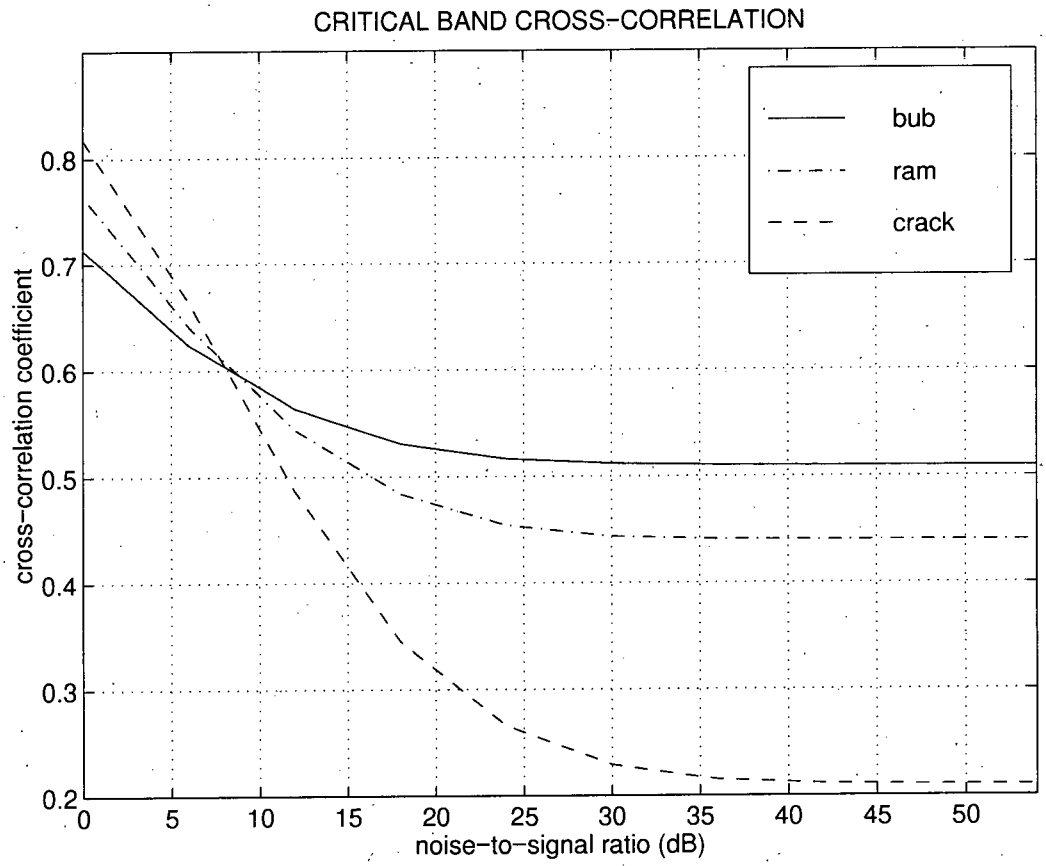


Figure 5.3: Critical band cross-correlation of the beluga vocalization and mixtures with bubbler system noise, ramming noise and natural icecracking noise. The noises change order at an nsr of 8 dB.

5.4 Visual Spectrogram Discrimination

A method which has the potential of visually illustrating the call detection process in noise is visual spectrogram discrimination. The spectrograms of successive mixtures of the beluga vocalization with the three noises in increasing noise-to-signal ratio were plotted. Thresholds at which the characteristic features of the call can no longer be found in the mixture were determined by eye. Examining the plots for bubbler noise in Fig. 5.4, the

call can easily be seen in the mixture up to a noise-to-signal ratio of 6 dB. Knowing where to look for the features, one can also identify the 2 kHz component of the first two pulses and the 3 kHz component of the fourth and fifth pulse in the 12 dB mixture. The threshold was therefore estimated to lie slightly above a noise-to-signal ratio of 12 dB. For ramming noise (Fig. 5.5), the call can still just be seen in the 18 dB mixture. For natural icecracking noise (Fig. 5.6), the call can be detected up to a noise-to-signal ratio of 24 dB. It is no longer obvious in the 30 dB mixture. The threshold was estimated to lie around 27 dB. Comparing with Aurora's thresholds which were at an nsr of 15.4 dB for bubbler noise, 18.0 dB for ramming noise and 29.0 dB for icecracking noise, visual spectrogram discrimination "models" her detection abilities very well. Noises are in the same order; relative distances among individual thresholds are similar; absolute thresholds are slightly shifted to lower nsr's.

Visual spectrogram discrimination is a subjective method as a person has to examine the plots by eye. I had 10 people estimate thresholds from Figs. 5.4, 5.5 and 5.6. Results were consistent. However, changing the nsr in steps of 6 dB is fairly rough. If intermediate nsr's were plotted, thresholds could be expected to change from individual to individual. Furthermore, discrimination ultimately depends on the gray- or colourscale used.

Although visual spectrogram discrimination yields thresholds close to those of the whale, it is a very time consuming method, because all the spectrograms have to be plotted and examined by one or more persons. An automatic technique which finds call features in spectrograms is preferable and can be realized as an artificial neural network.

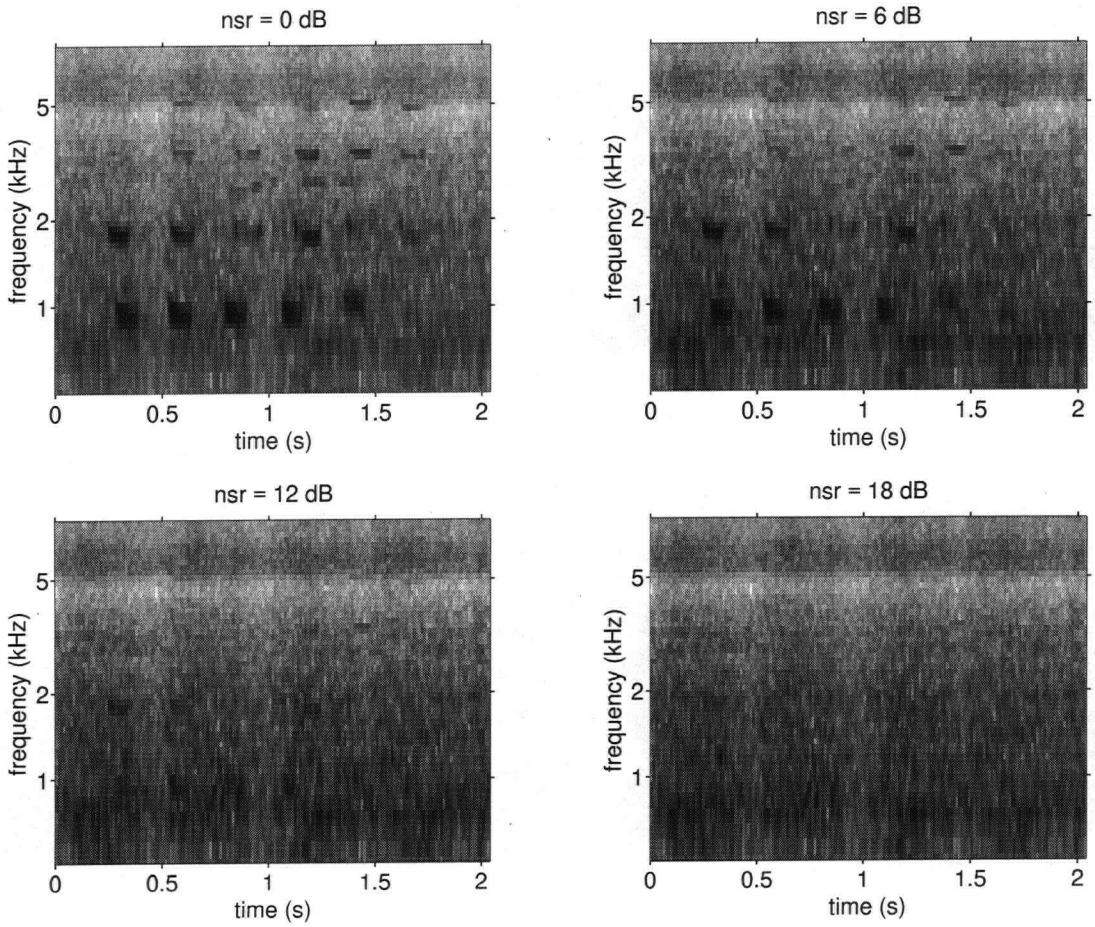


Figure 5.4: Visual spectrogram discrimination for bubbler system noise. The characteristic features of the call can just be seen up to a noise-to-signal ratio of 12 dB.

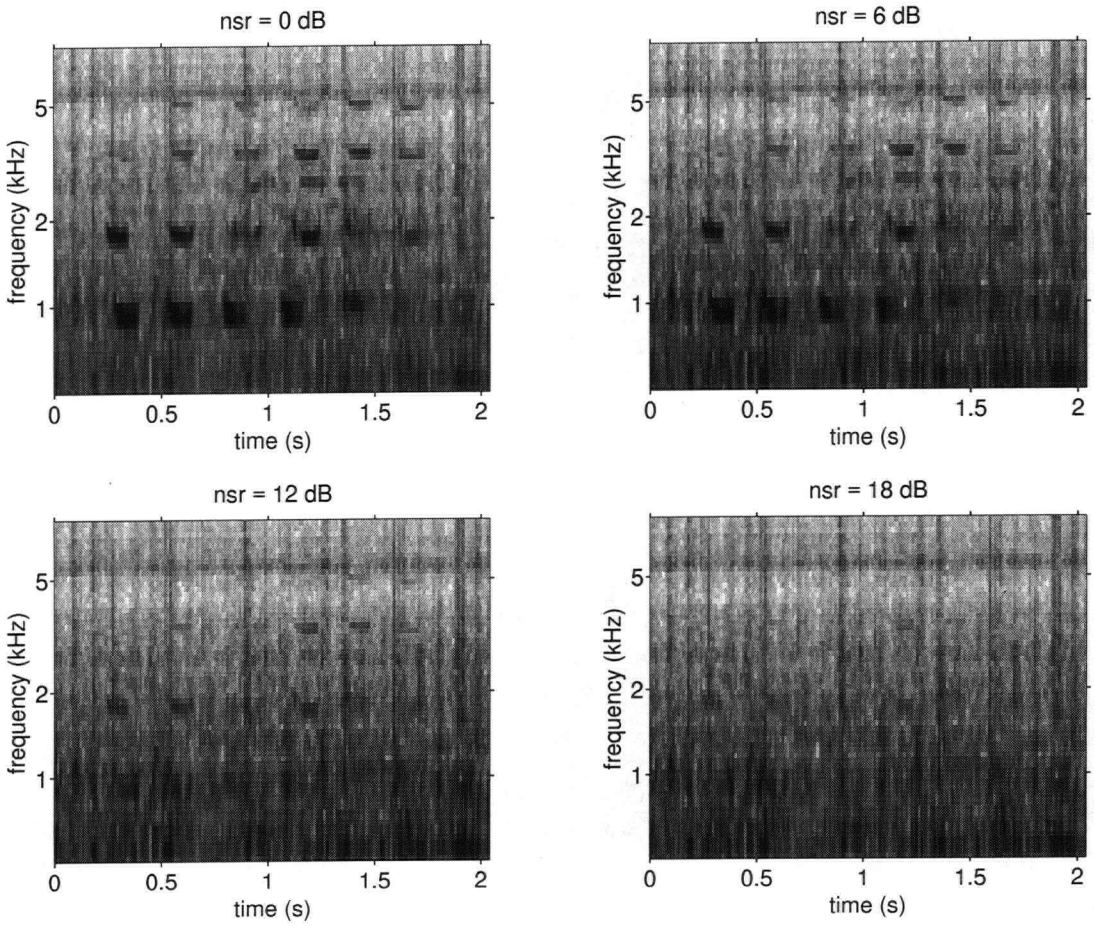


Figure 5.5: Visual spectrogram discrimination for ramming noise. Here, the call can just be detected in the 18 dB mixture.

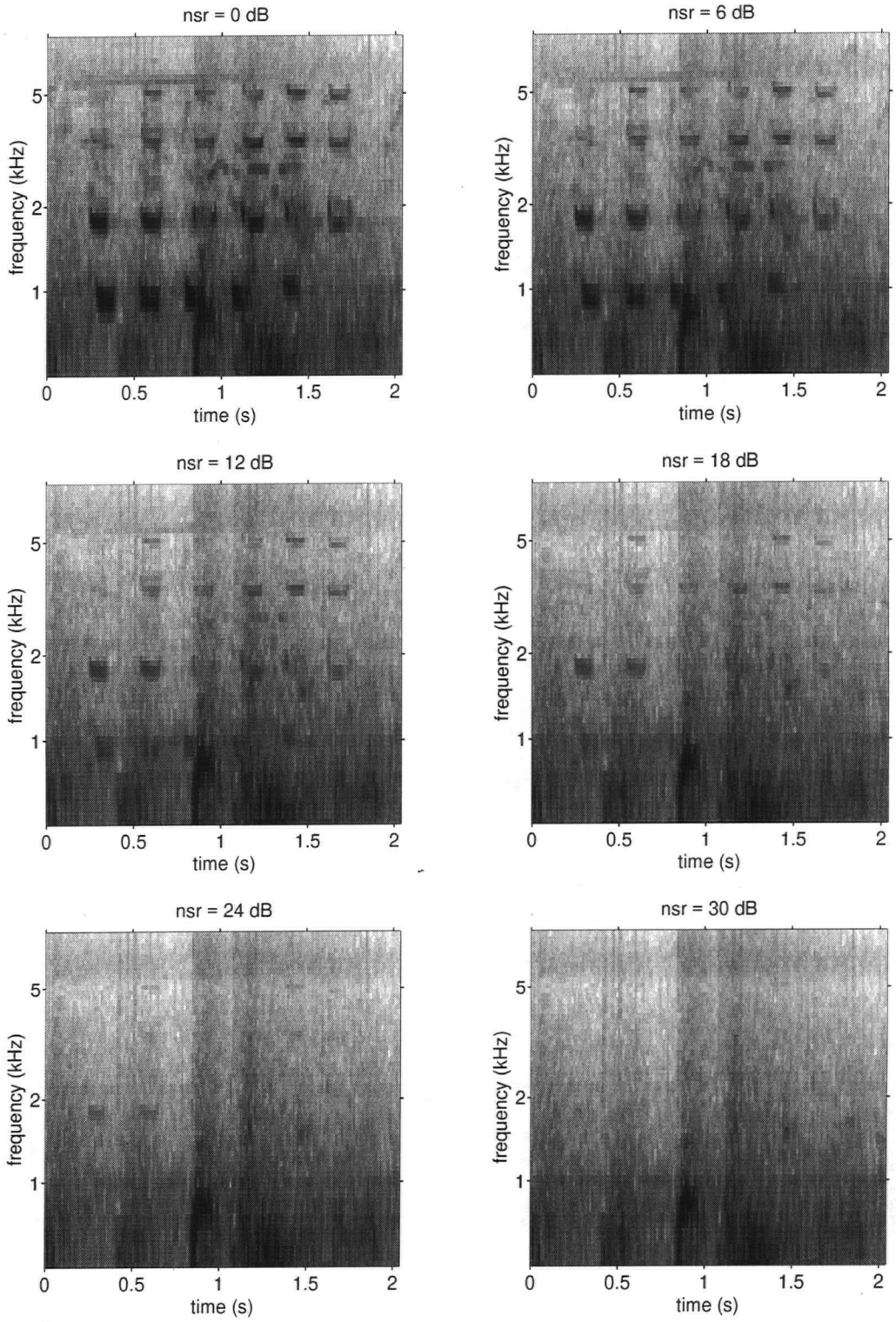


Figure 5.6: Visual spectrogram discrimination for natural icecracking noise. The last plot in which the call can be detected corresponds to an nsr of 24 dB. The threshold lies between 24 dB and 30 dB.

Chapter 6

Artificial Neural Network Models

I can't give you a brain, but I can give you a diploma.

–The Wizard of Oz to the Scarecrow

6.1 The Biological Neural System

In order to understand the idea behind artificial neural networks, it is necessary to look at some basic anatomical facts about the biological analogue.

The purpose of the nervous system is to receive and process information from the external and internal environments and to regulate the body's somatic and visceral motor functions. Throughout the animal kingdom, nervous systems greatly vary in complexity. For example, the simple nervous system of a sea anemone does not do much more than detect food or danger and cause retraction of tentacles and constriction of the body,

whereas in humans, the nervous system enables such extraordinary functions as storing and retrieving bits of information as memory, initiating thought processes and regulating emotions and behaviour.

In simple animals like the sea anemone or other lower invertebrates, the nervous system is a netlike arrangement of nerve cells which are uniformly distributed through the body. For higher animals like worms or jellyfish, the information processing parts of the nervous system are increasingly centralized in aggregates known as ganglia. Complex invertebrates have a distinct head that bears special sensory organs and larger ganglia than those in more posterior parts of the body. These creatures may be said to feature a brain. In even more complex animals, the vertebrates, the nervous system is anatomically divided into the central nervous system CNS and the peripheral nervous system PNS. The former comprises the brain and spinal cord. The latter consists of nerves transmitting information from the sense organs and sensory receptors to the CNS and nerves transmitting information from the CNS to the muscles and glands of the body.

The principal cellular elements of nervous systems are identical in all animals as different as squids and humans. They can be divided into two groups: glial cells and neurons. The former provide nutrition or support in form of a scaffolding to the latter. The number of nerve cells varies greatly from species to species. Humans have about 10^{11} neurons and perhaps 10 times as many glial cells in their brain [Nolte 1981, p. 1]. It is interesting to note that most neurons are present at birth or shortly thereafter. As the brain continues to grow during the postnatal period, the number and complexity of interneuronal

connections increase.

Though they vary greatly in size and shape, neurons have a common structure (Fig. 6.1). They consist of the inner nucleus and outer cell body. A treelike net of nerve fibres, called dendrites, surrounds each neuron. Through these dendrites, the neurons receive signals from other neurons or sensory cells. Furthermore, most neurons have one long output line, the so-called axon, through which they send signals to other neurons. The axon splits at its end into single nerve fibres, the axon terminals, that connect to dendrites of other neurons or muscle cells. The connections are formed by synapses.

Neurons exchange information by sending voltage pulses along the nerve fibres. The inside of a neuron usually maintains a negative voltage of -70 mV compared to its surrounding. This resting potential is achieved by complex biochemical processes at the cell membrane. The main features are so-called ion pumps that constantly move positive sodium ions (Na^+) from the inside of the cell to its outside leaving a negative voltage behind. A neuron permanently sums up the input signals of its dendrites. If at a certain time, the integrated voltage exceeds a threshold of -60 mV, ion channels in the cell membrane open. The cell membrane becomes permeable for the ejected sodium ions. They fall back into the cell body hence further decreasing the potential difference between the cell's in- and outside. In fact, the inside temporarily becomes more positive than the outside. An action potential of $+40$ mV may be created across the membrane (Fig. 6.2). Within one or two milliseconds, the ion pumps have reestablished the resting potential of -70 mV. The short voltage spike on the cell membrane gives rise to electrical currents flowing to

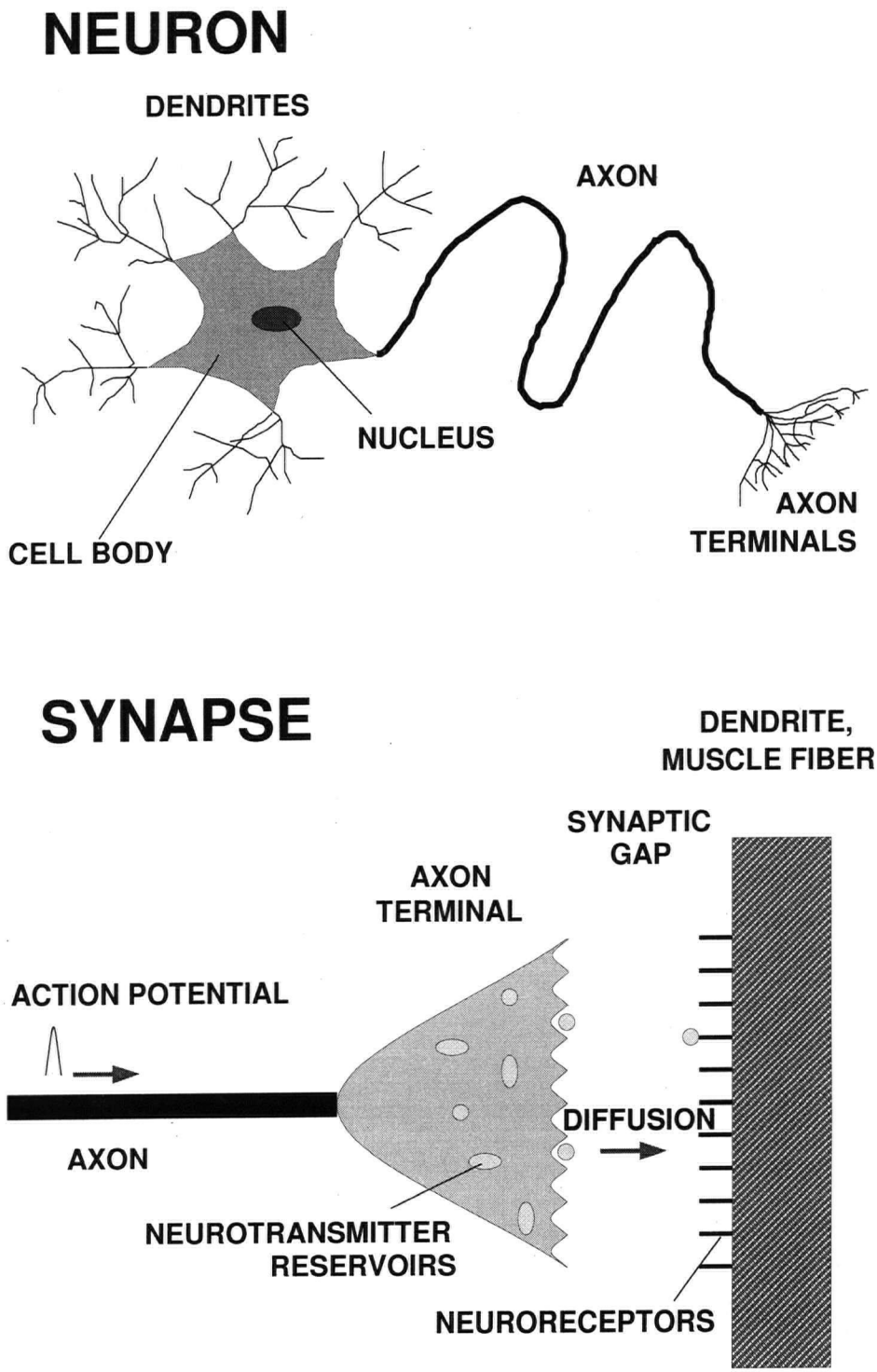


Figure 6.1: Sketch of a Neuron and a Synapse.

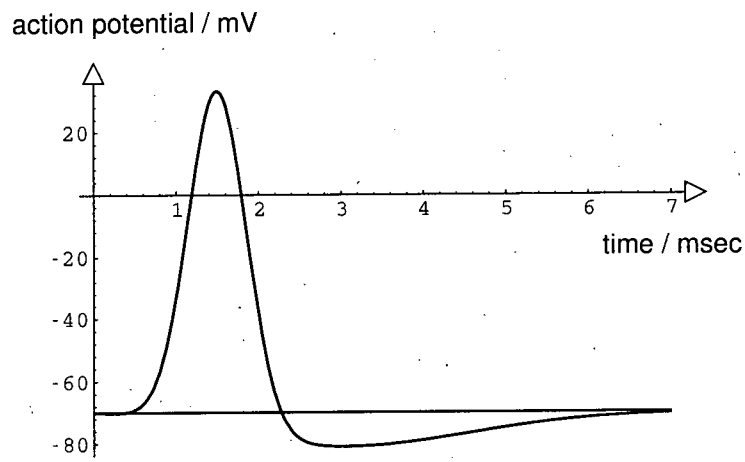


Figure 6.2: Action Potential.

adjacent areas of the axon membrane. Like the cell body, the axon is equipped with ion pumps and ion channels. The axon membrane becomes locally permeable for expelled sodium ions. They stream back into the axon hence creating the same action potential as the previous cell membrane. Again electrical currents are generated and travel further down the axon where they open ion channels. In this way, just like the domino effect, the action potential travels along the axon and continuously regenerates itself.

At the end of the axon, the electrical signal is stopped by the synaptic gap between the axon terminals and the dendrites of the next neurons or cells of muscle fibers (Fig. 6.1). Information can only cross the gap by the usage of chemical messenger molecules called neurotransmitters. When the action potential arrives at the presynaptic side, these neurotransmitters are released into the synaptic gap and drift to the postsynaptic side. Here they bind to neuroreceptors. In a biochemical reaction, a voltage pulse is generated that causes sodium channels to open. Hence a new action potential is created and travels to

the next neurons or muscle cells. Along dendrites, the action potential does not regenerate itself. It constantly decreases. Therefore, synapses close to or even directly at the cell body carry more weight in the input signal integration of neurons.

The neurotransmitters on the receptor side are usually destroyed by enzymes such that the synaptic gap is cleared for transmission of the next signal. The purpose of the synaptic gap is to prevent a short-circuit among the neurons and to determine the direction of signal flow. As in an electrical rectifier, signals can only travel from one side to the other, not backwards. Furthermore, by using different neurotransmitters and given the fact that different synapses have different neuroreceptors, synapses may either amplify or dampen the transmitted signals. One talks of differing synaptic strengths or weights.

6.2 The Artificial Neural System

An artificial neural network is a very simple model of the biological system. In Fig. 6.3, the information received by a neuron through its many dendrites is represented by an R-element input vector $\vec{p}$. Each of these inputs is multiplied by a corresponding synaptic weight from the weight vector $\vec{w}$. The neuron sums up all these weighted inputs and reacts according to its activation function f. Four commonly used activation or transfer functions are plotted in Fig. 6.5. One also accounts for an activation threshold or bias b which is simply added to the sum of weighted inputs. The neuron's output is a scalar a. Neural networks generally consist of more than one single neuron. Fig. 6.4 displays a

complex neural network, in this case a fully-connected two-layer feedforward network. It has an R-element input vector $\vec{p}$ that connects to each of the S1 neurons in the so-called hidden layer. The weight vector becomes a weight matrix $\mathcal{W}$ where the subscript denotes the layer, the first number counts the neurons in that layer and the second number counts the input elements. The outputs of these first- or hidden-layer neurons serve as inputs to the neurons of the second layer. Note that a neuron in this picture is represented by one single box which symbolizes the summation box plus the activation function box plus the threshold from Fig. 6.3. The dimension of the second weight matrix is S2xS1, with S2 being the number of neurons in the second layer and S1 being the number of neurons in the first layer which is equal to the number of outputs of the first layer and equal to the number of inputs to the second layer. The output of the entire neural network is no longer a scalar a but a vector $\vec{a}$ with as many elements as there are neurons in the last layer, also called the output-layer.

A neural network like the brain is a complex, massively parallel information-processing system which learns from experience and stores its knowledge. Learning is subject to a learning algorithm which modifies the synaptic weights and activation thresholds. In the particular case of pattern recognition, a neural network is usually trained with fixed input-output pairs. Weights and biases are adjusted until the network responds to its inputs with the desired outputs within an acceptable error. Afterwards, the network can perform on new inputs, for example, noisy versions of its training vectors. This mode of operation is called generalization. Weights and biases are kept constant and the network reacts to

its new inputs according to what it learned before, i.e. according to the information it stored in its parameters during the training phase.

In the so-called batch-mode, all the training vectors are presented to the net at once, rather than one after the other. The input vector becomes an input matrix $\mathcal{P}$ and the output vector becomes an output matrix $\mathcal{A}$. This is often a more efficient way of training a network.

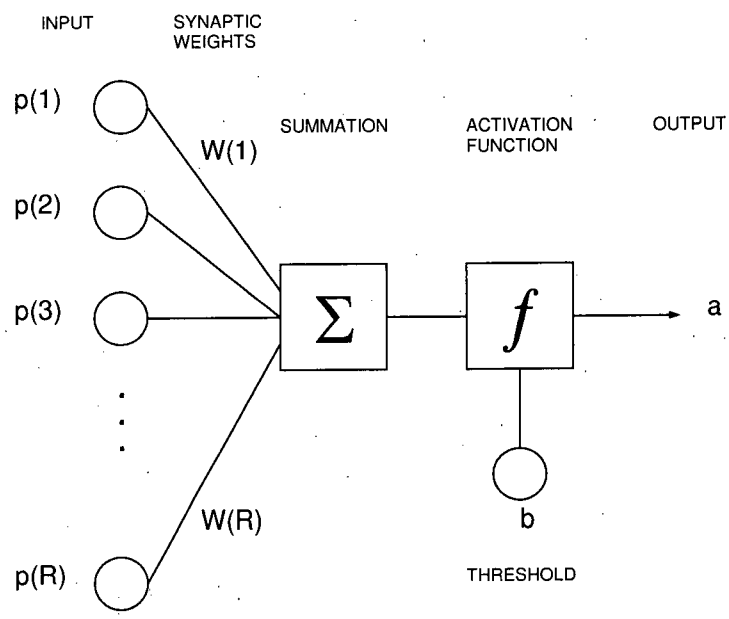


Figure 6.3: Artificial Neuron.

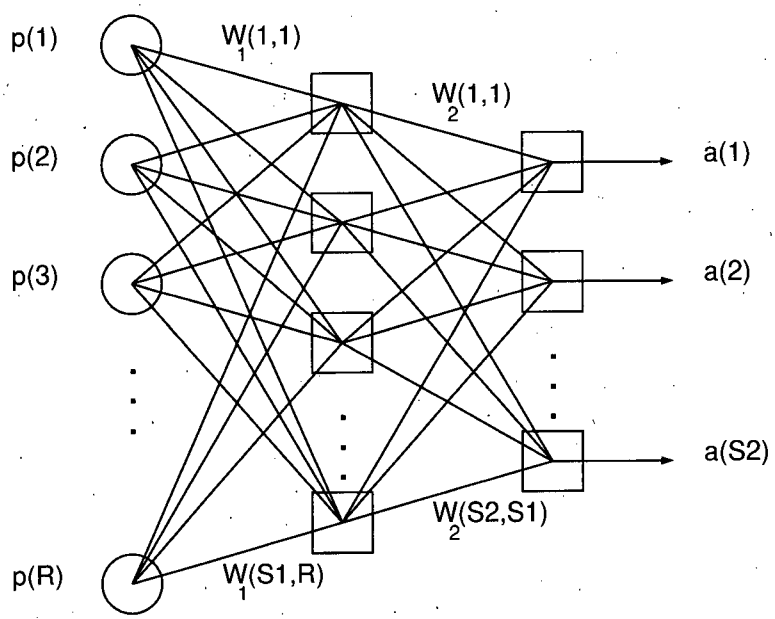


Figure 6.4: Artificial Neural Network.

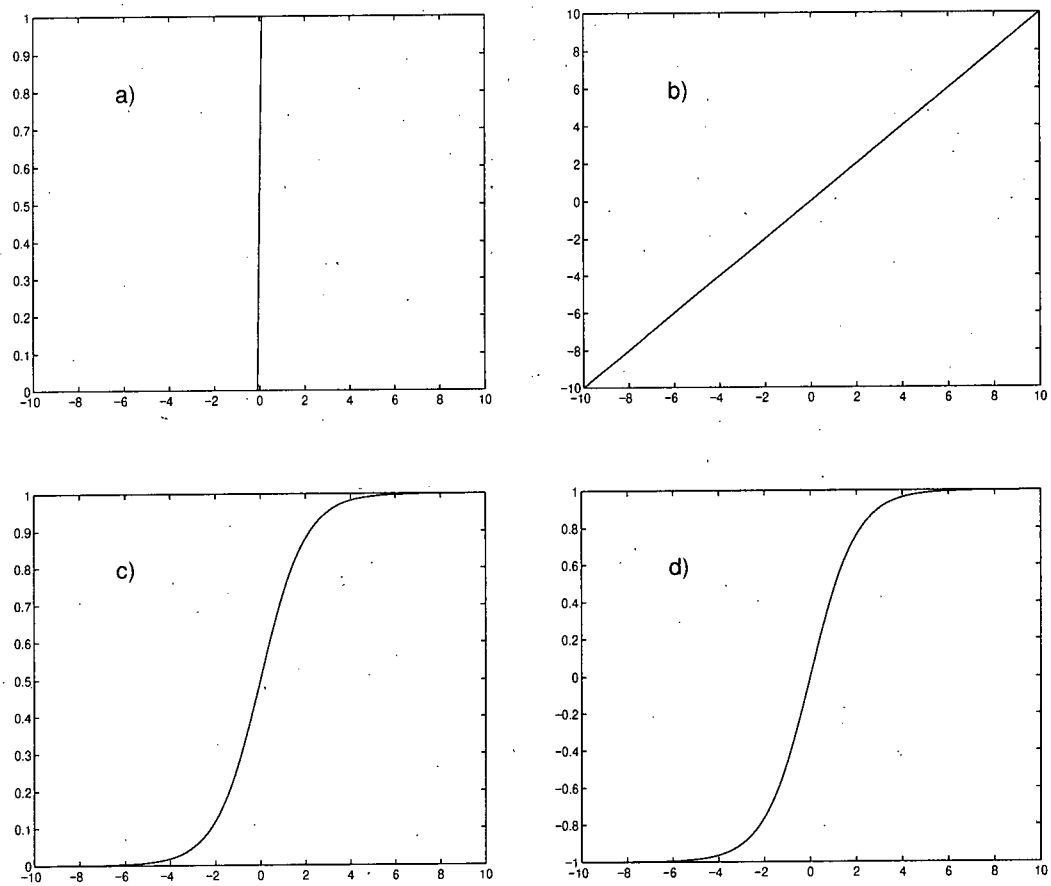


Figure 6.5: Transfer Functions for Artificial Neurons.

a) hardlimit function: $f(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases}$

b) linear function: $f(x) = x$

c) logistic sigmoid function: $f(x) = \dfrac{1}{1 + e^{-x}}$

d) hyperbolic tangent function: $f(x) = \tanh\left(\frac{x}{2}\right) = \dfrac{1 - e^{-x}}{1 + e^{-x}}$

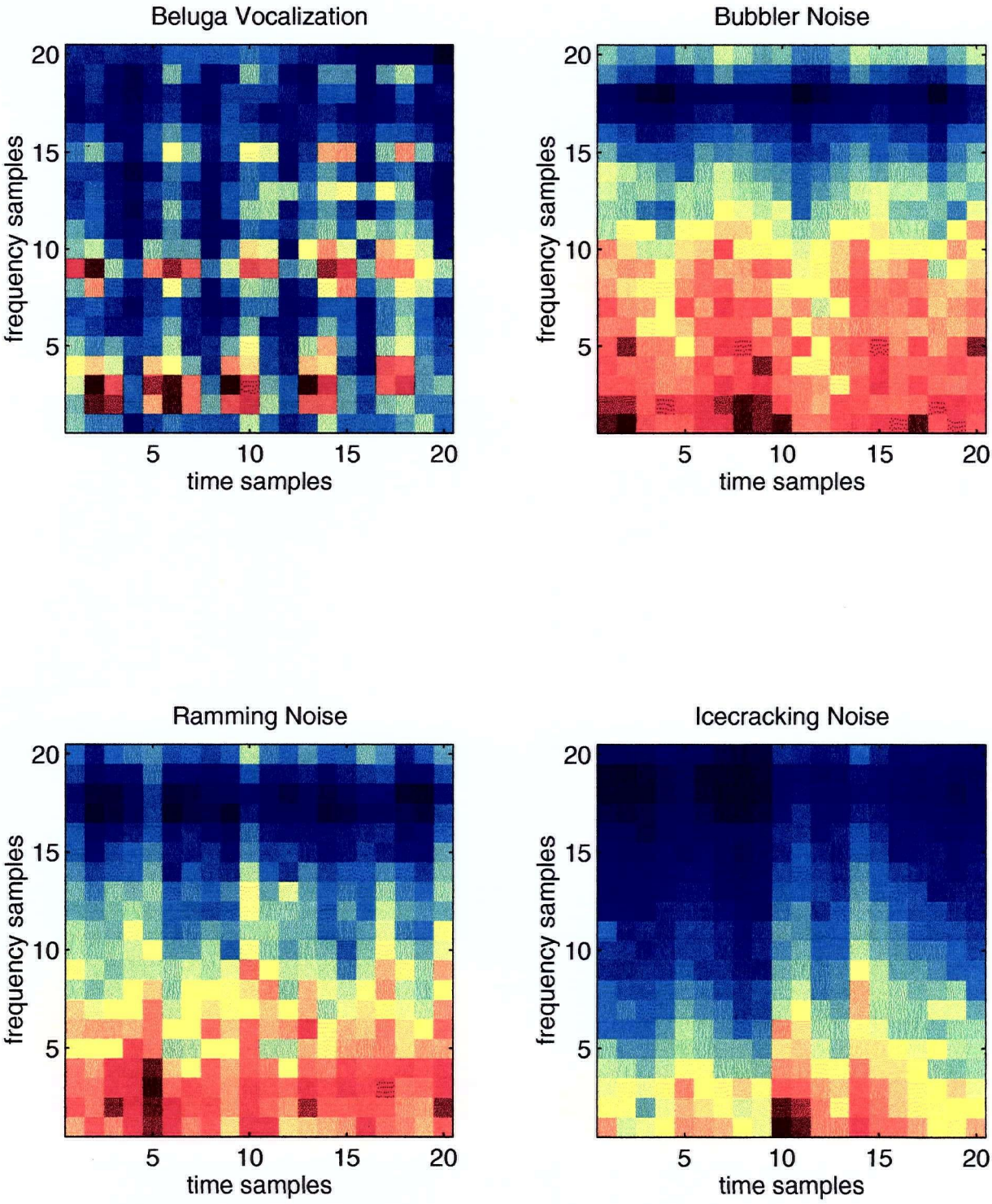


Figure 6.6: Input Vectors for Neural Network Analysis.

6.3 Input Vector Creation for Neural Net Modeling

In the mammalian ear, the cochlea performs "instantaneous" Fourier transformations of an incoming acoustic signal. Auditory nerve fibres transport a time series of Fourier transformations to the brain. In analogy, spectrogram matrices, such as Figs. 2.14, 2.11, 2.12 and 2.13, were used as inputs to the following neural networks. Neural network computation is based on matrix multiplications, which can lead to enormous computation times and easily exceed computer memory and swap space. Data reduction and compression prior to neural network modeling is essential. For data reduction, I discarded the first and the last 200 ms of the vocalization in the spectrogram Fig. 2.14. Furthermore, I 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, I averaged the energy in the spectrograms into square grids of 20 time steps and 20 frequencies. Fig. 6.6 shows these 20x20 matrices for the beluga vocalization, bubbler noise, ramming noise and icecracking noise. Frequency and amplitude were kept logarithmic to account for the logarithmic nature of mammalian hearing. Table 1.1 indicates 43 auditory bands between 700 Hz and 6 kHz. By averaging the spectrograms into 20 frequency bands, I chose a filter array half as fine. The length of one averaged box along the time axis is 82 ms. This is much coarser than the reported critical time interval of 200-300 μs for bottlenose dolphins (Section 1.3.1). Input vectors for the neural nets were finally created by reshaping the averaged spectrogram matrices into 400-element-long column vectors.

6.4 The Perceptron

The perceptron is one of the simplest and oldest neural networks. In its basic form it consists of one single neuron with a hardlimit transfer function. It can be used to classify input patterns which are linearly separable. In two dimensions, linear separability means that a straight line can be drawn through a set of input points which divides the points into two groups. In n dimensions, input patterns have to lie on opposite sides of a hyperplane.

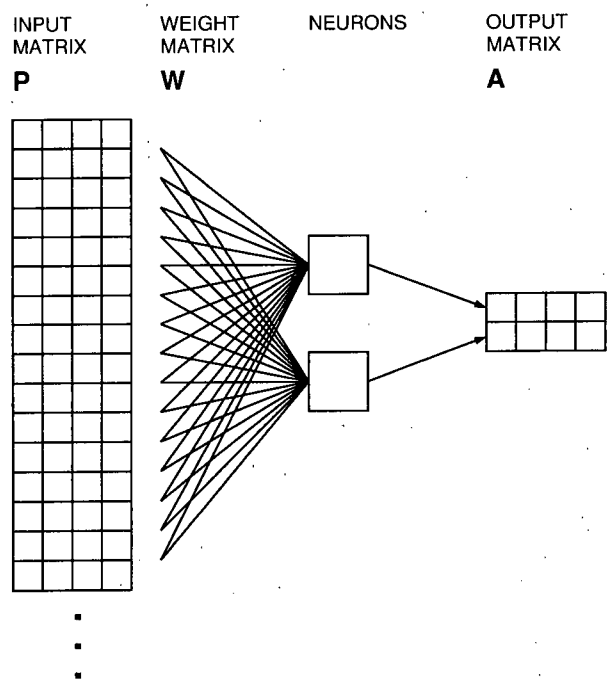


Figure 6.7: The Perceptron.

I designed a perceptron which had 2 parallel neurons, Fig. 6.7. Input to each neuron was a 400x4 input matrix $\mathcal{P}$. This matrix consisted of the four averaged spectrograms corresponding to the beluga vocalization, bubbler noise, ramming noise and icecracking noise. The output of the perceptron was a 2x4 matrix $\mathcal{A}$ which consisted of four 2-element

output vectors. Each neuron provided one output element, either 1 or 0. The four possible combinations were assigned to the four signals in such a way that the desired output for the beluga vocalization was a vector (1,0), the desired output for bubbler noise was (1,1), for ramming noise (0,0) and for icecracking noise (0,1). One neuron can divide the input space into two categories, two neurons can divide the input space into four categories. Therefore, the perceptron can successfully classify the four input signals.

Initially, the weights and biases of the two neurons were set to random values between -1 and +1. Training proceeded in such a way that the input matrix $\mathcal{P}$ and the desired output matrix $\mathcal{D}$ were presented simultaneously to the net and the actual output matrix $\mathcal{A}$ was calculated. The difference between the desired and the actual output was computed and the weights and biases were adjusted according to the perceptron learning rule [Rosenblatt 1962]:

$$\Delta\mathcal{W} = (\mathcal{D} - \mathcal{A})\mathcal{P}^T = \mathcal{E}\mathcal{P}^T$$

$$\Delta\vec{b} = (\mathcal{D} - \mathcal{A})\vec{1} = \mathcal{E}\vec{1} \tag{6.1}$$

$\mathcal{E}$ is the matrix of current errors and the superscript T denotes transposed matrices or vectors. The input matrix $\mathcal{P}$ and desired output matrix $\mathcal{D}$ were presented again and the actual output $\mathcal{A}$ was calculated with the updated weights and biases. This training process is usually repeated until the error is zero or a previously fixed, maximum number of training epochs is exceeded. If the input vectors are linearly separable, the perceptron will always find a zero-error solution [Haykin 1994].

In this particular case, the perceptron always found a zero-error solution after two

or three epochs. It successfully categorized the beluga vocalization and three noises into four separate groups. However, the solution which the perceptron finds is not unique. There are an infinite number of solutions. Final weights and biases depend on the initial conditions. Furthermore, the four acoustic signals usually don't lie in the centre of their assigned spaces. The perceptron learns in such a way that by making only slight changes to its hyperplane parameters, the zero-error solution is found as soon as the hyperplanes just pass by the points very closely. For the mode of generalization this means, that if the perceptron performs on noisy input patterns, where the noise is a short vector added to the pattern in the direction towards the dividing plane, then the point can easily fall onto the other side of the plane hence being in the wrong hyperspace. From the experiments with the whales and the human test persons, it is known that slightly noisy signals were always classified correctly and never fell into a wrong category. For mixed signals with varying nsr, the perceptron is thus no good generalizer, i.e. interpolator between the pure call and the pure noises.

A scenario more similar to the experiments with the whale and the humans is the classification of signals into two groups only: vocalization and noise. None of the individuals were asked to distinguish between the three noises. An appropriate perceptron has only one neuron, assigning an output of 1 to the vocalization category and an output of 0 to the noise category. A perceptron trained with these desired outputs also managed to converge to zero error. This means that the vocalization is linearly separable from the three noises. However, the solution is still not unique. The perceptron finds different dividing planes

depending on the (random) initial weights. Therefore, during the generalization phase, when the perceptron is presented with averaged spectrograms of call/noise mixtures in varying nsr, the switch of its output from 1 to 0 (the critical nsr) depends on the initial conditions. This is not acceptable for a proper model of the animal's response.

6.5 The Linear Neural Net

A linear neural network has linear transfer functions rather than hardlimit transfer functions as in perceptrons. Such a network can output any value not just 1 and 0. It can therefore interpolate between desired outputs. The linear neural network I designed, had the same structure as the perceptron in Fig. 6.7.

The problem posed is equivalent to *optimum Wiener filtering.* One linear neuron represents a *finite impulse response (FIR) filter* which has the filter components $\vec{w}$ and outputs a scalar $a = \vec{w}^T \cdot \vec{p}$ when presented with an input vector $\vec{p}$. The filter output a is an estimation of the desired output d with an error $e = d - a$. The optimum Wiener filtering problem is to find an optimum set of filter coefficients $\vec{w}$ for which a so-called cost function, which depends on the error, is minimum. In general, the cost function is chosen to be the mean square error (MSE):

$$J = E[|e|^2] = E[e^2] \qquad \text{for real values.}$$

E is the statistical expectation operator. A multidimensional plot of the cost function J versus the filter coefficients $\vec{w}$ constitutes the error-performance surface. It is bowl-

shaped and has a single, global minimum. This minimum is the point of optimum filter parameters. Mathematically speaking, the minimum lies at the stationary point of the cost function J, which is found by differentiating the cost function with respect to the filter components and setting the gradient equal to zero:

$$\vec{\nabla}_w J = \vec{0} \quad ,$$

The solution is given by the so-called *Wiener-Hopf equations*:

$$\vec{r} = \mathcal{R}\vec{w}_o \quad , \tag{6.2}$$

where $\vec{r}$ is the cross-correlation vector between the input vector and the desired output; $\mathcal{R}$ is the auto-correlation matrix of the input vector. Solving for the optimum Wiener filter coefficients $\vec{w}_o$ requires a Levinson recursion, which may be time consuming for large matrices. Another approach is the method of steepest descent.

The *method of steepest descent* is an iterative method in that starting from an initial set of filter parameters somewhere on the error-performance surface, the method of steepest descent moves the parameters gradually towards the global minimum. Filter parameters are upgraded from iteration to iteration by adding a vector which is proportional to the gradient of the error-surface at that very point, but points into the opposite direction,

$$\Delta\vec{w} = -\frac{1}{2}\mu\vec{\nabla}_w J \quad .$$

The positive constant μ is called the learning rate parameter. Inserting the mean square error for the cost function, yields

$$\Delta\vec{w} = \mu(\vec{r} - \mathcal{R}\vec{w}) \quad . \tag{6.3}$$

Strictly speaking, the cross-correlation vector and the auto-correlation matrix are ensemble averages, taken over an ensemble of spatial filters of identical design but with different inputs drawn from the same population. They are thus unknown (in general) and must be estimated. Assuming that the input signals and desired responses are jointly ergodic, the ensemble averages can be replaced by the more easily available time averages. In this sense, the method of steepest descent minimizes the sum of squared errors (SSE), summed over all iterations of the algorithm, but for a particular realization of the linear filter.

The *least mean square (LMS) algorithm* [Widrow and Hoff 1960] is an approximation of the steepest descent algorithm in that it replaces the time averages for the cross-correlation vector and the auto-correlation matrix with the instantaneous values. Thus, the simplest choices of estimators for $\vec{r}$ and $\mathcal{R}$ are:

$$\vec{r} = \vec{p} \cdot d \qquad \text{and} \qquad \mathcal{R} = \vec{p} \cdot \vec{p}^T \quad .$$

Substituting this into the steepest descent algorithm leads to the new simpler updates for the filter coefficients:

$$\Delta \vec{w} = \mu \vec{p} e \quad . \tag{6.4}$$

In the case of my linear two neuron network, I chose an initially random weight matrix $\mathcal{W}$ and two random biases with values between -1 and +1. In each iteration, the weights were updated by

$$\Delta \mathcal{W} = \mu \mathcal{E} \mathcal{P}^T \quad .$$

The biases are simply weights with constant inputs equal to 1 and are therefore updated

by

$$\Delta\vec{b} = \mu\mathcal{E}\vec{1} \quad .$$

Using the same 400x4 training matrix P and 2x4 output matrix D, I calculated the optimum Wiener filter components directly using the Wiener-Hopf equations and recursively using the least mean square algorithm. Solving the Wiener-Hopf equations gave a zero-error solution. The least mean square algorithm was run ten times with 40,000 iterations. Depending on the initial conditions, the final SSE lay between 0.01 and 0.001. The net successfully assigned four different output vectors to the four input signals, a task which the perceptron had also fulfilled. But, if the linear neural network was presented with mixed signals of the beluga call and one of the three noises, the neurons were not forced to output either 1 or 0. They could output intermediate values which contained more detailed information. For example, if, after training had been completed, the net was presented with a mixture of the beluga call and bubbler noise in a noise-to-signal ratio of 1, the net's output was a vector $\vec{a}$ =(0.96,0.23). The fact that this vector was not equal to one of the desired outputs, "proved" that the net had been presented with a noisy version of one of its training vectors. Furthermore, the output vector was closest to the output vector for the pure beluga call (1,0), indicating that the net recognized the strong call content. If presented with a mixture of the call and bubbler noise in a noise-to-signal ratio of 16, the net's output became $\vec{a}$ =(0.99,0.97). This is closest to (1,1), the desired output for the pure bubbler noise. The net therefore recognized the strong bubbler noise content.

As designed, the neural network can to some extent specify which signals were mixed and how strongly, but the distances between the output vectors depend on the initial network parameters. The freedom of the linear neural net is the sum of all its weights and biases. Each neuron has a 400-element weight vector and 1 bias. The number of variables of the neural net is hence 802. Constraints are applied to the net in terms of input-output pairs. With four input-output pairs and two elements in each desired output, the total number of constraints is 8. The system would be exactly determined and uniquely solvable if the number of variables was equal to the number of constraints. The current linear net is underdetermined; there is an infinite number of solutions. The error-performance surface does not have one global minimum point, but a global minimum plane containing an infinite number of equally low points.

A linear network with just one neuron, classifying input signals into either "call" or "noise" as the second perceptron did, has 401 variables and 4 constraints. The degree of freedom, equal to the number of variables minus the number of constraints, is smaller than for the 2-neuron network, however, still too large for a unique solution. The degree of freedom can further be reduced by increasing the number of training vectors (input-output pairs).

I created 500 random white noise vectors and mixed them with the beluga vocalization in the time domain in noise-to-signal ratios of 1. In the spectrograms, the call could easily be identified visually. These mixed spectrograms were averaged, added to the training matrix $\mathcal{P}$ and assigned a desired output of 1 equal to that of the pure call. The

corresponding pure white noises were included in $\mathcal{P}$ as well and assigned a "noise" output of 0. The neural network still managed to classify the 1004 input patterns into either "call" or "noise" as desired. The generalization, though, seemed as arbitrary, i.e. dependent on the initial weights, as previously. This was because after averaging the spectrograms, all the 500 white noises looked alike; each box in the 20x20 grid had the same intermediate amplitude. The degree of freedom of the neural network had thus not been reduced, because the extra input-output pairs were identical.

Differing training vectors were subsequently created by adding white noise to the averaged spectrogram of the beluga vocalization. Each of the boxes in the 20x20 grid was altered by adding a random number between $\pm\frac{1}{5}$th of the mean call value. With these training vectors, the neural network did not converge. The two categories were no longer linearly separable.

In conclusion, a linear neural network proved to be a poor design when applied to an interpolation problem between pure noise and a pure call. Few training vectors result in an underdetermined net which converges properly, though generalizes poorly. Many training vectors, on the other hand, increase the probability of the categories not being linearly separable, in which case the network does not converge to a desired small error.

A different neural network, which is usually a better interpolator and which can classify linearly non-separable categories by visually "wrapping" the dividing surface around the categories, is a backpropagation network which will be dealt with later.

6.6 Adaptive Noise Cancellation

One important feature of the least mean square algorithm is that it can operate in stationary as well as non-stationary environments. In a non-stationary environment, the statistics change with time. The optimum Wiener solution varies, too. The LMS algorithm therefore not only has to seek the minimum point of the error surface but it also must track it. This constitutes a linear adaptive filtering problem. Adaptive filters have been employed successfully in such diverse fields as radar, sonar, seismology, biomedical engineering and communications. This section discusses whether the mammalian auditory system is likely to perform adaptive noise cancellation to enhance signal discrimination in noise.

An adaptive filter can be realized as one linear neuron which receives the time series (not the spectrogram) of a pure noise signal as input and the time series of a mixed signal as desired output (Fig. 6.8). The data are presented not all at once but time step by time

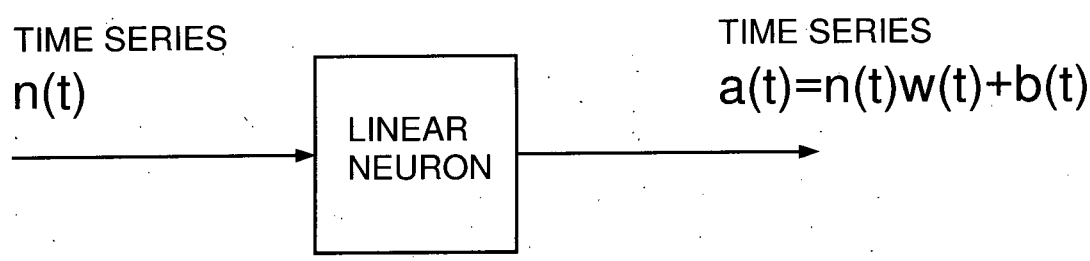


Figure 6.8: Adaptive Noise Cancellor.

step. If the mixed signal is a linear combination of the vocalization and the noise, the neural net will learn to predict the noise content in the mixture. Here the mixed signal

or desired output is a time series

$$d(t) = s(t) + \alpha n(t) \quad ;$$

with s being the vocalization, n being the noise and α being the noise-to-signal ratio. In an even more general case, one can also account for a DC offset:

$$d(t) = s(t) + \alpha n(t) + DC \quad .$$

The output of a linear neuron is

$$a(t) = w(t) \cdot n(t) + b(t) \quad ,$$

with $w(t)$ being the only weight and $b(t)$ being the only bias. The momentary error of the net is

$$e(t) = d(t) - a(t) = \big(\alpha - w(t)\big) \cdot n(t) + s(t) + \big(DC - b(t)\big) \quad .$$

Minimizing the sum squared error leads to $w(t)$ approximating the noise-to-signal ratio α and $b(t)$ converging to the DC offset. The error of the net thus converges towards the vocalization $s(t)$.

Fig. 6.9 shows the time series of the pure beluga call. It consists of 90,112 samples taken at a sampling rate of 44,100 samples per second. The call is therefore about 2 s long. Figs. 6.10, 6.11 and 6.12 show the time series of the bubbler noise, ramming noise and icecracking noise. All data were normalized to a root-mean-square amplitude of 0.23 which yielded amplitudes in the range of -1 to +1.

Without any training as in the previous neural net examples, the adaptive filter was immediately presented with a time series of a mixed signal. The mixture consisted of the beluga call and bubbler noise in a noise-to-signal ratio of 2. Fig. 6.13 shows a plot of $e(t) - s(t)$, i.e. the net's approximation to the vocalization minus the true vocalization. The plot hence shows the accuracy with which the net modeled the vocalization. One sees that the error started out fairly large with an amplitude of 0.8, which is the order of magnitude of the input signal. It converged fast, within less than 0.2 seconds. Whenever the pulses of the beluga vocalization occurred, the net's tracking behaviour was suddenly distorted; the error jumped to a higher value of up to 0.2 and then decreased again. Fig. 6.14 shows the net's performance on a time series with an nsr of 16. The initial error started out higher, at an amplitude of 8, but the net converged in exactly the same time, less than 0.2 seconds as with the less noisy vocalization. Once it had converged, the remaining error was the same as before, less than 0.2. Testing the neural net with various nsr's, the higher the nsr, the greater is the initial distortion. However, the time needed for convergence (0.2 s) and the final error (0.2) are independent of the nsr.

Results of the adaptive filter's performance with mixtures of the beluga vocalization and ramming noise in an nsr of 2 are shown in Fig. 6.15. As with the bubbler noise, the net converged within less than 0.2 seconds. The remaining error was of the order 0.2. The higher the nsr, the greater was the initial distortion. Tests with icecracking noise mixtures looked similar. Fig. 6.16 shows the convergence of the net for an nsr of 2.

Summarizing the results, no matter which noise the net had to deal with and no matter

what the noise-to-signal ratio was, the linear adaptive filter always converged within less than 0.2 seconds. Furthermore, its remaining errors were all of the same order, always less than 0.2. As looking at time series of acoustic signals is sometimes not very informative, I converted the neural net's outputs to analogue acoustic signals and listened to the performance of the adaptive filter. Apart from the initial "crashes" during convergence and quieter, minor "crashes" at the beginning of each beluga call pulse, the output of the net sounded perfectly like the pure undistorted beluga vocalization. One could not tell which noise the adaptive filter was cancelling out at the time. There was no audible difference between the outputs during bubbler noise, ramming noise or icecracking noise filtering. These are very important findings. They show that if the auditory systems of the whale and the human test persons make use of adaptive noise cancellation at all, then it is not the primary method. Data collected from the acoustic experiments show that recognition of a particular beluga call is sensitive to the type of noise it is distorted with and the noise-to-signal ratio. This dependence cannot be reproduced by adaptive filter theory.

One problem with the adaptive filter theory is that the filter needs the pure noise signal for reference. In nature and in my experiments, the whale and the humans only had the first signal but no reference signal available. However, one could imagine that the animal takes the reference signal from its memory. We know that solo musicians play entire concerts by heart, every single note of a Beethoven violin concerto is anchored in their mind. Similarly, a whale that grows up in a noise polluted ocean could store

representations of various underwater noises in its mind and recall them for filtering purposes later on. Another problem for adaptive filtering by the auditory system could be that the pure noise and the noise in the mixed signal must be in phase. But considering the solo musician again, one knows that if he or she gets lost during the concert, he/she will always immediately find the right "phase" to join the rest of the orchestra again. Phase matching of an incoming acoustic signal and a signal from memory should thus not pose a problem. One could also argue that the whale and the human test persons were not used to the types of noise I used in my experiments. However, the noise signals were very short, all less than 2 s long and on top of that of periodic nature. A typical noise sample containing all the necessary information could therefore be as short as half a second. The whale was trained with the same noise signals for about 6 months and should easily have been able to memorize the noise characteristics. Two of the human test persons were also very familiar with the sound of the noises before the experiments.

Even if our brains were "wired" as adaptive filters, Fig. 6.17 explains why such a filter cannot be successful when operating on time series. Plotted are the auto-correlation coefficients of the three noises and Gaussian distributed white noise. In each case, a 2s noise sample was correlated with a 3s noise sample, and lags between 0s and 1s were plotted. For zero lag, the coefficients are equal to 1. For all other lags, they oscillate around 0. The noises are hence uncorrelated with themselves. If an individual had a "typical" noise sample stored in its brain, it could not successfully use this sample to filter the same type of noise out of an incoming acoustic signal. This is not surprising considering the

underlying physical process that generates the current noises. The sounds are related to random physical processes such as the bursting of bubbles of widely distributed size and at random times. Therefore, even if two samples of the same noise sound very similar, their time series are greatly uncorrelated.

Adaptive filter theory requires that the reference noise signal is correlated with the noise contained in the mixed time series. Again, remembering that we don't hear time series but something more similar to spectrograms, can the mammalian brain compute adaptive noise cancellation on Fourier transformed signals? Fig. 6.18 shows the auto-correlation of the noise spectrograms. The correlation coefficients were computed by multiplying each spectrogram matrix with itself and summing up all the products. In the plot for bubbler system noise, a peak occurs whenever 2 pulses of the bubbler system overlap. Two pulses per second exist (Fig. 2.11). The plot for ramming noise identifies the blade frequency of the icebreaker's propeller: 11 Hz (Fig. 2.12). The plot for icecracking noise indicates that the ice burst about $2\frac{1}{2}$ times per second. Overall, the correlation coefficients of the non-white noises are fairly large.

However, even spectrograms with linear amplitude and frequency distribution are not strictly a linear transform of the underlying time series. The Fourier transform itself is linear, but a spectrogram plots the magnitude of the Fourier components. During this process, phase information is lost. An adaptive filter performing on spectrograms did not converge, because the call spectrogram cannot be produced by subtracting a noise spectrogram from the spectrogram of the mixed signal. I therefore conclude, that

adaptive noise cancellation is unlikely to occur in the mammalian auditory system. More likely are methods of pattern recognition as will be shown in the following section on backpropagation networks.

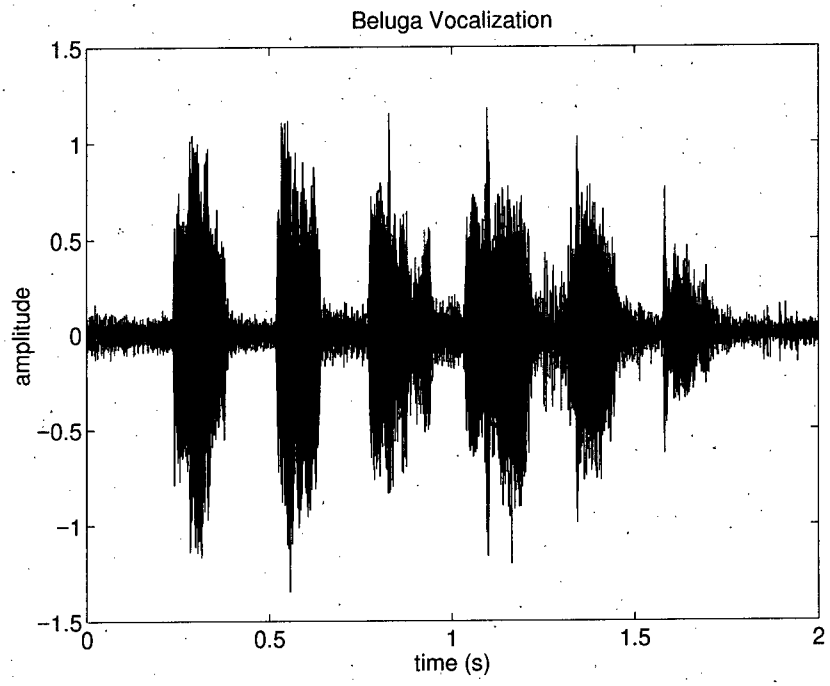


Figure 6.9: Time Series of the Beluga Vocalization.

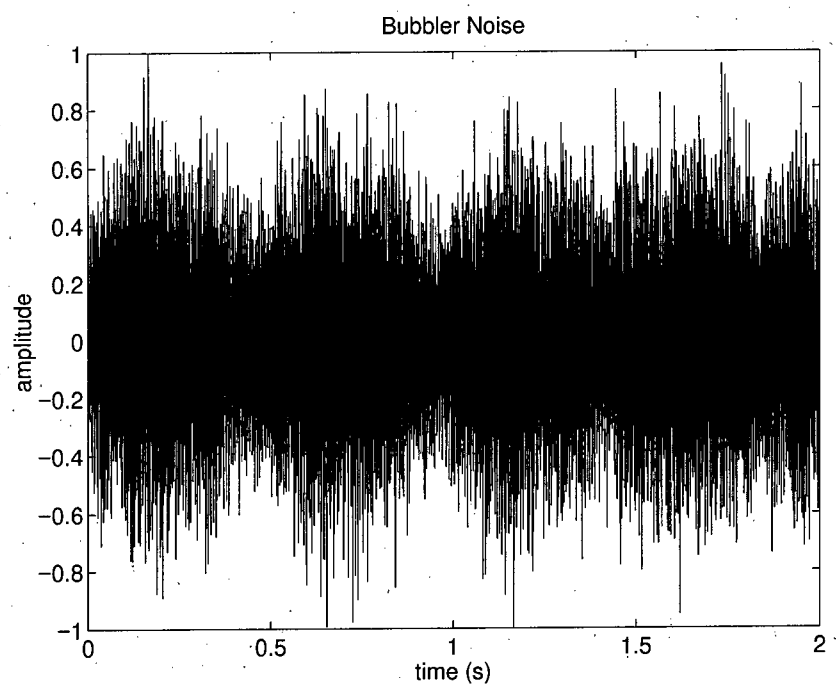


Figure 6.10: Time Series of the Bubbler Noise.

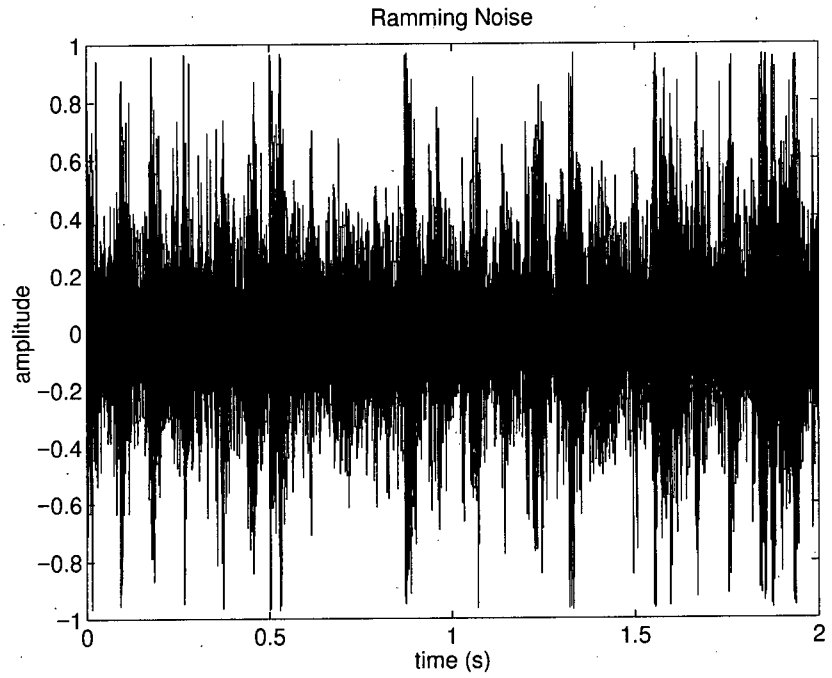


Figure 6.11: Time Series of the Ramming Noise.

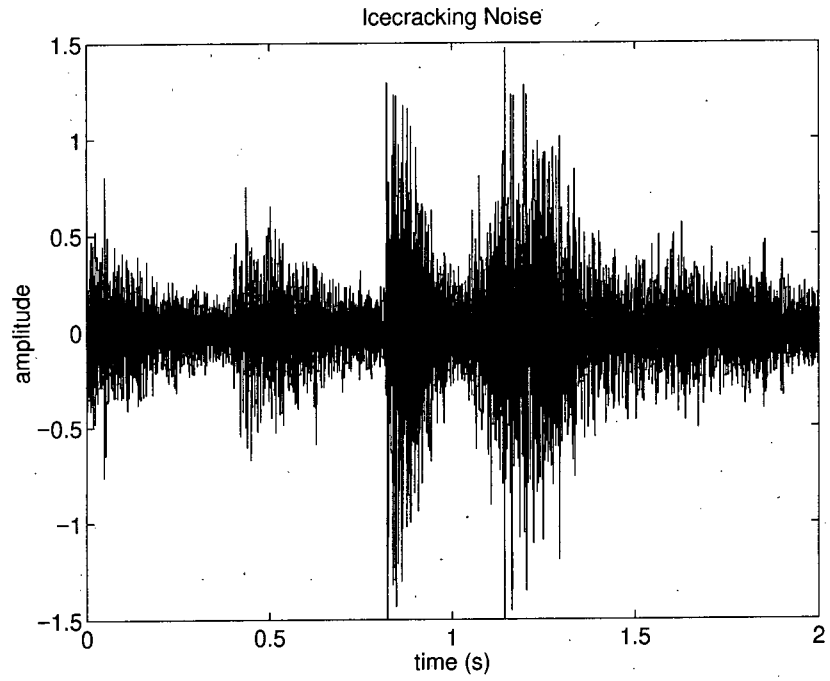


Figure 6.12: Time Series of the Icecracking Noise.

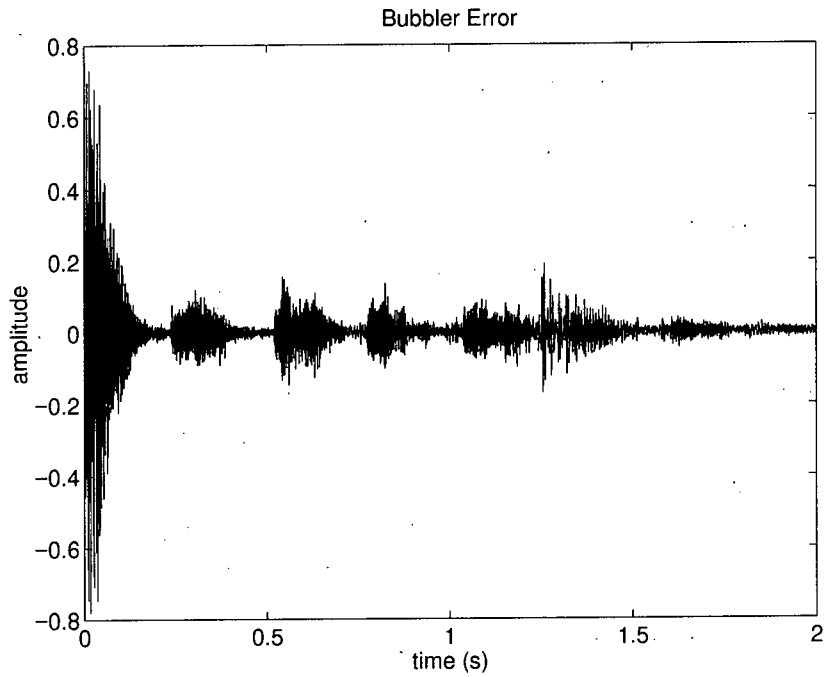


Figure 6.13: Error of the Adaptive Filter in Approximating the Beluga Vocalization, nsr=2.

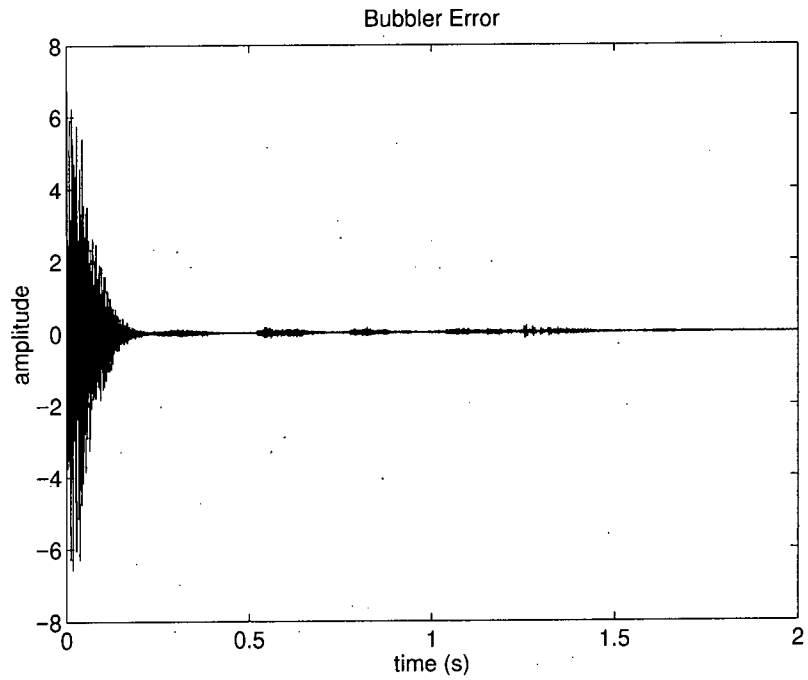


Figure 6.14: Error of the Adaptive Filter in Approximating the Beluga Vocalization, nsr=16.

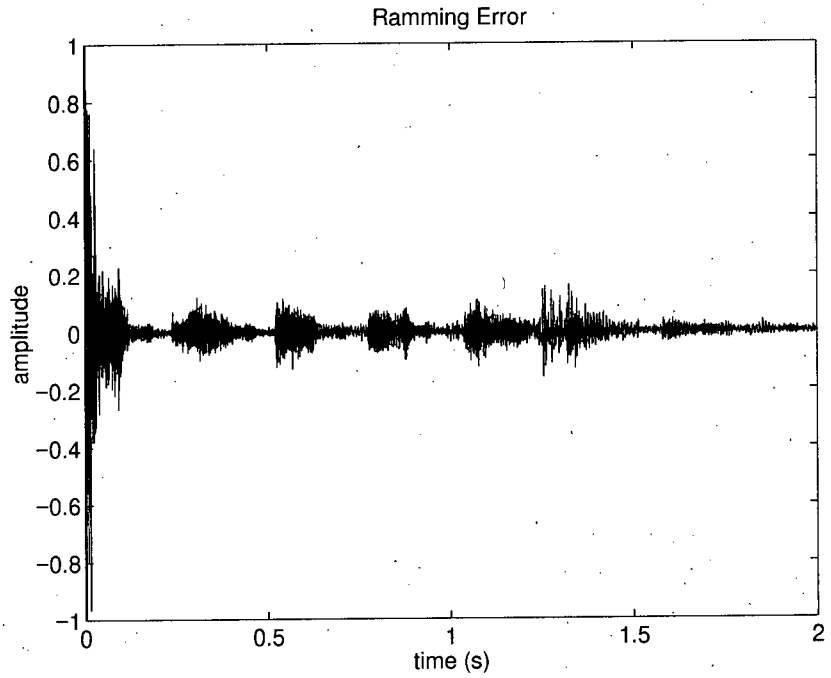


Figure 6.15: Error of the Adaptive Filter in Approximating the Beluga Vocalization, nsr=2.

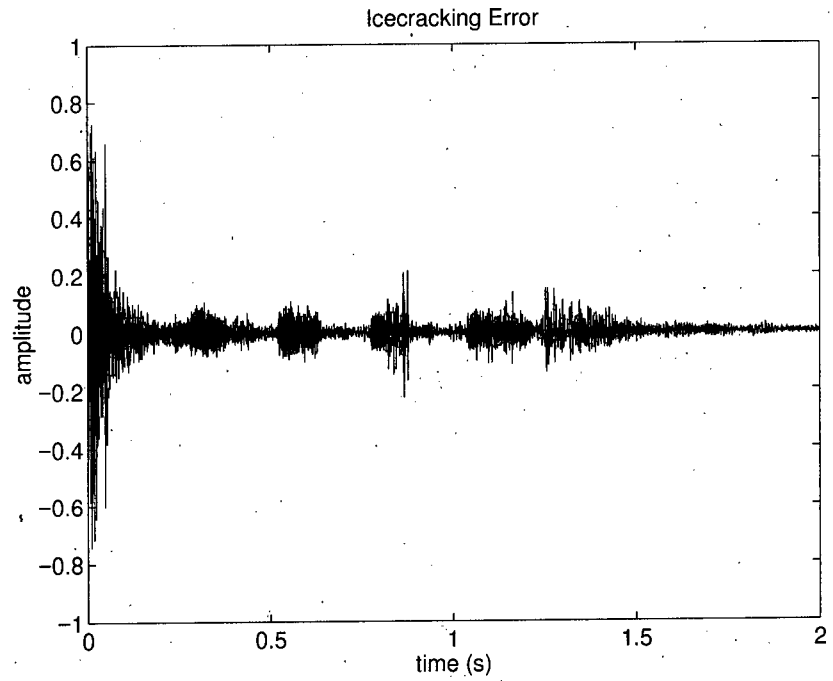


Figure 6.16: Error of the Adaptive Filter in Approximating the Beluga Vocalization, nsr=2.

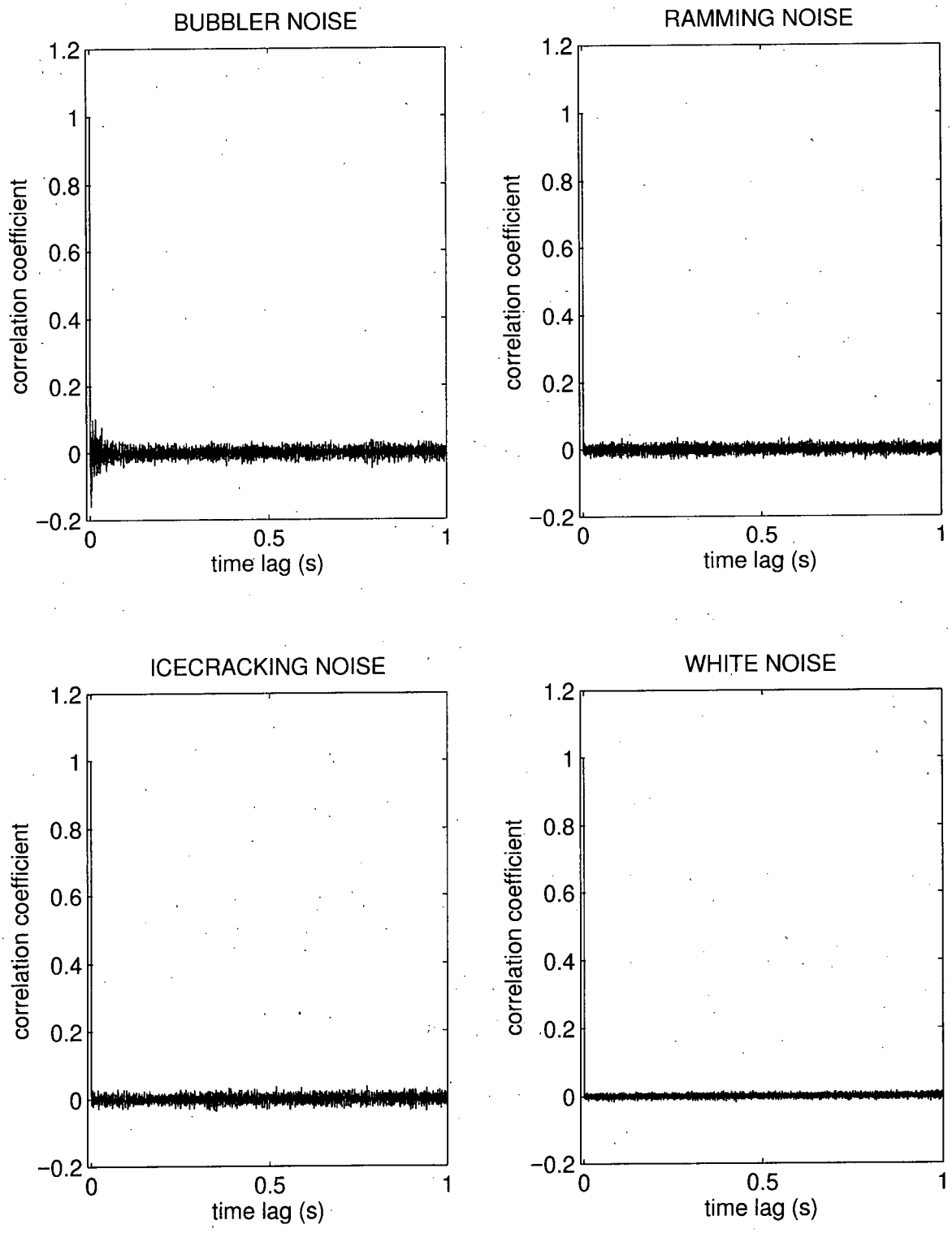


Figure 6.17: Auto-Correlation of the Noise Time Series.

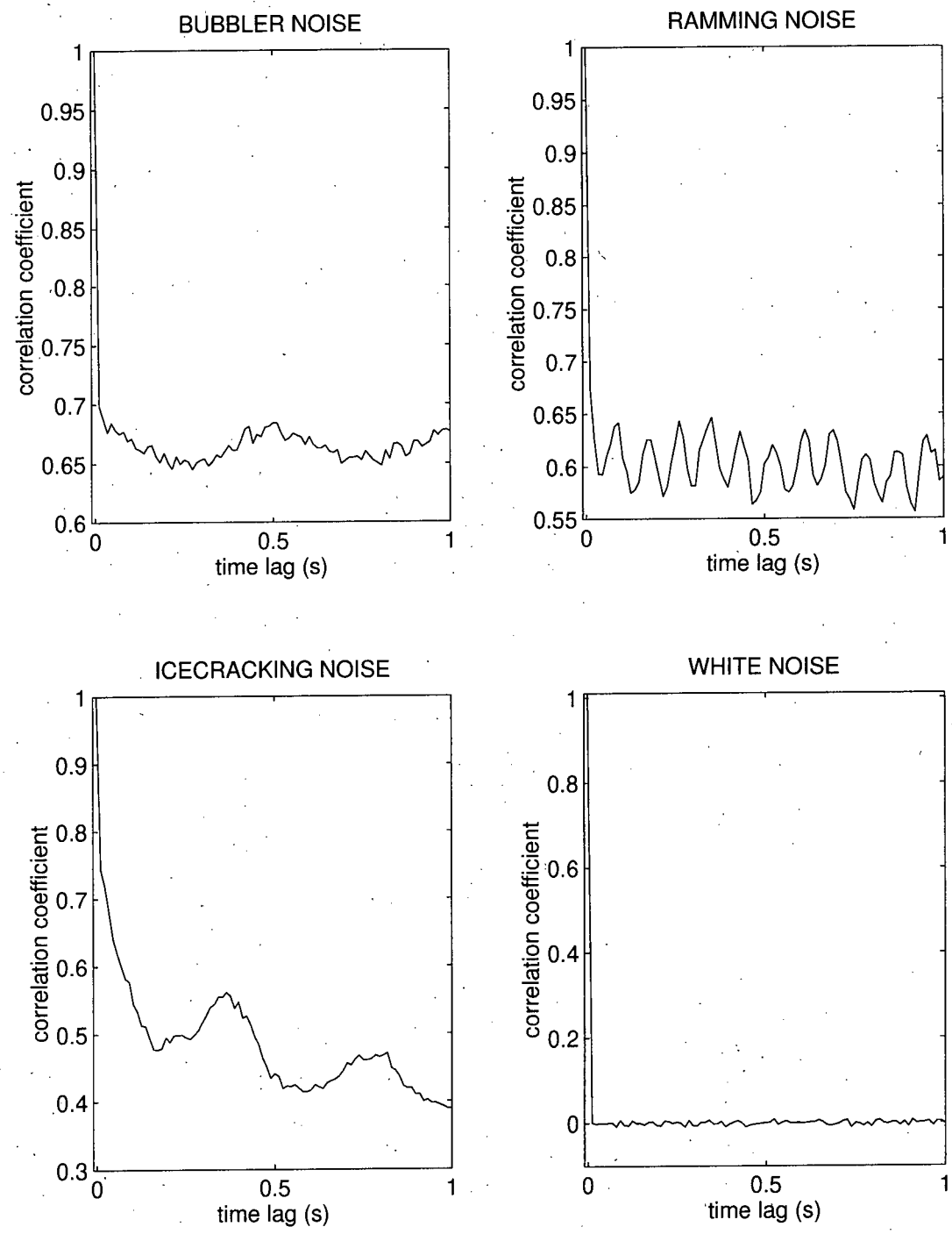


Figure 6.18: Auto-Correlation of the Noise Spectrograms.

6.7 A Multilayer Backpropagation Network

The *backpropagation* learning rule can be regarded as a generalized least mean square algorithm for multiple-layer networks and nonlinear differentiable transfer functions [Rumelhart *et al.* 1986a]. The network's weights and biases are adjusted from iteration to iteration such that the sum squared error of the network is minimized. This is done by gradually changing the network's parameters in the direction of steepest descent. Multilayer backpropagation networks have two distinct modes of computation. Starting with initial weights and biases, the output of the first layer is computed and taken as the input for the second layer. The output of the second layer becomes the input of the third layer and so forth. That way the total output of the neural net is calculated by propagating from left to right through the neural net. This pass is called the feedforward pass. The net's output is compared to the desired output, the error is calculated and backpropagated through the network from right to left while adjusting the weights and biases layer by layer. This pass is called the backpropagation pass. The rules for updating the weights and biases are similar to the LMS rules with the only exception that the error matrix $\mathcal{E}$ is replaced by the matrix of error derivatives $\mathcal{S}$ with respect to the weights. This is because of the nonlinearity of the transfer functions.

$$\begin{aligned} \Delta\mathcal{W} &= \mu\mathcal{S}\mathcal{P}^T \\ \Delta\vec{b} &= \mu\mathcal{S}\vec{1} \end{aligned} \tag{6.5}$$

During the following feedforward pass, the updated weights and biases are kept constant and the network's total output is computed. The error matrix is calculated, the deriv-

atives are calculated and by backpropagation through the network, weights and biases are updated again. This iteration continues until the sum-squared-error (SSE) reaches a required minimum or a maximum number of epochs is exceeded.

Due to the nonlinearity of the transfer functions, the error performance surface of a backpropagation network does not have one unique minimum anymore but exhibits a number of local minima. One common problem with backpropagation is that the method gets trapped in one of the local minima and never finds the absolute minimum. If the SSE at the local minimum is lower than required, this is not a problem. Often however, the SSE's at the local minima are too large. One way of avoiding being trapped in a local minimum is the *method of momentum* [Rumelhart *et al.* 1986b]. Momentum allows the network to respond not only to the local gradient of the error performance surface but also to recent trends. Weights and biases are updated by a weighted sum of the last update and the newly suggested update.

$$\begin{aligned} \Delta\mathcal{W} &= c\Delta\mathcal{W}_{\text{old}} + (1-c)\mu\mathcal{S}\mathcal{P}^T \\ \Delta\vec{b} &= c\Delta\vec{b}_{\text{old}} + (1-c)\mu\mathcal{S}\vec{1} \end{aligned} \tag{6.6}$$

The method of momentum gets its name from picturing a ball sliding down the error surface. Momentum helps the ball to roll through a local minimum and overcome the little hill on the other side. One wants to avoid momentum being such as to push the parameters out of a deep valley. Therefore, a maximum error ratio is introduced as well. If the ratio of new error to old error exceeds a fixed value, weight and bias changes are rejected.

On top of introducing momentum and a maximum error ratio, I found an adaptive learning rate useful. During iteration, if the new error exceeded the old error by a predefined ratio (1.04 as before), not only were the suggested weight and bias changes discarded but also the learning rate μ was decreased (by multiplying it with 0.7). If the new error was smaller than the previous error, the learning rate was increased (by multiplying by 1.05).

Many different backpropagation networks were tested. They differed in the number of layers (2 or 3), in the number of neurons used per layer, in the transfer functions and the training matrix. As far as the number of neural layers and the number of neurons per layer is concerned, generally, the fewer neurons, the faster the computation of one epoch; the more neurons, the faster the convergence of the neural network (in fewer epochs). There is a trade-off between the computation time per epoch and the number of epochs needed to reach the minimum error. I eventually chose not to search for the optimum number of neurons as determined by minimum computation time, but to settle for the minimum number of neurons, even if this was at the expense of slower convergence. The reason was that the performance of the neural network during generalization seemed more stable, i.e. less dependent on the initial conditions.

As far as the transfer functions are concerned, I did not find major performance differences for sigmoid or hyperbolic tangent transfer functions. I 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 [Pineda 1988].

The training matrix for the neural network was created by adding noise to the call not in the time series but in the averaged spectrogram (subscript s for spectrogram domain). In particular, 800 reshaped noise vectors of the form

$$n_s[t,f] = \sin\left(\frac{2\pi\omega_t t}{T} + \phi_t\right) \cdot \sin\left(\frac{2\pi\omega_f f}{F} + \phi_f\right) + 0.1 \cdot \mathrm{rand}(t,f) \tag{6.7}$$

were computed with

$$T = 20, \quad F = 20, \quad \omega_t, \omega_f \in [0, \frac{1}{2}, 1, \ldots, 6] \quad .$$

This equation describes a two-dimensional sine wave overlapped by random values between 0 and 0.1. The phases ϕ_t and ϕ_f were chosen randomly between 0 and π. Another 50 vectors of entirely random values were created as well.

The root-mean-square amplitudes of the call and the 850 noise vectors were calculated in the spectrogram domain; signals were then mixed according to

$$x_s[t,f] = s_s[t,f] + \alpha_s \cdot n_s[t,f] \quad , \tag{6.8}$$

with α_s varying between 0 and 1. The training matrix was the assembly of 850 noisy call spectrograms (averaged and reshaped) and 850 pure noises. The desired output was 1 for the noisy calls and 0 for the pure noises.

A fully-connected 2-layer neural network of the type sketched in Fig. 6.4 with three neurons in the hidden layer and one output neuron was trained in batch mode. The

backpropagation net managed to converge to a sum-squared-error of 0.001 in about 1000 iterations. I want to emphasize at this stage, that the neural net did not "see" any of the ocean noises or mixtures during training. It only saw somewhat arbitrary noisy versions of the vocalization and the corresponding artificial noises. In order to check proper generalization after completed convergence, I presented the net with the pure vocalization and the pure bubbler, ramming and icecracking noise. Its output was 1.0 for the call and 0.0 for the three noises, just as desired.

The network was then presented with the original mixtures of the call with the three noises in (time series) noise-to-signal ratios of 0 dB, 6 dB,...,30 dB. The net's output is shown in Fig. 6.19. The 50 % thresholds are 1.6 dB for bubbler noise, 4.7 dB for ramming noise and 14.0 dB for icecracking noise. The standard deviations are 0.9 dB for bubbler and ramming noise, and 1.4 dB for icecracking noise. These were calculated after running the neural network ten times with different initial conditions. Therefore, not only the order of the noises from strongest to weakest masking is the same as for the whale (and the humans), but also the relative degrees of masking are the same. Subtracting the net's thresholds from Aurora's, yields an offset of 13.8 dB for bubbler noise, 13.2 dB for ramming noise and 15 dB for icecracking noise. Thus shifting the net's thresholds an average of 14.0 dB to higher noise-to-signal ratios, gives Aurora's results with a maximum error of 6 %.

This "calibration" of the neural network to Aurora's performance is necessary, because the neural net did not get any species or individual specific input data. I tried a

normalization of training and generalization signals with respect to the beluga audiogram (Fig. 1.1). However, as the vocalization used was limited to between 700 Hz and 6 kHz, where the audiogram is basically linear, this normalization did not affect the neural network's performance. For more broadband signals though this audiogram normalization might become important and a good means of including biological background data in the computer modeling.

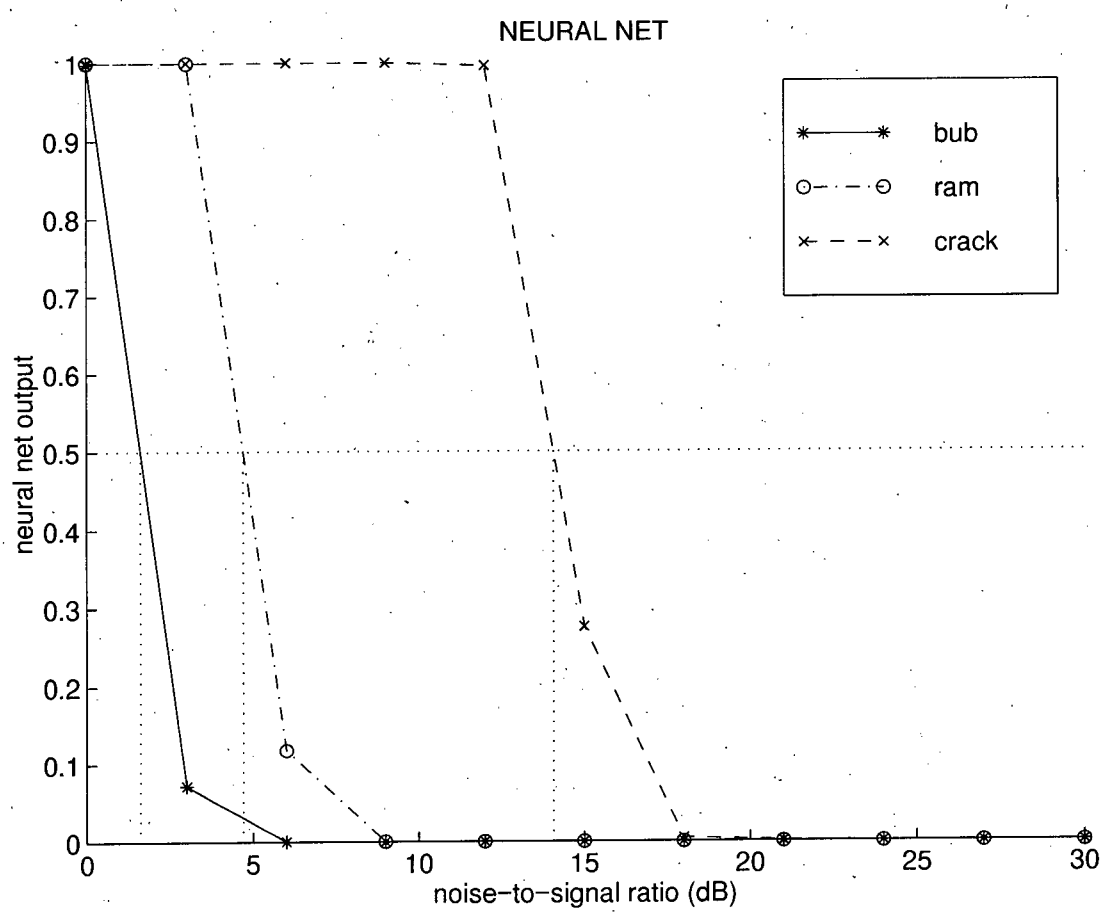


Figure 6.19: Masked hearing thresholds of the neural network in bubbler system noise, ramming noise and natural icecracking noise. The same explanations apply as in Fig. 3.15. Defining the hearing threshold at 50 % yields the following critical noise-to-signal ratios: 1.6 dB for bubbler system noise, 4.7 dB for ramming noise and 14.0 dB for natural icecracking noise. With an offset of 14.0 dB, the network models Aurora's thresholds within 6 %.

6.8 Data Summary

Fig. 6.20 summarizes the data from the whale experiment, human experiment and computer modeling for comparison. Plotted are the hearing thresholds taken at the 50 % probability of detection. In cases where a 50 % mark could not be defined, e.g. in the case of Kuan-Neng, whose icecracking curve did not fall below 50 %, the results were marked as N/A.

The three human listeners classified the three noises in the same order as the whale, Aurora. Bubbler system noise was the strongest masker, natural icecracking noise the weakest. Thresholds were slightly higher for the humans than for the whale, indicating that the humans were better at detecting the call in the noise than was Aurora. Matched filtering did not produce different thresholds for the different noises. Spectrogram cross-correlation showed an improvement by splitting up the individual noises. However, thresholds were in exactly the opposite order to the whale. Critical band cross-correlation exhibited a cross-over of curves with the wrong order of noises for high noise-to-signal ratios. The curve for bubbler noise did not drop below 50 %. Visual spectrogram discrimination, on the other hand, managed to order the noises correctly and also gave thresholds similar to those measured from Aurora. Adaptive filtering did not produce different thresholds for the three noises, nor was the filter's call detectability dependent on the noise-to-signal ratio. The backpropagation neural network was the only automated technique which classified the noises in the right order and with good relative thresholds.

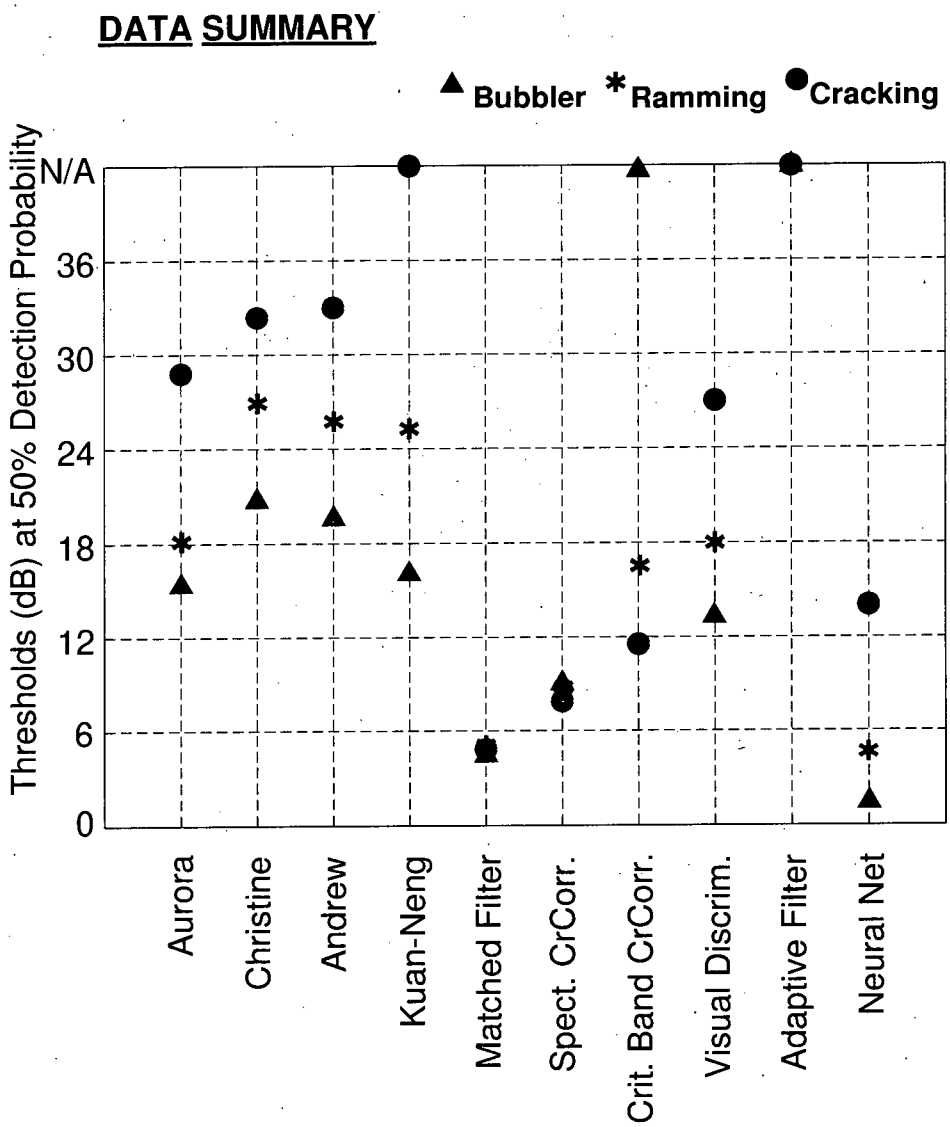


Figure 6.20: Summary of critical noise-to-signal ratios. The human listening experiments, the visual spectrogram discrimination and the backpropagation network are the only methods classifying the noises in the same order and with similar relative thresholds to the whale, Aurora.

Chapter 7

Modified Masking Experiments

If your experiment needs statistics,

you ought to have done a better experiment.

–Ernest Rutherford

7.1 Masked Hearing Thresholds of a Beluga in Continuous Noise

As explained in Section 4.2, the initial masking experiments were subject to psychoacoustic "traps". When the call always happened at the same time in the noise, the listeners started to imagine the call even if it wasn't present at all. In order to circumvent this problem, I redesigned the experiments. The setup was still very similar to the previous one outlined in Fig. 3.13. The difference was that I used two portable computers with two

identical soundcards. One computer stored 2 s long data files of the beluga vocalization at varying volume. The second computer stored a 15-minute version of the noise. The analog line-outs of the two soundcards were connected to the two microphone line-ins of the Radio Shack amplifier. They were mixed at a volume ratio of 1:1. The amplifier was then connected to the J9 projector as before.

In order to "control" the noise-to-signal ratio, I created the 15-minute noise files by taking the 2 s samples used in the early experiments and by repeating them for 15 minutes. This way, no matter when the call happened in the noise, the noise-to-signal ratio defined as the ratio of rms voltages was constant over the 2 s duration of the call.

This modification required a retraining of Aurora. At the beginning of each session, Aurora would station with the first whale trainer. I would start transmission of the 15-minute noise recording. Aurora then had to approach the J9 while it was playing continuous noise. She was extremely wary for weeks. Once she stationed properly, the second whale trainer would shout "steady" and within 30 seconds thereafter, I would play a pure beluga vocalization. In the same fashion as previously, Aurora would break away from the J9 upon detection of the call in the noise. Otherwise she would hold station until the first trainer recalled her.

Data were collected only in a standard titration. The other two methods, the reversed titration and the transmission of random volumes, did not produce different results previously, and were therefore discarded here. We did between 5 and 10 titrations per noise. Instead of changing the call volume in steps of 6 dB as before, I computed twice as many

steps 3 dB apart. For the sound pressure level this means that instead of halving it from step to step, it was divided by the square root of 2. Apart from the previously studied noises, the bubbler system noise, the propeller cavitation noise and the natural icecracking noise, I used a fourth type of noise: artificially created Gaussian white noise with a bandwidth of 22 kHz.

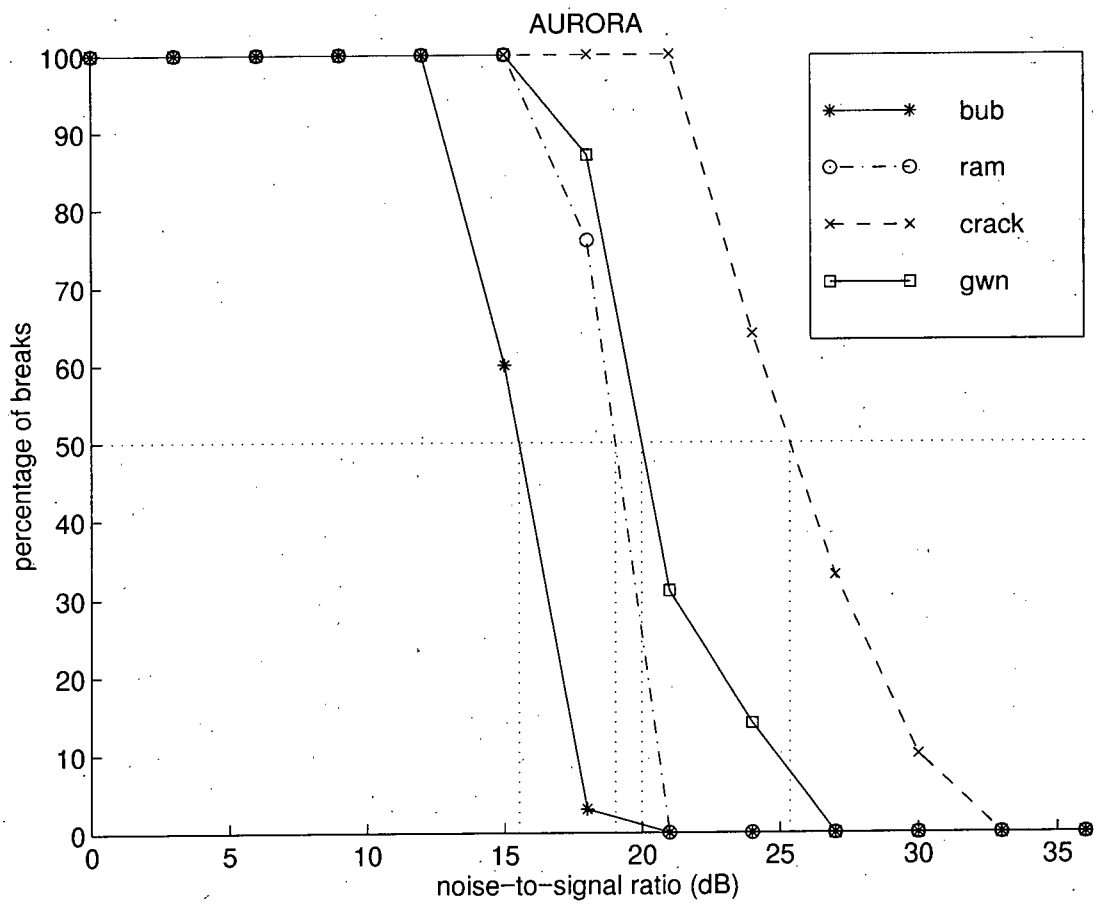


Figure 7.1: Masked hearing thresholds of the beluga whale, Aurora, in bubbler system noise, ramming noise, natural icecracking noise and Gaussian white noise. The x-axis denotes the logarithmic pressure noise-to-signal ratio in dB. The y-axis indicates in percent how often Aurora heard the call in the noise, i.e. how often she broke away from the target. Defining the hearing threshold at 50 % yields the following critical noise-to-signal ratios: 15.5 dB for bubbler system noise, 19.0 dB for ramming noise, 20.0 dB for Gaussian white noise and 25.4 dB for natural icecracking noise.

Results are shown in Fig. 7.1. In the early experiment (Fig. 3.15), every mixture was played exactly 20 times. This time, as data were only collected with the standard titration, mixtures close to the threshold were played more often than those further away. In order to illustrate how Aurora's breaking decreased with increasing noise-to-signal ratio, I plotted all the mixtures into the same plot. This way, Figs. 3.15 and 7.1 can directly be compared. Again, taking the threshold at the 50 % probability of breaking, the critical noise-to-signal ratios are: 15.5 dB (linear pressure nsr: 6) for bubbler system noise, 19.0 dB (linear 9) for ramming noise, 20.0 dB (linear 10) for Gaussian white noise and 25.4 dB (linear 19) for natural icecracking noise.

I had initially expected the artificial white noise to be the strongest masking one, as it leaves no gaps through which the whale might detect pieces of the call. However, one has to consider that the white noise has evenly distributed energy up to 22 kHz (for the sampling frequency of 44 kHz used for all the noises studied). The other three noises have most of their energy in the low-frequency range, which is also occupied by the call. They are hence more effective in masking this particular vocalization than is the artificial white noise. Natural icecracking noise, however, still is the exception, with its low masking effect due to the highly irregular temporal structure, leaving quiet gaps through which animal calls can emerge. The inclusion of artificial white noise compared to the previous experiment with only three recorded noises, has therefore nicely exemplified the interesting and important interplay between the temporal and frequency structure of both call and noise during interference.

Recalling the thresholds from the early experiment, 15.4 dB compared to 15.5 dB for bubbler system noise, the threshold is less than 1 % higher now. For ramming noise, the difference between 18.0 dB and 19.0 dB corresponds to an increase by 5 %. For icecracking noise, 29.0 dB compared to 25.4 dB shows a decrease by 12 %. Reasons for the differences can be founded in the "ghost detection" problem with the early experiment. Furthermore, the call would have occurred at different times in the noise during the modified experiment. For pulsed noises such as the ramming noise and the icecracking noise, call detection can depend on the time lag between the call pulses and the noise pulses. This time lag is random in the modified experiment, therefore the results represent an average over the possible lags. Also, Aurora, who was slightly liberal for icecracking noise during the early experiment (Fig. 3.17), might have adopted a more conservative attitude during the modified experiment. Tight experiment times at the aquarium did not permit a proper measurement of catch trials. Therefore, ROC curves cannot be computed for this experiment.

In conclusion, I consider the masked hearing thresholds from the modified experiments more reliable for two reasons. First, this experiment did not exhibit any psychoacoustic "pitfalls". Second, the resolution was twice as good as the noise-to-signal ratio was changed in steps of 3 dB rather than 6 dB.

7.1.1 Absolute Masking

In Section 1.3, I defined the *absolute masking* M to be the difference in dB between the sound pressure level of a signal at threshold in the presence of noise β_n and its threshold level in the absence of noise β_0 (Equation 1.2). In short, masking is the threshold shift in dB. The following figures (7.2, 7.3, 7.4, 7.5) illustrate this idea.

The grey shaded area is the area underneath the pure tone hearing thresholds of beluga whales. Comparing with Fig. 3.12, it consists of the mean of the 6 published audiograms AND the four hearing thresholds measured with Aurora. The top solid line is the spectrum of the noise as it was played continuously for 15 minutes in the experiment pool. The broadband sound pressure level of all the noises is 160 dB re 1 μPa at 1 m, i.e. at the location of the whale. The x-axis denotes the center frequencies of the characteristic filters in the beluga auditory system. The plotted spectra thus are 12th octave band levels. The units on the y-axis are dB re 1 μPa. The bottom solid line is the spectrum of the call at threshold in the absence of noise. It has a broadband sound pressure level of 108 dB. It is the same spectrum plotted in Fig. 3.12. The middle spectrum is the call spectrum at detection threshold in the presence of the noise as plotted in this figure. The value of absolute masking can therefore be read as the upwards shift of the masked call spectrum compared to the call spectrum. Values are 36.5 dB for bubbler system noise, 33.0 dB for ramming noise, 32.0 dB for Gaussian white noise and 26.6 dB for natural icecracking noise.

These plots offer a great opportunity to verify the critical bandwidth theory and the

equal power assumption initially proposed by [Fletcher 1940] (Section 1.3). A noise is supposed to just mask a signal when both have equal power in the corresponding critical band. The spectra plotted in this section are the mean spectra over 2 s long signals. The spectrum at the time of a pulse in the beluga vocalization is about 6 dB higher than the 2 s spectrum. From the experiments with the whale and the humans and also from the analysis of visual spectrogram discrimination, it has become obvious that for pulsed signal and noise, masking greatly depends on when the call happens in the noise. The animal does not need to hear the entire call over 2 s; it can recognize it from much shorter samples. Looking at Fig. 7.2 for bubbler noise, if the masked call spectrum is shifted up by 6 dB in order to represent the spectrum at the time of a characteristic pulse, the three major harmonics just rise above the noise spectrum. The call detection threshold hence lies at an equal power of call and noise in the three corresponding critical bands.

For propeller cavitation noise (Fig. 7.3), the situation is very similar. Again, the whale fails to discriminate the call when the major peaks drop below the noise spectrum. For Gaussian white noise (Fig. 7.5), the 800 Hz component of the masked call (after adding 6 dB for any particular pulse) surpasses the noise spectrum by about 7 dB; the 1.7 kHz component just touches the noise spectrum. I conclude that this again is an indication for the whale not recognizing the call from one frequency. It needs to hear at least one further harmonic to identify the vocalization (see Section 3.3.2).

Natural icecracking noise is of very irregular temporal structure. Most of its energy is concentrated in broadband bursts occurring about twice a second according to the

recordings available to me. The spectrum in Fig. 7.4 is the 2 s mean and has hence "lost" this information crucial for masking. As the signal played in the pool was normalized to a total sound pressure level of 160 dB, spectrum levels dropped to about 120 dB in the 500 Hz-1 kHz range in between the major bursts. This made it easy for the whale to hear the call peaks above the noise.

I would like to point out here, that ambient icecracking noise recorded in the Arctic had a similar spectrum to the one plotted in Fig. 7.4 though shifted about 60 dB down (Section 2.3). It thus falls almost entirely into the shaded area. This means that the whale will not hear it except for the high frequency components above 16 kHz. It furthermore falls below the call spectrum at threshold in the absence of noise. It is 30 dB below the call level in the major bands of the call and therefore does not contribute to masking (Chapter 8).

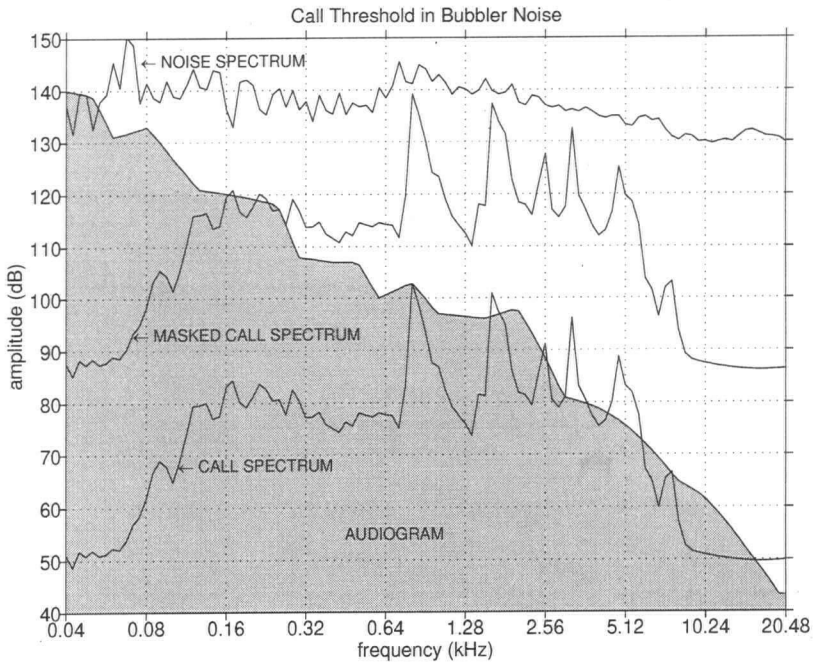


Figure 7.2: Absolute Masking of Bubbler Noise, Threshold Shift 36.5 dB.

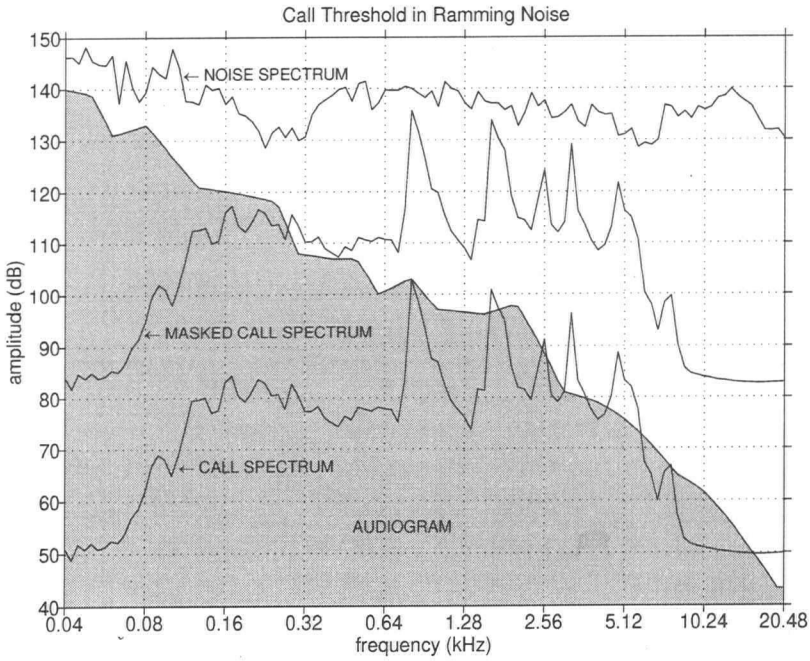


Figure 7.3: Absolute Masking of Ramming Noise, Threshold Shift 33.0 dB.

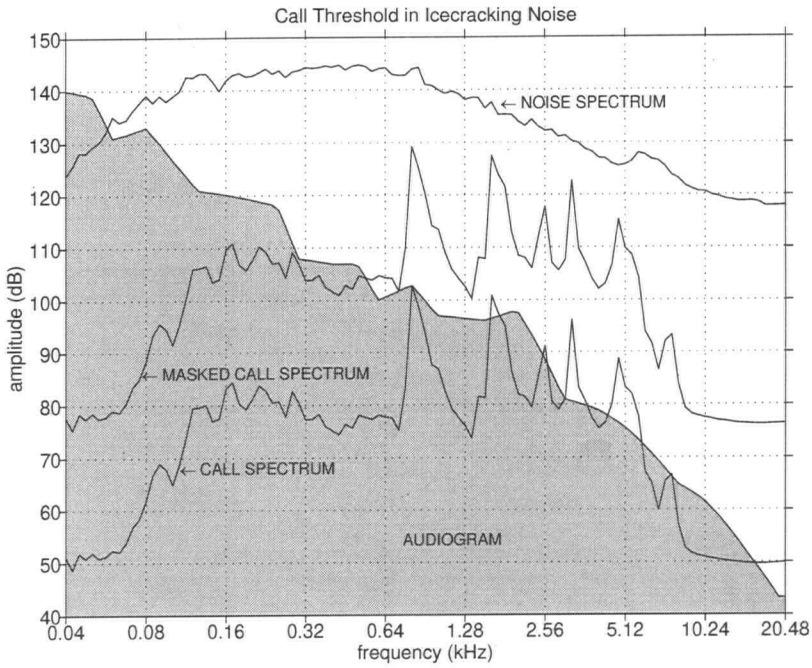


Figure 7.4: Absolute Masking of Icecracking Noise, Threshold Shift 26.6 dB.

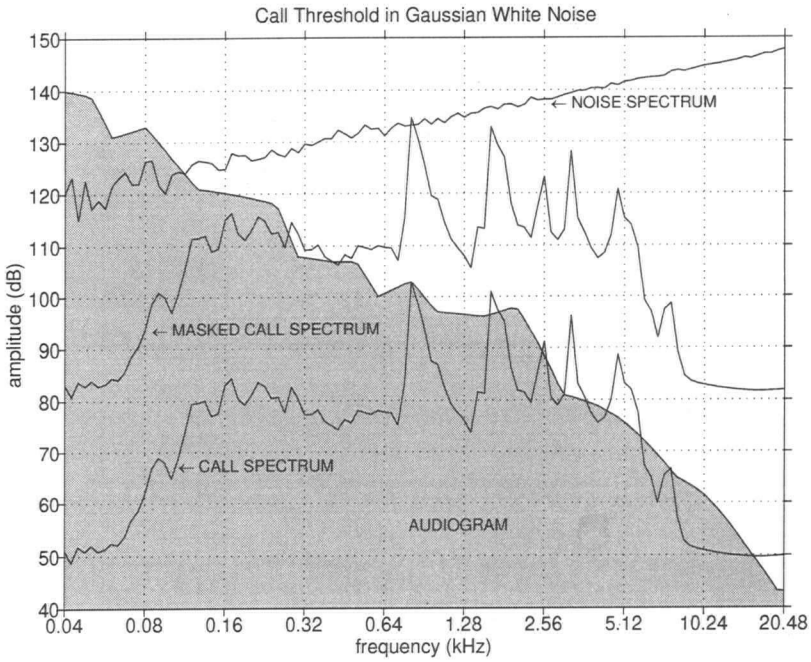


Figure 7.5: Absolute Masking of Gaussian White Noise, Threshold Shift 32.0 dB.

7.2 Masked Hearing Thresholds of Humans in Continuous Noise

The modified experiments were also carried out with humans, this time with only two subjects instead of three. We did not find any psychoacoustic traps. The only strong impression we got from the new experiments was that white noise is extremely stressful to listen to over long periods of time such as a couple of minutes. We both thought this might be because of our brain desperately trying to identify some structure in it which was bound to fail. Natural icecracking noise feels like a really soothing pleasure afterwards. Such is the beauty of nature...

Results are plotted in Figs. 7.6 and 7.7. Comparing my results to the previous ones, for bubbler noise the threshold is 20 % lower, for ramming noise 27 % lower and for icecracking noise 12 % lower. Looking at Andrew's results, the threshold for bubbler noise decreased by 26 %, for ramming noise by 22 % and for icecracking noise by 5 %. The thresholds from the modified experiment are all slightly lower than from the previous experiment for the two human subjects. Deviations in percent are larger than for the whale. Both Andrew and I said afterwards that we often thought "Oops, was this a call?". Not knowing when to expect a call during the modified experiment, we responded "No." in these cases of uncertainty. This is simply because by the time you've asked yourself whether this might have been a call or not, the 2 s reaction time is over already. We both think that during the early experiment, we would have responded "Yes." in these

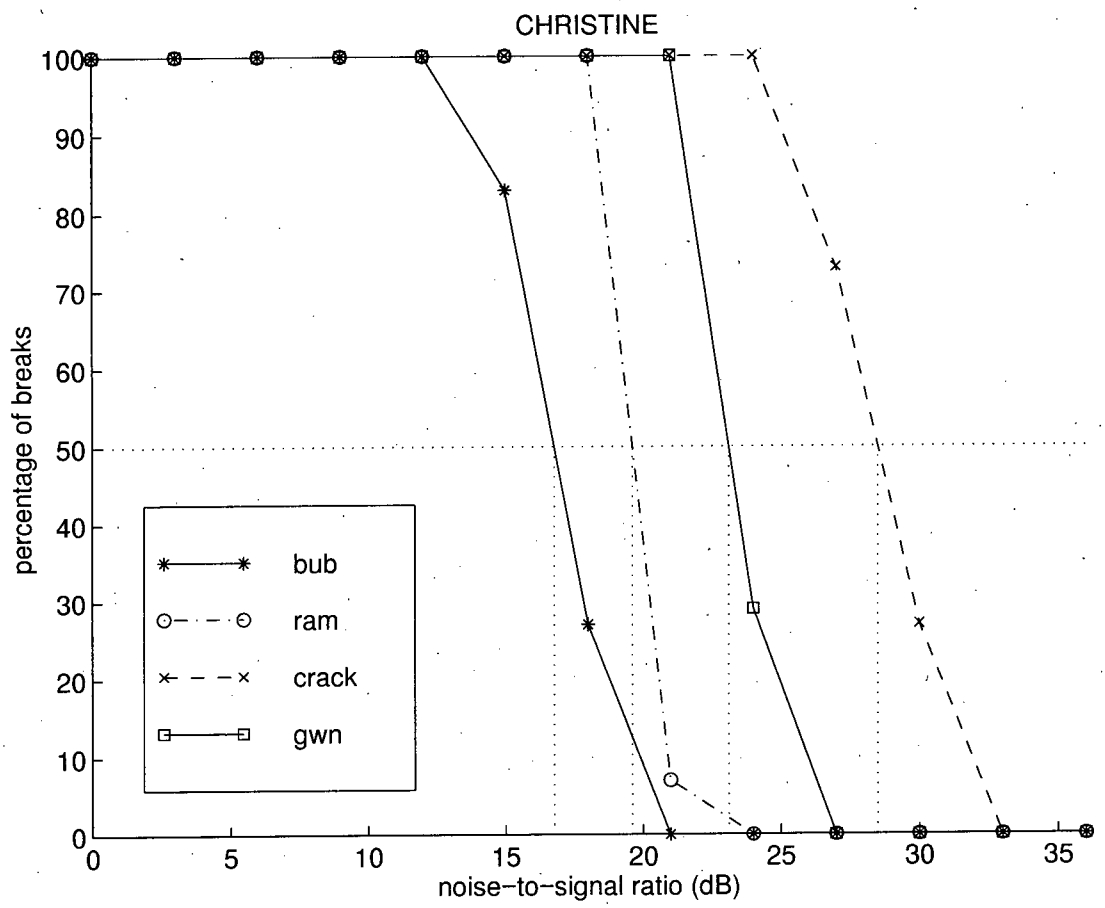


Figure 7.6: Masked hearing thresholds of myself, in bubbler system noise, ramming noise, natural icecracking noise and Gaussian white noise. Taking the hearing threshold at 50 % yields the following critical noise-to-signal ratios: 16.7 dB for bubbler noise, 19.6 dB for ramming noise, 23.1 dB for Gaussian white noise and 28.5 dB for icecracking noise.

cases of uncertainty. Our attitude thus shifted towards more conservative, which explains lower critical nsr. The individual variation between us two humans can reach up to 10 %. What all the tested subjects, the whale and the humans, have in common and which did not change from the early experiment to the modified one, is the order of noises. Bubbler system noise exhibits the strongest masking effect, followed by propeller cavitation noise. Natural icecracking noise shows the weakest interference. Gaussian white noise falls in

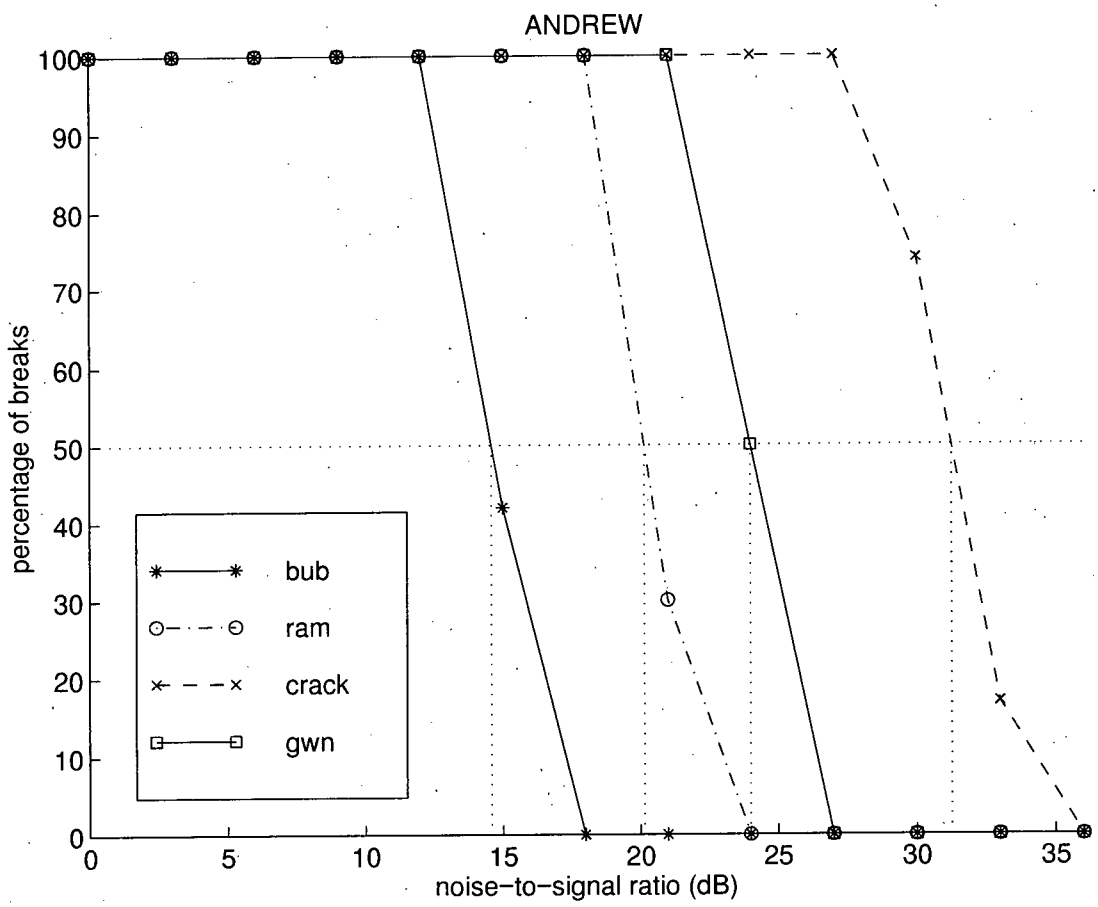


Figure 7.7: Masked hearing thresholds of Andrew, in bubbler system noise, ramming noise, natural icecracking noise and Gaussian white noise. Taking the hearing threshold at 50 % yields the following critical noise-to-signal ratios: 14.6 dB for bubbler noise, 20.2 dB for ramming noise, 24.0 dB for Gaussian white noise and 31.3 dB for icecracking noise.

between the cavitation noise and the icecracking noise.

7.3 Modified Neural Networks

The same backpropagation network as described in Section 6.7 was used to model the continuous noise experiments. However, I included slightly higher noise-to-signal ratios in

the training matrix. The maximum nsr was 1.5 compared to 1.0 previously. This resulted in slightly higher critical nsr's than previously. During the generalization phase, the neural network was presented with call/noise mixtures for bubbler, ramming, icecracking and Gaussian white noise in the same nsr's as the whale and the humans. A total of 40 test matrices was calculated by shifting the time series of the noises over the call in steps of 40 ms. This way, the "phase lag" between the call and the noise was varied. Critical nsr's were computed as the 50 % probability of call recognition. Figure 7.8 shows these

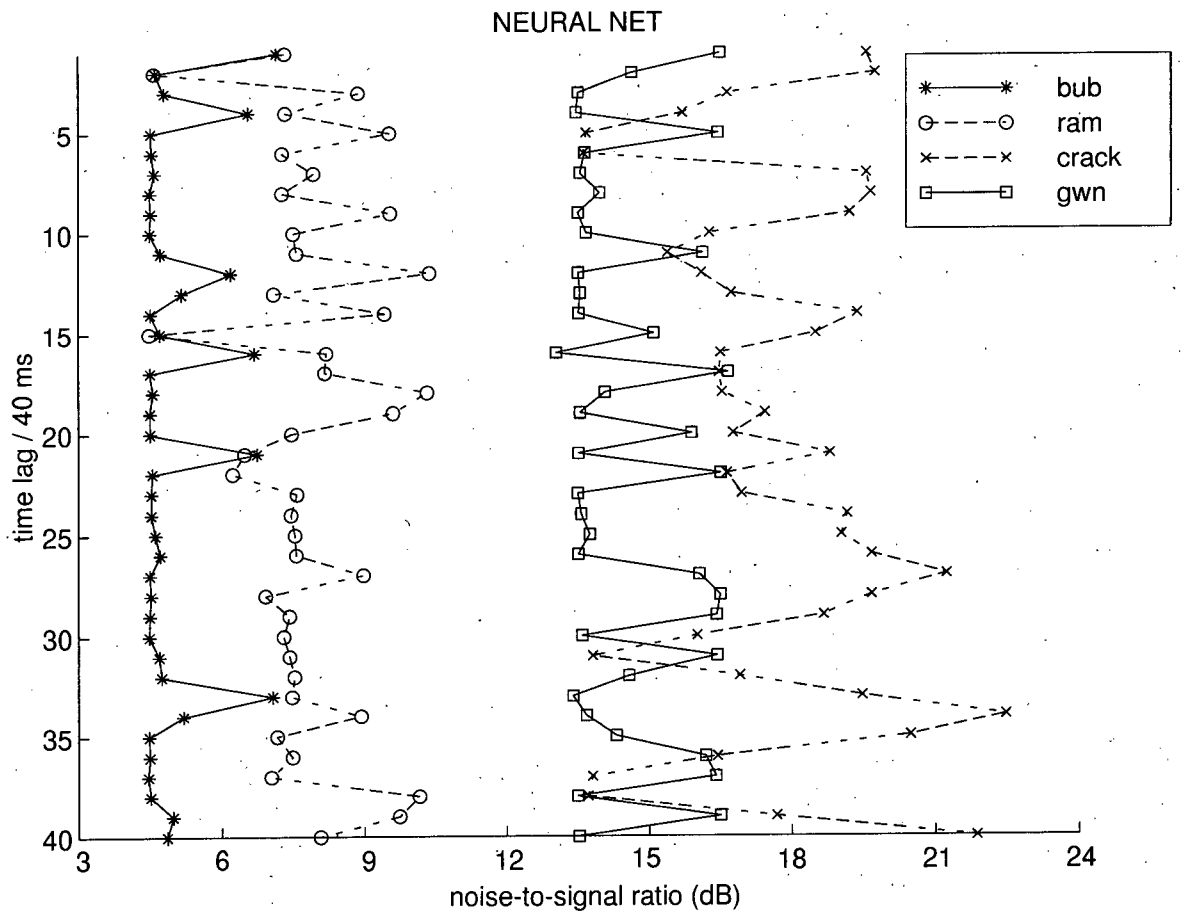


Figure 7.8: Critical noise-to-signal ratios of the neural network for bubbler system noise, ramming noise, natural icecracking noise and Gaussian white noise. The time lag between the call and the noises was changed in steps of 40 ms.

critical nsr's as they change with the phase lag. The mean values for bubbler, ramming, Gaussian white noise and icecracking noise are respectively: 4.9 dB, 7.9 dB, 14.6 dB and 17.7 dB. The standard deviations are 0.8, 1.3, 1.3 and 2.3 respectively. Bubbler system noise is the most independent of phase lag, because of its consistent, unchanging time structure. Ramming noise and white noise have slightly greater variances than bubbler noise. The variance of icecracking noise is greatest. The curves in Fig. 7.8 can be wrapped around the x-axis, i.e. the phase lag at 41 is equal to the phase lag at 1. Therefore, six distinct maxima can be identified in the icecracking noise thresholds, occurring every 6-7 time lags. The spectrograms in Figs. 2.14 and 2.13, show that the two major pulses in the icecracking noise are about 250 ms apart, which is the pulse interval of the beluga vocalization. The six maxima in the icecracking threshold plot correspond to the times when the two noise pulses lie just in between two of the call pulses.

7.4 Data Summary

Table 7.1 summarizes the results of the continuous noise experiment for Aurora, Christine, Andrew and the neural net. Fig. 7.9 shows them graphically. On average, the neural net thresholds are 8.7 dB lower than Aurora's, 11.3 dB lower than Andrew's and 10.7 dB lower than mine. These are individual and/or species specific offsets which are unknown to the neural net, but can be used to calibrate a neural network's predictions for new vocalizations and/or noises. Maximum deviations are given in percent. From these, it seems as if the neural network modeled the human data better than the whale's data.

	Bubbler	Ramming	White	Cracking	
Aurora	15.5	19.0	20.0	25.4	
Andrew	14.6	20.2	24.0	31.3	
Christine	16.7	19.6	23.1	28.5	
Neural Net	4.9	7.9	14.6	17.6	
	Distance from Neural Net				Mean
Aurora	10.6	11.2	5.4	7.7	8.7 (38 %)
Andrew	9.7	12.3	9.4	13.6	11.3 (20 %)
Christine	11.8	11.8	8.5	10.9	10.7 (21 %)

Table 7.1: Masked Thresholds (dB) of the Continuous Noise Experiment.

However, these are statistics based on only 4 data points.

In conclusion, the order of the noises is the same for the neural network, the whale and the humans. This indicates that a backpropagation network could be a suitable tool for modeling auditory masking. In order to further test the applicability of backpropagation models to auditory masking, it will be necessary to include more vocalizations and more noises, as well as more whale (and human) subjects. Thresholds can vary from individual to individual. Circumstances at the aquarium, however, did not permit training of more than one animal. Furthermore, thresholds can vary in one individual from session to session. Each nsr mixture was played only about 20 times to the whale and humans, a number which might be too small for a proper statistical analysis. Furthermore, the

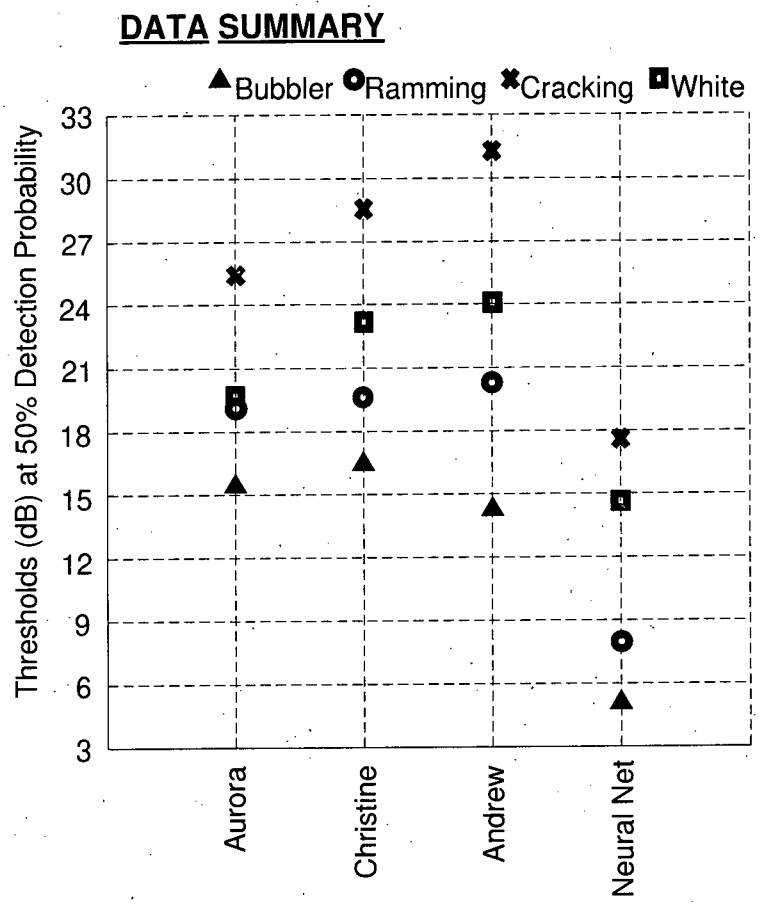


Figure 7.9: Masked Thresholds (critical noise-to-signal ratios) of the Continuous Noise Experiment.

performance of catch trials and subsequent computation of ROC curves would help in separating an individual's attitude from its sensitivity.

Chapter 8

Zones of Masking in the Field

The vocal repertoire of white whales (Delphinapterus leucas)

simmering in Cunningham Inlet, Northwest Territories.

–UNIX Spell Checker,

suggested correction to Sjare's and Smith's paper

Experiments with animals and computer models would be meaningless if they couldn't be converted to accessible parameters in the wild. So far I have studied masked hearing thresholds of a beluga vocalization in four noises. The noises were normalized in the sense that they all had the same root-mean-square amplitude, or the same broadband sound pressure level. Furthermore, their spectrum was taken as it would be 1 m away from the sound source. As pointed out many times throughout this thesis, sound spectra change along their traveling path in the wild. This means that at any location in a 3-dimensional ocean, the masking could be different depending on the exact location of the noise source,

the calling whale and the listening whale. How can one apply my measurements to the wild?

8.1 Call Recognition with Distance

Unfortunately, I couldn't find any source levels of beluga vocalizations in the literature (Section 2.4). Comparing with published source levels of other odontocete species [Richardson *et al.* 1995, Ch. 7.2], I chose source levels between 190 dB (84 W) and 140 dB re 1 μPa (0.8 mW) for the subsequent analysis.

In Fig. 8.1, the top solid curve shows the 12th octave band levels of the beluga vocalization with a broadband sound pressure level of 160 dB re 1 μPa at 1 m. This is exactly the spectrum which has been used in all the previous analysis. All the spectra below it were calculated using the SONAR Equation 2.15. They illustrate the sound spectra at increasing distance R from the source. I accounted for transmission loss due to spherical and cylindrical spreading with a transition radius equal to a water depth of $H = 300$ m. Furthermore, I included frequency dependent absorption using Equation 2.14 for the absorption coefficient under ice. Spectra are plotted up to a distance of 10 km. For such a low-frequency signal, the frequency dependent absorption becomes obvious only for distances greater than 500 m. Compare the bottom two spectra. They are about 10 dB apart at 40 Hz, but 20 dB apart for 20.48 kHz. Comparing this series of spectra with Fig. 3.12, the one positioned most closely to the measured call spectrum

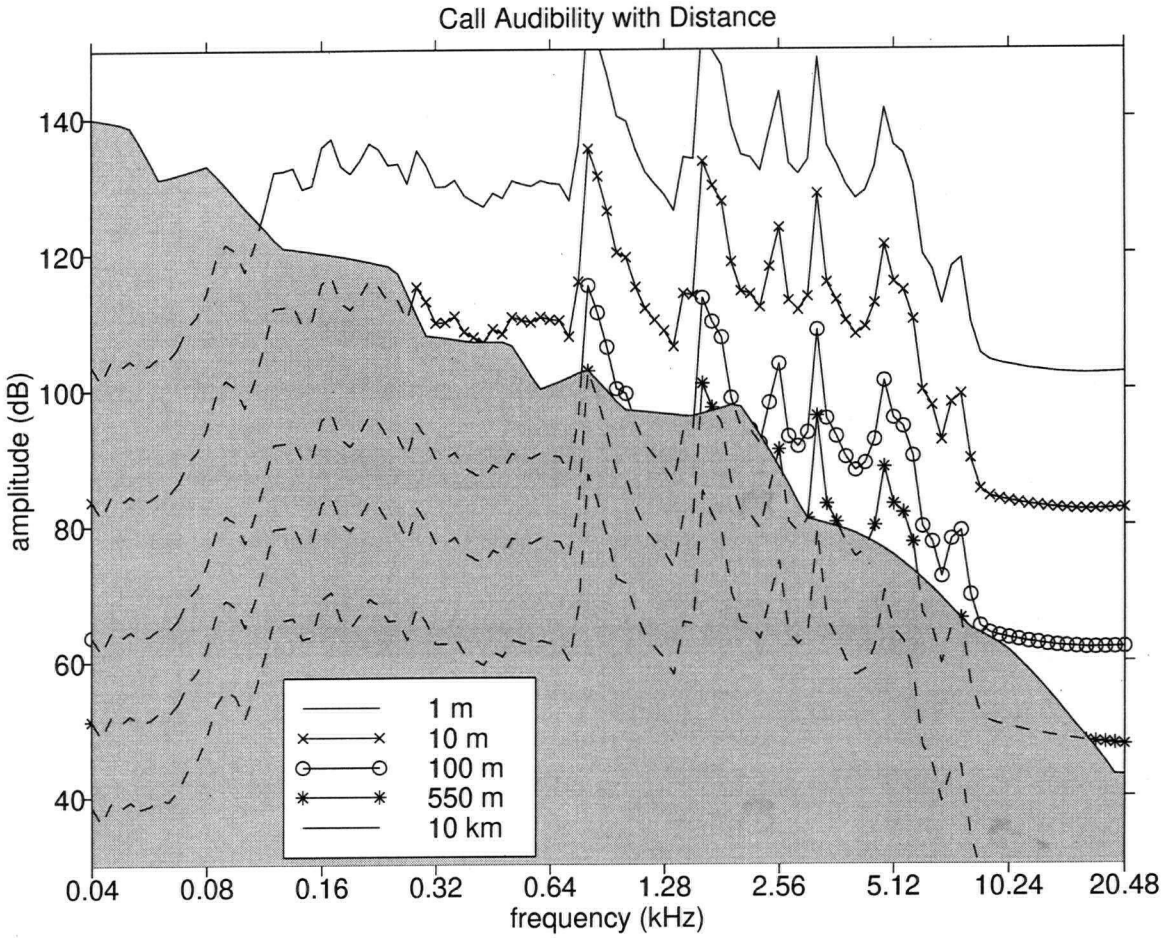


Figure 8.1: Plotted are the call spectra for increasing distance, calculated with a frequency dependent SONAR equation. The shaded area indicates the beluga audiogram. The source level is 160 dB. Spectra are 12th octave band levels in dB re 1 μPa. The last recognizable call corresponds to a distance of 550 m.

at threshold corresponds to 550 m. This is the maximum distance over which the beluga vocalization can be recognized if it has a source level of 160 dB re 1 μPa. I particularly want to stress that this is a "recognition" distance rather than an "audibility" distance, because the whale should be able to detect the high frequency call components over much larger ranges. Remember that it was shown that the whale (trained to break away from the target upon detection of the beluga call) stopped breaking while all major harmonics

of the call still "stood above" the audiogram (Section 3.3.2).

The *effective call level* at 550 m is the total energy of this call which lies above the audiogram, i.e. it is the area underneath the spectrum curve and above the audiogram. As the plotted curves are band levels already, the area can easily be calculated by adding the spectrum levels poking out above the audiogram and by subtracting the audiogram levels. This must be done on a linear amplitude (intensity) scale, one cannot integrate decibel units. Only after summation can one convert back to decibels. The effective call level at 550 m hence is 101.6 dB re 1 μPa.

Fig. 8.2 shows the sound pressure level and the effective sound pressure level of the beluga call as a function of distance. The source level is 160 dB re 1 μPa at 1 m. The call spectrum was propagated according to the SONAR equation. At every distance sample, the SPL was calculated as the total area underneath the spectrum. The effective SPL was calculated as the "white area" underneath the spectrum and above the audiogram. The SPL thus shows up as a smooth logarithmic drop-off; the effective SPL has "little dents" whenever one of the major call harmonics falls underneath the audiogram. As the area underneath both curves tends towards zero for increasing distance, the logarithm, i.e. the amplitude in dB, approaches minus infinity. The reason why the effective SPL curve suddenly stops is that the next distance sample already got assigned a value of $-\infty$ which wasn't plotted. Knowing from the experiments, that an effective SPL of 101.6 dB is the quietest call which Aurora could detect, one identifies again a maximum recognition distance of 550 m.

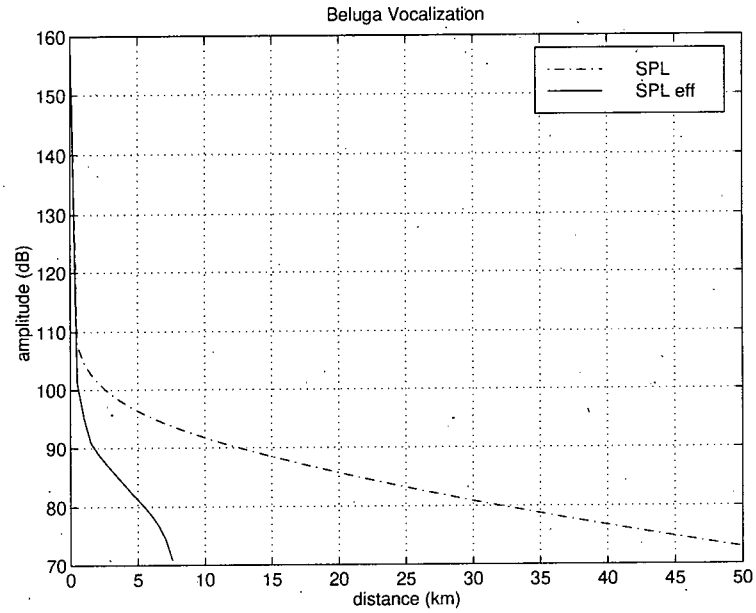


Figure 8.2: Decrease of Sound Pressure Level and Effective SPL with Distance for the Beluga Call, Source Level 160 dB.

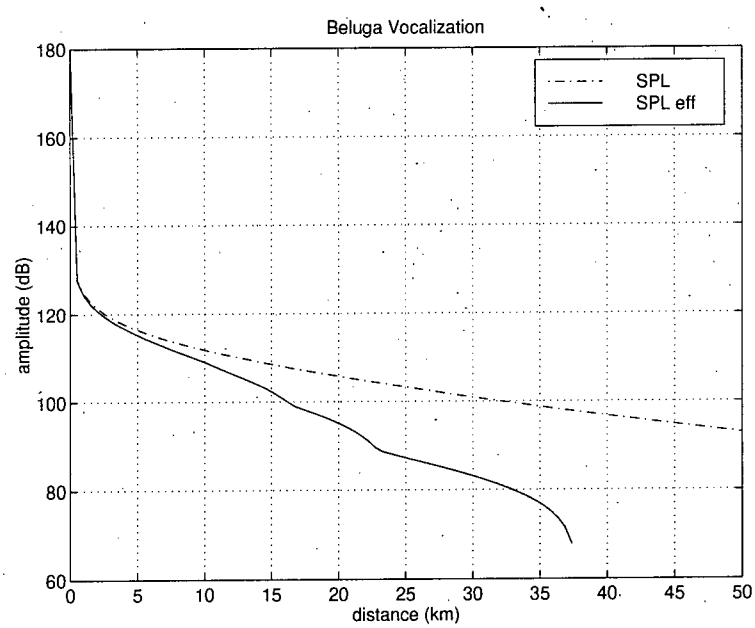


Figure 8.3: Decrease of Sound Pressure Level and Effective SPL with Distance for the Beluga Call, Source Level 180 dB.

Fig. 8.3 shows the same curves for a call at 180 dB source level. This time, the maximum recognition distance at an effective SPL of 101.6 dB is 16 km.

The SONAR equation used for sound propagation in this chapter is a very simple approach to SPL estimation. In particular for an animal vocalization of short duration, spreading will in general be spherical along its entire propagation path. This will lead to lower received levels than calculated with the SONAR equation and hence shorter recognition distances than quoted above. For continuous sound sources such as ship noise, the conversion from spherical to cylindrical spreading is more justified as cylindrical waves can be regarded as a superposition of multiple ray paths in a waveguide. In any case, if the discussion presented in this chapter is transfered to a particular location in the wild, sound propagation parameters such as water depth, bottom composition and topography, surface cover (waves or ice) should always be obtained and incorporated in an appropriate sound propagation model such as a parabolic equation, for example (Section 2.1.3). This will account for location specific effects such as scattering, absorption, diffraction and sound channeling.

8.2 Noise Audibility with Distance

For the noises the minimum SPL at which the animal can still recognize the noise was not measured, because it plays no role in masking. I therefore talk about audibility distances rather than recognition distances. Audibility is assumed to stop when the effective noise level approaches 0.

For bubbler noise (Fig. 8.4), the maximum audibility distance hence becomes 42 km. One can also assign a *maximum masking distance* to each noise. Recalling Fig. 7.2, which plotted the noise spectrum and the call spectrum at masked hearing threshold, I define the effective noise-to-signal ratio as the effective noise level divided by the effective call level. Thinking in terms of area, the effective noise-to-signal ratio is the "white area" underneath the noise spectrum and above the audiogram divided by the "white area" underneath the masked call spectrum and the audiogram. For bubbler noise, this effective nsr becomes 15.4 dB. Knowing that the call is only audible up to an effective sound pressure level of 101.6 dB, the bubbler noise will mask the call as long as it is 15.4 dB louder. $101.6 + 15.4 = 117.0$, the maximum masking distance of bubbler noise is 15 km.

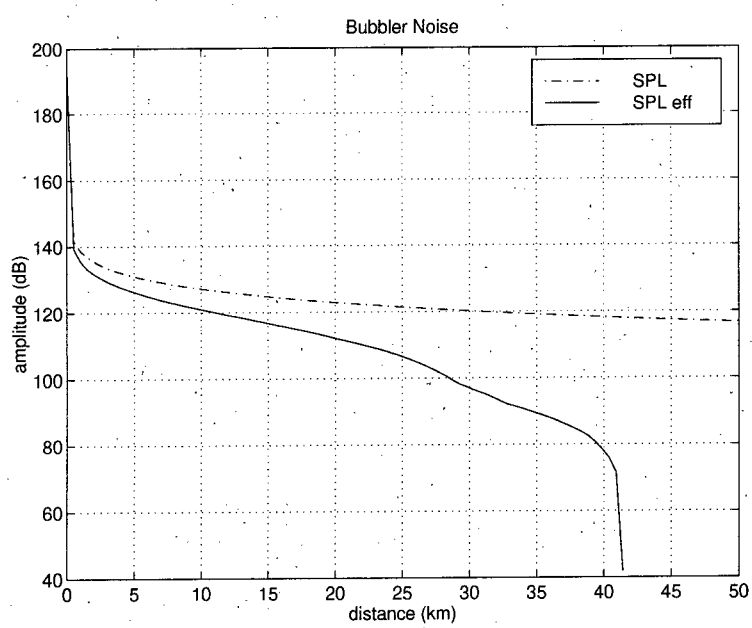


Figure 8.4: Decrease of Sound Pressure Level and Effective SPL with Distance for Bubbler Noise, Source Level 194 dB (211 W).

According to Fig. 8.5, ramming noise with a source level of 203 dB (1680 W) is audible over 60 km. The effective noise-to-signal ratio can be calculated from Fig. 7.3. It is 18.9 dB. $101.6 + 18.9 = 120.5$, the maximum masking distance of ramming noise is where the effective SPL drops below 120.5 dB, i.e. at 22 km.

Looking at Fig. 8.6, icecracking noise resulting from a single breaking event with a source level of 147 dB (4 mW) is audible over merely 1.4 km. The effective noise-to-signal ratio as calculated from Fig. 7.4 is 25.5 dB. $101.6 + 25.5 = 127.1$, the maximum masking distance of icecracking noise becomes 8 m, not km. This result is amazing. Only if a listening beluga whale is within 8 m of a natural icecracking event, can the resulting noise mask a vocalization uttered by a second whale. This is the maximum zone of masking. It is essential to understand that this maximum radius is independent of the source level of the vocalization and independent of the location of the speaker! It only depends on the no-noise detection threshold of the pure vocalization and the critical noise-to-signal ratio.

Ambient Arctic noise levels based on thermal or pressure icecracking depend greatly on time and location. Mean source levels of about 87 dB were measured (Section 2.3). This type of ambient noise does therefore not mask low frequency vocalizations such as the beluga call studied here.

For Gaussian white noise with a source level of 190 dB (84 W), Fig. 8.7 shows a maximum radius of audibility of 41 km. From Fig. 7.5, the effective noise-to-signal ratio is 20.2 dB. Adding this to the call level at threshold in the absence of noise, yields a minimum masking level for this type of noise of 121.8 dB. This corresponds to a distance

of 5 km. For Gaussian white noise with a source level of only 160 dB (84 mW), the maximum radius of audibility is 13 km, the maximum radius of masking is only 8 m.

Again, I want to emphasize that a location specific sound propagation model should always be used for an analysis of masking in the wild. Due to the reflecting properties of the ice cover, sounds originating at the surface in the Arctic (such as icecracking noise and ship noises) have a dipole character. Most of the energy is radiated downwards. Therefore, noise detectability and masking will be increased in the vertical and decreased in the horizontal direction.

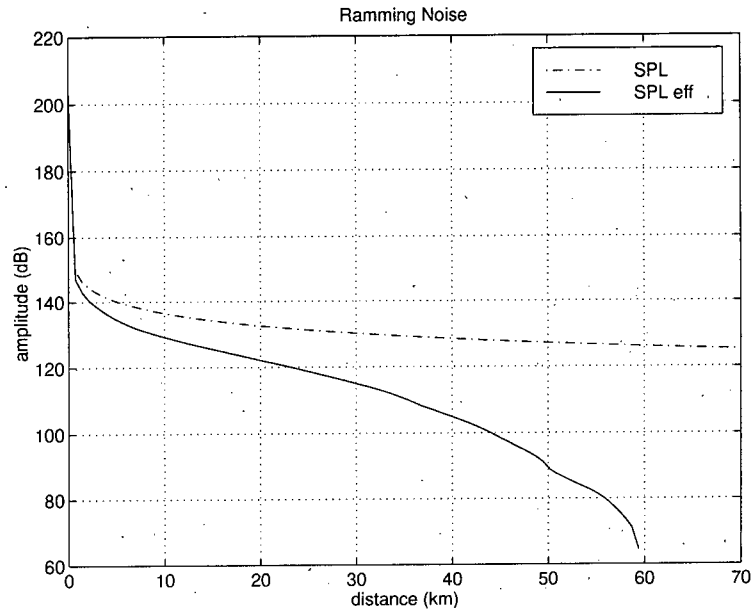


Figure 8.5: Decrease of Sound Pressure Level and Effective SPL with Distance for Ramming Noise, Source Level 203 dB (1680 W).

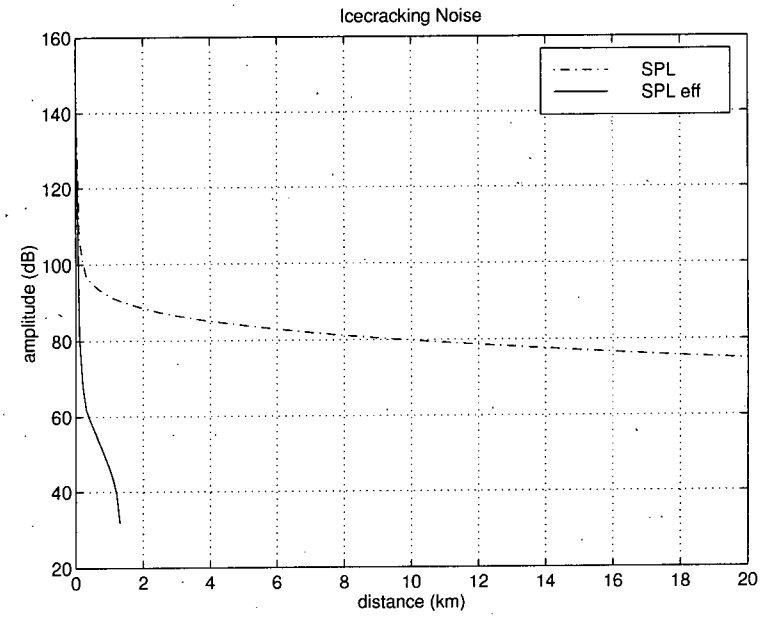


Figure 8.6: Decrease of Sound Pressure Level and Effective SPL with Distance for Icecracking Noise, Source Level 147 dB (4 mW).

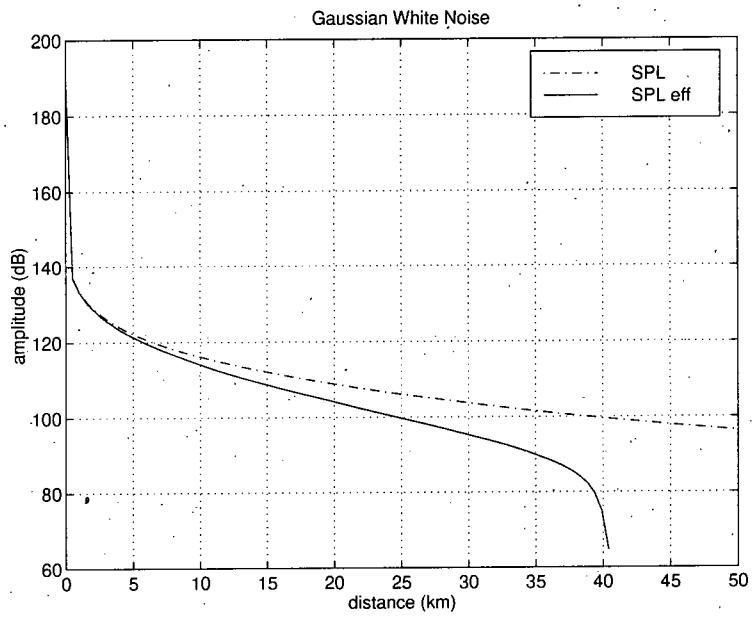


Figure 8.7: Decrease of Sound Pressure Level and Effective SPL with Distance for Gaussian White Noise, Source Level 190 dB (84 W).

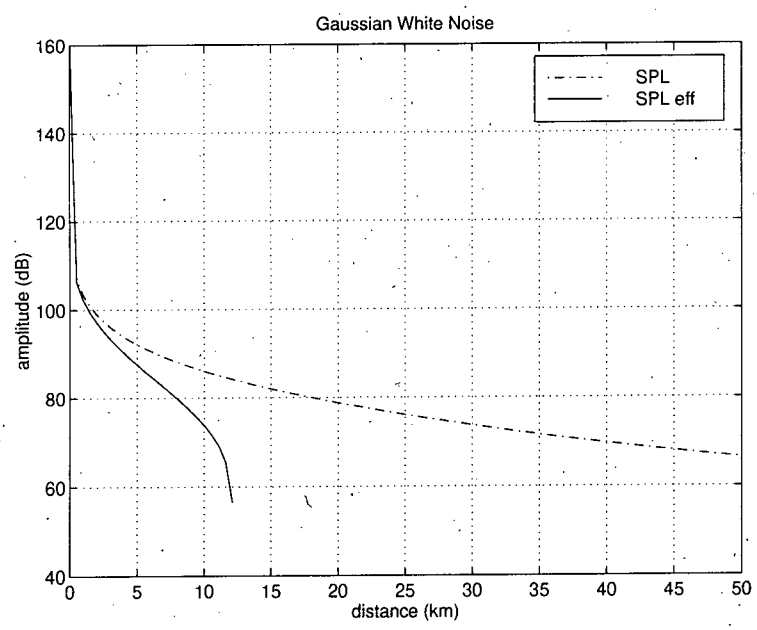


Figure 8.8: Decrease of Sound Pressure Level and Effective SPL with Distance for Gaussian White Noise, Source Level 160 dB (84 mW).

8.3 Maskograms

8.3.1 What is a Maskogram?

A *maskogram* is a colourplot which I designed to illustrate zones of masking as a function of two distances: the distance between a noise source and a listening whale and the distance between the noise source and a calling whale. In a three dimensional ocean, the distance between the calling whale and the listening whale is the vector subtraction of the other two distances. A maskogram contains a lot of information which is summarized in the following list. Fig. 8.9 shows a model maskogram.

- The noise source, illustrated here as a ship, is located along the y-axis of the plot. Its location is fixed; it does not move.
- A calling whale is located along the diagonal. Its distance from the ship forms the units on the y-axis. At the origin, the calling whale sits right at the same place as the ship. Moving up along the y-axis, the distance between the ship and the whale increases. As the plot is square, the ship-caller distance can also be read as the horizontal distance between the y-axis and the diagonal.
- The x-axis denotes the distance between the ship and a listening whale.
- The distance between the calling whale and the listening whale is the length of the horizontal connecting the diagonal with the current location of the listener.

- Sound pressure levels are represented by rainbow colours. The red end corresponds to the loudest signals, the purple end to the quietest.

- The effective SPL of the vocalization uttered by the caller is plotted wherever it is audible in the absence of noise. Its peak amplitude forms the diagonal. The effective SPL drops off fairly fast (see Figs. 8.2 and 8.3) in the direction perpendicular to the diagonal. The limiting purple lines (labelled (1) in Fig. 8.9) on the left and the right of the diagonal correspond to an effective SPL of 101.6 dB, which was the measured call recognition threshold in the absence of noise.

- The maximum call detection distance (labelled (2) in Fig. 8.9) is the horizontal length between the red diagonal and the parallel, darkest purple line on either side of the diagonal.

- The effective SPL of the ship is plotted in the background. It decreases in vertical lines from left to right. The last plotted line is the last one with a SPL $> -\infty$. Depending on the chosen sampling interval, this value changes. The rule is that the last line before the white area should always be read as an effective noise level of $-\infty$. This line marks the maximum noise detection distance on the x-axis (label (3)).

- In the white area, the listening whale does not hear either the noise nor the calling whale.

- In the black area, the noise masks the call.

Imagine a listening whale swimming across the maskogram from right to left. As long as it is in the white area it does not hear anything, neither ship nor call. As soon as it reaches the vertical purple line, it hears a far-away ship. Proceeding towards the left, the ship noise becomes louder. Suddenly, when the whale hits the first diagonal purple line, it recognizes a calling conspecific (another beluga). I particularly say "recognizes" because the first purple line corresponds to the maximum recognition distance, not audition distance (Section 8.1). Only the purple line of the noise level represents the maximum audition range.

Continuing towards the left, the two whales can give each other a belly rub on the diagonal. Even further to the left, the call becomes quieter. The ship noise has become louder and louder all the way. When the listening whale hits the black area, the noise masks the call, i.e. the whale only hears the ship.

At the height of the black arrow I drew, the masking area stops and the whale enters red ship area. The whale does not "notice" this transition. Its sensation is that of steadily increasing ship noise and no audible vocalization as soon as it hits the black area from the right. I defined the zone of masking as the area where the call would be recognizable in the absence of noise but not in the presence of noise. Thus, behind the black masking zone, the call wouldn't be detectable in any case and only the SPL of the ship is plotted.

If the whale travels across the maskogram at a lower horizontal, it encounters two zones of masking, with detectability of the vocalization in between.

Although the maskogram is only shown in the first quadrant of the coordinate system, it can easily be extended into the other three quadrants. The upper left quadrant with negative x and positive y would include the case of the ship being between the listening and the talking whale. The lower two quadrants can be obtained from the upper two by symmetry about the origin.

As a last point, the maximum range of masking by a noise source can be read as the furthest ship-listener distance to which the tips of the black masking areas reach (label (4)).

8.3.2 Maskograms for the Noises studied

Bubbler System Noise

Fig. 8.10 shows four maskograms for bubbler system noise. The maximum effective SPL of the noise is 194 dB re 1 μPa at 1 m. It is the same in each of the four plots. The SPL of the call changes from plot to plot. The colourscale was always clipped at the maximum call level in order to enhance the call structures.

In the first plot, the call has a maximum effective SPL of 190 dB. Comparing with Fig. 8.4, the maximum audibility distance of bubbler system noise is 42 km, that's where the white area starts. The very loud call is detectable over 35 km. This is the intersection of the first purple call line with the x-axis. Masking only exists in between the ship and the calling whale. To the right of the calling whale, bubbler system noise is not loud

enough to produce a second zone of masking. The maximum range of masking for bubbler system noise corresponds to the outmost black tip at a ship-caller distance of 15 km.

In the second plot, the call has a source level of 180 dB. Its maximum range of detection is 16 km as measured in Fig. 8.3 already. There is again only one zone of masking with a maximum range of 15 km.

In the third plot, the call source level drops to 160 dB. The call is detectable only over 550 m (Fig. 8.2). It thus appears as a very narrow line. At this level, two zones of masking exist, one to the left and one to the right of the calling whale.

For an even quieter call with a source level of 150 dB as depicted in the fourth plot, the detection range drops to about 100 m. The zones of masking have become very narrow. It is only because of the chosen pixel number of 100x100, that the two zones don't reach their full distance of 15 km. For the same reason, the bottom zone seems to stop sooner than the top zone. As will become apparent in the subsequent pictures, the two zones converge differently but to the same maximum masking range.

Propeller Cavitation Noise

Propeller cavitation noise (Fig. 8.11) has a source level of 203 dB. The maximum audibility distance is at 60 km, which was already measured in Fig. 8.5. The first plot shows the same call as in the first picture of the bubbler maskograms. The maximum recognition distance is 35 km, where the first purple call line intersects the x-axis. Masking exists only between the ship and the calling whale. The maximum range of masking this propeller has is 22 km.

The second plot illustrates the different shapes of the two masking zones. Considering a listening whale traveling along a horizontal line, the masking zone is narrower to the right of the calling whale than it is to the left. Thus the zone in between, where the listener can successfully discriminate call from noise, is wider to the right side of the diagonal than to the left. The upper masking zone converges very gradually, the bottom zone very rapidly to the same maximum masking distance of 22 km.

In the third picture, the zone of call detectability and hence the two zones of masking have become very narrow, because the source level of the call is "only" 160 dB. The fourth picture is a magnification of the area between 20 and 22 km, and corresponds to a call source level of 150 dB.

As can be seen from these pictures, the existence of one or two zones of masking only depends on the source level of the call. If the maximum range of call detectability is greater than the maximum range of masking, then only one zone of masking exists, and this one lies in between the noise source and the calling whale. If, on the other hand, the maximum range over which the call is recognizable is smaller than the maximum range of masking, two zones will arise.

Natural Icecracking Noise

As learned from Fig. 8.6 already, the range of audibility of a single icecracking event with a source level of 147 dB is very short, 1.4 km. In the first of the four plots of the icecracking maskograms in Fig. 8.12, the icecracking noise thus shows up as a narrow band close to the y-axis. A call of 180 dB source level is quite "overpowering" and successfully

detectable in most of the area. The noise isn't even audible on the right side of the call. Masking appears as a narrow line of maximally 8 m width at the left end of the picture.

In the second plot, the call level has dropped to 160 dB. The noise is now audible on either side of the calling whale. Masking again, is just a narrow line on the y-axis. The third plot has the same parameters, the call volume didn't change. This plot is merely a magnification of the bottom left corner in the second picture.

In the last maskogram, the call has a source level of 140 dB. There is still only one zone of masking with a maximum range of 8 m.

The interesting feature this set of maskograms exhibits, is a vertical shift of the masking zone along the y-axis. This is not just a matter of pixel size. In the first picture, the call is so loud that there is no masking at the origin. The listening whale has to be in between the cracking ice and the calling whale and a minimum of 2 km away from the caller for masking to take place. The quieter the caller talks, the closer the masking zone gets to the origin. In the second and third plot, the listener has to be about 100 m away from the caller for masking to occur; in the fourth plot, the quiet 140 dB vocalization is masked at a distance of 10 m.

Gaussian White Noise

The last set of maskograms (Fig. 8.13) shows artificially created, Gaussian white noise with a source level of 190 dB in the first two plots, and 160 dB in the last two plots. For 190 dB, maximum noise audibility occurs at 41 km (see Fig. 8.7). The maximum range

of masking is 5 km. There is one zone of masking for a call of 180 dB source level, and there are two zones for a call of 160 dB.

Referring to Fig. 8.8, for Gaussian white noise with a source level of 160 dB, noise audibility stops at 13 km; the maximum range of masking is about 8 m. The maximum range of masking thus is the same as for natural icecracking noise. However, Gaussian white noise has a larger range of audibility with 13 km compared to 1.4 km for icecracking. The reason for this is manifold. First of all, the two noises have a source level difference of 13 dB. Therefore, the effective SPL of the icecracking noise reaches $-\infty$ faster than the level of the white noise. If the icecracking noise was louder, its zone of audibility would be larger, but also its zone of masking would increase. The more important reason is that the white noise has much more energy at high frequencies than has the icecracking noise. The beluga whale's hearing is more sensitive at high frequencies than at low frequencies, therefore the effective SPL of the white noise is greater than that of the icecracking noise. Frequency dependent absorption only starts to diminish this advantage over traveling ranges of a few km and more. Last but not least, the reason why two structurally very different noises can have the same distance of masking is founded in the interplay of frequency characteristics and time characteristics in masking. The masking zone of icecracking noise is so small, because of this noise's irregular time structure. As pointed out earlier, the whale manages to recognize the call from short pulses which emerge through quieter gaps in the noise field. The masking zone of Gaussian white noise, on the other hand, is so small, because it has relatively little energy in the frequency bands

occupied by the call. In a sense, the white noise "wastes" its energy on high frequencies where the call cannot be masked.

A new feature in this set of maskograms is the emerging gap in masking at the bottom of plot 4, between 1 and 20 m. This picture could be a fifth plot in the icecracking maskogram series. If the call volume of the fourth plot in Fig. 8.12 decreased even further, the masking zone which had been shifted away from the origin along the y-axis, would reach 0 again. For even lower call source level, a second zone of masking would emerge, however, not starting from the origin but offset horizontally along the x-axis. This phenomenon that the masking zones do not touch the origin, only occurs if the source level of the call is greater than the source level of the noise plus the critical noise-to-signal ratio. It therefore does not occur for the loud ship noises.

Data Summary

Table 8.1 summarizes the results of the maskogram analysis. For the beluga vocalization, the listed radius of audibility is actually the maximum radius of recognition.

When all noises were played at equal source level during the acoustic experiments, the order of the noises from strongest masking to weakest masking was: bubbler noise - ramming noise - icecracking noise. In the wild, at their characteristic source levels, however, ramming noise masks furthest, followed by bubbler noise, then icecracking noise.

	Source Level	Source Power	max. radius of audibility	max. radius of masking
Beluga Vocalization	180 dB	8 W	16 km	-
Beluga Vocalization	160 dB	84 mW	550 m	-
Bubbler Noise	194 dB	211 W	42 km	15 km
Ramming Noise	203 dB	1680 W	60 km	22 km
Icecracking Noise	147 dB	4 mW	1.4 km	8 m
White Noise	190 dB	84 W	41 km	5 km
White Noise	160 dB	84 mW	13 km	8 m

Table 8.1: Summary of the Maskogram Analysis.

8.4 Discussion

There are a few important findings based on the maskogram analysis, which will be pointed out in this section.

First, noise sources in the ocean have a maximum range of masking which is independent of the source level of the masked vocalization and independent of the distance between the noise source and the calling whale.

Second, the zone of masking is not equal to the zone of audibility. Prior to this study, the masking of complex animal vocalizations by man-made noise had never been studied experimentally. Zones of masking had only been dealt with in a hypothetical manner.

The general assumption was that the maximum radius of audibility of a noise source was equal to the maximum radius where it might mask other sounds. [Richardson *et al.* 1995, Ch. 10.5] gives an excellent summary on "theoretical zones of masking". I wish this book had been around two years earlier. This general assumption is only true for pure tones or to be more accurate, for noise and signal occupying the same critical band. Furthermore, both have to be temporally continuous. For real animal vocalizations and real noises, the situation becomes very complex. Recalling Fig. 7.5, Gaussian white noise is continuous in the time and frequency domain. It has increasing band levels with frequency, and will thus be audible over long ranges, because of its energy located at high frequencies where the beluga auditory system is most sensitive. The range of masking is much shorter, because its band levels at the low frequencies of the call are lower to start with and quickly drop below the call levels.

Icecracking noise is reasonably continuous in the frequency domain, though irregular in the time domain. The whale will hear the noise until the peaks of the bursts drop below the audiogram. The intermittent quieter bits of icecracking sound will disappear much sooner and hence set a maximum range to masking which is smaller than the audibility range.

Therefore, the equal-power assumption for masking is valid for call and noise occupying the same critical bands and being temporally continuous. The zone of masking will be equal to the zone of audibility of the noise in this case. If either of these criteria is invalidated, the range of masking will be shorter than the range of audibility. The

exception is the case of continuous noise and a pulsed call occupying the same critical band(s); the two zones will be equal again.

In a maskogram, if the zone of masking was equal to the zone of audibility, the tips of the black masking areas would stretch out up to the intersection of the last purple call lines with the last purple noise line. Quite obviously, this is not the case for any of the noises and the particular vocalization studied here.

Third, in hypothetical discussions of masking in the field, masking has previously been considered limited by ambient noise in two ways. It would stop if the call dropped below the ambient noise or if the man-made noise plus the ambient noise dropped below the call. The maskogram software I wrote includes these criteria in its computation of the black zones. However, it turns out that without the two criteria, the masking zones are the same. Ambient Arctic noise, if it is predominantly icecracking noise, therefore plays no role in masking. This was indicated in Section 7.1.1 already. The mean spectrum of ambient Arctic noise based on icecracking has a broadband sound pressure level of 60 dB less than the single event studied in the maskograms of Fig. 8.12. Only the high-frequency components above 16 kHz are audible to belugas. At the major frequencies of the call studied, the ambient noise spectrum lies about 30 dB below the audiogram and hence below the call level at hearing threshold in the absence of noise. It therefore cannot mask the call. We know that a small contribution to masking can be related to interband effects compared to intraband masking for signal and noise in the same critical band. In this sense, however, high-frequency noise is less efficient at masking low-frequency signals

than vice versa (Section 1.3). The bottom line is that the call falls below its audibility threshold before it falls below the ambient noise. As the noises studied here have to be a critical noise-to-signal ratio (of minimally 16 dB for bubbler noise) louder than the call at threshold in the absence of noise, ambient noise will not add to their masking effect. This rules out both ambient noise criteria previously considered in masking analysis.

I want to emphasize that ambient Arctic noise can be based on a variety of sources, such as wind and waves, rain, and sounds from other animals. None of these were studied. My analysis is limited to icecracking noise. Ambient noise of other sources might well contribute to masking. Apart from being location-dependent, ambient Arctic noise also exhibits great variability with time of day and season [Greene 1981, Verrall 1981].

All the beluga vocalizations I looked at had their energy below 10 kHz. As the recordings had been sampled at 44 kHz, frequencies up to 22 kHz would have shown up. [Ford 1977] measured beluga communication signals between 400 Hz and 12 kHz. The frequency range of odontocete communication signals has generally been considered to have an upper limit of 20 kHz [Richardson *et al.* 1995, Ch. 7]. Higher frequencies would only be used for echolocation. Many of these recordings were hardware-limited to low frequencies. Recent work with broadband recording equipment has shown that some species communicate up to 50 kHz [Lammers and Au 1996]. If this is the case, ambient icecracking noise has the potential to mask if the call is close to its no-noise threshold.

Fourth, as outlined in Section 1.3, directional hearing abilities of belugas will not improve the signal-to-noise ratio at such low frequencies as occupied by the call I studied.

The maskogram software though can easily be modified to include directional hearing for higher frequencies if necessary. The SONAR Equation 2.15 forms the basis of the maskogram. The sound pressure level of a noise source at the location of a listening whale is:

$$SPL = SL - TL_{spread} - TL_{abs} \quad .$$

For an omnidirectional source but a directional receiver, the sound pressure level perceived by the whale can be calculated as:

$$SPL = (SL - DI) - TL_{spread} - TL_{abs} \quad . \tag{8.1}$$

The *directivity index* DI (Equation 1.1) is the logarithmic ratio of the SPL received by an omnidirectional receiver and the SPL received by a directional receiver. It is thus greater than 0. A "complete" SONAR equation giving the signal-to-noise ratio at a whale listening to high-frequency sound is:

$$SNR = (SL - TL_{spread} - TL_{abs})_C - (SL - TL_{spread} - TL_{abs} - DI)_N - (NL - DI) \quad . \tag{8.2}$$

The first term with subscript C is the received sound pressure level of the animal call. For simplicity, I assumed that the source level SL of the call is not an omnidirectional source level but the level in a focussed beam which would "match" the directivity of the receiving system such that all of the energy is received. The second term is the received source level of the assumed omnidirectional man-made noise corrected for the directivity of the receiving system. The third term is the amount of ambient Arctic noise received by the whale, with NL being the source level of this noise.

Taking receiving directivity indices measured for bottlenose dolphins at high frequencies [Au and Moore 1984], and extrapolating to low frequencies, yields no directivity below 10 kHz. Masking at low frequencies therefore does not depend on the exact location of the noise source, the calling and the listening whale in vector space but only on their relative distance. The ocean is a 3-dimensional space; three points, however, always fall into a (2-dimensional) plane. If directional hearing played a role for the signals I studied, the maskograms would have to be read as ship, caller and listener, all lying on one straight line. As directional hearing does not exist at the frequencies involved, the maskograms plotted in this section can be read as ship, caller and listener lying anywhere in a plane.

Finally, I would like to point out again that if masking is studied at a particular location in the wild, the simple SONAR equation which I used should always be replaced by the appropriate sound propagation model as discussed in Section 2.1.3. In particular if the noise source and the two whales are at different depths, maskogram analysis will become 3-dimensional. This means that more than one maskogram will need to be computed, each representing a particular cross-section through the 3-dimensional ocean. Lines which are straight in the pictures presented in this chapter will bend. Convergence zones will show up as areas of increased sound pressure levels, shadow zones will have decreased sound pressure levels. Zones of masking will depend on the local signal-to-noise ratio. Altogether, local sound propagation modeling will make the maskograms look much more interesting!

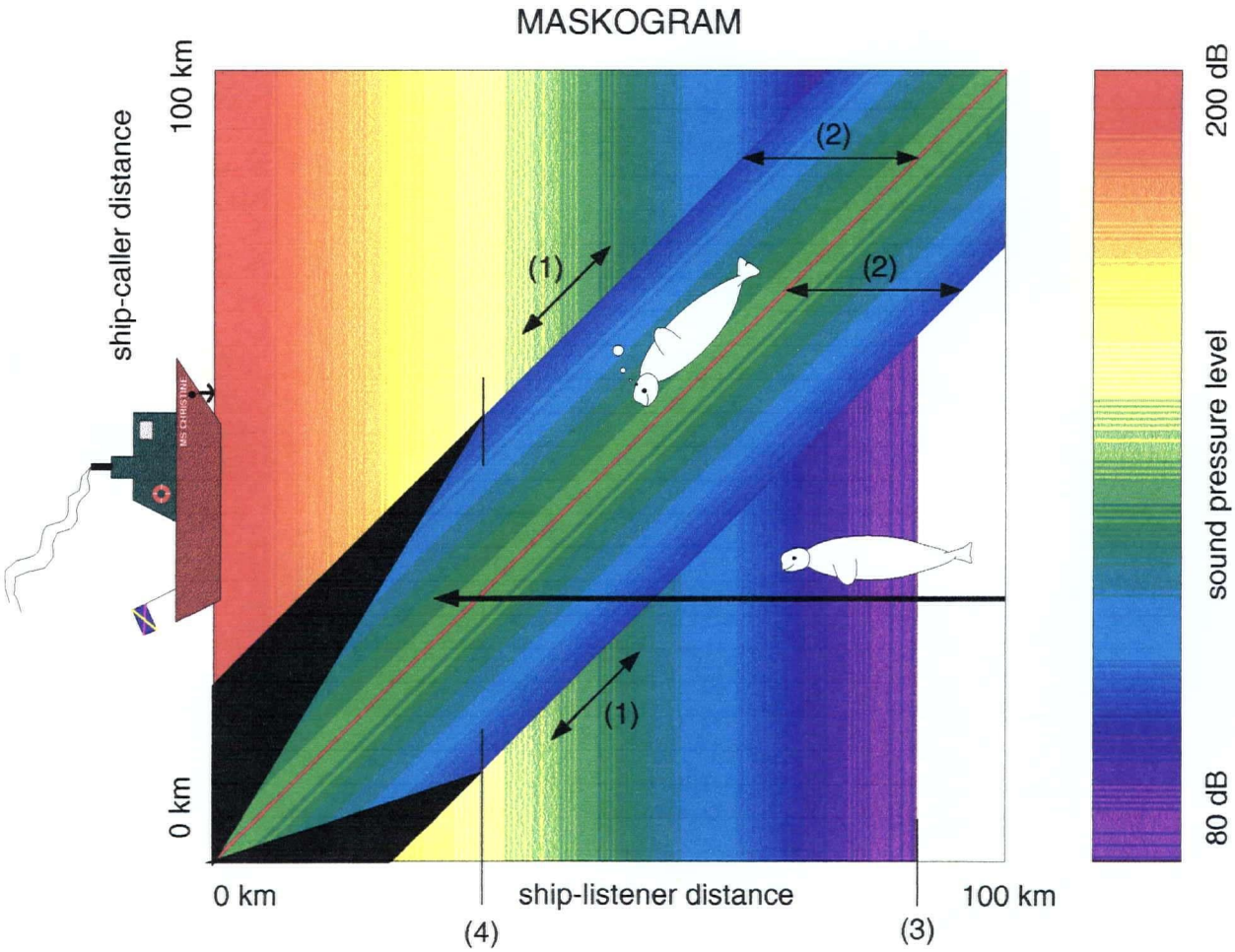


Figure 8.9: Model Maskogram.

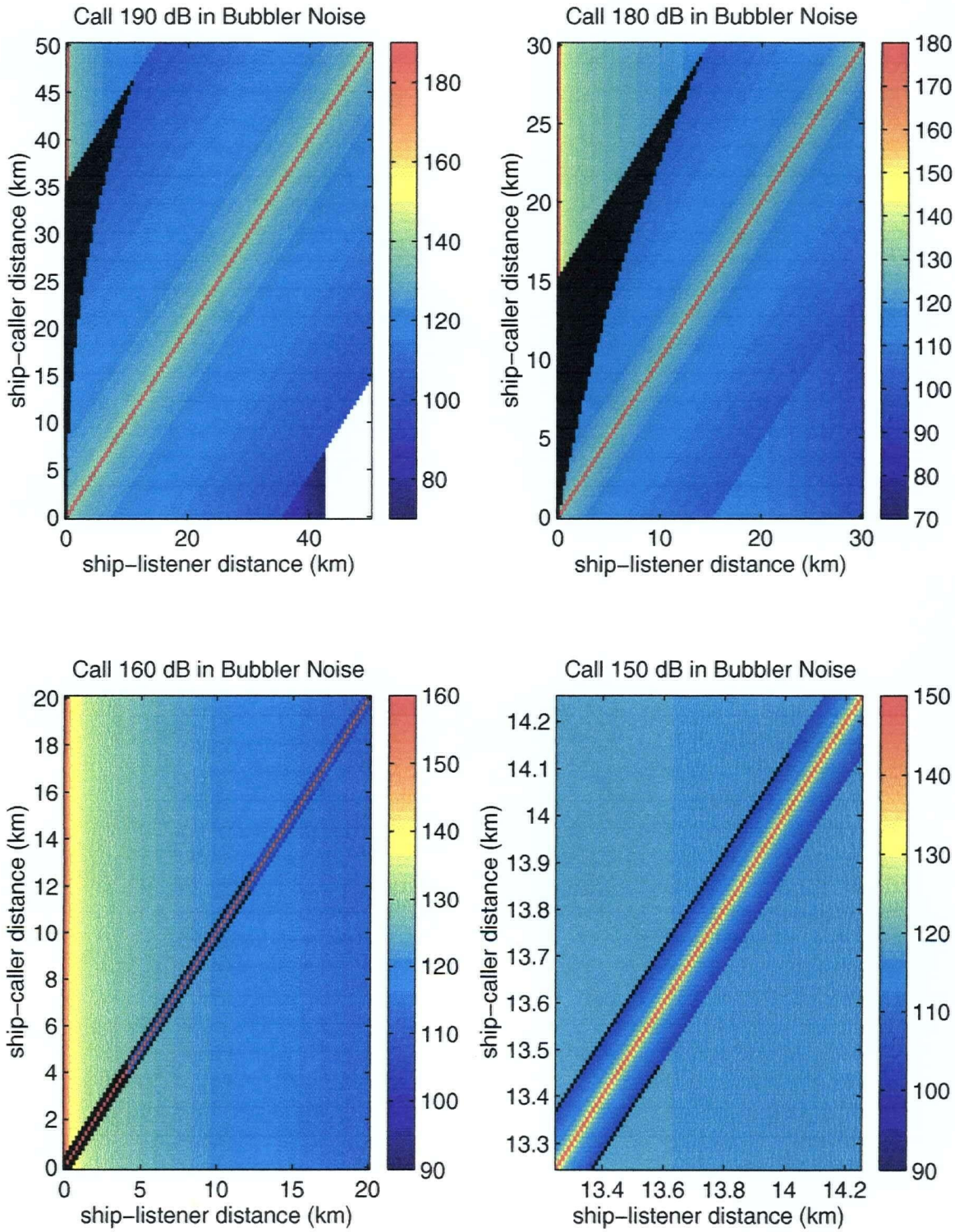


Figure 8.10: Maskograms for Bubbler System Noise.

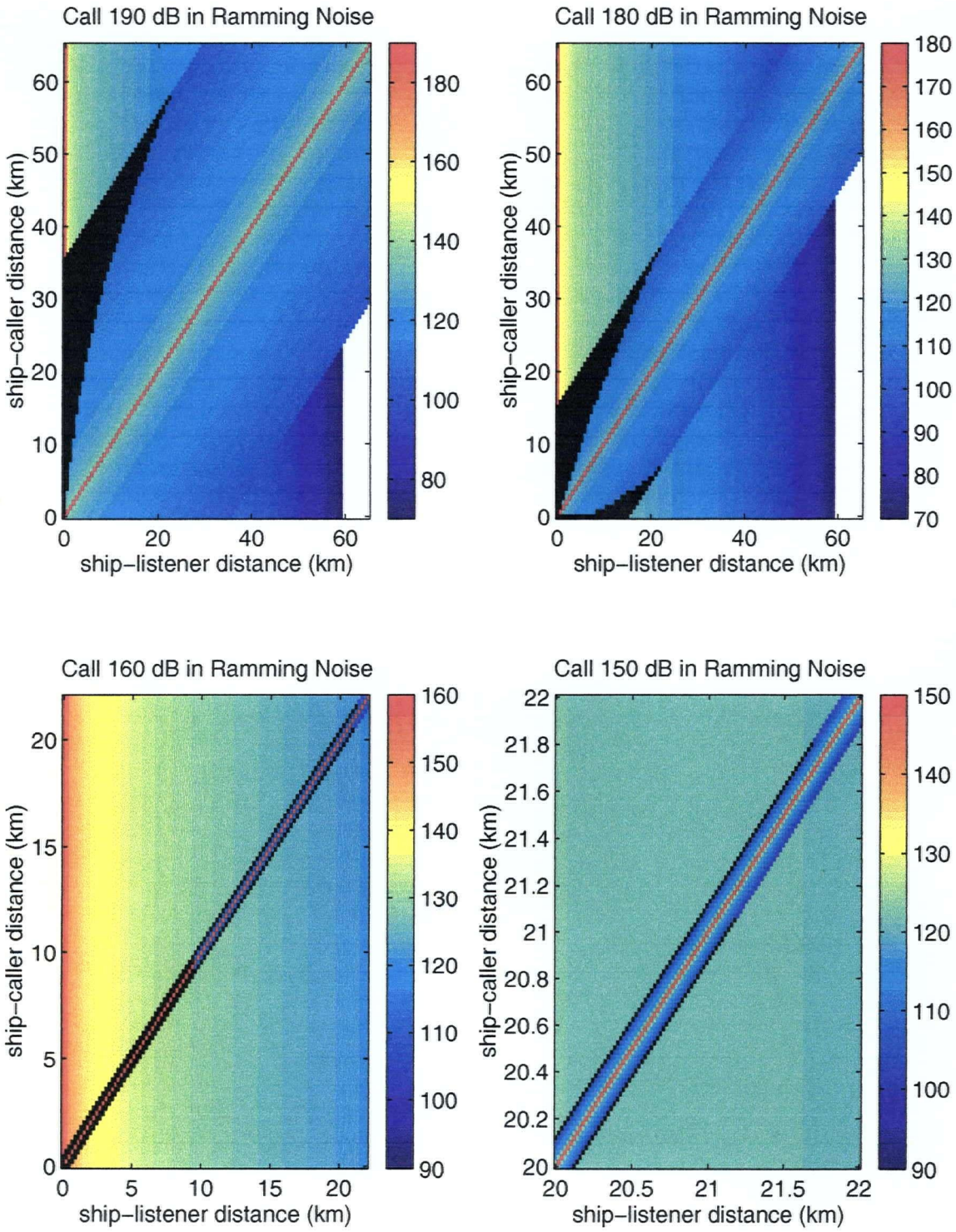


Figure 8.11: Maskograms for Ramming Noise.

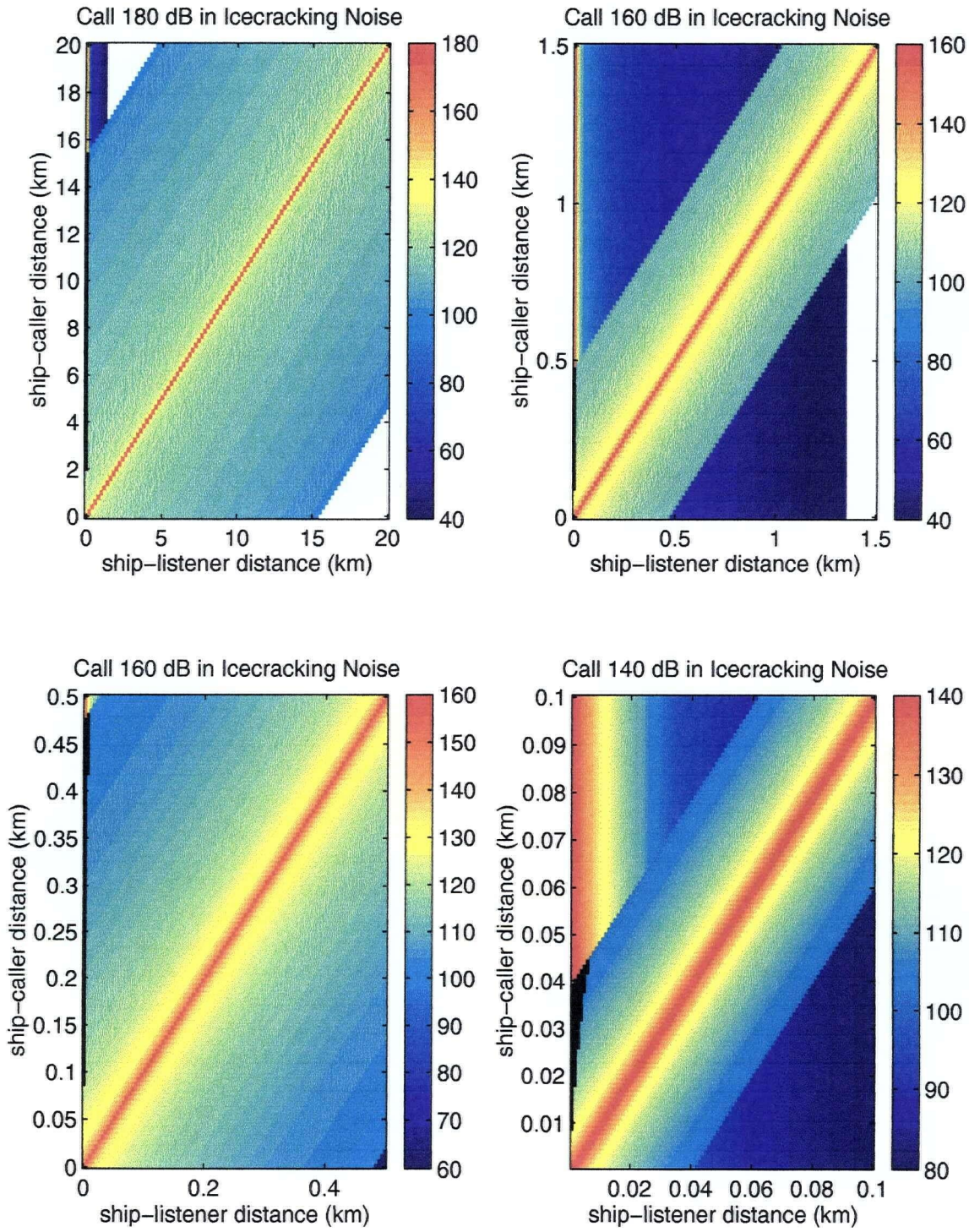


Figure 8.12: Maskograms for Icecracking Noise.

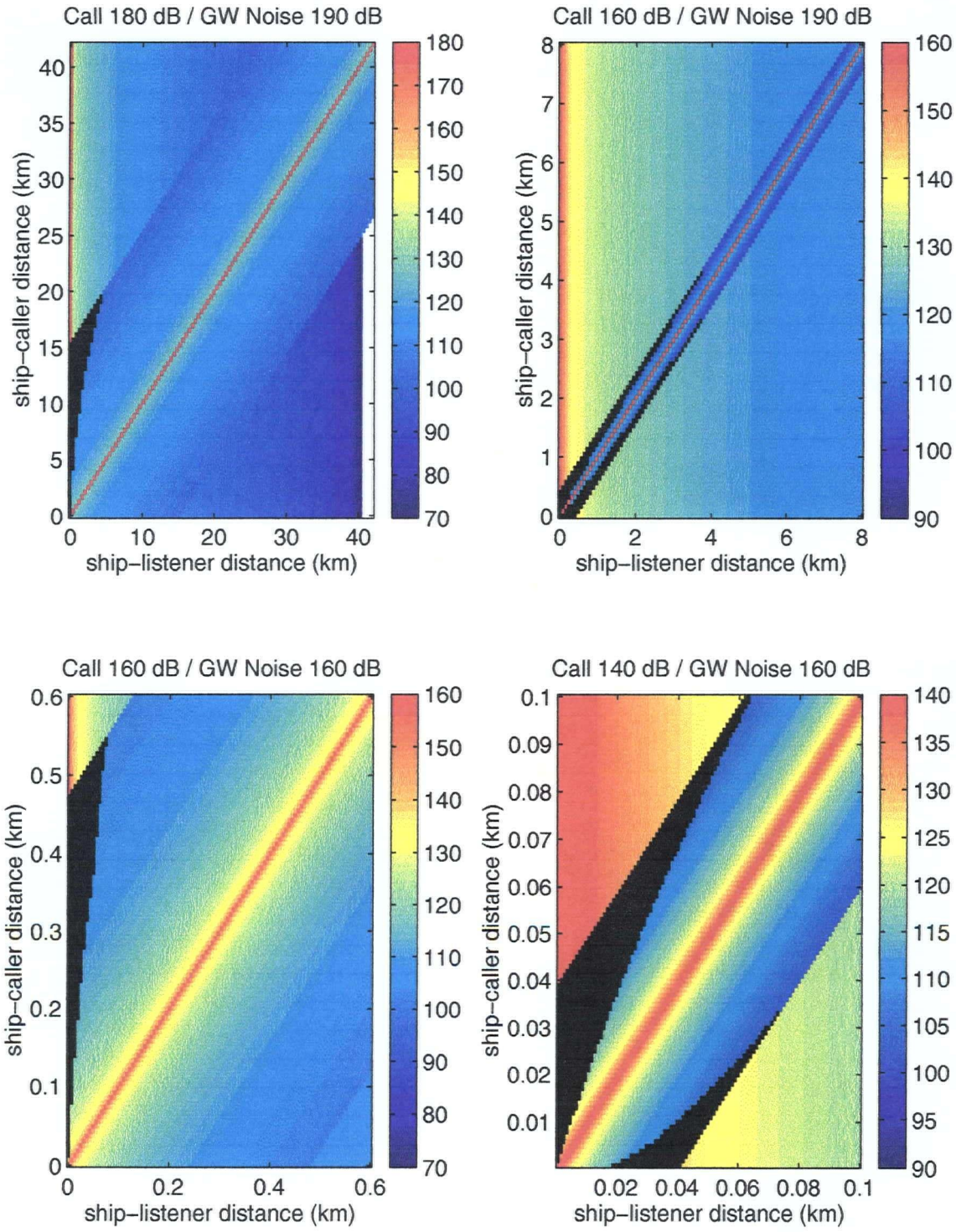


Figure 8.13: Maskograms for Gaussian White Noise.

Chapter 9

Summary and Conclusion

The more we humans learn about our nonhuman relatives in the wild, the more closely we approach them in spirit; the more sharply we sense the old links between their bloodlines and ours; the more easily we share their troubles.

–Victor B. Scheffer

In times of great concern for the world's marine mammals, this thesis dealt with threats posed by man-induced noise pollution. It was the first study looking at the masking of complex animal vocalizations by real underwater noise. Even though the analysis focussed on beluga whales for convenience, the techniques designed and presented can readily be applied to other species.

I have shown that a combination of animal experiments and subsequent computer modeling provides an integrated tool to assess the degree of masking a particular noise

has on a particular vocalization. Experiments with trained animals are very time and cost consuming and often impractical. They are, however, necessary to develop appropriate models, i.e. to test invented models for their accuracy and to calibrate them if necessary. Once such a model has passed this test, it can be used to predict the effect of new noises on the same or potentially different vocalization. The backpropagation network I designed needs to be calibrated to the beluga's hearing by shifting its thresholds 8.7 dB to higher values (Table 7.1). It then models the whale's response with a maximum deviation of 38 %. Comparing the neural network to the other tested detectors for a beluga vocalization in noise, human listening experiments and visual spectrogram discrimination also managed to simulate the whale's data closely. Both, however, involve the participation of humans and are thus more time consuming hence inefficient than the neural net. They also exhibit a subjective component. All other techniques, such as matched filtering, spectrogram cross-correlation, critical band cross-correlation, adaptive filtering and neural networks other than of backpropagation paradigm, did not manage to reproduce the whale's data appropriately.

Future work should include further testing of the backpropagation network. Although it performed well on the four noises and the beluga vocalization chosen, a generalization for beluga communication cannot be done without measuring the masking of different vocalizations in both animal experiments and computer modeling. Further vocalizations should preferably be of very different frequency and time structure. In addition, more than one individual animal should be tested. This can indicate age, gender or individual

variance within a population. I highly recommend that data be collected with a proper inclusion of catch trials in order to plot ROC diagrams. This is necessary to estimate the animal's decision bias. If time permits, ROC curves could be computed by actively changing the bias. This way, the animal's attitude can be separated from its sensitivity.

This thesis was the first that allowed the calculation of zones of masking around a noise source. Assumptions previously made in hypothetical estimations of masking ranges were proven to be inadequate for the signal and noises studied here. In my analysis, ambient Arctic noise, which was solely based on natural icecracking, did not add to the masking effect of other noise such as ship noise. Furthermore, it did not limit the range of masking of a noise source to radii smaller than the one where the vocalization would drop below the ambient noise in the absence of additional noise. I showed that the beluga vocalization became unrecognizable, i.e. its major harmonics dropped below the beluga audiogram, before being masked by ambient icecracking noise. In fact, it appears that beluga whales don't even hear ambient icecracking noise except for frequencies above 16 kHz.

The development of *maskograms*, colourplots of overlapping sound fields around a noise source and a calling whale, proved to be a useful technique to illustrate zones of masking in the ocean. From the maskogram analysis, I derived three important conclusions. First, a noise source has a maximum range of masking which is independent of the source level of the masked vocalization and independent of the distance between the noise source and the calling whale. Second, for most underwater noises, which are of temporal incoherence and which are not limited to exactly the same critical bands occupied by the vocalization under

study, the zone of masking is smaller than the zone of audibility. Third, the maskogram analysis re-confirmed that ambient icecracking noise does not add to the masking of ship noise as assumed in previous studies.

Future research might include vocalizations of higher frequencies where directional hearing abilities might allow an animal to increase the received signal-to-noise ratio. In this case, I gave constructive hints at how the maskogram technique can account for directivity.

In closing, Arctic beluga populations are not yet on the endangered species list. The extinction of the St. Lawrence population seems to be inevitable, although the prime reason is chemical pollution. Noise pollution in the extremely busy estuary certainly exists, however, the extent to which it plays a role in marine mammal mortality is unknown.

Noise sources in the Arctic are largely related to oil and mineral exploration. In its various steps, exploration includes acoustical, geophysical surveys, offshore construction, dredging, drilling and increased ship traffic (icebreakers, drill ships, tankers, lay barges for pipelines, supply vessels etc.). Another major noise source is fishing. Alone in the Eastern Canadian Arctic between Baffin Island and Greenland, about 100 fishing vessels, including large fishing factory vessels, can be found during the summer months. Furthermore, there sail merchant tankers supplying secluded villages and large vessels, cargo vessels, ore carriers, scientific research vessels, various tugs and passenger ships. Last but not least, Arctic marine mammals have to cope with ever increasing tourism.

I hope that my thesis will be a step towards preventing Arctic beluga whales from one

day being listed as endangered. Ideas and techniques derived could be used to estimate the degree of masking of existing industrial noise sources and new technologies and thus support control regulations for noise emissions.

Overall, I hope that I have managed to increase awareness of human impact on nature and that my thesis can contribute to the conservation of our fascinating relatives, the whales and dolphins. By putting all our efforts together, step by step and piece by piece, I would like to believe that we can protect the environment and conserve the immense diversity of life on Earth for our children.

Bibliography

Altes, R.A. (1979) Models for echolocation. pp. 625-671. In: R.-G. Busnel and J.F. Fish (eds.), *Animal Sonar Systems.* New York: Plenum Press. 1135 pp.

Angiel, N.M. (1997) The vocal repertoire of the beluga whale in Bristol Bay, Alaska. M.Sc. Thesis, University of Washington, USA. 78 pp.

Aroyan, J.L. (1996) Three-dimensional numerical simulation of biosonar signal emission and reception in the common dolphin. Ph.D. Thesis, University of California at Santa Cruz, USA. 184 pp.

Au, W.W.L., and P.W.B. Moore (1984) Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncatus. J. Acoust. Soc. Am.* 75(1):255-262.

Au, W.W.L. (1990) Target detection in noise by echolocating dolphins. pp. 203-216. In: J.A. Thomas and R.A. Kastelein (eds.), *Sensory Abilities of Cetaceans: Laboratory and Field Evidence.* New York: Plenum Press. 710 pp.

Au, W.W.L. (1993) *The Sonar of Dolphins.* New York: Springer Verlag. 277 pp.

Awbrey, F.T., J.A. Thomas and R.A. Kastelein (1988) Low-frequency underwater hearing sensitivity in belugas *Delphinapterus leucas. J. Acoust. Soc. Am.* 84(6):2273-2275.

Bain, D.E. (1992) Hearing abilities of killer whales *Orcinus orca.* Rep. for National Marine Mammal Laboratory, National Marine Fisheries Service, Seattle WA, USA. 19 pp.

Brekhovskikh, L., and Yu. Lysanov (1982) *Fundamentals of Ocean Acoustics.* Berlin: Springer Verlag. 250 pp.

Caldwell, M.C., and D.K. Caldwell (1965) Individualized whistle contours in bottlenosed dolphins *Tursiops truncatus. Nature* 207:434-435.

Carwardine, M. (1995) *Whales, Dolphins and Porpoises.* London: Dorling Kindersley. 256 pp.

Clay, C.S., and H. Medwin (1977) *Acoustical Oceanography: Principles and Applications.* New York: Wiley. 544 pp.

Cosens, S.E., and L.P. Dueck (1988) Responses of migrating narwhal and beluga to ice-breaker traffic at the Admiralty Inlet ice-edge, N.W.T. in 1986. pp. 39-54. In: W.M. Sackinger and M.O. Jeffries (eds.), *Port and Ocean Engineering under Arctic Conditions.* Fairbanks, Alaska: Geophysical Institute, University of Alaska. 111 pp.

Cosens, S.E., and L.P. Dueck (1993) Icebreaker noise in Lancaster Sound, N.W.T., Canada: implications for marine mammal behavior. *Mar. Mam. Sci.* 9(3):258-300.

Dahlheim, M.E. (1987) Bio-acoustics of the gray whale *Eschrichtius robustus.* Ph.D. Thesis, University of British Columbia, Canada. 315 pp.

Department of Fisheries and Oceans, Canadian Government (1991) The Beluga. Underwater World Factsheets, Communications Directorate, DFO, Ottawa ON, Canada. 12 pp.

Egan, J.P., and H.W. Hake (1950) On the masking pattern of a simple auditory stimulus. *J. Acoust. Soc. Am.* 22:622-630.

Etter, P.E. (1996) *Underwater Acoustic Modeling.* 2nd ed. London: E & FN Spon. 344 pp.

Farmer, D.M., and Y. Xie (1989) The sound generated by propagating cracks in sea ice. *J. Acoust. Soc. Am.* 85(4):1489-1500.

Finley, K.J., G.W. Miller, R.A. Davis and C.R. Greene (1990) Reactions of belugas *Delphinapterus leucas* and narwhals *Monodon monoceros* to ice-breaking ships in the Canadian High Arctic. *Can. Bull. Fish. Aquatic Sci.* 224:97-117.

Fletcher, H. (1940) Auditory patterns. *Rev. Mod. Phys.* 12(1):47-65.

Ford, J.K.B. (1977) White Whale–Offshore Exploration Study. Rep. by F.F. Slaney & Company Ltd. for Imperial Oil Ltd., Calgary, Canada. 41 pp.

Ford, R.D. (1970) *Introduction to Acoustics.* New York: Elsevier Publishing Company. 154 pp.

French, N.R., and J.C. Steinberg (1947) Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 19(1):90-119.

Green, D.M., and J.A. Swets (1966) *Signal Detection Theory and Psychophysics.* New York: Wiley. 455 pp.

Greene, C.R. (1981) Underwater acoustic transmission loss and ambient noise in arctic regions. In: N.M. Peterson (ed.), *The Question of Sound from Icebreaker Operations.* Workshop Proceedings, Feb. 23-24, 1981, Toronto, Canada. Calgary, Canada: Petro-Canada. 360 pp.

Hamilton, E.L. (1980) Geoacoustic modeling of the sea floor. *J. Acoust. Soc. Am.* 68(5):1313-1340.

Haykin, S. (1994) *Neural Networks.* New York: Macmillan College Publishing. 696 pp.

Helweg, D.A., A.S. Frankel, J.R. Mobley, Jr., and L.M. Herman (1992) Humpback whale song: Our current understanding. pp. 459-483. In: J.A. Thomas, R.A. Kastelein, A.Y. Supin (eds.), *Marine Mammal Sensory Systems.* New York: Plenum Press. 773 pp.

Herman, L.M., and W.N. Tavolga (1980) The communication systems of cetaceans. pp. 149-209. In: L.H. Herman (ed.), *Cetacean Behavior: Mechanisms and Functions.* New York: Wiley. 463 pp.

Hoyt, E. (1990) *The Whales of Canada.* Camden East ON, Canada: Camden House Publishing. 128 pp.

Jensen, F.B., W.A. Kuperman, M.B. Porter and H. Schmidt (1994) *Computational Ocean Acoustics.* Woodbury NY, USA: American Institute of Physics. 612 pp.

Johnson, C.S. (1967) Sound detection thresholds in marine mammals. pp. 247-260. In: W.N. Tavolga (ed.), *Marine Bio-Acoustics.* Vol. 2. Oxford, UK: Pergamon Press. 353 pp.

Johnson, C.S. (1971) Auditory masking of one pure tone by another in the bottlenose porpoise. *J. Acoust. Soc. Am.* 44:965-967.

Johnson, C.S., M.W. McManus and D. Skaar (1989) Masked tonal hearing thresholds in the beluga whale. *J. Acoust. Soc. Am.* 85(6):2651-2654.

Karl, J.H. (1989) *An introduction to digital signal processing.* San Diego CA, USA: Academic Press. 341 pp.

Ketten, D.R., J. Lien and S. Todd (1993) Blast injury in humpback whale ears: Evidence and implications. *J. Acoust. Soc. Am.* 94(3)Pt.2:1849-1850.

Ketten, D.R. (1995) Estimates of blast injury and acoustic trauma zones for marine mammals from underwater explosions. pp. 391-407. In: R.A. Kastelein, J.A. Thomas and P.E. Nachtigall (eds.), *Sensory Systems of Aquatic Mammals.* Woerden, Netherlands: De Spil Publ.

Kryter, K.D. (1985) *The Effects of Noise on Man.* 2nd ed. Orlando FL, USA: Academic Press. 688 pp.

Lammers, M.O., and W.W.L. Au (1996) Broadband recording of social acoustic signals of the Hawaiian spinner and spotted dolphins. *J. Acoust. Soc. Am.* 100(4)Pt.2:2609 (Abstract).

LePage, K., and H. Schmidt (1994) Modeling of low-frequency transmission loss in the central Arctic. *J. Acoust. Soc. Am.* 96(3):1783-1795.

LGL and Greeneridge (1986) Reactions of beluga whales and narwhals to ship traffic and icebreaking along ice edges in the eastern Canadian high arctic: 1982-1984. Environ. Stud. No. 37. Indian & Northern Affairs Canada, Ottawa ON, Canada. 301 pp.

LGL and Greeneridge (1995) Acoustic effects of oil production activities on bowhead and white whales visible during spring migration near Pt. Barrow, Alaska–1991 and 1994 phases: Sound propagation and whale responses to playbacks of icebreaker noise. OCS Study MMS 95-0051. Rep. for U.S. Minerals Management Service, Herndon VA, USA. 539 pp.

Ljungblad, D.K., P.O. Thompson and S.E. Moore (1982) Underwater sounds recorded from migrating bowhead whales *Balaena mysticetus* in 1979. *J. Acoust. Soc. Am.* 71(2):477-482.

Morse, P.M., and K.U. Ingard (1968) *Theoretical Acoustics.* Princeton NJ, USA: Princeton University Press. 927 pp.

Nolte, J. (1981) *The Human Brain: An Introduction to its Functional Anatomy.* St. Louis MO, USA: C.V. Mosby. 322 pp.

Pickles, J.O. (1988) *An introduction to the physiology of hearing.* San Diego CA, USA: Academic Press. 367 pp.

Pineda, F.J. (1988) Generalization of backpropagation to recurrent and higher order neural networks. pp. 602-611. In: D.Z. Anderson (ed.) *Neural Information Processing Systems.* New York: American Institute of Physics. 871 pp.

Pippard, L. (1985) Status of the St. Lawrence River population of beluga *Delphinapterus leucas. Can. Field Nat.* 99(3):438-450.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992) *Numerical Recipes in C*, 2nd ed. New York: Cambridge University Press. 994 pp.

Purves, W.K., G.H. Orians and H.C. Heller (1992) *Life: The Science of Biology.* Sunderland MA, USA: Sinauer Associates. 1218 pp.

Renaud, D.L., and A.N. Popper (1975) Sound localization by the bottlenose porpoise *Tursiops truncatus. J. Exp. Biol.* 63(3):569-585.

Richardson, W.J., M.A. Fraker, B. Würsig and R.S. Wells (1985) Behaviour of bowhead whales *Balaena mysticetus* summering in the Beaufort Sea: Reactions to industrial activities. *Biol. Conserv.* 32(3):195-230.

Richardson, W.J., B. Würsig and C.R. Greene, Jr. (1986) Reactions of bowhead whales *Balaena mysticetus* to seismic exploration in the Canadian Beaufort Sea. *J. Acoust. Soc. Am.* 79(4):1117-1128.

Richardson, W.J., B. Würsig and C.R. Greene, Jr. (1990) Reactions of bowhead whales *Balaena mysticetus* to drilling and dredging noise in the Canadian Beaufort Sea. *Mar. Environ. Res.* 29(2):135-160.

Richardson, W.J., C.R. Greene, Jr., C.I. Malme and D.H. Thomson (1995) *Marine Mammals and Noise.* San Diego CA, USA: Academic Press. 576 pp.

Rosenblatt, F. (1962) *Principles of Neurodynamics.* Washington DC, USA: Spartan Books. 616 pp.

Rumelhart, D.E., G.E. Hinton and R.J. Williams (1986) Learning internal representations by error propagation. pp. 318-362. In: D.E. Rumelhart and J.L. McClelland (eds.) *Parallel Distributed Processing.* Vol. 1. Cambridge MA, USA: MIT Press. 547 pp.

Rumelhart, D.E., G.E. Hinton and R.J. Williams (1986) Learning representations by back-propagating errors. *Nature* 323:533-536.

Ross, D. (1976) *Mechanics of Underwater Noise.* New York: Pergamon Press. 375 pp.

Sayigh, L.S., P.L. Tyack, R.S. Wells and M.D. Scott (1990) Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: Stability and mother-offspring comparisons. *Behav. Ecol. Sociobiol.* 26(4):247-260.

Schusterman, R.J., and B.W. Johnson (1975) Signal probability and response bias in California sea lions. *Psychol. Rec.* 25:39-45.

Schusterman, R.J., R. Barret and P.W.B. Moore (1975) Detection of underwater signals by a California sea lion and a bottlenose porpoise: variation in the payoff matrix. *J. Acoust. Soc. Am.* 57(6)Pt.2:1526-1532.

Schusterman, R.J. (1976) California sea lion underwater auditory detection and variation of reinforcement schedules. *J. Acoust. Soc. Am.* 59:997-1000.

Sergeant, D. (1986) Present status of white whales *Delphinapterus leucas* in the St. Lawrence Estuary. *Nat. Can.* 113(1):61-81.

Sivian, L.J., and S.D. White (1933) On minimum audible sound fields. *J. Acoust. Soc. Am.* 4(4):288-321.

Sjare, B.L., and T.G. Smith (1986) The vocal repertoire of white whales *Delphinapterus leucas* summering in Cunningham Inlet, Northwest Territories. *Can. J. Zool.* 64:407-415.

Thiele, L. Underwater noise from the icebreaker MS *Voima.* Rep. by Ødegaard & Danneskiold-Samsøe K/S for Greenland Fish. Invest., Copenhagen, Denmark. 35 pp.

Thiele, L., A. Larsen and O.W. Nielsen (1990) Underwater noise exposure from shipping in Baffin Bay and Davis Strait. Rep. by Ødegaard & Danneskiold-Samsøe ApS for Greenland Environmental Research Institute, Copenhagen, Denmark. 92 pp.

Tyack, P. (1981) Interactions between singing Hawaiian humpback whales and conspecifics nearby. *Behav. Ecol. Sociobiol.* 8(2):105-116.

UNESCO Technical Papers in Marine Science 40 (1987) *International Oceanographic Tables* Vol. 4. Paris: UNESCO.

Urick, R.J. (1975) *Principles of Underwater Sound.* New York: McGraw-Hill. 423 pp.

Verrall, R. (1981) Acoustic transmission losses and ambient noise in Parry Channel. In: N.M. Peterson (ed.), *The Question of Sound from Icebreaker Operations.* Workshop Proceedings, Feb. 23-24, 1981, Toronto, Canada. Calgary, Canada: Petro-Canada. 360 pp.

Wegel, R.L., and C.E. Lane (1924) The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.* 23:266-285.

Wenz, G.M. (1962) Acoustic ambient noise in the ocean: Spectra and sources. *J. Acoust. Soc. Am.* 34(12):1936-1956.

White, M.J., Jr., J. Norris, D. Ljungblad, K. Baron and G. di Sciara (1978) Auditory thresholds of two beluga whales *Delphinapterus leucas.* Rep. by Hubbs/Sea World Research Institute for Naval Ocean System Center. Rep. 78-109. San Diego CA, USA. 35 pp.

Widrow, B., and M.E. Hoff, Jr. (1960) Adaptive switching circuits. *IRE WESCON Convention Record.* pp. 96-104.

Xie, Y., and D.M. Farmer (1991) Acoustical radiation from thermally stressed sea ice. *J. Acoust. Soc. Am.* 89(5):2215-2231.

Yelverton, J.T., D.R. Richmond, E.R. Fletcher and R.K. Jones (1973) Safe distances from underwater explosions for mammals and birds. DNA 3114T. Rep. from Lovelace Foundation for Medical Education and Research, Albuquerque NM, for Defense Nuclear Agency, Washington DC, USA. NTIS AD-766952. 67 pp.

Zaitseva, K.A., V.P. Morozov and A.I. Akopian (1980) Comparative characteristics of spatial hearing in the dolphin *Tursiops truncatus* and man. *Neurosci. Behav. Physiol.* 10(2):180-182.