ACOUSTIC COMMUNICATION AND VOCAL LEARNING IN BELUGAS (*DELPHINAPTERUS LEUCAS*)

by

VALERIA VERGARA

B.Sc. (Honours), Trent University, 1992
M.Sc. Trent University, 1996

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

February 2011

© Valeria Vergara, 2011
ABSTRACT

Belugas (*Delphinapterus leucas*) are highly vocal cetaceans, but the function of their calls, repertoire ontogeny, and role of learning in vocal behavior are poorly understood. This dissertation examines these issues, focusing on a captive beluga group at the Vancouver Aquarium. First, I investigated vocal development in a beluga calf, longitudinally throughout his first year of life, and later opportunistically. The first sounds after birth were low energy, broadband pulse-trains, which increased in pulse repetition rate with age. He incorporated rudimentary whistles at two weeks. His mixed calls, which became consistent at four months, became progressively stereotyped, increasingly like his mother’s “Type-A” call, a presumed contact call. Six months after he was first exposed to his father’s calls, he developed a call type similar to one of his father’s. I discuss these findings in light of theories of sound production mechanisms, developmental stages of vocal acquisition, and vocal learning.

Secondly, I examined context-specific use of call types recorded from the beluga group, with particular focus on the Type-A call. This signal constituted 24-97% of the vocalizations during isolation, births, deaths, presence of external stressors, and re-union of animals after separation. In contrast, it represented 4.4% of the vocalizations during regular sessions. I identified five Type-A variants subjectively and quantitatively. I used these findings to generate hypotheses about the usage of these signals by wild belugas, verified the existence of A-calls in the repertoire of St. Lawrence herds, and documented their usage by two wild individuals from different populations in contexts that supported their contact function.

Finally, I investigated contextual vocal learning in trained tasks in adult belugas, focusing on the ability of a female beluga to respond to playbacks of two categories of beluga calls with matching vocalizations; pulse-trains are a natural category, and screams an artificial class shaped by training. The subject successfully matched only pulse-trains, the class that is part of this species’ natural repertoire. Her poor performance on matching screams might be partly explained by a difficulty to perceive categorically a signal that lacks a function in the natural repertoire of belugas.
PREFACE

All analysis and writing presented in this thesis are my own.

Chapter two is a revised version of the following paper:


I designed the study, analyzed all data and wrote the paper. L.G. Barrett-Lennard provided support and advice throughout the study.

Chapter 3 is a revised version of the following paper:


I initiated and planned the project, analyzed the data, and wrote the paper. R. Michaud provided recordings and invaluable logistical support for the field portion of the study. Both R. Michaud and L.G. Barrett-Lennard assisted with field work and provided support and advice throughout the study.

The Animal Care Committee has examined and approved the use of animals (captive whales) for this dissertation (A06-0241-R004).
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................................................... ii

PREFACE........................................................................................................................................ iii

TABLE OF CONTENTS ..................................................................................................................... iv

LIST OF TABLES ............................................................................................................................. vii

LIST OF FIGURES ........................................................................................................................... viii

GLOSSARY ......................................................................................................................................... xi

ACKNOWLEDGMENTS .................................................................................................................... xiii

DEDICATION ...................................................................................................................................... xvi

1. GENERAL INTRODUCTION.......................................................................................................... 1
   1.1 Vocal learning ......................................................................................................................... 2
     Mimicry and vocal learning........................................................................................................... 6

   1.2 What do we know about vocal learning in belugas? ............................................................... 7

   1.3 What do we know about beluga social structure? .................................................................. 9

   1.4 What do we know about the acoustic behaviour of belugas? ............................................... 12

   1.5 Thesis objectives and organization ....................................................................................... 18

2. VOCAL DEVELOPMENT IN A BELUGA CALF (*DELPHINAPTERUS LEUCAS*) .................... 19
   2.1 Introduction ........................................................................................................................... 19

   2.2 Methods .................................................................................................................................. 21
     Social group ................................................................................................................................. 21
     Sampling regime and observation area ..................................................................................... 22
     Acoustic recordings .................................................................................................................... 23
     Identification of vocalizing individuals ..................................................................................... 24
     Acoustic analysis .......................................................................................................................... 25
     Vocal development after new auditory input ............................................................................ 27
     Statistical analyses ..................................................................................................................... 27
2.3 Results ........................................................................................................................................... 29
  General vocalization types ................................................................................................................. 29
  Pulse-train development ...................................................................................................................... 30
  Whistle development .......................................................................................................................... 32
  Mixed calls .......................................................................................................................................... 35
  Acquisition of a novel call type ........................................................................................................ 41

2.4 Discussion ........................................................................................................................................ 43
  Early vocalizations and repertoire development ............................................................................... 43
  Babbling in belugas ............................................................................................................................. 45
  Maturational processes ........................................................................................................................ 47
  Vocal learning ....................................................................................................................................... 48

2.5 Postscript ......................................................................................................................................... 50

3. WHAT CAN CAPTIVE WHALES TELL US ABOUT THEIR WILD COUNTERPARTS? 
IDENTIFICATION, USAGE, AND ONTOGENY OF CONTACT CALLS IN BELUGAS ............... 52

3.1 Introduction ....................................................................................................................................... 52
  Study approach ..................................................................................................................................... 54

3.2 Captive component: methods ......................................................................................................... 54
  Time frame and social group ............................................................................................................... 54
  Sampling regime and observation area .............................................................................................. 56
  Acoustic recordings ............................................................................................................................. 56
  Call type categorization ....................................................................................................................... 57
  Discriminant Function Analysis (DFA) ............................................................................................... 57
  Contextual use of call types ................................................................................................................. 59
  Vocal exchanges ................................................................................................................................... 61

3.3 Captive component: results ............................................................................................................ 62
  Context specific use of the Type A call ............................................................................................... 62
  Vocal exchanges ................................................................................................................................... 64
  Variants of the Type A call .................................................................................................................. 65
  Usage of the Type A variants in isolation ........................................................................................... 68
  Development of the Type A1 variant .................................................................................................... 70

3.4 Wild component ................................................................................................................................. 70
  Overview ................................................................................................................................................ 70
  Nelson River Estuary ............................................................................................................................. 71
  St. Lawrence Estuary ............................................................................................................................ 72
3.5 General discussion ............................................................................................................. 75
    Future research .................................................................................................................. 81

4. CALL USAGE LEARNING IN BELUGAS: CATEGORICAL MATCHING ......................... 84
   4.1 Introduction ..................................................................................................................... 84
   4.2 Methods ......................................................................................................................... 88
       Subjects, training regime and general procedure .......................................................... 88
       Playback stimuli ........................................................................................................... 89
       Phase 1: Training procedure ......................................................................................... 90
       Phase 2: Categorical matching experiment .................................................................. 92
       Statistical analyses ..................................................................................................... 94
   4.3 Results ......................................................................................................................... 94
       Training trials ............................................................................................................... 94
       Transfer trials ............................................................................................................. 96
       Calibration trials ....................................................................................................... 97
   4.4 Discussion ..................................................................................................................... 98
       Perception of vocal categories .................................................................................... 99
       Categorical perception in relation to function ............................................................ 101
       Procedural limitations ................................................................................................. 102
       Motivation ................................................................................................................... 104
       Learning to learn ........................................................................................................ 104
       Conclusion .................................................................................................................. 106

5. GENERAL DISCUSSION ................................................................................................... 108
   5.1 Summary ..................................................................................................................... 108
   5.2 A sound-centered species in today’s noisy world ....................................................... 111
   5.3 Future work ................................................................................................................. 112
   5.4 Final thoughts .............................................................................................................. 116

REFERENCES .................................................................................................................... 118

APPENDIX A .................................................................................................................... 137
LIST OF TABLES

Table 1.1: Summary of studies that attempted a classification of the acoustic repertoire of belugas. ........................................................................................................ 14

Table 2.1: Acoustic parameters of the fully developed Type A1 call produced by the beluga calf, Tuvaq, at 32 months of age, and by his mother, Aurora, at the Vancouver Aquarium. ........................................................................................................ 41

Table 2.2: Acoustic parameters of call Type B2 produced by a beluga calf, Tuvaq, and by his father, Imaq, at the Vancouver Aquarium. ......................................................... 43

Table 3.1: Ages, sex and period of study for the belugas at the Vancouver Aquarium .......... 55

Table 3.2: Descriptive parameters for the variants of the Type A contact call identified in captivity, and for the structurally similar and probably functionally equivalent broad band pulsed calls identified in two capture and tagging operations and one dead calf situation. ............................................................................... 77
LIST OF FIGURES

Figure 2.1: Genealogy of the Vancouver Aquarium belugas at the time of this vocal development study. ................................................................. 22

Figure 2.2: Diagram of the beluga pool and the adjacent medical holding pool at the Vancouver Aquarium. ......................................................... 23

Figure 2.3: Time line of the different compositions of a beluga calf’s social group at the Vancouver Aquarium throughout the duration of this study, and husbandry isolation events during which the calf (Tuvaq) or his father (Imaq) were recorded................................................................. 28

Figure 2.4: Changes in the usage of different vocalization types by a captive male beluga calf during his first year of life (n = 2185). ........................................ 29

Figure 2.5: Examples of a pulse-train emitted during the first day of life of a beluga calf born at the Vancouver Aquarium (A) and of a vocal exchange between the calf and his mother on day five (B). ................................................. 30

Figure 2.6: Mean pulse repetition rate per session (± SD) in the pulsed calls of a beluga calf at the Vancouver Aquarium, plotted as a function of age. ....................... 31

Figure 2.7: Mean dominant frequency per session (± SD) of pulse-trains emitted by a beluga calf at the Vancouver Aquarium, plotted against his age........................................ 32

Figure 2.8: First recorded whistle, at the end of a pulse-train produced by a beluga calf at the Vancouver Aquarium at age 13 days ........................................ 32

Figure 2.9: Representative examples of whistle development in a captive beluga calf during his first year of life......................................................... 34

Figure 2.10: Mean whistle dominant frequency per session (± SD) plotted against age of a beluga calf at the Vancouver Aquarium. ........................................ 35

Figure 2.11: Examples of the calf’s mother’s stereotyped mixed call Type A1 at the Vancouver Aquarium, recorded at 44 kHz (A) and at 96 kHz (B). ................................. 36

Figure 2.12: Example of a bout of the three classes of Type A calls produced by an adult female beluga (Aurora) in the presence of her calf, at the Vancouver Aquarium................................................................. 37

Figure 2.13: Natural sequences of three calls from a beluga calf at the Vancouver Aquarium, each a representative example of the development of the stereotyped mixed call Type A1........................................................................ 39

Figure 2.14: Parameters of the mixed call Type A1, for a beluga calf, Tuvaq (TU), at the ages when this call type was recorded, and for his mother Aurora (AU) for comparison. ................................................................. 40

Figure 2.15: Two call types, B1 and B2, produced by an adult male beluga, Imaq, at the Vancouver Aquarium ................................................................. 42
Figure 2.16: B2 calls produced by an adult male beluga (Imaq) and two randomly selected renditions by his son (Tuvaq) at 2 years of age, 6 months after their first contact...........................................................................42

Figure 3.1: Genealogy of belugas held at the Vancouver Aquarium. Black arrows indicate parent-offspring. Ages of the animals and whether captive-born or wild caught are listed on Table 1..........................................................55

Figure 3.2: Example of the pulse repetition rate measurements at the beginning, middle, and end of a call (PRR 1, PRR 2, PRR 3)...........................................................59

Figure 3.3: Histogram of the time intervals between adjacent (black bars) and overlapping (grey bars) Type A calls made by different whales..................................61

Figure 3.4: Context specific use of all calls recorded at the Vancouver Aquarium during special sessions (bars A to H) and regular sessions (I)..................................62

Figure 3.5: Average frequency (number per individual per hour) of Type A calls and of all other call types combined during special sessions (A through H) and regular sessions (I)..........................................................................63

Figure 3.6: Spectrogram illustrating a call matching exchange between two adult female belugas, Aurora, who in this case was producing A1 variants (see next section), and Qila, producing overlapping A3 variants.........................................................65

Figure 3.7: Representative spectrograms (bottom) and oscillograms (top) of the five variants of the Type A call produced by the Vancouver Aquarium belugas.................66

Figure 3.8: Pulse repetition rate pattern for each of the five variants of the Type A call produced by the Vancouver Aquarium belugas.......................................................66

Figure 3.9: Discriminant function plot showing the separation of the five variants of the Type A call based on the most important discriminative acoustic parameters, pulse repetition rate and peak frequency..........................................................68

Figure 3.10: Proportional use of Type A variants and of all other call types combined (white) for each Vancouver Aquarium beluga that was recorded in forced or voluntary isolation..........................................................................69

Figure 3.11: Sample oscillograms (top) and spectrograms (bottom) of the broadband pulsed calls that resemble the captive Type A calls both in structure and context, identified in two different wild beluga populations: Nelson River estuary (left) and St Lawrence estuary (right)......................................................74

Figure 3.12: Adult female and dead calf in the St. Lawrence estuary, 1999......................74

Figure 4.1: Oscillograms and spectrograms of the two signal types used in training, the scream and the pulse-train...............................................................89

Figure 4.2: Proportion of categorical matches by a female beluga, Qila, at the Vancouver Aquarium, in response to the scream (grey squares) and pulse-train (white circles) playbacks with which she was trained, and overall success rate for both playback types combined (red triangles).................................95
Figure 4.3: Proportion of categorical matches by a female beluga, Qila, at the Vancouver Aquarium, in response to seven novel exemplars of the scream playbacks (grey squares) and six novel exemplars of the pulse-train playbacks (white circles), and overall success rate for both playback types combined (red triangles). .......................................................... 96

Figure 4.4: Proportion of correct vocal responses by a female beluga, Qila, at the Vancouver Aquarium, to the scream hand signals (grey squares), the pulse-train hand signals (white circles), and to both signal types combined (red triangles). ............... 97
GLOSSARY

**Bout:** I use this term throughout the thesis to refer to series of vocalizations of the same type that are produced in sequence, with brief intervals being relatively common and long ones relatively rare - as defined by Slater and Lester (1982) for behavioural bouts.

**Bridge:** A signal that indicates when an animal has done something correct. It bridges the gap in time between when the animal did something correct and when it will receive positive reinforcement. Some examples of bridges are whistles, the word “good” or a touch. The bridge becomes associated with positive reinforcement and becomes a secondary, or conditioned, reinforcement: the animal has learned that a clicker, whistle, or particular word indicates it has done something correctly and can expect a food reward.

**Capturing Behaviour:** A method of training new behaviours, which consists of catching an animal in the act of presenting a particular behaviour. If the behaviour is immediately followed by a positive reinforcer, the animal may repeat the behaviour. This process may lead to the behaviour being offered frequently. At this point a cue (e.g. a hand signal) can be inserted prior to the presentation of the behaviour, so that the animal begins to associate the behaviour with the cue.

**Least Reinforcing Stimulus (LRS):** Also termed Least Reinforcing Scenario, this is a short “time-out”, when all interaction or reinforcement is withheld for a period of time (generally a few seconds) following an incorrect response or a lack of response.

**Operant Conditioning:** The modification of the strength of a response by manipulating the consequences of the response. A desirable consequence to a behaviour will increase the likelihood of the behaviour being repeated, while an undesirable consequence will decrease the likelihood of the behaviour being repeated. In other words, operant conditioning is a form of learning in which the consequences of the behaviour lead to changes in the probability of its occurrence.
**Phoneme:** The smallest segmental unit of speech sound that is distinguished by the speakers of a particular language. Phonemes are sound units that are employed to form meaningful contrasts between utterances.

**Primary Reinforcer:** An *unconditioned reinforcer* that naturally reinforces behaviour because it satisfies a biological need, and that does not need to be learned (in contrast to the conditioned reinforcer, or bridge, which is associated with a primary reinforcer). Food, water, sleep and sex are primary reinforcers.

**Shaping Behaviour:** A commonly used method to train a new behaviour, or to refine a behaviour that is already known (for example, encouraging louder vocalizations). It consists of reinforcing small behavioural modifications that lead to the final behaviour in successive approximations. Behaviour that deviates from the desired objective is ignored. In this way, the animal’s behaviour is slowly molded, or shaped, into the final desired behaviour.

**Stereotyped vocalizations:** Signals in a repertoire that are regularly repeated and have little variability in acoustic parameters between utterances. Highly stereotyped calls in a vocal repertoire are usually easier for humans to identify and classify than variable calls.
ACKNOWLEDGMENTS

I have been looking forward to writing this section. It is my opportunity to thank the numerous people that have helped in one way or another to make this study possible, and that have provided support, encouragement, friendship and joy throughout the years. In attempting to name each one, I run the very real risk of excluding someone, so I apologize in advance to anyone I omitted.

First and foremost, I am profoundly indebted to my co-supervisors, Lance Barrett-Lennard and Lee Gass. I thank them for taking a leap of faith and choosing to believe in me, taking me on as a graduate student, and for continuing to have confidence in me when I doubted myself. Lance and Lee are inexhaustible critical thinkers and I have learned a great deal from them over the years. Lance made it possible for me to study the Vancouver Aquarium belugas, and made the cetacean research lab a wonderful workplace. I value the many stimulating conversations throughout the years, and the fact that he did not tire of challenging me with alternative interpretations of my results. I am thankful to him for coming to the field with me, for teaching me everything I know about hydrophones, for his careful edits of my writing, and for his wonderful sense of humor – which allowed him, I am sure, to take my tempestuous self with a thick grain of salt! Thank you Lance, above all, for understanding that the key to having a pleasant graduate student around was to never let me go hungry.

Lee has been a source of inspiration, a true mentor. His website’s stories and musings about everything from creativity to risk-taking never failed to inspire me. Lee reminded me to think outside the box, helped me become a better writer, and always, always encouraged me to talk about a beluga as a “he” or a “she” and not an “it”. I am grateful to Lee for his high expectations of me – which forced me to push myself more than I thought I could – and for the care with which he reviewed each draft of each chapter, not only handing back the manuscripts riddled with comments, but accompanying them with thoughtful, deep narratives about one thing or another, stories that always had a point, always hit home. I am sorry, Lee, for finding some of them funny when you intended to be serious.

I am very grateful to the rest of my thesis committee: Volker Deecke, whom I am fortunate to be able to call a friend, has been extremely generous and patient with his time and advice since the very beginning of this project. His thorough knowledge of acoustic communication, experimental design of playback studies, and biology in general assisted greatly to improve several aspects of this study. Jessica de Villiers has been an encouraging and supportive friend, with an unfailingly positive outlook. I feel fortunate to have had the interdisciplinary insight of someone so knowledgeable in human linguistics. Darren Irwin agreed to be recruited into my committee when he probably already had a tremendously heavy work-load with his own students, and I am very thankful for that, for his acutely practical and insightful comments during every committee meeting, and for reviewing my work on short notice. Thank you guys, you have been a wonderful team, the suggestions and comments from each of you have greatly improved the quality of this thesis. I would also like to thank my external examiner, Laela Sayigh, for the exhaustive and wonderful list of suggestions that contributed to a much more polished final version of this dissertation.

I am greatly indebted to the Vancouver Aquarium for allowing me access to the animals. I thank Brian Sheehan, Marcy Tarvid, and all the marine mammal trainers that have assisted in one way or another throughout the years. The marine mammal staff managed to accommodate the lengthy
training required for my experiments within their extremely busy schedule, and I am very thankful for that. Thank you also to the Engineering and Information Technology Departments for contributing to the smooth functioning of the recording system by drilling, bolting, cable feeding, sound testing, and the like. Funding for the captive and field expenses were covered with a Vancouver Aquarium beluga research fund, the BC Wild Killer Whale Adoption Program, an NSERC scholarship, and an Animal Behaviour Society Student Grant.

Thank you to the past and present members of the Cetacean Research Lab: Doug Sandilands helped more than once with graphs, taught me almost everything I know about the physics of sound, and probably added a few years to my life by dragging me out of the office for great jogs during lunch breaks. Harald Yurk provided valuable input on several aspects of my research. Colin MacLeod and Meaghan McKillop gathered most of the data on the latest beluga calves born at the Aquarium, Tiqa and Nala. Alana Phillips spent a long afternoon with me patiently teaching me how to exploit the drawing tools in Word to make a great figure. Thanks also to Caitlin Birdsall, Nick Dedeluk, Charissa Fung, Katie Kuker, Cara Lachmuth, Nancy Marcus, Judy McVeigh and Chris Tucker for their camaraderie and support.

A special thank you goes to everyone who made the field aspects of this study possible. Pierre Richard and Jack Orr, of the Arctic Aquatic Research Division of Fisheries and Oceans Canada, Winnipeg, Manitoba, facilitated the Nelson River Estuary study. They kindly invited me to tag along with my hydrophones during the beluga tagging operation, an unforgettable experience. Jack and his wonderful team were extremely accommodating to facilitate my recordings amidst what was an often frantic rush to instrument and release the whales as quickly and smoothly as possible. Robert Michaud went above and beyond what I could ever have hoped to expect, to make the St.Lawrence field portion of this study possible. He provided invaluable logistic help, accommodation, hospitality, and long and stimulating brainstorming sessions. He also provided many hours of recordings of St.Lawrence belugas, data that proved invaluable for my thesis. Robert and the rest of the GREMM\(^1\) team were fantastic. I will always remember Michel Moisan and Renaud Bindiaux with fondness, and am very thankful for their hard work. Their knowledge of the St. Lawrence belugas never ceased to amaze me -- they could individually identify a beluga that would surface for all of 2 seconds, by a little scratch mark on the dorsal, with astounding accuracy. It was a truly delightful experience to work with these guys.

I would also like to thank all the folks who lent a hand whenever needed. Alistair Blachford programmed the old T-80, my initial attempt at standardizing and speeding behavioural sampling. Rob Harrison helped with set up of computers and hydrophones. Mike Melnychuk helped immensely with stats and graphs. No one knows excel and R (programming language) like he does! Lisa Walker came to the field with me and took wonderful pictures, assisted with the insane recording schedule during two of the beluga births, and shared many intellectually stimulating conversations with me over the years. Elaine Leung, Cheryl Smith, and Andrea Coombs spent many hours gathering valuable behavioural data. Ari Shapiro and Sofie Van Parijs graciously provided narwhal and beluga sound clips. Ari also provided the inspiration for the categorical matching study, and shared many insightful conversations with me about training and vocal learning in marine mammals. He has always been a joy to talk with.

I feel very fortunate to have had great support and encouragement from my family over the years. I am especially thankful to my parents, for identifying and encouraging my love for and

---

\(^1\) Group for Research and Education on Marine Mammals, Tadoussac, Quebec
curiosity about animals since I was a kid, and for always cheering me on to follow my passion, even when that meant moving to the other end of the continent at the age of 19 to study biology in Canada. I began this journey shortly after my father passed away, but I know he would have been exuberantly proud. A huge thank you goes to my sister María, for being supportive and understanding of a lifestyle and career choice that has kept me so very far away from home. I thank my aunt Julia, who always believed in me and helped me immensely at different times in my life. And a very special thank you, of course, to Martina, my wonderful daughter, my greatest delight, and to Mike, without whom I could not have accomplished this. Thank you, Mike, for being the best father any child could ever wish for. Thank you Martina, my pichoncita: the happiness I feel because you are in my life makes anything possible. Your birth really put this thesis in perspective: I am afraid to say that it got bumped back more than a few notches in priority!

A heartfelt thank you for existing goes out to my chosen family, my life-long friends, you know who you are. Without your constant presence in my heart, without all our adventures and shared moments, I might not have found the strength and perseverance to get this thesis finished! Thanks to all of you, without forgetting anyone. There are two women, however, that I must mention in particular, because their words to me were like beacons in the mist: Christa, thank you for your words that night by the river nearly a decade ago, encouraging me to embark on this journey. Thank you for saying “do it because you can” and for reminding me that there is no greater privilege than learning, and that a PhD would be a good thing for its own sake. Line, thank you for reminding me to enjoy the process of writing this thesis! Those key words spoken with your usual vehemence only a few months ago were a turning point for me.

I will remember these past few years of my life as rich learning years. I learned, of course, about belugas and about how to study these amazing animals. I learned to sail and discovered the immensely satisfying live-aboard life. I learned to drum – and, in the process, I became familiar with music and rhythm, areas that I had been naive about and that enhanced, I believe, the way I listened to belugas. And, most importantly, I had the tremendous fortune and joy to learn about being a mom. All these diverse aspects of my life intersected in one way or another, and helped me keep my balance and my sanity. So there are a few remaining acknowledgements I would like to make: Thank you to all those wonderful and patient people that came aboard Pinta and taught me, without laughing at me too much, how to sail and maintain a boat that I moved into when I barely knew how to tie the enormous thing to the dock. Thank you to all my incredible and talented drumming teachers for awakening and nurturing the drummer in me these past four years. Drumming kept me focused especially during the intense writing months. I had my favorite drum, an African djembe, strategically positioned next to my writing chair, and when thoughts got cloudy I would turn to it and drum my heart out, giving way to my right brain and feeling the immediate relaxation that comes with that. After a few minutes, I would revert to thesis mode with renewed concentration. I am not sure why this worked, but it did! And last but not least, a whale-sized thank you to Aurora, Allua, Imaq, Kavna, Qila, Tiqa, Nala, and Tuvaq, for allowing me the privilege of learning about them. If we really must have these intelligent, long lived, far-ranging animals in captivity, it is my hope that, by studying them, their captive situation will at least serve some purpose by helping us to understand, love and protect their wild counterparts.
DEDICATION

I dedicate this thesis to my daughter, Martina, who is teaching me more than I could ever put into words.

Aurora, Allua and Tuvaq
1. GENERAL INTRODUCTION

Acoustic communication is central to the socioecology of cetaceans. In an underwater environment, where sight and smell are much less efficient than on land, this is hardly surprising. The olfactory sense of cetaceans is underdeveloped, and, since light travels only short distances before it is absorbed or scattered (DOSITS 2010), underwater vision has limited utility for detecting anything further than a few tens of meters at best. Sound, by contrast, propagates so well in water that it makes sense for cetaceans to rely primarily on acoustic signals to communicate with one another over short and long distances, to navigate, and to locate prey (Tyack 2008a). Vocal signals are especially important for maintaining contact between mothers and their dependent but precocious calves, who could easily swim out of sight (Janik 2009).

Belugas (*Delphinapterus leucas*), nicknamed “sea canaries” by early whalers, are among the most vocal cetaceans, producing what has been suitably described as a “bewildering array of sounds” (Finley et al. 1990). As in other social, far-ranging odontocetes, vocal communication probably plays an important role in mediating complex social interactions. Yet, as we shall see in the following sections, we know little about the functional significance of their calls, the ontogeny of their extensive repertoire, and the role of vocal learning.

To address these gaps, I studied the beluga social group housed at the Vancouver Aquarium. Despite the differences between the captive and wild environments, captivity affords opportunities such as longitudinal sampling of continuous sequences of behaviour of known individuals, and uninterrupted observations of social interactions at close range concurrent with sound recordings, along with knowledge of the ages, sexes, and social relationships of the animals. Such opportunities - logistically difficult, if not prohibitive, in the wild environment - afford a more transparent window into the function of beluga calls. Moreover, research on captive animals can reveal capabilities and behavioural flexibility (Shapiro et al. 2004) which may not be directly apparent through studies in the wild. Captive studies can thus guide our investigation of vocal usage in wild belugas by providing details that can help us uncover important questions and make informed predictions. The implicit objective of my work is thus

---

2 A version of this chapter will be submitted for publication as a review paper on beluga vocal communication studies.
to evaluate how studying vocal usage in captive belugas is relevant to understanding how these animals use vocal signals in their natural environment.

From a theoretical standpoint, this study contributes to comparative assessments of continuity in vocal communication processes across species, including humans – as is often the case with production-based research on animal vocalizations (Rendall 2003). Studies of any faculty involved in language, such as vocal learning, are considered relevant to broader investigations of human language evolution and worthy of comparative study, an approach advocated, for example, by Hauser et al (2002) and Fitch et al (2005). Thus, this study also adds to the growing body of work on vocal learning in nonhuman animals, leading to an integrative study of this trait. From a practical standpoint, identifying the functional significance of beluga vocal behaviour has management and conservation implications. The need for bio-acoustical research on species-specific vocal patterns that could be used for welfare assessment has been identified in the last decade (Weary and Fraser 1995). There is concern that the proliferation of human generated noise in beluga habitat poses adverse effects on their population, including the disruption of communication processes (which may in turn induce a displacement from favoured habitats, disruption of breeding or nursing, etc.). By determining which calls are indicative of stress or of loss of acoustic contact, we will be in a better position to evaluate the impacts of human activities. For example, a significant increase in the rate of contact calls – identified in this dissertation - in a noisy environment could indicate increased stress and the possibility that communication is being compromised.

In this chapter I introduce the theoretical background for this study, review the state of knowledge on the acoustic and social behaviour of belugas, and delineate the objectives of each of the chapters in this dissertation.

1.1 Vocal learning

Vocal learning is an aspect of social learning and has been shown to mediate and reflect important aspects of social relationships in species in which it has been proven to exist (see Tyack 2008b for examples). The expression “vocal learning” has traditionally been applied to the concept of production learning, whereby individuals modify their signals through experience with those of other individuals, leading to signals that are either similar or dissimilar to the model (Janik and Slater 1997). This definition has been expanded to incorporate examples of any kind of signal modification based on auditory input. In simple terms, vocal production learning is “the capacity to modify what one says based on what one hears” (Tyack 1993 p.129),
a rare trait amongst non-human terrestrial mammals. Janik and Slater (1997; 2000) distinguished *production* learning from *contextual* learning, a more common form of learning that affects the usage and comprehension of sounds, whereby individuals learn to associate existing signals with their appropriate context.

Vocal development need not be socially guided. Acoustic deprivation experiments have demonstrated that learning and auditory feedback seem to play no part in normal song ontogeny in suboscine birds; at least for the three studied species (alder flycatcher, *Empidonax alnorum*, and willow flycatcher, *Empidonax traillii*: Kroodsma 1984; eastern phoebe, *Sayornis phoebe*: Kroodsma and Konishi 1991). In contrast, it is well established that all studied species of oscine songbirds learn at least some aspects of their species-specific song (e.g. Marler and Mundinger 1971; Kroodsma 1982; Slater 1989; Baptista 1996). Social experiences can influence call learning not only in young birds, but in adult songbirds (e.g. great tits, *Parus major*: McGregor and Krebs 1989; cedar, *Serinus canaria*: Lehongre et al. 2009). Several species of parrots (e.g. Todt 1975; Farabaugh et al. 1994; Vehrencamp et al. 2003) and hummingbirds (e.g. Baptista and Schuchmann 1990) are also on the growing list of flexible vocal learners.


What similarities do the diverse taxa of vocal learners share? They tend to form large social groups that are often fission-fusion groups (whereby individuals regularly split from the larger group or form small subgroups); they are highly mobile and can lose sight of one another
easily (especially in three-dimensional space such as water or air), typically communicating across long distances; they live in high background-noise environments (e.g. parrots, marine mammals, bats, and humans), and they tend to have individual or group specific calls that facilitate the identification of individuals or groups by their call alone (summarized in Crockford et al. 2004).

The prevailing speculation to explain the widespread occurrence of vocal production learning in marine mammals is that this trait facilitates individual or group recognition in aquatic environments, which is fundamental for the maintenance of the individual-specific relationships common in social odontocetes (Tyack 1991; Tyack and Clark 2000). Bottlenose dolphins are a classic example. The fission-fusion nature of their society, in which individuals form strong long term bonds with conspecifics but within very fluid groupings (Connor and Smolker 1985; Wells et al. 1987; Smolker et al. 1992) is thought to be intimately related to their vocal plasticity and system of learned signature whistles for individual recognition. Most terrestrial animals can probably recognize each other acoustically simply through the individual voice characteristics caused by subtle morphological differences in their vocal tracts. In diving mammals, by contrast, these involuntary structural characteristics of calls are not likely to be as reliable for individual recognition (Tyack 1991; 1997). The high background noise of the marine environment can mask those minor differences between calls (Janik 1999a).

Furthermore, when an animal dives, depth-induced changes in the gas-filled vocal tract can overshadow the subtle differences in voice characteristics (Tyack 1991; 1997). For instance, Ridgway (1997) documented changes in the frequency spectra of whistles produced by individual belugas calling at varying depths. Thus, if diving mammals rely upon individually distinctive calls, they may need to create them by learning to modify the acoustic characteristics of their vocalizations (Tyack 1991).

Regardless of the evolutionary origins of vocal learning, both production and contextual vocal learning have considerable functional consequences in the social environment. For instance, individuals could address conspecifics through imitation (e.g. vocal matching in bottlenose dolphins: Janik 2000), or referentially label objects, food, conspecifics, or danger with acoustically different vocalizations. Predator alarm calls of vervet monkeys, *Cercopithecus aethiops* (Seyfarth et al. 1980) are an example of the latter, although data are stronger for contextual learning than production learning (see Seyfarth and Cheney 1986).

Despite the strong evidence for vocal learning in marine mammals, Tyack (1997) noted a dearth of studies of vocal development in this mammalian group, a gap that is still apparent
today. Ideally, disentangling the role of learning in repertoire development would involve longitudinal studies of the vocalizations produced by individual animals as they age, along with recordings of the vocalizations that these animals hear from known adults in the social group. Additionally, in order to separate inheritance from experience we would have to know the genetic relationship between the animals in the social group. For cetacean species in the wild, there are notable logistic difficulties in determining the age, sex, and genetic relationships of individuals and resighting them reliably for months or years (Tyack 1997). A capture-release program of individual bottlenose dolphins near Sarasota, Florida, has made it possible to collect such kinds of data, allowing Sayigh (1992) to study the development of individually distinctive signature whistles in this species. Her data suggest that learning plays a role in determining whistle structure. More recently, Sanvito et al. (2007) documented the vocal development of 29 identified male elephant seals, and showed that the relative frequency of vocalization types changed over eight years, as young peripheral males adopted the vocal types of older more successful breeders, strongly suggesting vocal learning.

In general, however, longitudinal studies of repertoire development are easier to conduct in captivity than in the wild (e.g. Caldwell and Caldwell 1979; McCowan and Reiss 1995). One of the advantages of studying vocal development in captive cetaceans is that calves are typically exposed to some genetically unrelated individuals from different geographic locations or pods. McCowan and Reiss (1997) suggested that for bottlenose dolphins, this artificial mix of individuals facilitates the documentation of non-inheritable whistle acquisition. In fact, Tyack (1997) noted that cross-fostering experiments (isolating subjects from conspecific vocalizations by having adults of another species care for the infants, a common research approach with songbirds) may occur naturally with captive bottlenose dolphins when mothers fail to provide proper care and infants are raised by non-related individuals. Tyack describes a successful cross-fostering of a 1-2 month old bottlenose dolphin calf who adopted the signature whistle of the adult female who took on most of the roles of a dolphin mother except for nursing.

Evidence of vocal production learning in adult animals derives from cases of imitation of unusual signals by animals, from studies documenting the convergence of acoustic features of calls among adults (Tyack 2008b), and from studies of contingency learning -- the modification of behaviour through selective reinforcement (Schusterman 2008). According to Tyack (2008b), evidence of production learning should also include cases of animals adapting the acoustic features of their calls to communicate in noise. Imitation is believed by some to constitute the
clearest evidence of production learning (see Janik and Slater 1997; Egnor and Hauser 2004; Schusterman 2008; and Tyack 2008b for reviews).

Mimicry and vocal learning

*I cannot doubt that language owes its origin to the imitation and modification, aided by signs and gestures, of various natural sounds, the voices of other animals, and man’s own distinctive cries.* (Darwin, 1871, p.56).

The importance of mimicry in vocal learning has been well established in the acquisition of human language (e.g. Bloom et al. 1974) and in the development of avian song (Marler 1970; Slater 1989). The most extensive experimental evidence for the relationship between mimicry and vocal learning in marine mammals comes from a series of studies pioneered by John Lilly in the early 1960s. Lilly (1961; 1965; 1967) demonstrated that dolphins could learn to change their vocal emissions from natural patterns, by training captive bottlenose dolphins to emit sounds from the blowhole, in air, that closely mimicked model human sounds.

Lilly’s studies were based on his belief of a true language in the natural communication system of dolphins. He promoted this hypothesis despite a lack of scientifically valid evidence, and was extensively criticized for this (e.g. Andrew 1962; Herman and Tavolga 1980; see Samuels and Tyack 2000 for a thorough discussion of the merits, shortcomings and scientific and social consequences of Lilly’s research). Yet, the controversy surrounding his work acted as a catalyst for extremely productive dolphin mimicry research in the next 30 years (Herman and Thompson 1982; Richards et al. 1984; Herman 1986; Sigurdson 1989; Reiss and McCowan 1993; Sigurdson 1993). For example, Richards et al. (1984) trained a dolphin to mimic a variety of computer-generated sounds with high reliability, and later to use them to label a variety of objects by producing the sounds upon presentation of the objects.

Evidence that marine mammals can learn new sounds that are not part of their natural repertoire can also be found in reports of un-trained imitation. For example, two male harbour seals at the New England Aquarium imitated human speech sounds spontaneously (Ralls et al. 1985); one juvenile killer whale mimicked the barks of California sea lions (*Zalophus californianus*; Foote et al, 2006); and two bottlenose dolphins raised at the Miami Seaquarium in close social contact with their trainers incorporated an accurate rendition of their trainer’s whistle into their vocal repertoire (Tyack and Sayigh 1997; see also Miksis et al. 2002). Of most
relevance to my study is the anecdotal evidence of vocal imitation in belugas (Eaton 1979; Ridgway et al. 1985), detailed in the next section. The significance of mimicry experiments and anecdotes of marine mammals imitating foreign sounds rests on the fact that they permit us to target taxa suitable for the systematic study of vocal learning (Tyack 1993). As we shall see below, belugas are clearly one such species.

1.2 What do we know about vocal learning in belugas?

The only evidence in nature of belugas modifying their signals based on auditory input is a report of significant upward shifts in the mean frequencies of calls used by belugas in the St. Lawrence estuary in the presence of low frequency vessel noise (Lesage et al. 1999). Given the consistency in the direction of the frequency shifts, and the fact that they occurred during the short periods of exposure to high noise levels, the frequency shifts were unlikely explained by chance alone (Lesage et al. 1999). According to Tyack (2008b), when animals alter the parameters of their calls in this way to communicate in noise, they show production learning, because they are altering their vocal output, producing a modified signal, based on what they hear (noise). Tyack reasons that since noise compensation mechanisms that modify vocal behaviour may have played a role in the evolution of vocal learning, this ubiquitous phenomenon (e.g. great tits, Parus major: Slabbekoorn and Peet 2003; nightingales, Luscinia megarhynchos: Brumm 2004; right whales, Eubalaena australis: Parks et al 2007; manatees, Trichechus manatus: Miksis-Olds and Tyack 2009) should be considered in discussions of vocal production learning in order to understand the evolution of this trait. While the point brought up by Tyack undoubtedly has merit, some caution seems warranted. In this broader view of vocal learning, this trait could no longer be considered an aspect of social learning, because an animal that produces a higher pitch vocalization in response to noise does not necessarily acquire it by hearing other individuals vocalizing with higher pitch. It seems reasonable to include this phenomenon in discussions of vocal learning, but it might be more productive to call it phenotypic plasticity in sound production, and save the term vocal learning for true acquisition of new sounds from other individuals.

More robust evidence of vocal production learning in belugas in the traditional sense of the term comes from reports of captive whales mimicking human speech and other sounds. According to Eaton (1979), a 15-year-old beluga at the Vancouver Aquarium produced utterances that resembled human speech. Another 9-year old male beluga, held at San Diego
Bay, imitated the sound of human conversations, and did so with his head out of the water (Ridgway et al. 1985). This behaviour occurred frequently enough that the trainers were able to solicit it with a fish reward, so that his vocalizations could more easily be recorded and analyzed.

Similarly, an unpublished beluga mimicry study conducted at the Chicago Shedd Aquarium, by Gregory Dye (personal communication, November 2002)\(^3\), provides a compelling account of vocal production learning in this species. The idea for Dye’s project originated from a hearing threshold study that involved training a Pacific white-sided dolphin to perform a series of tasks when played particular tones. A hand cue indicated to the dolphin to “station under a hoop”, a task that she quickly understood and performed well. She was then taught to remain under the hoop until she heard a computer generated tone. But quite suddenly the dolphin began responding as though she had heard a tone when none had been generated by the computer. After several failed data collection sessions, the trainers put a hydrophone in the pool to determine if the broadcasting of the tone underwater was malfunctioning. They discovered that the belugas that shared the enclosure with the dolphin (separated by an acoustically transparent mesh) had learned the tone and were imitating it throughout the day, likely confusing the dolphin. This episode sparked Dye’s interest in vocal mimicry in belugas, and he initiated a controlled study. He first gathered 40 hours of baseline data on the vocalizations produced by the four captive belugas. He found that some of the sounds were good imitations of some bird calls in a CD of Pacific Northwest Birds (e.g. a loon), which the whales were able to hear daily as it played loudly near their pool. Dye then generated three unique artificial tones with a computer, and played them in the beluga exhibit every 90 seconds for two hours every day, for a total of 10 sessions each sound. He introduced the tones one at a time. The whales soon copied each tone, eventually producing all three tones consecutively in the same vocalization (Dye, unpublished data).

These observations of spontaneous imitations in captive belugas raise the question about the selective pressures responsible for such a proclivity for vocal mimicry. There is, undoubtedly, a historical legacy: vocal learning in belugas fits well with the growing evidence of social learning and cultural transmission in this taxonomic group (e.g. Rendell and Whitehead 2001). In cetaceans, social learning coupled with a tendency for vocal plasticity is evident in species with the following characteristics: strong mother-infant bonds, extended maturation

\(^3\) Gregory Dye is currently Director of Operations for the Duke Lemur Center.
period, long life spans, consistent social groupings where repeated interactions with the same individuals occur, and multi-generation family groups (Brodie 1969; Smolker et al. 1993; McCowan and Reiss 1997; Boran and Heimlich 1999; Mann and Smuts 1999). The next section describes the socio-ecology of belugas and sheds some light on the question of how this particular species’ environment and social organization provide a context for vocal learning.

More specifically, we may ask: what is the role of vocal learning in the development of beluga calls? The ability of adult individuals to imitate sounds that are not part of their repertoire suggests that they may use vocal learning in developing their natural vocalizations (Tyack 1993). Longitudinal studies of repertoire development to determine the role of learning in the emergence of beluga calls and contingency learning experiments to understand the vocal flexibility of adults may contribute greatly to this area of inquiry. In this dissertation I initiated the systematic study of vocal learning in this species using both types of approaches.

1.3 What do we know about beluga social structure?

“The different structures of cetacean societies create different kinds of problems of social living, and there appears to be a close connection between the structure of a cetacean society and the kinds of social communication that predominate in it” Tyack (1998, p. 199).

A clear understanding of beluga sociality continues to be elusive. We know that they are highly gregarious odontocetes with a complex social structure and a correspondingly rich vocal repertoire. The latter is not surprising, considering the general relationship between high sociality and communicative complexity (Marler 1977; Blumstein and Armitage 1997) and the key role that acoustic communication plays in aquatic environments. A patchwork of information has provided us with a picture – if dim – of beluga society. I will next summarize some elements of the socio-ecology of belugas that bear on their vocal behaviour.

Belugas are long lived, highly gregarious, circumpolar odontocetes. In Canada, there are seven populations, according to the most recent available evidence (COSEWIC 2004): (1) the St. Lawrence Estuary population which is at the southern limit of their distribution; (2) the Ungava Bay population; (3) the Eastern Hudson Bay population; (4) the Western Hudson Bay population; (5) the Eastern High Arctic – Baffin Bay population; (6) the Cumberland Sound population; and (7) the Eastern Beaufort Sea population. In all cases belugas migrate from overwintering areas of broken pack ice to spring and summer calving and feeding areas, usually
shallow river estuaries (COSEWIC 2004). Genetic evidence indicates strong philopatry of females and their calves (Brennin et al. 1997), with longitudinal data on cow-calf pairs remaining together for many years (Michaud, unpublished data). It may be important for young belugas to travel with their mothers for a period of several years to learn the migration route, the overwinter areas, where to eat, and where to spend the summer (Brennin 1992), amongst other skills (e.g. maternal skills). Summer aggregations are in general separated into herds of males and large nursery groups of females, calves, and juveniles occupying slightly different areas (Michaud 1993; Smith et al. 1994; Smith and Martin 1994). A long term photo-identification study of the St. Lawrence estuary belugas in their summering range has provided the first detailed portrait of social structure and behaviour in this species, revealing a fission-fusion grouping pattern, and long-lasting male-male associations and alliances (Michaud 2005).

Michaud (2005) found a difference in association patterns between the sexes, and refers to this social-system model as a mix of constant companions (males) and casual acquaintances (females). He documented two instances when a pair and a trio of males chased and “captured” a female for a few hours and were observed to copulate with her repeatedly, and three additional instances of groups of males herding a female, where penises and genital slits were observed but copulation was less clear. Michaud attributes this apparent alliance formation between males to a number of conditions that are met in belugas: a high population density, a clumped distribution of females and a high operational sex ratio (resulting from slow reproductive schedule) all tend to increase the rate of encounter between males that are potential competitors for mates.

Twenty years of studies of belugas in the Russian White Sea and the Amur River Estuary also suggest that groups of belugas stay together over at least several years (Bel’kovitch and Sh’ekotov 1993). In this region, beluga aggregations separate into herds of males and large nursery groups of females, calves, and juveniles, which occupy slightly different areas during the summer months. Small groups of belugas ranging from 2 to 8 individuals temporarily separate from the larger aggregation, which is reminiscent of the fluid fission-fusion structure of bottlenose dolphin societies. These smaller groups are likely family units representing more than one generation – females with immature animals of different ages – and are often resighted in the observation region during the same and different seasons (unfortunately, Bel’kovitch and Sh’ekotov 1993 do not specify how often these groups are resighted). During migration, groups reunite in larger aggregations of up to several hundred individuals (Bel’kovitch and Sh’ekotov 1993).
The mother-calf bond in belugas is probably as strong and long lasting as in bottlenose dolphins. There are a few reports of the length of the suckling period in belugas (e.g. Brodie 1971; Drinnan and Sadleir 1981). Brodie (1971) inferred that lactation lasts approximately 24 months, based on inspection of reproductive tracts, and on the associations of calves and adult females in capture nets. However, in captivity beluga whales are known to nurse for much longer than that. At the Vancouver Aquarium, Qila, captive-born daughter of Aurora (see genealogy in Chapters 2 and 3), was observed nursing until she was six years of age (B. Sheehan, Curator of Marine Mammals, Vancouver Aquarium, personal communication). Tuvaq, a male calf born in 2002, nursed until he was three years old not only from his mother Aurora, but also from his half sibling, Qila, and from an unrelated female, Allua, both of whom began lactating despite not having calves of their own (Leung et al. 2010).

Altruism and cooperation by belugas has been observed in the form of mutual assistance in cases of danger (whaling operations; Bel’kovich and Sh’ekotov 1987, cited in Bel’kovich and Sh’ekotov 1993), cooperative feeding strategies, the presence of “scouts” (Bel’kovich and Sh’ekotov 1993) and alloparental care (captive belugas: Drinnan and Sadleir 1981; Leung et al. 2010; wild belugas: Bel'kovich and Sh'ekotov 1993). The reported cooperative feeding formations are named “carousels” and are characterized by a high level of acoustic activity (Bel’kovich and Sh’ekotov 1993). This hints at the important role that acoustic communication may play to facilitate the coordination of such formations in the ocean, although this has not been demonstrated.

Behavioural differences between groups of belugas in different areas have also been reported. Bel’kovich and Sh’ekotov (1993) observed different hunting strategies for White Sea and Amur belugas, which they attributed to dissimilarities in the type and the behaviour of prey. Such differing hunting strategies were coupled with use of different patterns of echolocation signals, which they classified into series types. The relative proportion of the main echolocation series types differed between the areas, an indication of vocal flexibility. Additionally, White Sea juveniles are quite independent, with larger distances between juveniles and females than Amur belugas. Only White Sea belugas form “kindergarten” groups composed of juveniles of different ages. Such groups in which calves are tended only by immature animals for long periods of time (2-3 hours), are unique among odontocetes (Bel’kovich and Sh’ekotov 1993).

In sum, it appears that beluga society has qualities that would foster vocal learning and plasticity: strong mother-infant bonds, long lives and extended maturation periods (traits that provide the young time to acquire many of their abilities through learning, Janik 1999), mixed
age class groups, long term relationships in the context of fluid structures, cooperation, and marked behavioural variability.

1.4 What do we know about the acoustic behaviour of belugas?

The astoundingly diverse sound repertoire of belugas is composed of the predominant sound types among toothed whales: (1) whistles, or narrow-band, frequency modulated vocalizations, believed to be social signals, and (2) pulsed sounds, or trains of broad band pulses. The latter can in turn be divided into two functional categories: click trains, where each click is of very short duration, used largely for echolocation, and burst pulse sounds, or bursts of pulses with rapid pulse repetition rates, which may sound to the human ear like grunts, squawks, screams, whines and even whistles. Some researchers (e.g. Faucher 1988, Karlsen et al. 2002; Belikov and Belkovitch, 2003; Vergara and Barrett-Lennard 2008) identified mixed calls in belugas, consisting of either a whistle and a pulsed component, or two pulsed sounds with different pulse repetition rates, produced synchronously in the same vocalization.

The mixed vocalizations are likely produced by the two twin phonating apparatuses described by Cranford et al. (1996; Chapter 2). The bird literature refers to this production of two simultaneous sounds by two independent sound sources as “two-voiced” (Greenwalt 1965, Suthers 1990, Zollinger 2008) to distinguish this phenomenon from biphonation, when a single source produces two simultaneous signals. In the cetacean literature, however, no distinction has been made between two-voiced and biphonic sound production (Brown 2008), and calls showing two independent fundamental frequencies simultaneously in a call spectrum are referred to as “biphonic” despite the possibility that they can originate from two sound sources (killer whales: Filatova et al. 2009; killer whales and right whales, *Eubalaena glacialis*: Tyson et al. 2007). Zollinger et al (2008) emphasized that although the physical basis of biphonic and two-voiced vocalizations is different, spectrographic analysis alone cannot always reliably identify the production mechanism. In this dissertation I thus refer to these vocalizations with the neutral term “mixed calls”, which is used in the beluga literature (Faucher 1988, Recchia 1994, Angiel 1997; Belikov and Belkovitch 2003; “combined calls” in Karlsen 2002).

Despite the fact that some cetaceans, such as sperm whales (*Physeter macrocephalus*), use click trains for both navigational and social purposes (Whitehead and Weilgart 1991), such social use of click trains is largely unexplored in belugas and other odontocetes. Furthermore, in belugas, as in most odontocetes, echolocation clicks and burst pulsed sounds merge into the
other, which makes it difficult to know when they are using clicks for echolocation or for communication. Au and Hastings (2009) noted that the distinction between the two kinds of sound may be best found in the inter-click intervals. At least in dolphins, click trains used for echolocation have an inter-click interval that is longer than the two-way travel time for the signal to travel from the animal to the target and back, whereas burst pulse sounds tend to have shorter inter-click intervals. But the distinction is hazier in belugas, which can emit “packets” of clicks with inter-click intervals less than the two-way travel time to the sonar target (Turl and Penner 1989).

The echolocation abilities of belugas have been extensively studied (Gurevich and Evans 1976; Au 1985; Penner et al. 1986; Au et al. 1987; Turl et al. 1987; Turl and Penner 1989; Turl et al. 1991; Lammers and Castellote 2009), revealing a sophisticated sonar system. Belugas have better abilities than bottlenose dolphins to discriminate targets in clutter (Turl et al. 1991) and in the presence of masking noise (Turl et al. 1987). The individual pulses that make up echolocation click series can extend from 100 Hz to 120 kHz, concentrating on the range between 20 and 60 kHz (Gurevich and Evans 1976; Au 1985), beyond human hearing.

In contrast to the echolocation abilities of belugas, their communicative system is little understood. This is especially noticeable in comparison to the impressive body of work on the communication system of other odontocetes such as killer whales, bottlenose dolphins (see Janik 2009 for review on delphinids), and sperm whales (e.g. Weilgart and Whitehead 1997; Jaquet et al. 2001; Madsen et al. 2002; 2003; Schulz et al. 2008). Schevill and Lawrence (1949) were the first to present a cursory description of the acoustic repertoire of the St Lawrence belugas, without attempting a formal classification. They emphasized the loquaciousness of this species, using terms such as whistles, squeals, mewing, chirps, bell-like calls, and sounds “reminiscent of a string orchestra tuning up” and of “a crowd of children shouting in the distance” (p. 143). Since then, there have been a number of attempts to classify the vocal repertoire of belugas, summarized in Table 1. It is immediately apparent, when one examines this body of data, that the classification of the vocal repertoire of belugas into principal call types has suffered, as have other cetacean studies, from classification inconsistencies between researchers, rendering anywhere from nine (Fish and Mowbray 1962) to 52 call types (Belikov and Bel'kovich 2006; 2007; 2008).
<table>
<thead>
<tr>
<th>Authors</th>
<th>Location</th>
<th>Time frame</th>
<th># calls analyzed</th>
<th>Classification method</th>
<th># Call types classified</th>
<th>Pulsed types</th>
<th>Whistle types</th>
<th>Mixed types</th>
<th>Observations</th>
<th>Advances on function of call types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish &amp; Mowbray (1962)</td>
<td>Captive. Three animals from Bristol Bay, Alaska.</td>
<td>24 hours</td>
<td>206</td>
<td>Aural impressions, visual inspection of spectrograms</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>Large variation within single sound categories produced by one individual and different individuals.</td>
<td>Not enough data to correlate behaviour with vocalizations.</td>
<td></td>
</tr>
<tr>
<td>Sjare &amp; Smith (1986b; 1986a)</td>
<td>Cunningham Inlet, Nunavut, Canada</td>
<td>Multi-year study</td>
<td>1243</td>
<td>Visual inspection of spectrograms</td>
<td>21</td>
<td>5</td>
<td>16</td>
<td>Pulsed call categories include click series, pulsed tones, and noisy calls. Graded system.</td>
<td>General relationship between usage rates of broad call categories and group activity: resting, directive swimming, social interactions, and alarm situations.</td>
<td></td>
</tr>
<tr>
<td>Faucher (1988)</td>
<td>St. Lawrence Estuary, Canada</td>
<td>One summer</td>
<td>689</td>
<td>Aural impressions, spectral analysis</td>
<td>24</td>
<td>8</td>
<td>16</td>
<td>Commonly heard mixed calls (included under “pulsed call types”). Graded continuum of whistles and pulsed sounds.</td>
<td>General relationship between usage rates of broad call categories and behavioural, social and environmental contexts.</td>
<td></td>
</tr>
<tr>
<td>Bel'kovich &amp; Sh'ekotov (1993)</td>
<td>White Sea and Amur Estuary, Russia</td>
<td>Multi-year study</td>
<td>26,000</td>
<td>Aural impressions, spectral analysis</td>
<td>35</td>
<td></td>
<td></td>
<td>Same call types in both study areas, but different proportional use. Within category sub-types also described. Not always specified if call type was pulsed or whistle.</td>
<td>Listed relative proportions of main signal classes during a) searching, hunting, resting, excited groups, nursery groups; and b) ‘close communication’ and ‘distant communication’.</td>
<td></td>
</tr>
<tr>
<td>Recchia (1994)</td>
<td>Captive. Three aquaria. Twelve belugas from Hudson Bay, Canada, one captive-born.</td>
<td>Two and a half months (March - May)</td>
<td>1134</td>
<td>Aural impressions, spectral analysis, and discriminant function and cluster analysis</td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>Graded system, with general call types shifting into each other, at least to human perception.</td>
<td>No inferences about the biological significance of the vocalizations.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1: Summary of studies that attempted a classification of the acoustic repertoire of belugas. The last column indicates in what way - if any - each study contributed to understanding the function of beluga call types (continues on next page).
<table>
<thead>
<tr>
<th>Authors</th>
<th>Location</th>
<th>Time frame</th>
<th># calls analyzed</th>
<th>Classification method</th>
<th># Call types classified</th>
<th>Pulsed types</th>
<th>Whistle types</th>
<th>Mixed types</th>
<th>Observations</th>
<th>Advances on function of call types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiel, 1997</td>
<td>Bristol Bay, Alaska</td>
<td>12 days</td>
<td>1288</td>
<td>Aural impressions, spectral analysis, and cluster analysis</td>
<td>24</td>
<td>12</td>
<td>12</td>
<td></td>
<td>Mixed calls with a pulsed and a tonal component were classified as pulsed. Showed that there is a significant difference among two independent observers in call categorization, and that cluster analysis provides a less refined categorization than that of a human.</td>
<td>No attempt to correlate behaviour with vocalizations.</td>
</tr>
<tr>
<td>Karlsen et al. (2002)</td>
<td>Svalbard, Norway</td>
<td>19 days (August)</td>
<td>1694</td>
<td>Aural impressions, spectral analysis, and cluster analysis</td>
<td>21</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>Based on only two encounters with vocal whales, all-male groups in both cases.</td>
<td>General relationship between overall vocalization rate and three classes of behaviour: travelling, milling and joining.</td>
</tr>
</tbody>
</table>

**Table 1.1 (cont.):** Summary of studies that attempted a classification of the acoustic repertoire of belugas. The last column indicates in what way - if any - each study contributed to understanding the function of beluga call types.
A number of factors can explain the observed classification inconsistencies. Belikov and Bel’kovitch (2008) attribute such inconsistencies to the fact that some researchers do not separate calls into specific, discrete types, but only distinguish broader groupings of signals with similar elements. Another source of discrepancies may be the large difference in total number of calls sampled by different studies, and the fact that the sampled vocalizations come from diverse beluga populations in different geographical areas, and from different group sizes and compositions. Also, if the vocal behaviour of belugas varies seasonally, as my years of recordings at the Vancouver Aquarium suggest, studies conducted at different times of the year would render different results. Comparisons across studies are further hindered by differences in terminology. Some researchers (e.g. Sjare and Smith 1986a) use the term “whistles” only for true tonal signals (those with harmonics that are integral multiples of the fundamental frequency), while other researchers use aural impressions of tonality (e.g. Fish and Mowbray 1962; Recchia 1994), including both true tonals and signals that have a sideband structure characteristic of burst-pulsed signals (for a thorough explanation of the difference see Watkins 1967). Finally, the fact that the vocal repertoire of belugas tends to be described as graded (Sjare and Smith 1986a; Faucher 1988; Recchia 1994; Karlsen et al. 2002), also makes it difficult to classify calls and to compare classes across studies. Graded repertoires exhibit continuous variation between signal classes, with no clear acoustic demarcation between call types (Hammerschmidt and Fischer 1998). The great variability within signal classes means that authors “struggle to capture the essence of a given category with one or two sample spectrograms” (Recchia 1994, p. 180).

Although the repertoire of belugas appears to be of a generally graded nature, some studies (including this dissertation) have found repeated series of stereotyped sounds (see glossary). Faucher (1988) found that some whistle types of St. Lawrence belugas tended to be repeated consecutively (and were stereotyped), and others did not. Sjare and Smith (1986a, b) also described stereotyped whistle series from Arctic belugas. Series of stereotypic sounds present, on the one hand, considerable analytical challenges because determining the unit of analysis becomes problematic (reviewed in Hauser 1996), and on the other hand, certain advantages, since they provide multiple, identifiable replicates of a sound pattern that can be used to test methods of categorization (Janik 1999; Deecke and Janik 2006).

The study of the social functions of beluga calls has been hampered by a key methodological problem: our difficulty in localizing sound sources underwater, which must be
overcome to determine which animal in a group is producing a sound (Tyack and Recchia 1991). Efforts to address this problem in studies of other cetacean species (e.g. dataloggers: Tyack and Recchia 1991; hydrophone arrays: Miller and Tyack 1998; Thomas et al. 2002; Schulz et al. 2006), have not been applied to beluga communication research. Thus, inferences about the function of beluga calls have been thus far very general. Some of the studies listed in Table 1 have made progress in correlating vocalization rates and broad call classes with general behavioural states. For example, “chirps” (a broad category that can encompass many different call types) have been found to be a good acoustic indicator of sexual and social behaviour (Belikov and Bel'kovich 2003). Even more broadly, studies in several regions found that belugas are most vocal during social interaction (e.g. Cunningham Inlet, Canada: Sjare and Smith 1986a; Svalbard, Norway: Karlsten et al 2002; White Sea, Russia: Belikov and Bel’Kovich, 2003); others reported a decrease in overall vocal activity and an increase in the repetition of specific calls when disturbed or threatened (Sjare and Smith 1986a; Finley et al, 1990; Lesage et al 1999).

Two studies attempted to understand the biological significance of beluga calls without categorizing their entire repertoire of sounds (thus not listed in Table 1), but using instead more focused approaches. Morgan (1979) used conspecific playbacks to both captive and free-ranging belugas in the St. Lawrence estuary. Playbacks to free-ranging whales caused a decrease in vocal activity, and playback of one particular stereotyped captive call type to the same individual from which it was originally recorded elicited a call-matching response. The author had no prior knowledge of the general contexts in which the call types selected for playback tended to occur, and thus could not test hypotheses about the function of these sounds by predicting particular responses. Van Parijs et al. (2003) investigated the sounds produced by individual whales during capture events in Svalbard, Norway, and identified various call types that may serve as “contact calls”, including harmonic sounds for a solitary calf, and broadband “crooning” click trains for the mother in a mother-calf pair.

In sum, the body of data presented here provides some indication that the sounds vary with behavioural and group context, and suggests geographic variation in signal use among populations. In general, however, the studies have been primarily descriptive, with no major efforts to provide a functional analysis of specific call types. The biological significance of the communication signals of belugas remains poorly understood.
1.5 Thesis objectives and organization

The literature review presented here indicates that the function of beluga calls, the ontogeny of their repertoire, and the role of learning in vocal behaviour are inadequately understood. This dissertation addresses these gaps in our knowledge of the vocal behaviour of belugas. It concentrates primarily on a captive beluga group at the Vancouver Aquarium, and uses some of the findings as a springboard for generating ideas for research in the wild.

The dissertation consists of four chapters, in addition to the present one. In Chapter 2 I examine the development of calls of one male beluga calf at the Vancouver Aquarium, systematically throughout his first year of life and opportunistically through his second and third years, and include recent preliminary findings from two additional calves. This is important given the lack of knowledge about the ontogeny of this species’ extensive repertoire, and the role of learning in vocal development.

In Chapter 3 I address the poorly understood function of beluga vocalizations by considering contact calls, ubiquitous in social birds and mammals. I investigate the context-specific use of call types in a captive beluga social group at the Vancouver Aquarium to identify candidates that would function as contact calls. I use my findings to generate testable predictions about the usage of these signals by wild belugas, and attempt to test some of these predictions, opportunistically, in the wild environment.

In Chapter 4 I address vocal learning in adult belugas. The ability to modify and shape the structure and context of vocalizations through learning plays a key role in the social interactions of many species. Given the beluga’s reputation as vocal mimics, and the dearth of published studies addressing their vocal plasticity, I investigate categorical matching, an aspect of call usage learning. I use operant conditioning (see glossary) with food reinforcement to test the ability of belugas to respond to playbacks of different types of beluga calls with calls of the same type, in an effort to gain insight into their vocal control, acoustic perception and categorization of sounds.

Chapter 5 summarizes and briefly discusses the findings of this dissertation and proposes future work.
2. VOCAL DEVELOPMENT IN A BELUGA CALF (*Delphinapterus leucas*)

2.1 Introduction

The highly vocal nature of many species of toothed whales and the importance of sound to mediate their complex social interactions are well known. Yet, knowledge of the ontogeny of vocal repertoires in young odontocetes is scant, being mostly limited to one species, the bottlenose dolphin (Caldwell and Caldwell 1979; Lindhard 1988; Reiss 1988; Sayigh 1992; McCowan and Reiss 1995; Killebrew et al. 2001; Fripp et al. 2005; Morisaka et al. 2005), with the exception of a killer whale study (Bowles et al. 1988) and a description of sound production in neonate sperm whales (Madsen et al. 2003).

Parallels in vocal development between humans, other primate species, and songbirds have been amply documented. For instance, the “babbling” stage of human infants (Locke 1993), consisting of adult-like segments that are not fully formed phonemes, is similar to a stage in infant pigmy marmosets (*Cebuella pygmaea*; Elowson et al. 1998; Snowdon and Elowson 2001). In human infants, babbling is thought to provide necessary practice for language use, and to increase social exchanges with caregivers (Locke 1993; Elowson et al. 1998; Oller 2000). Similarly for marmosets, babbling might be a form of vocal practice and a way of attracting attention from group members (Snowdon and Elowson 2001). The subsong of songbirds, characterized by long, irregular and variable components, is also often compared to human babbling, and is believed to help shape the stereotyped crystallized song (Marler and Peters 1982). Bottlenose dolphin data are ambiguous as to whether they undergo an analogous phase to human babbling (McCowan and Reiss 1995; Tyack 1997).

The importance of vocal learning in the development of human speech (e.g. Locke and Snow 1997) has striking parallels in other vocal systems (e.g. songbirds: Marler and Mundinger 1971; Doupe and Kuhl 1999; bottlenose dolphins: McCowan and Reiss 1995; Tyack 1997). For instance, vocal learning has been shown to be an important factor in the ontogeny of signature whistles (Sayigh 1992; Tyack 1997) and of other whistle types (McCowan and Reiss 1995) in bottlenose dolphins. Both production and contextual learning, defined by Janik and

---

4 A version of this chapter has been published. Vergara, V. and Barrett-Lennard, L. G. (2008). Vocal development in a beluga calf (*Delphinapterus leucas*). Aquatic Mammals 34: 123-143
Slater (2000) and described at length in Chapter 1, can play key functions in the development of an ability to interact vocally with conspecifics.

Belugas are one of the most vocal odontocete species, producing highly varied communication calls (Schevill and Lawrence 1949; Fish and Mowbray 1962; Sjare and Smith 1986a, b; Karlsen et al. 2002), as well as possessing an unparalleled echolocation system (Au 1985; Turl et al. 1987; Turl et al. 1991; see Chapter 1 for further detail). There is much to be gained by understanding how beluga vocal development maps onto what is known of the acquisition of vocal repertoires in other mammals and in birds, especially the role of vocal learning in repertoire development. Descriptions of repertoire ontogeny may illuminate the role of learning in the development of calls and should also provide clues about the physical mechanisms of sound production in odontocetes: an unresolved area of inquiry (Cranford 2000a).

To examine the issues outlined above, I studied the vocal development of a male beluga calf, Tuvaq, at the Vancouver Aquarium, longitudinally from the moment of his birth, throughout his first year of life and opportunistically through his second and third, providing the first account of the sequencing and timing of vocal acquisition in a beluga whale. I identify his mother’s predominant call type, used primarily to maintain contact with her calf, and document the calf’s gradual development of the same call type.

Disentangling learning from other mechanisms in the development of vocal repertoires can be challenging. Robust evidence for vocal learning requires quantification of the acoustic environments and social settings of more than one infant to determine how variation in these characteristics affects vocal development (e.g. Fripp 1999). This was not possible in the current study. However, two adult belugas, one of them the calf’s father, were introduced into the calf’s social group eighteen months after his birth. I was able to isolate the vocalizations of only one of these two animals, the calf’s father, from whom I recorded unique sounds never recorded from any of the other adults. In view of the genetic relatedness of father and son, changes in the calf’s calls in relation to his new auditory contact with his father may not be enough to invoke production learning. Further, maturational change must always be considered. Yet, any new call acquisition after first auditory contact with his father may at least suggest usage learning, in the sense given by Janik and Slater (1997; 2000): that vocal usage learning arises when animals learn to use existing vocalizations in specific contexts based on experiences of other individuals’ use of such signals.
Thus, I also asked whether after first acoustic contact with his father, Tuvaq adjusted his vocalizations to increase their similarity to his father’s sounds. Given the data, reviewed in Chapter 1, that beluga society has qualities that would foster vocal learning and plasticity and the anecdotal evidence of adult vocal learning in this species, I predicted that Tuvaq would learn to use his father’s call types.

The general objective of this study is to shed light on the following four important questions: 1) What is the significance of the changes that occur over the course of repertoire development? 2) Is early vocal production in belugas analogous to babbling in humans and other primates and subsong in birds? 3) What is the role of physical maturation of the vocal apparatus in the structural development of the calls? 4) How might learning mediate vocal development?

2.2 Methods

Social group

The main subject of this study was a male beluga calf, Tuvaq, born on July 20, 2002 at the Vancouver Aquarium. Tuvaq was an immature animal throughout the duration of this study. At the last sampling session at 32 months of age, he was 2.83 m, far from achieving full size (belugas grow to an average length of 4.57 m, and males are larger than females), he had the uniform grey coloration typical of calves, and was still nursing frequently.

Tuvaq was housed only with his 15 year old mother, Aurora, until his third month of life, when two other females, Qila (Aurora’s eight year old daughter, and half-sibling to Tuvaq) and Allua (unrelated, 19 years old) were re-introduced into the group. Tuvaq’s father, Imaq (15 years old), and a 35 year-old unrelated female, Kavna (captive since 1976), were kept in a separate pool in an entirely different area of the facility (thus acoustically isolated from the main pool), for husbandry reasons (this pool is referred to as the ‘back pool’). Imaq and Kavna were introduced into Tuvaq’s social group when he was 18 months old. Except for Tuvaq and Qila, who were captive born, all the Vancouver Aquarium belugas originate from Hudson Bay, near Churchill, Manitoba. Figure 2.1 illustrates the genealogy of the Vancouver Aquarium belugas at the time of this vocal development study.
**Figure 2.1**: Genealogy of the Vancouver Aquarium belugas at the time of this vocal development study. Between brackets are the sex and the age of each individual when Tuvaq was born. Thick black arrows indicate “parent-offspring”.

**Sampling regime and observation area**

The animals were kept in an outdoor pool (18 m x 29 m; depth 6 m) connected to an adjacent 3 m deep smaller medical holding pool (3 m x 3 m). They were observed and recorded from an underwater window with a good view of the entire larger pool, and occasionally from a smaller window with a view of the medical holding pool (Figure 2.2).

I performed 14 hours of continuous recording and observation immediately after birth. Thereafter sessions were 3-4 hours long, at least 3 times a week, for the first 2 months, and later 1-3 hours, twice a week, at variable times of day. After month 12, sessions in the large pool were less regular, and the calf was recorded opportunistically whenever he was in voluntary or involuntary isolation in the medical holding pool (see below).
**Figure 2.2**: Diagram of the beluga pool and the adjacent medical holding pool at the Vancouver Aquarium. A narrow shallow passageway connects both pools, and the animals can voluntarily swim from one to the other. A rope gate (that allows visual and acoustic contact) is used when an animal needs to be temporarily isolated in the medical holding pool.

**Acoustic recordings**

Underwater audio recordings were made using two hydrophones installed permanently in the adjacent tanks (Figure 2.2): an Offshore Acoustics hydrophone in the main pool and a Brüel & Kjaer 8101 hydrophone in the medical holding pool. The calls were recorded digitally on two Pentium IV computers (each connected to a hydrophone) using *Avisoft SASLAB-PRO* software (Avisoft Bioacoustics) and *Cool Edit 2000* at a sampling rate of 44.1 kHz. The system had a response of 0.02 kHz – 22 kHz ± 1 dB. I occasionally had access to a sound card that sampled at 96 kHz, which assisted in determining whether high frequency components were being missed in the 44.1 kHz recordings (the fundamental frequencies were not missed). The noises of the pool filtration system sometimes masked the frequencies below 2 kHz. This did not affect the narrow-band whistle recordings, since only two whistles had a dominant frequency below 2 kHz. For the broad-band pulse-trains and mixed calls, only the energy above 2 kHz was measured for the analysis of energy distribution.

Recordings were made throughout the observation sessions, and all records were saved for later analysis. The time of the sound, its social and behavioural context, and the identity of the vocalizer, if possible, were simultaneously dictated on a mini-cassette recorder. The sounds
were detected aurally via the speakers and visually by examining the running spectrograms on both computers.

**Identification of vocalizing individuals**

An inherent difficulty in cetacean communication research is to identify the vocalizing individual, as cetaceans do not always produce visible signs when they vocalize and localizing underwater sound sources is difficult for humans. I addressed this problem in the following ways:

1) **Sound source localization.** Both hydrophones recorded simultaneously, so it was often possible to discern if a sound originated in the medical holding pool or in the main pool by comparing the amplitude of the same sound on the real-time spectrograms on the two computers because sounds originating in one of the pools were extremely attenuated, or even inaudible, from the other pool’s hydrophone. This allowed for occasional identification of the vocalizing individual when the animal was alone in one of the pools (voluntarily, or for husbandry reasons). Since sounds originating near the passageway between the pools were often clearly audible from both hydrophones, I excluded any identification of an individual in those sounds.

2) **Bubble-streams.** Emissions of streams of bubbles concurrent with sound production in bottlenose dolphin calves facilitate research on vocal development (McCowan & Reiss 1995). They occur most frequently in infants, perhaps due to a lack of full control over the vocal apparatus, although the physiological mechanisms underlying them are inadequately understood. Tuvaq also initially consistently emitted streams of bubbles while vocalizing. Fripp (2005) identified several problems associated with using bubble-streams and cautioned that whistles associated with bubble-streams do not adequately represent a dolphin’s full whistle repertoire. Considering this caveat, I used bubble-streams in conjunction with other methods, and I do not assume that I documented the entire repertoire. I view this study as probing the development of the known, rather than the full repertoire.

3) **Adults feeding or interacting with trainers.** Often the adults had their heads out of the water while being fed by trainers while Tuvaq continued to swim around the pool, sometimes vocalizing. Odontocetes produce sound in their nasal cavities, and project sound waves through the fatty melon (Cranford *et al*., 1996). While feeding and interacting with trainers, both the
nasal cavity and the melon are above the water line, and I assumed that the adults could not produce calls audible underwater at these times (further, the adults were generally silent when being fed, and I was notified by the trainers if any sound had been produced above the water that could have been picked up by the recording system).

During the first few months, concurrent bubble-streams allowed me to discern many of Tuvaq’s vocalizations. As Tuvaq learned to produce sounds without emitting bubbles, I relied more on the other methods of identification, and therefore analyzed fewer vocalizations in later months.

Aurora was the only adult in the pool during the first three months of her calf’s life, thus it was easy to discern her vocalizations from those of her calf by her location in the tank relative to the two hydrophones, or by the acoustic quality of her calls. Two months of baseline recordings prior to the birth of the calf, during which time the remaining four adults were in the pool, familiarized me with the adults’ vocalizations. These have a very different acoustic quality than the calf’s vocalizations, as I shall illustrate in the results section. Further, the calf’s calls (concurrent with bubble streams) often partly overlapped Aurora’s, which allowed me to clearly distinguish between both. When the rest of the animals were introduced, the only adult sounds analyzed for this vocal development study were produced in voluntary or forced isolation (although all adult sounds were recorded and categorized, see Chapter 3). They were thus not representative of the complete adult vocal repertoire, but that was not the focus of the current study.

**Acoustic analysis**

Call parameters were measured automatically with *Raven 1.2* (Cornell Lab of Ornithology), and spectrograms displayed in this dissertation were generated with *Avisoft SASlab Pro* (Avisoft Bioacoustics), using an FFT size of 512 (for pulse-trains) or 1024 (for whistles and mixed calls), a frame-size of 100% with a 75% overlap between frames, and a Hamming window. This resulted in a frequency resolution of 86 Hz and a time resolution of 2.9 ms for pulse-trains, and of 43 Hz and 5.8 ms for whistles and mixed calls. Sounds were classified into four major categories: whistles, pulsed calls, mixed calls and variable calls.

**Pulsed calls**

Due to the large number of pulsed calls recorded, I randomly sub-sampled 20 of the
pulsed calls produced by Tuvaq per month for analysis. To estimate changes in pulse repetition rate (PRR) with age, I calculated mid-call PRR by counting the number of pulses in the middle 0.2 seconds of each call and multiplying by 5 to give pulses per second. I selected the middle of the call for consistency, because PRR was not always constant across calls, the middle of the call usually having the highest PRR. When PRR was so high that pulses did not resolve into vertical bars on a spectrogram, I estimated the parameter from the sideband interval (Watkins 1967). The energy content of the pulse-trains at different ages was also examined to evaluate whether high frequency components were more prevalent later in life. This analysis was limited to the frequencies above 2 kHz to account for the noise of the filtration system, and below 22 kHz (the upper frequency limit of my recording system). Dominant frequency was defined as the frequency of peak energy in the power spectrum (Raven displays the average power spectrum of a selected portion of a sound, computed over its duration). If there were energy peaks of the same magnitude at more than one frequency, the lowest frequency at which peak energy occurred was considered.

Whistles

All whistles that were too faint were excluded from the analysis. As for the pulse-trains above, the estimate of dominant whistle frequency was defined as frequency of peak energy in the average power spectrum of the call. I verified this by examining the whistle contour in the spectrogram, which is of a darker shade at the frequency of peak energy. To measure parameters such as duration, and maximum and minimum frequency, only the fundamental, carrying the dominant frequency, was considered.

Mixed calls

Mixed calls consist of both a whistle and a pulsed component, overlapping in the same vocalization. I measured some parameters of those calls that acoustically resembled Aurora’s stereotyped mixed call type (A1, described later), to help understand Tuvaq’s development of these calls over time. The parameters measured were: PRR of the pulsed component, duration, dominant and start frequencies of the whistle component, and number of inflection points along the whistle component as a measure of its steadiness. I also measured these parameters in 51 Type A1 calls attributable to Aurora (I selected the best quality A1 calls, in terms of signal-to-noise ratio and spectrographic clarity, recorded in different observation sessions from 20 July 2002 to 23 October 2002).
Variable calls

Non-stereotyped chirps, trills and noisy calls (signals having no discernible fundamental frequency and harmonic structure) were assigned to the variable category.

Vocal development after new auditory input

Through the calf’s second year, I examined his vocal development in relation to the introduction of new tank mates, Imaq and Kavna, with different acoustic repertoires. I conducted 11 recording sessions of the two adults that were kept isolated in the back pool, seven to five months before their introduction into Tuvaq’s social group, to identify any call types that I had not recorded from the calf’s acoustic environment. After their introduction when Tuvaq was 18 months, I recorded Tuvaq’s father during six husbandry isolation sessions in the medical holding pool, lasting from 20 minutes to 2 hours, to identify any stereotyped vocalizations attributable to him alone (Kavna, the other new adult, was never isolated during this study, which meant that I could not confirm her call types). Figure 2.3 illustrates the different compositions of Tuvaq’s social group during this study, and the husbandry isolation events during which Tuvaq or his father Imaq were recorded. Allua and Qila were never recorded in isolation prior to their re-introduction in Tuvaq’s pool. In essence, then, I could only look at Tuvaq’s vocal repertoire before and after new auditory contact with his father, which was the only re-introduced animal for whom I had isolated distinct vocalizations.

Statistical analyses

Statistical analyses were performed with SPSS v. 14.0 (SPSS, Chicago, Illinois, U.S.A.). All tests were two-tailed. Measurements are reported here as mean ± standard deviation, except for the number of inflection points in the whistle component of mixed calls, which are reported as median and quartile distance because the number of inflection points constitutes a discrete variable. The similarity of Aurora’s and Tuvaq’s mixed call type A1 was assessed by discriminant function analysis (DFA) using a randomly selected 70% of the data set to compute the discriminant function and cross-validating the results with the remaining 30%. The DFA extracted the parameters that best predicted assigning call type A1 to the appropriate animal.
Figure 2.3: Time line of the different compositions of a beluga calf’s social group at the Vancouver Aquarium throughout the duration of this study, and husbandry isolation events during which the calf (Tuvaq) or his father (Imaq) were recorded (including 11 recording sessions in the back pool in 2003).
2.3 Results

General vocalization types

I recorded a total of 2185 sounds from the calf during his first year of life (1385 pulsed calls, 359 whistles, 350 mixed calls, and 91 variable calls; Figure 2.4). He produced mostly pulse-trains during his first month. His whistle production increased substantially during his second month, and he began to use mixed calls consistently at four months. Tuvaq’s acoustic repertoire broadened as he aged, and non-stereotyped chirps, trills and noisy calls were recognized and assigned to the “variable” category. Although rare, mixed calls in which two pulsed components of drastically different pulse repetition rates overlapped in the same utterance were characterized. Due to their scarcity in the record (4 during 9 months of age, and 2 during 12 months of age), these are not represented here.

Figure 2.4: Changes in the usage of different vocalization types by a captive male beluga calf during his first year of life (n = 2185).
Pulse-train development

The first sounds from Tuvaq were heard half an hour after birth. These were barely audible low energy pulse-trains. Bubble streams were not apparent, but the calls were easily distinguishable from Aurora’s loud, continuous, stereotyped calling, which often partly overlapped the calf’s pulse-trains. The initial pulse-trains had a mean PRR of $17.5 \pm 4.2$ pulses per second, and were of variable duration with a mean of $2.2 \pm 1.7$ s ($n = 10$). Similar low PRR’s characterized all his earlier pulse-trains. Figure 2.5 illustrates one of the earlier pulse-trains, produced the first day of life, and a series of pulse-trains produced on day five which clearly overlap his mother’s broadband, mixed vocalizations (Type A1, described later). Mean PRR increased significantly with age ($R^2 = 0.30$, $n = 89$, $p < 0.0001$, Figure 2.6). The calls progressively acquired the buzz-like aural quality of the adult pulsed calls. The variability of PRR in the pulsed calls (which translates into variability in their aural qualities) also increased with age, as reflected by the larger standard deviation in later months.

![Figure 2.5](image)

**Figure 2.5:** Examples of a pulse-train emitted during the first day of life of a beluga calf born at the Vancouver Aquarium (A) and of a vocal exchange between the calf and his mother on day five (B). Note the calf’s pulse-trains overlapping the broad-band maternal calls in B (FFT length: 512, window: Hamming, frame size: 100%, frequency resolution: 86 Hz, time resolution: 2.9 ms).
During the first few days of Tuvaq’s life, his pulse-trains contained little detectable energy above 10 kHz. The pulse-trains increased in bandwidth through the first month, until their energy was distributed relatively uniformly across the frequency range of the recording system, like the adult calls. An analysis of the energy content of the pulse-trains at different ages showed an increase in the mean dominant frequency of the pulse-trains per recording session during the first 12 months of Tuvaq’s life ($R^2 = 0.45$, $n = 89$, $p < 0.0001$, Figure 2.7). The range of dominant frequencies also increased with age, from 3.4 kHz (Month 1) to 18.6 kHz (Month 8 1/2). In fact, the variability was consistently higher after 6 ½ months of age. As this analysis was limited to frequencies below 22 kHz, real dominant frequencies, at least for older ages (more than 5 months), may be well above 22 kHz.

![Figure 2.6: Mean pulse repetition rate per session ($\pm$ SD) in the pulsed calls of a beluga calf at the Vancouver Aquarium, plotted as a function of age. Data points represent the average of each session (some sessions had only one sub-sampled pulsed call).](image-url)
Figure 2.7: Mean dominant frequency per session (± SD) of pulse-trains emitted by a beluga calf at the Vancouver Aquarium, plotted against his age. Data points represent the average of each session. For sessions with more than one sub-sampled call, the wide standard deviation bars at months 5 ½ to 12 reflect an increase, with age, of the range of dominant frequencies.

Whistle development

Thirteen days after his birth, Tuvaq began producing faint whistle-like elements immediately preceding or following his pulse-trains (Figure 2.8). These whistles were not clearly audible until day 23. Tuvaq may have produced whistle-like elements before day 13 which could have gone unnoticed. However, given my intensive recording schedule during Tuvaq’s first month, this is unlikely.

Figure 2.8: First recorded whistle, at the end of a pulse-train produced by a beluga calf at the Vancouver Aquarium at age 13 days (FFT length: 512, window: Hamming, frame size: 100%, frequency resolution: 86 Hz, time resolution: 2.9 ms).
Figure 2.9 illustrates whistle development with representative spectrograms. Tuvaq’s initial whistles had a low mean dominant frequency (4.34 kHz ± 1.88, n = 9) relative to his whistles in later months and to some adult whistles. They were unstereotyped, and were highly modulated but with irregular frequency modulation (a characteristic that can be described by the terms tremulous or unsteady). There was a weak positive relationship between the average dominant frequency of the whistles in a recording session and age, during the first 12 months of Tuvaq’s life ($R^2=0.10$, $n = 55$, $p = 0.016$) (Figure 2.10). However, this relationship was much stronger when considering only the first five and a half months of life ($R^2=0.81$, $n = 28$, $p < 0.0001$) (Figure 2.10), after which there was a drop in dominant frequency until month 10. The larger standard deviation in later months indicate that as Tuvaq aged, he continued to produce some low frequency whistles but he was also using a larger proportion of higher frequency whistles. In other words, there was an increase in the variability in dominant frequencies of the whistles produced.

Through his first year, Tuvaq’s whistles remained unstereotyped. Unpublished data on proportional use of sounds in the pool in relation to motivational state (agonistic vs. affiliative) suggest that the adult belugas in Tuvaq’s social group typically produce highly stereotyped series of short whistles during agonistic interactions (Figure 2.9, Adult), and unstereotyped high frequency (above 15 kHz) whistles of longer duration during affiliative interactions or non-aggressive states (although these differences in whistle production between agonistic and affiliative states have not been verified statistically at the time of writing). After his third month of life, Tuvaq’s whistles were often produced in series (Figure 2.9), but not until 10 months did the first series of whistles begin to resemble the stereotyped adult series (however, it was not produced, as are the adult stereotyped series, during aggression). After month 10, it was not possible to record a fully stereotyped whistle series attributable with certainty to Tuvaq, thus it remains unclear when he achieved full stereotypy in whistle production.
Figure 2.9: Representative examples of whistle development in a captive beluga calf during his first year of life (FFT length: 1024, window: Hamming, frame size: 100%, frequency resolution: 43 Hz, time resolution: 5.8 ms). Two typical adult whistle series are shown for comparison. During the first two months of life the whistles were low in frequency, of variable length, and rarely produced in series. By month three many of Tuvaq’s whistles were produced in un-stereotyped series. From month four on, the series consisted of long (1.97 ± 1.3 s, n = 29) variable whistles, a pattern that continued until month 10, when I recorded the first whistle series that resembled the adults’ stereotyped series.
Figure 2.10: Mean whistle dominant frequency per session (± SD) plotted against age of a beluga calf at the Vancouver Aquarium. Data points represent the average of each recording session. Note the two trend lines: from birth to 12 months, the $R^2$ is weak, but from birth to five months the relationship between whistle frequency and age is much stronger.

**Mixed calls**

Tuvaq began incorporating a whistle overlapping a few of his pulse-trains on his 20th day of life. However, mixed call production during his first three months of life was minimal. At four months Tuvaq’s mixed call production increased dramatically. Some of his mixed calls progressively resembled the predominant stereotyped mixed call of his mother, Type A1, described below.

**Aurora’s stereotyped Type A1 call**

Tuvaq was frequently exposed to the mixed call Type A1 of his mother, which she began to produce the day after Tuvaq’s birth, and later in every situation where there was an apparent need to regain or maintain contact with him (e.g., forced or voluntary separation, divers in the tank), and in vocal exchanges with her calf (Chapter 3 provides details about the usage of this call type). The Type A1 call aurally resembles a buzz, and is characterized by a pulse-train component with energy distributed more or less equally along the frequency range of the
recording system, and an overlapping tonal component (Figure 2.11). There is little variability in the dominant frequency of the tonal component, with a mean of $14.4 \pm 0.2$ kHz (CV: 1%, $n = 51$). The whistle also exhibits little frequency modulation (median number of inflection points: 1, range 0 to 2), with a simple contour that starts at a mean frequency of $8.8 \pm 2.2$ kHz,

![Figure 2.11: Examples of the calf’s mother’s stereotyped mixed call Type A1 at the Vancouver Aquarium, recorded at 44 kHz (A) and at 96 kHz (B). The dark areas that span the entire frequency range of the spectrograms represent the pulsed component. The tonal component is clearly delineated at about 15 kHz overlapping the pulsed component. Note in B that there is little energy in the first harmonic of the fundamental frequency of the whistle.]

followed by one shift to the dominant frequency of 14.4 kHz (in some renditions of the call, the whistle started directly at around 14.4 kHz). This call type was also occasionally recorded from Tuvaq’s half sibling, Qila (dominant frequency of the tonal component: $14.3 \pm 0.5$ kHz, 0 to 3 inflection points with a median of 1, whistle start frequency: $8.2 \pm 3.3$ kHz, $n = 18$). Aurora tended to produce this call type in series or bouts (see glossary), sometimes in combination with acoustically similar call types lacking the whistle component, A2 and A3 (Figure 2.12). The calls A1, A2 and A3, collectively referred to as “Type A” calls, composed 97% of Aurora’s underwater sounds during the three months between Tuvaq’s birth and the re-introduction of the rest of the whales (i.e., when Tuvaq and his mother were alone in the pool).
During the first two hours immediately after the calf’s birth, Aurora repeated the A3 pulsed call 588 times, while echelon swimming with her calf. The next day she began to incorporate the other two variations, A1 and A2, and from the 3rd day after the birth until the introduction of the rest of the social group, the mixed call type A1 composed 66.7% of her sounds.

Figure 2.12: Example of a bout of the three classes of Type A calls produced by an adult female beluga (Aurora) in the presence of her calf, at the Vancouver Aquarium. Note the whistle component of the three Type A1 mixed calls in this bout, at nearly 15 kHz (FFT length: 1024, window: Hamming, frame size: 100%, overlap: 75%, frequency resolution: 43 Hz, time resolution: 5.8 ms).

The Type A1 call was not recorded during the 18 sessions (1-2 h in length) performed through the two months preceding the birth, when all five adults were in the pool, nor was it heard during the daily continuous monitoring of Aurora throughout the five days immediately before the birth, when she was in isolation (she was completely silent).

Development of the Type A1 call

The apparent contact function of the Type A1 call made it methodologically easier to document its development. Since it was often used in isolation, I was able to record its production from Tuvaq well past the time when he stopped producing bubbles when vocalizing.

Figures 2.13 and 2.14 illustrate the development of Tuvaq’s renditions of the Type A1 call. The reduction in variability with age was clearest for the inflection points in the whistle
component and for its dominant frequency (Figure 2.14). The number of inflection points decreased from month four to month 12 of life ($R^2 = 0.78, N = 9, p < 0.0001$), until the whistle component attained the relatively constant frequency of his mother’s. The coefficient of variation in the dominant frequency of the whistle component also decreased with age ($R^2 = 0.81, N = 9, p < 0.0001$) indicating progressive stereotypy. When Tuvaq was 20 months old I opportunistically recorded a series of 12 calls of Type A1 produced during a brief voluntary isolation in the medical holding pool. These were stereotyped, with the contour of the whistle component clearly resembling that of his mother’s (see Figure 2.13), and little variability in its dominant frequency (1.2% CV), which matched his mother’s calls at 14.35 ± 0.2 kHz.

At 32 months of age, Tuvaq was kept in isolation for husbandry procedures for a period of two hours. He produced a series of 46 Type A1 calls. The parameters of these fully developed calls are strikingly similar to those of his mother’s Type A1 calls (Table 2.1). I performed a discriminant function analysis on Aurora’s 51 A1 calls (see section on “mixed calls” in methods) and Tuvaq’s 46 A1 calls, based on the following parameters: PRR, duration, dominant frequency and start frequency of the whistle component. A model with subset validation classified 83.3% of the calls correctly. The only variables with discriminant ability were the start frequency of the whistle (consistently lower for Tuvaq) and the call duration (consistently longer for Tuvaq). Removing these two variables, reclassification accuracy was much poorer (47.1%). This confirms that Tuvaq and Aurora’s calls are most similar in dominant frequency and PRR.
<table>
<thead>
<tr>
<th>Month 4</th>
<th>Weak whistle component, with irregular frequency modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 8</td>
<td>Louder whistle component, irregular frequency modulation</td>
</tr>
<tr>
<td>Month 11</td>
<td>The frequency modulation of the whistle component resembles that of his mother’s call Type A1 but is still irregular.</td>
</tr>
<tr>
<td>Month 20</td>
<td>Tuvaq’s A1 calls are stereotyped. The modulation of the whistle component is more regular and there is little variability in its dominant frequency.</td>
</tr>
<tr>
<td>Tuvaq’s mother’s series of Type A1 calls</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.13:** Natural sequences of three calls from a beluga calf at the Vancouver Aquarium, each a representative example of the development of the stereotyped mixed call Type A1 (FFT length: 1024, window: Hamming, frame size: 100%, overlap: 75%, frequency resolution: 43 Hz, time resolution: 5.8 ms). Representative examples were not randomly selected, but hand-picked based on low signal-to-noise ratio and spectrographic clarity. Note the similarity between the whistle contour from the calf at 20 months and from his mother (bottom spectrogram).
Figure 2.14: Parameters of the mixed call Type A1, for a beluga calf, Tuvaq (TU), at the ages when this call type was recorded, and for his mother Aurora (AU) for comparison. For the dominant frequency, start frequency and duration, the circles represent the means, the thick error bars represent standard deviations, and the thin bars represent the range, excluding the outliers which are shown as triangles outside the bars. For inflection points, which form a discrete distribution, the circles are the medians, the thick error bars represent the interquartile range, and the thin bars are the range.
Table 2.1: Acoustic parameters of the fully developed Type A1 call produced by the beluga calf, Tuvaq, at 32 months of age, and by his mother, Aurora, at the Vancouver Aquarium.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>TUVAQ(^a)</th>
<th>AURORA(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whistle dominant frequency (kHz)</td>
<td>14.5 ± 0.2</td>
<td>14.4 ± 0.2</td>
</tr>
<tr>
<td>Whistle low frequency (kHz)</td>
<td>6.6 ± 2.1</td>
<td>8.8 ± 2.2</td>
</tr>
<tr>
<td>Call duration (seconds)</td>
<td>1.9 ± 0.6</td>
<td>1.4 ± 03</td>
</tr>
<tr>
<td>PRR of pulsed component (pulses/second)</td>
<td>94.6 ± 14.1</td>
<td>94.4 ± 12.7</td>
</tr>
<tr>
<td>Number of inflection points in the whistle component</td>
<td>Median: 1 (range 1 - 2)</td>
<td>Median: 1 (range 0 – 2)</td>
</tr>
</tbody>
</table>

\(^a\) n = 46  
\(^b\) n = 51

**Acquisition of a novel call type**

My recordings of the two belugas kept in a separate pool during Tuvaq’s first 18 months of life indicated that there were two distinct, stereotyped mixed vocalizations that I had never recorded from Tuvaq’s initial social group, labeled types B1 and B2 (Figure 2.15). After the re-introduction of Imaq and Kavna to the rest of the group, I confirmed that these two call types were produced by Imaq, Tuvaq’s father. I recorded 21 calls of type B1 and 12 calls of type B2 during three isolation events. Both mixed call types had an acoustically similar tonal component, a very audible trill, and an acoustically different pulsed component (that sounded like a trumpet in type B1, and like a clear train of pulses in type B2).

Six months after Tuvaq began to be exposed to his father’s calls, I recorded new mixed vocalizations from Tuvaq that were characterized by the most noticeable element (at least to the author) of Imaq’s mixed call types: a trill component. Eighty-two of these calls, recorded during seven events, acoustically and spectrographically resembled Imaq’s type B2 calls (Figure 2.16), and differed from any vocalization recorded from Tuvaq until then. A random sample of 12 of these calls (to match my sample of Imaq’s 12 calls) reveals that the dominant frequency of the trill component of these calls was the same as Imaq’s (9.6 kHz ± 0.2), with very little variability (Table 2.2). However, the trill was more irregular than his father’s, and the length of each trill segment more variable (CV 52.8%, Table 2.2). The remaining parameters also showed more variability (higher CVs) than Imaq’s calls.
Figure 2.15: Two call types, B1 and B2, produced by an adult male beluga, Imaq, at the Vancouver Aquarium (FFT length: 1024, window: Hamming, frame size: 100%, frequency resolution: 43 Hz, time resolution: 5.8 ms). Types B1 and B2 were occasionally produced in series of two or three, with irregular inter-call intervals. Note the different harmonic structures of the trill component in B1 and B2. Despite these spectrographic differences, the trill in both call types was aurally similar to the author.

Figure 2.16: B2 calls produced by an adult male beluga (Imaq) and two randomly selected renditions by his son (Tuvaq) at 2 years of age, 6 months after their first contact (FFT length: 1024, window: Hamming, frame size: 100%, frequency resolution: 43 Hz, time resolution: 5.8 ms).
Table 2.2: Acoustic parameters of call Type B2 produced by a beluga calf, Tuvaq, and by his father, Imaq, at the Vancouver Aquarium.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Tuvaq(^a)</th>
<th>CV</th>
<th>Imaq(^b)</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trill dominant frequency (kHz)</td>
<td>9.6 ± 0.2</td>
<td>2.1%</td>
<td>9.6 ± 0.4</td>
<td>4.2%</td>
</tr>
<tr>
<td>Trill segment length (seconds)</td>
<td>0.12 ± 0.06</td>
<td>52.8%</td>
<td>0.08 ± 0.01</td>
<td>12.5%</td>
</tr>
<tr>
<td>PRR of pulsed component (pulses/second)</td>
<td>31.4 ± 20.9</td>
<td>66.9%</td>
<td>58.3 ± 19.8</td>
<td>33.9%</td>
</tr>
<tr>
<td>Call duration (seconds)</td>
<td>1.6 ± 0.7</td>
<td>42%</td>
<td>0.77 ± 0.29</td>
<td>37.9%</td>
</tr>
</tbody>
</table>

\(^a\) n = 12
\(^b\) n = 12

2.4 Discussion

The first sounds that a male beluga calf, Tuvaq, produced after birth were exclusively low energy, broadband pulse-trains, which increased in pulse repetition rate with age – preliminary results from two additional calves support these findings. Tuvaq incorporated rudimentary whistles at two weeks. Whistle dominant frequency increased with age, and, at least in his first year, whistles did not attain full stereotypy. He began to use mixed pulsed/tonal calls consistently at four months. Some of those mixed calls became progressively stereotyped, increasingly like his mother’s predominant stereotyped mixed call “Type A1”, which was also used by his half sister. Six months after he was first exposed to his father’s vocalizations, he incorporated one of his father’s call types into his repertoire. I will next discuss these findings in light of current theories of sound production mechanisms in odontocetes, developmental stages of vocal acquisition, and vocal learning.

Early vocalizations and repertoire development

Tuvaq produced only pulse-trains before he incorporated rudimentary whistles at two weeks of age. Similarly, sound production of a neonate captive beluga whale born in 2006 at L’Oceanogràfic, Spain, consisted exclusively of low frequency, short duration broadband pulse-trains (Castellote et al. 2007). Despite differences in populations of origin (Canada vs. Russia), captive facilities, health (the L’Oceanogràfic calf was never vigorous and died at 1 month of age), and vocal context, the sound production observed in these two neonate whales suggests a species specific pattern of this developmental stage in vocal acquisition.
Burst pulse sounds were also the first vocalizations by a captive bottlenose dolphin (Killebrew et al. 2001). Bottlenose dolphins can emit whistles at birth (Caldwell and Caldwell 1979), but some individuals may not produce them until a few days after birth (Killebrew et al. 2001). Caldwell & Caldwell (1979) and Reiss (1988) described infant bottlenose dolphin whistles as “tremulous” and lacking a stereotyped frequency modulation. These descriptors also apply to Tuvaq, who produced similar early whistles. Like a bottlenose dolphin (Killebrew et al. 2001) Tuvaq also placed his first whistle-like elements at the beginning or end of pulse-trains.

Little is known of the development of species-specific stereotyped calls in most cetaceans. Neonate bottlenose dolphins (Caldwell and Caldwell 1979), sperm whales (Madsen et al. 2003), and killer whales (Bowles et al. 1988) produce unstereotyped sounds that are very different from the stereotyped calls of juveniles and adults. Production of stereotyped signature whistles by bottlenose dolphins can occur as early as one to two months after birth, or as late as 17 months (reviewed in Tyack 1997). My study suggests that beluga calves also progress towards more stereotyped calling.

Bowles et al. (1988) reported that the first adult-like calls of a captive killer whale female calf resembled the most common call type produced by her mother. Similarly in my study, the earliest call that structurally resembled an adult stereotyped call type was one of Tuvaq’s mother’s most frequent call types. In the resident population of dolphins in Sarasota, Florida, female bottlenose dolphin calves develop stereotyped signature whistles different from those of their mothers, while males’ whistles are like their mothers’, perhaps reflecting the different social roles of males and females in dolphin communities (Sayigh et al. 1990). Unlike bottlenose dolphins, Tuvaq’s female sibling, Qila, also produced her mother’s Type A1 call.

By 32 months, Tuvaq’s Type A1 calls were practically identical to those of his mother in the dominant frequency of the whistle component and in the pulse repetition rate of the pulsed component. Only the start frequency of their whistles and the duration of their calls differed consistently, so the most important identity information in the Type A1 call might be in these two parameters. However, we cannot forget that Tuvaq was involuntarily isolated and conceivably stressed when this series of calls was recorded. Frequency and temporal parameters may be influenced by motivational state. Bottlenose dolphins, for instance, produce longer signature whistles when stressed (Caldwell et al. 1990). Further, several more subtle differences were also apparent, such as the rounder inflection point of Tuvaq’s whistle.
component of this mixed call type. It is also possible that identity information can be found in the PRR pattern and pulse power spectrum of the call, not quantified in this study.

The Type A1 call may play an important functional role in mother-offspring recognition, highlighting the value of its early acquisition. Tyack (2003) pointed out that the combination of early mobility and extended dependence in bottlenose dolphin calves generates a strong need for early development of a mother-offspring recognition system. Signature whistles are believed to be important in maintaining contact between bottlenose dolphin mothers and calves (Tyack 1997). Fripp and Tyack (2008) investigated the role of maternal whistles in captive bottlenose dolphins during the first weeks postpartum and found that the rate of maternal whistle production increased significantly after calves were born. They hypothesized that this might facilitate imprinting, enabling the calves to recognize their mother’s whistles. Perhaps the Type A call, including variant A1, functions similarly in belugas, given the strong and long lasting mother-calf bonds (described at length in Chapter 1), and the fact that, in the wild, mothers and their highly mobile calves must recognize each other amongst large aggregations of belugas. In my study, Aurora’s Type A call production, including variant A1, was conspicuous after Tuvaaq’s birth. In fact, in Chapter 3 I show that the highest overall vocalization rate and the highest frequency of Type A calling per individual per hour occurred after the death of a calf and after birthing events.

The following observations hint at the role of the Type A1 call in mother-calf recognition: a) it was one of three acoustically similar call types that Aurora produced repeatedly during the first 24 hours after Tuvaaq’s birth (but not before birth), suggesting the possibility of strong acoustic imprinting by the calf; b) it was used predominantly when the two were separated, during tight synchronous swimming when disturbances occurred in the tank, and in response to each other’s calls (details in Chapter 3); and c) Tuvaaq’s longest, fully stereotyped series of this call type was produced in involuntary isolation. In Chapter 3 I investigate the context specific variation in the vocalizations of the six captive belugas at the Vancouver Aquarium, and examine in further detail the contextual uses of the different variants of the Type A call, including A1.

**Babbling in belugas**

Strong evidence of babbling exists for taxa as diverse as bats (e.g. Knörrnschild et al. 2006), primates (e.g. Elowson et al. 1998), and birds (e.g. bird subsong, Marler and Peters
1982), all highly vocal species, suggesting that babbling behaviour may have evolved in species in which juveniles must acquire complex vocal repertoires (Knörnschild et al. 2006). Given the importance of babbling in other vocal systems, it is reasonable to consider whether the first stage of vocal development in belugas might be analogous to babbling by humans and some other primates, or to the subsong of birds. As some of Tuvaq’s calls were becoming stereotyped, he was also widening his repertoire by including a larger variety of sounds. Trills, chirp series, noisy calls (signals having no discernible fundamental frequency and harmonic structure, as defined earlier), and other sounds became more prominent as he aged, hinting at a similarity between beluga vocal development and that of bottlenose dolphins. While the variability of some dolphin whistles is reduced as the young develop a stereotyped signature whistle, there is also an increase of highly variable sounds with age (Tyack 1997; 2003). Tyack (2003) noted that this pattern is dissimilar to the babbling and subsong stages of other taxa, which consist of a progressive narrowing of an initially overproduced repertoire. McCowan and Reiss (1995), on the other hand, postulated that whistle development in bottlenose dolphins could be reasonably compared to the ontogeny of human and bird vocalizations, since out of an unstereotyped repertoire emerges a subset of vocalizations. It is true that a subset of stereotyped vocalizations (e.g. signature whistles) might emerge gradually, but it is also important to note that bottlenose dolphins do not narrow their repertoire, but continue to produce a diverse repertoire throughout their lives (e.g. Cook et al. 2004). If we view babbling as “a subset of the phonetic units found in adult speech” (Oller and Eilers 1988), it is possible that Tuvaq’s initial unstereotyped pulsed calls were akin to human babbling. There is little doubt that Tuvaq’s initial repertoire of pulse-trains and whistles was only a subset of the adult repertoire (see Chapter 3 and Appendix A for a description of 28 identified adult discrete vocalizations, in addition to unclassified variable calls). As is the case of other vocal species (e.g. pigmy marmosets, Snowdon and Elowson 2001), babbling may be functionally important in belugas in order to practice for adult vocal behaviour, and to maintain acoustic contact with or attract the attention of other group members. During the first month of life, Tuvaq’s unstereotyped pulse-trains often elicited an approach and a vocal response from his mother (Chapter 3).

Despite Tuvaq’s variable repertoire at 12 months, I am unsure about the possibility of attrition later on because I do not have systematic month-to-month data after the one-year mark. Moreover, like bottlenose dolphins, belugas are long-lived, slow-maturing animals, and, accordingly, the vocal development stages in this species may be lengthier than I have
documented. My data suggest that in belugas, as in bottlenose dolphins (McCowan & Reiss 1995), acoustic development likely continues well past their first year of life.

**Maturational processes**

Determining the role of maturation of the vocal apparatus in vocal development is challenging. This is exacerbated by the fact that the precise mechanism of sound generation in odontocetes remains unclear, although much progress has been made in understanding it. Cranford et al. (1996) proposed a unified phonation mechanism for odontocetes. They described two structural complexes, each formed by a fatty bursa embedded in a pair of lips, associated with the upper nasal passages in all odontocetes (except sperm whales, which have only one). The most parsimonious working hypothesis is that sound is produced when air passes through the lips (termed “monkey lips”), causing them to open and slap together, creating vibrations in the monkey lip/dorsal bursa (MLDB) complex. This periodic opening and closing of the lips determines the pulse repetition rate, and may be regulated by muscle tension. It is likely that only the left lips generate whistles (Cranford et al. 1996; Cranford 2000a; 2000b).

Cranford et al (1996) suggest that simultaneous use of the two MLDB complexes may be necessary to achieve the overlapping but distinct sounds common to several odontocete species. Mixed vocalizations with overlapping tonal and pulsed components, such as those reported here, have been described in killer whales (Schevill and Watkins 1966), false killer whales, *Pseudorca crassidens* (Murray et al. 1998), spotted dolphins, *Stenella frontalis* (Herzing 2000), bottlenose dolphins (Killebrew et al. 2001) and pilot whales, *Globicephala melas* (Nemiroff and Whitehead 2009).

I occasionally recorded a different type of mixed vocalization, consisting of two overlapping pulsed components of different pulse repetition rates. Karlsen et al (2002) reported similar double-pulsed vocalizations in free ranging belugas off Svalbard, Norway. This is consistent with strong evidence that bottlenose dolphins can generate acoustic pulses by both sets of phonic lips, independently and simultaneously (Cranford 2000b).

In view of the current knowledge of sound generation in odontocetes, I can interpret the patterns observed as follows: 1) the marked increase in mixed vocalizations at four months suggests better control of the simultaneous use of the two MLDB complexes at this age. 2) The increase in pulse repetition rate with age may result from an increase in strength of the muscles.
that control the monkey lips, and from improved coordination of these muscles. 3) Whistle production is thought to require greater nasal air pressure than does pulse generation (Cranford 2000b). The delay in developing whistles documented here supports the notion that whistle production may entail finer muscular control and greater muscle strength than the production of pulse-trains. 4) Finer muscle control may also be required to produce high frequency pulse-trains and whistles, as suggested by the increase in dominant frequency of these vocalization types with age. For the whistle production, this relationship was quite strong only for the first five months of Tuvaq’s life (see Figure 2.10).

Increased vocal control with age probably has a learned component, although it is likely that the vocal apparatus of beluga calves also matures physically in the first year. To support the idea that physical maturation of the vocal apparatus explains developmental changes in the calf’s vocalizations would require demonstrating that more than one infant show similar changes at similar times in their development. Strong inter-individual variability would support the existence of other mediating mechanisms. That approach revealed that changes in the “coo” vocalization of rhesus monkeys (Macaca mulatta) during early development are attributable mainly to physical maturation factors (Hammerschmidt et al. 2000).

Vocal learning

There is growing consensus that many behavioural traits have a genetic component but are shaped to a smaller or larger degree by learning and experience. When infants develop vocalizations like those of their parents, it is difficult to separate the role of learning and genetic inheritance (Tyack 1997). The lack of his father’s Type B calls in Tuvaq’s vocal repertoire before their first acoustic contact, and his incorporation of less stereotyped versions of one of his father’s unique call types after contact, hints at production learning. However, the close genetic relationship between the two animals, and the time elapsed (six months) between first exposure to this vocalization and my first recordings of its production by Tuvaq, could indicate that the call was in fact genetically programmed, with an age threshold for its production. While my data are clearly not sufficient to distinguish between these ideas, it might be short-sighted to not entertain the possibility that learning played some role. Even if Tuvaq’s acquisition of this call type at two years of age had a genetic basis, and hearing his father’s rendition of it was not necessary for Tuvaq to develop his own, the possibility remains that he had to learn to use the vocalization by hearing his father’s use of the call (sensu Janik
and Slater 1997; 2000). This is in line with the idea that there is an important genetic component to learning, or an “instinct to learn” (Marler 2004). As Gould (1990) noted: “Even learning can be innately guided, so that a creature “knows” in advance how to recognize when it should learn something…” (pg. 84).

Tuvaq’s acquisition of his mother and half-sister’s (Qila, who was introduced into the pool when Tuvaq was 3 months old) Type A1 call is also revealing. Tuvaq’s Type A1 call did not emerge suddenly, but was slowly perfected towards full stereotypy. This gradual process could be the result of either production learning or slow maturation of the vocal apparatus, as discussed earlier. Although Tuvaq’s gradual development of his mother’s Type A1 call is not sufficient evidence that its production was predominantly learned, we cannot discard that at least its use has been learned. Further, ontogenetic changes in the acoustic structure of calls, such as the gradual changes shown by Tuvaq’s renditions of the Type A1 call, are at least a prerequisite for vocal production learning – if vocalizations were adult-like from birth, production learning would be unlikely (Egnor and Hauser 2004). Suboscine birds, which require no auditory feedback for normal vocal development, produce remarkably normal songs from the time of first production. For example, Kroodsma and Konishi (1991) deafened four eastern phoebes at 35 days of age, before any attempts at song had been made (i.e. before the birds had any chance to practice song production). The four birds produced normal songs, no different in duration, tempo, or variability, from the songs of wild birds and songs of laboratory reared but intact birds. My data are clearly not strong enough to claim that they offer proof of vocal production learning. I cannot say whether the calf copied the sounds of the two unrelated females introduced into the pool, because I was unable to identify the majority of their calls (except for three series of Type A calls produced by Allua, see Chapter 3, and Qila’s various series of Type A variants, Chapters 2 and 3). However, the fact that neither Tuvaq’s Type A calls (his mother’s call type) nor his Type B calls (his father’s) were normal from the time of first production is at least a condition for vocal development to be experientially guided.

That learning plays some role in vocal ontogeny in belugas would make sense in view of their life history and social parameters. In chapter 1 I delineated the similarities shared by the diverse taxa that are vocal learners. I explained that in cetaceans, social learning and vocal plasticity are evident in species with certain characteristics: strong mother-infant bonds, extended maturation period of the young, long life spans, consistent social groupings where repeated interactions with the same individuals occur, and multi-generation family groups (Brodie 1969; Smolker et al. 1993; McCowan and Reiss 1997; Boran and Heimlich 1999;
Mann and Smuts 1999), characteristics that appear to be shared by beluga society (evidence for this is summarized in Chapter 1).

Furthermore, these animals seem capable of incorporating new sounds into their vocal repertoire as adults, as indicated by evidence of vocal imitation in belugas (Chapter 1). The significance of anecdotes of marine mammals imitating human words and other sounds rests on the fact that these anecdotes *may signal what may exist in the wild*. Tyack (1993) stated it quite clearly: “If an animal can imitate a sound that is not normally part of its repertoire, then it must have learned to modify its normal vocalizations to match the model. Animals that have evolved this rare capability may use it in developing their natural vocalizations” (p.130).

In summary, it is unlikely that one mechanism alone completely explains developmental changes in repertoire variability and call characteristics. More likely, a combination of mechanisms mediates the ontogeny of calls, including maturational processes, social influences, and genetic inheritance. As beluga births are becoming more common in captive facilities, studies of vocal development comparing more than one subject could validate the patterns observed here. Furthermore, careful descriptions of the ontogeny of usage and responses to sounds, in relation to the social context, will be required (e.g. Snowdon et al. 1997 for New World primates) before we can claim to understand the stages of vocal development in belugas and other cetacean species.

2. 5 Postscript

After this study was published (Vergara and Barrett-Lennard 2008), two female beluga calves were born at the Vancouver Aquarium. Qila, captive-born daughter of Aurora, gave birth to Tiqa in June 2008, and Aurora gave birth to Nala in June 2009. Like Tuvaq, the first sounds that both calves produced after birth were low energy pulse-trains with low pulse repetition rates (relative to later pulse-trains). A study of Tiqa’s vocal production from the time of her birth until she was 6 months old revealed that both the bandwidth and the PRR of her pulse-trains increased with age (MacLeod 2009). The first recorded whistles attributed to Tiqa appeared only after 50 days of life, a month later than Tuvaq’s first recorded whistles. Analysis of Nala’s vocalizations is pending, but it is clear that her pulse calls also progressively acquired the buzz-like aural quality of the adult pulsed calls, suggesting a similar increase in PRR with age. These data support my findings that beluga calves produce only pulse-trains at birth, progressively increasing their PRR and dominant frequencies so that the pulse-trains begin to
resemble the loud, broad-band pulsed calls of the adults. Further, data on these three captive-born calves indicate that beluga calves do not appear to be able to produce whistles until later in life, perhaps because these may require finer control over the sound producing apparatus.
3. WHAT CAN CAPTIVE WHALES TELL US ABOUT THEIR WILD COUNTERPARTS? IDENTIFICATION, USAGE, AND ONTOGENY OF CONTACT CALLS IN BELUGAS

3.1 Introduction

The inquiry into the acoustic system of belugas faces two problems that are intimately related and are persistent stumbling blocks in the study of animal communication: First, great variability in the physical features of the sounds, with general call types grading into each other, often introduces great uncertainty in sound categorization. Secondly, it is inherently difficult to categorize sounds that are biologically meaningful without testing how belugas themselves perceive or use these sounds (e.g. Tyack and Clark 2000). To exacerbate these problems, belugas, like all cetaceans, rarely produce visible signs when they make sounds, making the identification of the phonating individual in a group, so necessary for the study of function, problematic (in Chapter 2 I outlined how I addressed this problem to study the vocal development of a beluga calf).

How, then, do we begin to understand the function of the calls produced by this highly vocal species? One approach is to look for call types that are predictably produced in particular, identifiable circumstances. The signals commonly referred to as “contact calls,” ubiquitous in social birds and mammals, are a good place to start because we can predict the circumstances when we would hear them. These signals are used to mediate group cohesion and coordinate movements in social animals, and to facilitate contact between particular social companions, including mothers and dependent young. When these signals are used by animals that are losing or have lost contact with one another, they are sometimes termed “isolation calls” (Tyack 2000). They are particularly advantageous in mobile species that inhabit environments where conspecifics can easily lose sight of one another, such as the marine environment.

Contact calls have been widely studied in birds (e.g., budgerigars, *Melopsitacus undulatus*: Farabaugh et al. 1994; several New World parrot species: Bradbury 2003; orange-fronted parakeets, *Aratinga canicularis*: Cortopassi and Bradbury 2006; communal nesting

---

5 A version of this chapter has been published. Vergara, V., Michaud, R., and Barrett-Lennard, L.G. (2010). What can captive whales tell us about their wild counterparts? Identification, usage and ontogeny of contact calls in belugas (*Delphinapterus leucas*). International Journal of Comparative Psychology 23: 278-309.
long-tailed tit, *Aegithalos caudatus*: Sharp and Hatchwell 2006), terrestrial mammals (e.g. baboons, *Papio cynocephalus*: Cheney et al. 1996; African elephants: McComb et al. 2003; vampire bats, *Diaemus youngi*: Carter et al. 2008; marmosets, *Callithrix jacchus*: Chen et al. 2009) and marine mammals (e.g. killer whales: Ford 1989; bottlenose dolphins: Sayigh et al. 1990; Janik and Slater 1998). Depending upon the species’ social system, it may suffice for those calls to be shared at the level of the group, or they may need to be individually distinctive (Tyack 2000; Tyack and Clark 2000). A system of reliable, individually distinctive vocal signatures would make it easier for animals to keep track of one another in species with high mobility and long-term yet fluid relationships in an aquatic environment (Tyack 2003). This is the case for bottlenose dolphins (Caldwell and Caldwell 1965; Caldwell et al. 1990; Sayigh et al. 1998; Tyack 2003; Watwood et al. 2005; for a differing view see McCowan & Reiss, 2001), which are known to live in fission-fusion societies in which group composition may change rapidly, yet which have strong and stable bonds between individuals (Wells et al. 1987; Connor et al. 2000; Tyack and Clark 2000).

We would expect that, as a species with high mobility and important long-term associations between individuals, belugas would use a system of discrete contact calls. Furthermore, given their fission-fusion social system (Chapter 1), we might expect selection to have favoured individually distinctive recognition signals. In my previous study of the vocal development of a beluga calf (Chapter 2), I identified three variants of a broadband, pulsed call type, labeled “Type A,” used predominantly by the mother after a birthing event. I hypothesized that at least one of the variants, A1, produced by the mother, her juvenile daughter, and gradually developed by a male calf, is a contact call that may play an important role in mother-offspring recognition.

The present study further investigates the context-specific use of the Type A call in a captive beluga social group at the Vancouver Aquarium to ascertain its function as a contact call. I predicted that this call type would be recorded more often in situations that would elicit a need to regain or maintain contact, such as the presence of external stressors (divers), changes in the group composition (re-introductions, separations, births, deaths), and forced or voluntary isolation – the latter ameliorating the problem of identifying the individual vocalizer. This is important as a first step towards understanding the functional significance of beluga signals, and to aid in the categorization of sounds that are biologically meaningful to these animals.
Recchia (1994) noted that this approach, whereby one correlates frequencies of occurrence of classes of sound with different contexts, can be effective only if the sounds of interest are not graded, but can be separated into discrete categories (whereby precise boundaries can be demarcated for each signal, Hauser 1996). The sounds must be reliably identified by their acoustic features. Despite the fact that many sounds in the beluga repertoire have been considered graded (see Chapter 1), the calls of interest (Type A calls) form a discrete and clearly identifiable category.

Despite the clear advantages afforded by studies of captive animals (Chapter 1), there always remains the question of whether their wild counterparts behave the same way. In this study I use analyses of signal structure and usage in captivity to construct testable hypotheses to guide investigations of sound usage in wild belugas. I then use opportunistic recordings from the wild environment as a field test of some of the hypotheses derived from the captive study.

**Study approach**

My general study approach has been to exploit the ease with which the animals can be observed underwater in a captive setting in order to begin to investigate call function, and then use the results and ideas generated from this research to direct research in the wild. I will thus first present the methodology and results of the captive component of the study. I will then outline a set of predictions about wild belugas generated from my captive work, and describe two attempts to test some of these predictions in the wild. I will discuss the captive and wild results jointly.

**3.2 Captive component: methods**

**Time frame and social group**

This study took place at the Vancouver Aquarium (British Columbia, Canada) between 2002 and 2006, with additional data collected during two beluga births in 2008 and 2009, and one death in 2010. A total of 8 captive belugas, four captive born, and four captured in Hudson Bay near Churchill, Manitoba (Canada), have been observed and recorded since 2002. Their social group had varying compositions throughout the length of this study. Figure 3.1 illustrates the genealogy of the studied animals. The ages of individuals at the time of writing, whether captive-born or wild-caught, and the periods when they were observed and recorded are summarized in Table 3.1.
**Figure 3.1:** Genealogy of belugas held at the Vancouver Aquarium. Black arrows indicate parent-offspring. Ages of the animals and whether captive-born or wild caught are listed on Table 1.

**Table 3.1:** Ages, sex and period of study for the belugas at the Vancouver Aquarium

<table>
<thead>
<tr>
<th>Individual</th>
<th>Wild-caught or captive born?</th>
<th>Year born or captured</th>
<th>Age and Sex</th>
<th>Observed and Recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaq</td>
<td>Wild caught</td>
<td>1990</td>
<td>Adult M</td>
<td>2002-2006</td>
</tr>
<tr>
<td>Allua</td>
<td>Wild caught</td>
<td>1985</td>
<td>Adult F</td>
<td>2002-2005</td>
</tr>
<tr>
<td>Kavna</td>
<td>Wild caught</td>
<td>1976</td>
<td>Adult F</td>
<td>2002-2006</td>
</tr>
<tr>
<td>Tuvaq</td>
<td>Captive born</td>
<td>2002</td>
<td>Calf M</td>
<td>2002-2005</td>
</tr>
<tr>
<td>Tiqa</td>
<td>Captive born</td>
<td>2008</td>
<td>Calf F</td>
<td>June-July 2008</td>
</tr>
<tr>
<td>Nala</td>
<td>Captive born</td>
<td>2009</td>
<td>Calf F</td>
<td>June 2009</td>
</tr>
</tbody>
</table>

a. Imaq and Kavna were transferred to another pool from July 2002 (Tuvaq’s birth) until January 2004, when both animals were re-introduced with the rest of the social group. Both animals were transferred again in June 2008 (Tiqa’s birth) and they remain in a separate pool at the time of writing.

b. Allua was transferred to another aquarium in January, 2005.

c. Allua and Qila were transferred to another pool for a period of three months in July 2002 (Tuvaq’s birth) and re-introduced in October 2002.

d. Tuvaq died at 3 years of age, in July 2005

e. Tiqa is currently a subject of an ongoing vocal development study. Only data related to her mother’s vocalizations during her first month of life are included here.

f. Nala was a subject of a vocal development study. I only include here data related to her mother’s vocalizations during the 2 hours after her birth, and during a 1 hour session after her recent death at 1 year of age.
Sampling regime and observation area

The animals were kept in an outdoor pool (18 m x 29 m; depth 6 m) connected to an adjacent 3 m deep smaller medical holding pool (3 m x 3 m; see Figure 2.2). A netted gate that allowed acoustic and visual contact between the whales occasionally separated both pools. The animals were observed and recorded from an underwater window with a good view of the entire larger pool, and occasionally from a smaller window with a view of the medical holding pool.

Observation/recording sessions varied in length from 15 min to 2 h (and longer for the initial 2 months of the vocal development study, see Chapter 2) at variable times of day. The frequency of the observation sessions was variable throughout the entire study period, occurring as often as three times weekly, and as infrequently as once per month, depending on the time allocated to other aspects of the beluga research.

Acoustic recordings

I made underwater audio recordings throughout each observation session using two hydrophones installed permanently in the adjacent pools: an Offshore Acoustics hydrophone in the main pool and a Briel & Kjaer 8101 hydrophone in the medical holding pool. The calls were recorded digitally on two Pentium IV computers (each connected to a hydrophone) using Avisoft SASlab Pro software (Avisoft Bioacoustics), Cool Edit 2000, and Raven (Cornell Lab of Ornithology) at a sampling rate of 44.1 kHz. The system had a frequency response of 0.02 kHz – 22.0 kHz ± 1 dB. The recording of the 2010 session following the death of a calf was made with a Marantz PMD660 solid state recorder at a sampling rate of 48 kHz with each hydrophone connected to one channel of the device. I occasionally had access to a sound card that sampled at 96 kHz (giving a recording bandwidth of 48 kHz), which allowed me to determine whether high frequency components were being missed in the 44.1 kHz recordings.

Since both hydrophones recorded simultaneously, it was often possible to identify the individual vocalizer by comparing the amplitude of the same recorded sound on the two computers (or stereo channels).

Spectrograms are displayed here with Avisoft SASlab Pro (Avisoft Bioacoustics). Call parameters were measured automatically with Raven 1.3.
Call type categorization

The efficacy of sound categorization methods that rely on aural and spectrographic comparisons is well demonstrated. For example, Ford (1984) compared the categorization of killer whale calls based on statistical comparisons of signal parameters, to the call categorization performed by ear and spectrographic examination, and found no significant difference between these methods. Deecke et al. (1999) compared the results of ratings of acoustic similarity of killer whale pulsed calls by a neural network with ratings made by human subjects and found that both methods perform similarly and quantify acoustic similarity between calls in a biologically meaningful manner. Janik (1999b) found that humans perform more reliably than computer methods when he compared the categorization of bottlenose dolphins whistles by human subjects with the performance of three computer methods.

In view of the above, I categorized recorded call types by ear and by visual inspection of spectrograms. I considered each discrete vocalization that was found 5 or more times in the repertoire a distinct call type. Vocalizations that were unstereotyped or presented a ‘graded’ structure (borders hard to define, unable to tell when a vocalization ended and another started) were assigned to the “variable” category. As did Recchia in her 1994 categorization of captive beluga calls, vocalizations that were intermediate in structure and did not clearly fall into any of the categories, or that were rare (less than 5) were assigned to a category labeled “other”.

Call categorization based on observer consensus (e.g. Watwood et al. 2004), on mean values and reliabilities of judges’ ratings (Sayigh et al. 1995) or on automated methods (e.g. Deecke and Janik 2006) was beyond the scope of this study, and categorization was instead performed by a single experienced observer (Vergara). Of relevance here is the distinctiveness of the Type A call in relation to the overall repertoire of 28 call types (see Appendix A for representative spectrograms of each call type). Several of these call types have variants that aurally fall into the same general call category but that show some spectrographical differences in duration, energy distribution, or pulse repetition rate. In this paper I consider only the five variants of the Type A call, A1 – A5.

Discriminant Function Analysis (DFA)

I used a DFA to verify my subjective classification of five variants of the Type A call, referred to as A1 to A5, using SPSS, version 16.0 (SPSS, Chicago, Illinois, USA). For this purpose, I randomly selected 60 cases of each variant from a total of 2835 cases. The DFA is a
classification procedure that assigns each call case either correctly to its appropriate variant
type, or incorrectly to another type, based on combinations of those acoustic parameters that
best separate the groups of cases. I used subset validation, whereby a random subsample of 40
cases of each variant type (for a total of 200 cases) was used to create the model, setting the
remaining 100 unselected cases aside to validate the analysis. To subsample the cases of each
variant type for the measurement of acoustic parameters, I used the online tool “Research
Randomizer” (http://www.randomizer.org/form.htm).

Acoustic parameters for the DFA

The following parameters were entered into the DFA for each call: maximum power,
average power (the value of the spectrogram’s power spectral density in each bin of the
spectrogram averaged over the entire call), peak frequency (the frequency at which maximum
power occurs, which, for the A1 mixed variant, coincides most often with the whistle
component), duration, and three measurements of pulse repetition rate (PRR; number of pulses
per second) along the time axis of each call: initial PRR, middle PRR and end PRR.

I used these three measurements of pulse repetition rate to describe the ‘inflection’ of the
five pulsed variants. These measurements were obtained by counting the pulses in the 0.2
seconds at the beginning, middle, and end of each call, and multiplying each count by 5 for an
estimate of pulses per second at three locations of each call (Figure 3.2). The mean PRR of a
call was estimated by averaging these three numbers. I used an analyzing filter bandwidth of
200 Hz and a FFT size of 512 samples in order to discriminate pulses more readily. I also
counted the pulses by ear, to corroborate the visual counts, by playing back each 0.2 seconds
clip at 3% to 10% of the original speed.

The PRR of some pulsed sounds can be high enough to confer a tonal character to the
call (generally classified as burst-pulsed sounds), with harmonically related frequency sidebands
on a spectrogram (Watkins, 1967). When the PRR was too rapid to discriminate individual
pulses, it was estimated through this harmonic interval, as described in Watkins (1967). Thus, I
measured the frequency of two harmonic bands at the same three locations on the time axis of
the call (i.e., beginning, middle, end), and subtracted one from the other to obtain the number of
cycles per second, or PRR, at each location. In such cases I reduced the filter bandwidth to 50-
70 Hz, increasing the FFT to 1024-2048 samples, for better resolution of the harmonic bands.
Figure 3.2: Example of the pulse repetition rate measurements at the beginning, middle, and end of a call (PRR 1, PRR 2, PRR 3). The bottom graph is a spectrogram and the top a waveform. The x-axis indicates time in seconds, while the y-axis denotes frequency in kilohertz (kHz) for the spectrogram view and amplitude in kilo units (kU) for the waveform view.

**Contextual use of call types**

I examined both the overall proportional usage of the Type A call in relation to all other recorded call types (including variable calls) and the call rate (call number/individual/hour) during recording sessions where maintaining contact was presumed to be important to the animals, termed for simplification purposes “special sessions” (described below), and during regular sessions (sessions where no isolation events, external stressors, separations and reunions, births, or deaths were taking place). I predicted that if the Type A call functions as a contact call, its usage would be higher during special sessions than during regular sessions.

I sub-sampled 18 regular sessions from the multi-year study using a random date generator ([http://www.lrs.org/interactive/randomdate.php](http://www.lrs.org/interactive/randomdate.php)). The special sessions were the following:

A. **Births.** I considered the 2 hours immediately following each of three beluga births at the Vancouver Aquarium “birth sessions”: 1) On July 20, 2002, Aurora gave birth to a male calf, Tuvaq. She was alone in the pool. No animals were held in the contiguous medical holding pool. 2) On June 9, 2008, Qila, captive-born daughter of Aurora, gave birth to a female calf, Tiqa. Aurora was held in the contiguous medical holding pool, from where she could see and hear the birthing mother and the newborn calf. 3) On June 7, 2009, Aurora gave birth to a female calf, Nala. Qila and
Tiqa (Aurora’s daughter and granddaughter) were held in the medical holding pool during the birth. For both the 2008 and 2009 births, only the vocalizations produced in the main pool were considered.

B. **Isolation of mother and calf.** Twelve sessions when Aurora was vocal, out of 27 recording sessions when Aurora and her calf Tuvaq were isolated from all other animals (from birth to 3 months of age).

C. **Separation of females.** Eleven sessions when Qila and Aurora were in the two contiguous pools, separated by a netted gate. The animals were in acoustic contact. Visual contact was limited (since the aperture between pools is small). Qila was in the larger of the two pools with her newborn calf, Tiqa, and Aurora was kept in the smaller pool for 1 month.

D. **Re-introduction.** Two sessions when first Allua (October 9, 2002) and then Qila (October 23, 2002) were re-introduced into the pool with Aurora and her calf, after 3 months of separation.

E. **Divers.** Seven sessions when 2 or more divers entered the pool for cleaning and maintenance purposes.

F. **Husbandry Procedures.** Six husbandry procedures involving temporary isolation in the medical holding pool (termed “gating”) and/or restraint of an adult, a calf, or an adult-calf pair (for the purpose of weighing, taking blood samples, etc). The gated animals were always within acoustic contact of the rest of the group.

G. **Voluntary Isolation.** Four sessions when a particular animal spent 80% or more of the session alone in one of the two pools. This could have occurred by active exclusion by the other animals. For example on July 3, 2003, Qila was prevented from entering the medical holding pool by the rest of the animals, who jointly rebuffed all her attempts to enter. “Voluntary” then, is relative, and is used here in the sense of an isolation event not imposed by humans locking a gate.

H. **Death of a calf.** On June 22, 2010, at 10:15 pm, Nala, Aurora’s one-year old calf, died suddenly of respiratory failure. I conducted a 1 hour recording session 2 hours after her death, once her body had been retrieved from the pool. At the time of this recording the only animals in the two contiguous pools were Aurora, Qila, and 2-year old Tiqa.

---

6 Since Nala’s death occurred when the data for this manuscript had been analyzed (and accepted for publication), I added these important data to the contextual analysis, but did not include these data in the analysis of variant types.
Vocal exchanges

I considered two in-sequence Type A calls by different individuals produced within 2 seconds of one another a “vocal matching exchange.” I assumed that 2 calls were from different individuals if they originated from the two contiguous pools. I chose the 2 s cutoff based on the distribution of the time differences between adjacent Type A calls made by different whales, which shows that most occurred within 2 s of each other (Figure 3.3; see Schulz et al. 2008).

A second-in-sequence Type A call was considered a potential answer to an initial Type A call if the onset occurred after the onset but before the termination of a preceding Type A call, termed “overlapping call” or if it occurred within 2 s but after the termination of the preceding Type A call, termed “adjacent call” (see Soltis et al. 2005).

Figure 3.3: Histogram of the time intervals between adjacent (black bars) and overlapping (grey bars) Type A calls made by different whales.
3.3 Captive component: results

Context specific use of the Type A call

Figure 3.4 shows the context specific use of call types (both discrete and variable) for a total of 8,214 recorded calls. The Type A call was clearly the predominant type – of 28 classified types, see Appendix A – during births, isolation of mother and calf, separation of females, re-introduction, divers in pool, veterinarian or husbandry procedures (forced isolation), voluntary isolation, and death of a calf, collectively termed “special sessions.” It constituted 24% to 97% of the vocalizations during those sessions. In contrast it composed 4.4% of the vocalizations produced during regular sessions.

![Proportion of total recorded calls](image)

**Figure 3.4:** Context specific use of all calls recorded at the Vancouver Aquarium during special sessions (bars A to H) and regular sessions (I). The numbers between brackets indicate the number of observation sessions for that session type. The total number of vocalizations recorded for each session type is shown to the right of the respective bar. During separation of adult females (C), diving events (E), and husbandry procedures (F), the animals were separated by an acoustically transparent net, i.e. they could see each other and hear each other while being held in different pools.
In order to standardize for number of animals in the pool and length of the observation session, I calculated the average vocalization rate per individual per hour, for each session type, for the Type A call, and for all other call types combined (Figure 3.5). There was a strong association between the frequency of use of the Type A call and special sessions, \( \chi^2 = 1253.262, n = 2858, p < 0.0001 \). Even during those special sessions when the animals vocalized little, such as isolation of mother and calf or separation of adult females (bars B and C of Figure 3.5), when they did vocalize they favoured Type A calls. Conversely, during regular sessions (bar I of Figure 3.5) the animals rarely produced Type A calls.

![Figure 3.5](image)

**Figure 3.5**: Average frequency (number per individual per hour) of Type A calls and of all other call types combined during special sessions (A through H) and regular sessions (I). The numbers in brackets indicate the number of observation sessions for that session type.

Both the highest overall vocalization rate and the highest frequency of Type A calling occurred after the death of a calf and after the birthing events. The high average vocalization rate/individual/hour for the birth sessions was due to Aurora’s high vocal production during the births of both Tuvaq (2002) and Nala (2008). Qila, on the other hand – a first-time, captive born
mother – produced no Type A vocalizations during the first 2 hours after the birth of her daughter Tiqa (her Type A call production began 5.5 hr after birth).

Aurora produced the majority of the calls during the recording session immediately after her year-old daughter’s death. She spent the majority of this session by herself in one of the contiguous pools, which permitted the attribution of her calls by comparing amplitudes. Out of the 586 recorded calls, 404 were Aurora’s, and were almost entirely (99%) Type A calls.

During the re-introduction and diving sessions, the Type A call production occurred while the animals swam rapidly around the pool in a tight and synchronous formation.

**Vocal exchanges**

**Call matching exchanges**

Of 1115 assigned Type A calls, 38% occurred in call matching exchanges (Figure 3.6), whereby two or more adults produced Type A calls in an antiphonal manner, using either the same or structurally different variants (described later). Considering only the second-in-sequence calls, a total of 20% of the assigned Type A calls occurred in the 2 s following the onset of a Type A call by a different whale, often overlapping.

**Mother-calf vocal exchanges**

During the first three months of Tuvaq’s life, when he and his mother (Aurora) were isolated from all other animals, 50% of Aurora’s Type A calls (n = 264) were preceded or followed by a calf’s call (an unstereotyped pulse-train) within 2 s. Thus, considering the dyadic vocal exchange as the unit of analysis, 50% of Aurora’s Type A calls were produced in vocal exchanges with her calf. This is a conservative estimate, as it does not take into account that the response unit may be an entire bout of several Type A calls, rather than one individual call; 29% of the maternal responses that occurred in the 2 s following the onset of a calf’s call were not isolated Type A calls, but bouts of 2 to 10 consecutive Type A calls within 2 s of each other.
Variants of the Type A call

I grouped 2835 Type A calls into five acoustically different variants labeled A1 to A5 (Figure 3.7). All five variants were typically produced in bouts, and consisted of a broad-band rapid pulse-train (the dark bands that span the frequency range of the spectrograms) of relatively long duration (1.2 to 1.9 s). They differed in the pulse repetition rates and in their energy distribution, translating into slightly different aural qualities. Variation in PRR along the time axis of a call confers each call variant a particular inflection (Figure 3.8). Types A1 and A4 overlap closely in this pattern of PRR, with an average PRR of $94.6 \pm 13.0$ and $115.0 \pm 26.1$ pulses per second respectively, and they sound practically the same – a loud buzz much like a door creaking, with a somewhat crooning intonation. A2, A3 and A5 have more of a bleating quality to their sound, and are closer to each other than to A1 and A4 in PRR pattern, with average PRRs of $328.9 \pm 36.4$ for A2, $306.4 \pm 42.4$ for A3, and $371.8 \pm 40.3$ for A5. These three variants fall into the burst pulsed sound category demarcated by Au and Hastings (2008): trains of pulses characterized by a pulse repetition rate higher than 300 pulses per second.

Figure 3.6: Spectrogram illustrating a call matching exchange between two adult female belugas, Aurora, who in this case was producing A1 variants (see next section), and Qila, producing overlapping A3 variants (FFT length: 1024, frame size: 100%, window: hamming, overlap: 75%, frequency resolution: 43 Hz, time resolution: 5.8 ms).
Figure 3.7: Representative spectrograms (bottom) and oscillograms (top) of the five variants of the Type A call produced by the Vancouver Aquarium belugas (FFT length: 1024, frame size: 100%, window: hamming, overlap: 75%, frequency resolution: 43 Hz, time resolution: 5.8 ms). The only mixed pulsed/tonal variant is A1 (note the overlapping whistle). The dark band across A2 is not tonal (as it can be resolved into pulses when manipulating the spectrogram FFT length), but a strong (and consistent) concentration of energy at 8 kHz. Variant A3 often has two overlapping pulse-trains of different PRR, one much more rapid than the other (the more rapid of the two overlapping pulse-trains was considered to be the characteristic PRR of this call, as it was the most acoustically conspicuous and the slower pulse-train was not always present).

Figure 3.8: Pulse repetition rate pattern for each of the five variants of the Type A call produced by the Vancouver Aquarium belugas. The two clear groups coincide with the aural similarity of these call variants. The PRR is much more rapid for A3, A2, and A5 than for A1 and A4.
Although, as noted above, A1 and A4 sound very similar to the author, there is different structural information in the calls: A1 is a mixed pulsed/tonal call, with a narrow band tonal component consistently at 14.6 ± 0.6 kHz \( (n = 559) \) overlapping the pulse-train, while A4 lacks the tonal component. The lack of clear audible differences between A1 and A4 may have to do with the relatively high frequency of the tonal component (the tonal component of A1 is clearly audible to me only when slowing down the call to half its normal speed). A3 can also be a mixed call, sometimes having a synchronous, slower pulse-train of the same duration. All variants may begin or end with a series of a few discrete pulses. I obtained recordings of three of the variants (A1, A4, and A5) with a 96 kHz sound card, and the broadband pulse-train still spans the frequency range of the system, with energy at 48 kHz in all three variants.

The DFA correctly classified 83% of the calls selected to create the model and 87% of the calls left aside for validation. The model assigned to the correct category 100% of the unselected A1 cases, 100% of the A2 cases, 75% of the A3 cases, 75% of the A4, and 84% of the A5 cases. The only variable that had no significant effect on the model was average power. The remaining variables contributed to the model, but maximum power and duration were the least likely variables to discriminate between groups. Predictably, the acoustic parameters that best describe these variants were: peak frequency and initial PRR, middle PRR and end PRR. For a visual representation of the discriminant space see Figure 3.9. The PRR pattern, represented by Function 1 on Figure 3.9, was better than peak frequency at overall discrimination between groups, and coincides with the pattern illustrated in Figure 3.8.
Figure 3.9: Discriminant function plot showing the separation of the five variants of the Type A call based on the most important discriminative acoustic parameters, pulse repetition rate and peak frequency. Function 1 represents the repetition rate pattern (initial PRR, middle PRR and end PRR were highly correlated with this function) and Function 2 represents peak frequency (strongly correlated with function 2). Type A1 and A4 tend to be at one end of Function 1 and A2, A3 and A5 at the opposite end, which means that PRR differentiates A1 and A4 from the other 3 vocal types. Although A1 and A4 are hard to distinguish from each other by their PRR only, A1 tends to be higher on dimension 2 (Peak Frequency) than A4. This makes sense, given that peak frequency often coincides with the tonal component of A1.

Usage of the Type A variants in isolation

I examined the use of the A call variants by each of 4 belugas recorded in isolation (mostly forced isolation, occasionally voluntary) in one of the contiguous pools (Figure 3.10). All four belugas favoured the Type A call over other call types when in isolation, using one or more of its variants. Allua, an unrelated female, used only variant A5. Aurora and her two offspring, Qila and Tuvaq, used three, four, and two of the remaining variants respectively, including A1 (the only mixed pulsed/tonal variant). Aurora and Qila used predominantly A3, and Tuvaq used predominantly A1. Data presented for Tuvaq are based on only two isolation
events, recorded at 20 and 32 months of age, shortly after his Type A calls had developed full stereotypy (see Chapter 2)

Although Aurora’s Type A calls as a whole composed 93% of her vocalizations during 26 isolation events from 2002 to 2010, the A1 variant accounted for only 18.6% of such calls. However, this mixed variant played a more predominant role during most of the vocal sessions when Aurora was alone with her calf after the 2002 birth. From the third day after Tuvaq’s birth until the reintroduction of the rest of the social group into the pool 3 months later, A1 composed 67% of Aurora’s vocalizations (Chapter 2). After the death of her year old calf (Nala), Aurora produced the same variants (A1, A2, A3) as immediately after the birth of her calves (Tuvaq and Nala), favouring A3, but at over double the rate (522 Type A calls per hour, compared to an average Type A calling rate of 229 calls per hour for Tuvaq and Nala’s birth sessions). Aurora was also heard producing Type A calls repeatedly after the death of her 3 year-old offspring Tuvaq in 2005 but those were not recorded (Vancouver Aquarium Beluga Trainers, personal communication, 2005).

Two or more variants could be produced in the same bout by one animal (Vergara and Barrett-Lennard, 2008), and vocal matching exchanges of the Type A call between two animals consisted of up to 4 different variants in a short (12 s) exchange.

Figure 3.10: Proportional use of Type A variants and of all other call types combined (white) for each Vancouver Aquarium beluga that was recorded in forced or voluntary isolation. The “n” between brackets on the X axis is the number of isolation events when the belugas were vocal. The numbers above the bars represent the total vocalization count for that individual in isolation.
Development of the Type A1 variant

My study of the vocal development of a beluga calf, Tuvaq, from the time of his birth until 3 years of age (Chapter 2) addressed the ontogeny of the mixed variant A1. Tuvaq’s mixed call production was minimal during the first three months of his life, and increased dramatically at four months of age. Some of his mixed calls increasingly resembled the stereotyped mixed call A1 of his mother Aurora (Figure 2.14, Chapter 2). Both the number of inflection points along the whistle component of his mixed calls and its dominant frequency showed a reduction in variability with age, indicating progressive stereotypy (Figure 2.14, Chapter 2). Opportunistic recordings of Tuvaq during 2 isolation events at 20 months and 32 months of age revealed that he was producing stereotyped Type A1 calls with parameters strikingly similar to those of his mother and half sister’s A1 calls (for details, refer to Chapter 2).

3.4 Wild component

Overview

The captive work described here has generated the following hypotheses about what we might expect in the wild environment: a) wild belugas utilize a specific broadband call type (a “contact call”) to maintain contact between individuals and promote group cohesion, and b) females and their offspring may share specific contact call variants which serve a recognition function in large aggregations. Some test predictions include the following: a) the use of contact calls increases when beluga groups are disturbed, broken-up, or when individuals are separated from the group, and b) all other factors being equal, contact call use is highest in nursery groups where it serves to maintain mother-calf contact.

General approaches to testing these hypotheses include recording the vocalizations produced by belugas temporarily isolated from the rest of the social group during capture and satellite tagging operations, analyzing recordings from wild beluga social groups, and using playbacks to further investigate the function of these calls. I will next describe two brief, opportunistic studies to validate the existence and function of this call type in wild beluga populations. The first study tests the disturbance/separation prediction. The second confirms the
existence of contact calls in wild nursery groups, without formally testing the prediction that this call type is predominant in such groups (the data examined were not collected with this test in mind).

**Nelson River Estuary**

The use of satellite telemetry and pectoral flipper band tags to study movements and diving behaviour of belugas in the last decade has required the temporary capture and release of many belugas (Orr et al. 2001). Capturing and tagging of wild belugas offers an ideal situation to address questions about contact calls: a controlled behavioural context (temporary restraint and separation from the group), a known primary variable (stress), and information on the sex and estimated age of the isolated individuals. The latter is key in light of the fact that the major challenge for studying the contact calls of free ranging cetaceans rests on the difficulty of identifying the individual vocalizer.

With this in mind, I joined a Department of Fisheries and Oceans (Canada) research team to record the isolation calls of temporarily restrained wild belugas in the Nelson River Estuary, Western Hudson Bay (Manitoba, Canada, 57 ° 02’ N, 92 ° 28’ W), between July 15 – August 2, 2005.

Belugas were captured by herding a small group into shallow water (about 2 m deep) using two Zodiacs, and then deploying a seine net from a fast moving jet boat so that it surrounded one of the whales. The animal was quickly disentangled from the net, a hoop net (1.2 m diameter) and a tail rope were placed on it, and the whale was taken carefully to water shallow enough for the capture team to handle and instrument it. The time elapsed from capture to release of each whale ranged from 20 to 35 min. For further details of the capture procedure, see Orr et al. (2001).

Acoustic recordings were made with an Offshore Acoustics hydrophone and a Sony TCD-D100 Digital Audio Tape recorder (this system has a frequency response of 0.02 Hz – 22 KHz ± 1 dB). I recorded continuously during the handling and release of each captured whale by placing a hydrophone about 0.5 m deep in the water approximately 2 m from the restrained animal. Simultaneous voice notes were dictated onto a mini-voice recorder.

Two of six restrained belugas vocalized, both producing only broad band, long duration pulsed calls that sounded to human listeners much like the Type A calls described for the captive belugas. Both were 2-3 year old juveniles, a male and a female and in both cases a herd
of belugas could be heard faintly in the distance. I obtained a recording of one of the two, the juvenile female. She was captured together with another young female, and began vocalizing only when the other female was released (she was still restrained when her companion was released), producing 43 broadband vocalizations in 21 min. Like the captive Type A calls, these signals consisted of loud broadband rapid pulse-trains with a mean duration of 1.8 ± 0.5 s and an average PRR of 103.5 ± 15 pulses per second (average of the PRR at the three locations of each call, see methods). Although the energy was distributed more or less equally throughout the frequency range of the spectrogram, there was average peak energy at 3.18 ± 1.6 kHz (n = 43). The interval between calls ranged from 0.7 to 24.1 s, with an average of 4.3 ± 5.9 s. Figure 3.11 shows an exemplar of one of these calls.

The restrained belugas that did not vocalized comprised an adult male captured with three other animals that escaped through an opening in the net (two adults and one calf), a juvenile female which was captured with the vocal juvenile female described above and which was released before the latter began vocalizing, a calf of undetermined sex (too young to tag, 1-2 year old, released 10 minutes after capture), and a juvenile male.

**St. Lawrence Estuary**

The St. Lawrence Estuary beluga population, recently estimated at 1100 individuals (Hammill et al. 2007) is genetically and geographically isolated from other populations (Brennin et al. 1997) and is at the southernmost limit of the species’ range. Summer aggregations are separated into herds of males and large nursery groups of females, calves, and juveniles occupying slightly different areas (Michaud 1993).

To verify whether the Type A call described in captivity is used in wild nursery groups, I reviewed recordings from St. Lawrence beluga social groups of both females and young and adult males obtained in June-September 1999, and June-September 2000. The 1999 recordings were made with an NRD-Q2 omnidirectional, long line hydrophone with a flat response from 0-60 kHz (±1dB), and a Sony TCD-D10 Pro II digital audio tape (DAT) recorder with a frequency response of 20-22,000 Hz (±1.0dB). In 2000, recordings were made with Vemco VHLF omnidirectional hydrophone with a flat response of 20 Hz-22 kHz (±1dB), and the same Sony digital tape recorder. Acoustic samples were recorded onto the DAT every 30

---

7 I helped instrument the juvenile male at the last minute, which involved getting in the water with a dry suit and wading a ways away from the recording gear, making the recording of this animal problematic.
min for a duration of 3 min in 1999 and 5 min in 2000, as long as whales were sighted within 300 meters from the research platform. I reviewed 112 three-minute sessions and 344 five-minute sessions, for a total of 34 h of recordings.

Type A calls are easily recognizable both by ear and in spectrograms. The tapes were reviewed by listening while examining the real-time spectrograms in 30 second segments, using Raven 1.3 (Cornell Lab of Ornithology), and this process was performed “blind”: the reviewer (Vergara) did not know the contextual behavioural details of the calls during the review process.

I identified 89 calls, in 16 different events, that resembled both acoustically and spectrographically the contact calls described for the captive animals. I defined as an “event” the presence of at least one Type A call in a 5-min (2000) or a 3-min (1999) recording session. Overall, Type A calls were identified in 12 out of 299 recording sessions of nursery groups (4%), on 2 of 106 recording sessions of adult male herds (1.9%), and in 1 out of 45 recording sessions of large mixed herds (adult males, females, and young) (2.2%). One event (of 1 call) was recorded in a session for which herd type was not noted.

Specific behavioural details beyond group composition existed for only one of the 16 events, identified in a 1999 recording of a herd of females and young. This event consisted of a series of 17 calls that had been produced by an adult female keeping a dead calf afloat, swimming around it, and pushing it along (Figure 3.12). DNA signatures of both animals are awaiting testing and will reveal if the female was the calf’s mother. Her Type A calls had a mean duration of $1.7 \pm 0.3$ s, an average pulse repetition rate of $188.6 \pm 7.6$ pulses per second, and average peak energy at $10.6 \pm 2.6$ kHz ($n = 17$). The inter-call interval ranged from 3.4 to 77.6 s, with an average of $20.1 \pm 18.7$ s. Figure 3.11 shows an exemplar of one of these calls.
Figure 3.11: Sample oscillograms (top) and spectrograms (bottom) of the broadband pulsed calls that resemble the captive Type A calls both in structure and context, identified in two different wild beluga populations: Nelson River estuary (left) and St Lawrence estuary (right) (FFT length: 1024, frame size: 100%, window: hamming, overlap: 75%, frequency resolution: 43 Hz, time resolution: 5.8 ms).

Figure 3.12: Adult female and dead calf in the St. Lawrence estuary, 1999. A series of 17 broadband “contact calls” were recorded from this female as she was staying close to the calf. Photo courtesy of GREMM.
3.5 General discussion

This study shows that captive belugas at the Vancouver Aquarium use a particular broadband pulsed call type, referred to as “Type A”, predominantly during isolation, husbandry and pool-cleaning procedures, re-union of animals after separation, immediately after births and deaths, and in mother-calf vocal exchanges. It is also used in vocal matching exchanges between adults and in some situations (divers in the pool, reunion of animals) during tight synchronous swimming. This call type therefore appears to be favoured in situations where establishing contact and group cohesion are important to the whales. There were variants of this call type within a common template, some of which were shared only by one mother and her two offspring. One of these offspring, the male calf, progressively developed the combined pulsed/tonal contact variant of his mother and half sister, decreasing variability and increasing stereotypy with age (Chapter 2).

Although my findings about the acoustic behaviour of captive belugas should be interpreted with caution given the small sample size, they warrant some hypotheses about the usage of these signals by their wild counterparts, critical to understanding vocal communication in the wild. They suggest that wild belugas utilize stereotyped broadband pulsed vocalizations as contact calls, that females and their offspring share variants that may serve for long-term recognition, and that calves slowly develop such variants. These hypotheses set the stage for future studies in the wild environment. I initiated such research by verifying the existence of this call type in the repertoire of groups of females and young and adult males in the St. Lawrence River estuary, and documenting its usage by two wild individual belugas from different populations, Hudson Bay and St. Lawrence Estuary, in contexts that further ascertain its function as a contact call.

By identifying the general function of the Type A call as a contact call we can be relatively confident that this category of sound is not a meaningless construct of an arbitrary human categorization scheme, but that it is a category important to belugas. This assertion is reinforced by the fact that Type A calls often occur in adult reciprocal call matching exchanges that fit the description of antiphonal calling for other species (e.g. cotton-top tamarins: Ghazanfar et al. 2002; Miller et al. 2005; African elephants: Soltis et al. 2005; sperm whales: Schulz et al. 2008). The vocal response to a contact call is classified as antiphonal if it consists of the same type of vocalization occurring within a few seconds of receiving the eliciting vocalization (Miller et al. 2005). Since by definition, a sound must be recognized as a specific
type of conspecific vocal signal to elicit an antiphonal response (Miller et al. 2005), this
behaviour further points to the categorical perception of this call type by belugas. There is,
however, one caveat: temporally associated signals are not necessarily communicative events
but may instead be produced independently of each other and be temporally associated by
change (Soltis et al. 2005). Further testing using methods such as randomization techniques
(Miller et al. 2004 for killer whales; Schulz et al. 2008 for sperm whales) is necessary to
determine if Type A calls of different adults are adjacent or overlapping by chance alone, or if
an adult is indeed more likely to produce a Type A call after a conspecific call of the same type.

This call type may have similar functional significance in other beluga populations. Van
Parijs et al. (2003) recorded sounds produced by belugas during capture events in Storfjorden,
Svalbard. The sounds produced by the mother of a captured mother-calf pair were
spectrographically similar to the contact calls described here. I listened to audio files kindly
provided by S. Van Parijs, and confirmed an aural similarity between the broadband pulsed
calls emitted by their captured mother (see Figure 2 in Van Parijs et al., 2003) and the contact
calls I describe here, particularly those produced by the St Lawrence female. Table 3.2 shows
the parameters of these calls, compared to the calls described here, based on a cursory analysis
of the first 10 good quality maternal calls from Van Parijs et al. (unpublished data).

The broadband, pulsed structure of beluga contact calls differs from the tonal signals that
serve as cohesion calls in bottlenose dolphins, the well studied signature whistles, recorded from
captive (e.g. Caldwell and Caldwell 1965; Janik and Slater 1998), wild (e.g. Smolker et al.
1993) and temporarily restrained animals (Sayigh et al. 2007). On the other hand, there is recent
preliminary evidence of broadband mixed pulsed/tonal contact calls in narwhals, Monodon
monoceros. Shapiro (2006) recorded the phonations of two adult male narwhals immediately
upon their return to deep water after they had been separated from their groups during a brief
period of capture. They both produced broadband combined tonal/pulsed signals, in addition to
whistles, that were individually distinctive and may have facilitated their reunion with group
members. Shapiro’s published spectrograms and parameters of these signals reveal the same
general template as the beluga contact calls identified here, although with longer inter-call
intervals (A. Shapiro, personal communication, November 30, 2009). Sound clips provided by
Shapiro allowed me to verify by ear that the pulsed/tonal signal types produced in particular by
one of the animals (individual mm226 in Shapiro, 2006) sound, at least to the author, similar to
### Table 3.2: Descriptive parameters for the variants of the Type A contact call identified in captivity, and for the structurally similar and probably functionally equivalent broad band pulsed calls identified in two capture and tagging operations and one dead calf situation. In all cases of this call type (captive and wild) the energy was distributed throughout the entire frequency range of the spectrograms (band limited to 22 kHz – the upper limit of the recording equipment). Note also the long duration of this call type, and the varying PRRs which translate into slightly different inflections.
the combined Type A1 call described here, but with a more audible tonal component overlapping the buzz. Shapiro (2006) reported that these signals had a mean duration of 1.2 s and were characterized by broad band pulses (48 kHz) with PRR of between 148 and 180 pulses/second, and an overlapping tonal component at a lower frequency than the Type A1 call.

That narwhals and belugas would favour similarly structured contact calls is not surprising, given their close taxonomic relationship, and the similarity of their environments. The structure of signals should reflect selection to facilitate effective transmission in the habitats that they have evolved to function in (e.g. Morton 1975; Richards and Wiley 1980; Blumstein and Turner 2005). The ice filled waters of the arctic create a highly noisy and reverberant environment. The clutter produced by the echoes and reverberation might make sound perception difficult, reducing a signal’s active space, the area in which an individual can detect the calls of a conspecific (Brenowitz 1982). Some features of the call may have evolved to increase this active space. For instance, the persistent energy at a wide frequency band may help to minimize masking by ambient noise. The simultaneous sounds that characterize the mixed structure of the contact call variants A1 (tonal/pulsed) and A3 (pulsed/pulsed) are likely produced with the twin sound sources described by Cranford et al. (1996; see Chapter 2). This “two-voiced” phenomenon (see Chapter 1) has a different physical basis than biphonation, the simultaneous generation of harmonically unrelated frequencies by a single sound source (Zollinger et al. 2008), but both may serve a similar adaptive function. Wilden et al. (1998) showed that subharmonics, biphonation, and chaos, conjunctively termed “non-linear phenomena” are common in mammalian vocalizations. They proposed that these phenomena can diversify and refine acoustic signals, and that they may have a function in individual recognition. Fitch et al. (2002) proposed that while such differences in call morphology between individuals, produced by non-linear phenomena in vocalizations, “are not necessarily adaptations in and of themselves, they may serve the adaptive purpose of allowing individuals to recognize one another more easily” (p. 414).

Just as non-linear phenomena can aid in individual recognition, the two-voiced co-occurrence of two sounds in the mixed contact call variants described here might also enable an individual to distinguish the signal of a familiar individual from signals of other conspecifics in the high-interference acoustic environment of the herd. This capacity in relation to the specific structure of contact calls has been well studied in several penguin species (e.g. Robisson 1992; Robisson et al. 1993; Lengagne et al. 2000; Jouventin and Aubin 2002; Aubin 2004). Of most relevance here is the finding that two non-nesting species of penguin, the emperor penguin
(Aptenodytes forsteri) and the king penguin (A. patagonicus), rely solely on acoustic cues, and not on landmarks, to reunite with their mate in a noisy colony of thousands of mobile birds, and use their two-voice system to recognize each other (Aubin et al. 2000; Lengagne et al. 2001). Their recognition calls consist of two simultaneous series of harmonically related bands of different frequencies produced independently with each of the two structures of the syrinx (Robisson, 1992). Individuality is partly encoded in the beat structure generated by the interaction of these frequencies (Aubin et al 2000; Lengagne et al, 2001). Given that the A1 variant I describe consists of two simultaneous signals (and often the A3 variant, when produced with two simultaneous, overlapping pulse-trains), it is conceivable that something akin to this might be at work, an idea that should be investigated further. This explanation, of course, would not apply to the variants that are not mixed.

In addition, vocal behaviour and signal design should reflect a trade-off between maximizing signal efficacy while minimizing conspicuousness to eavesdroppers (Bayly and Evans 2003). The risk of being overheard by their acoustically sensitive predators, killer whales, may also affect the features of beluga contact calls. High frequency sounds tend to be more directional than low frequency sounds. For example, there is evidence from two cetacean species that higher frequency elements of calls are more directional than lower frequency components (Miller 2002 for killer whales; Lammers and Au 2003 for spinner dolphins, Stenella longirostris). The relatively high frequency tonal component of the pulsed/tonal variant A1 might provide directional cues so that individuals could efficiently locate each other over short distances, but, since the higher the frequency the more rapid the attenuation of sound in water (Au and Hastings 2009), it would attenuate quickly, decreasing the chances of detection by distant killer whales. However, without proper propagation experiments, speculating further on the possible advantages of the kind of contact call structure I have described for belugas would be premature at this point.

Lastly, it is important to take into account that a considerable portion of the beluga vocal production consists of whistles (see Table 1.1). In fact, several studies from different geographical areas that have classified the beluga repertoire found that whistles were the most common signal type (Cunningham Inlet, Canada: Sjare and Smith 1986b; Bristol Bay, Alaska: Angiel 1997; Svalbard, Norway: Karlsen et al. 2002; White Sea, Russia: Belikov and Bel'kovich 2007). The contact calls I identified in this study are pulsed, but the possibility that the signal type described here is just one of several vocalization types that might play a role as cohesion calls should remain open.
Let us now turn to my finding that contact calls comprise clearly identifiable variants. This is in line with evidence of contact call variants in other species. For example, orange-fronted parakeets and budgerigars produce several different contact call variants, but favour one or two dominant types per bird (Farabaugh et al. 1994; Cortopassi and Bradbury 2006). Bottlenose dolphins are also known to produce more than one signature whistle type per dolphin. The prevailing explanation for this phenomenon, at least for male dolphins, is that individuals share whistle types with closely allied social partners (Smolker and Pepper 1999; Watwood et al. 2004).

The five variants of the Type A call identified in this study do not appear to have an individual signature function. Only one animal, Allua, an unrelated female, adhered to one variant type alone, but she was recorded in isolation only three times, so this is inconclusive. The remaining variants were shared between the three related individuals. The possibility remains, of course, that even if each particular variant is not an individual signature per se, each could nonetheless contain subtle identity information. Another possibility to be considered is that each animal does have its own signature Type A call, and the other animals occasionally might copy this signature, which seems to be the case when bottlenose dolphins use each other’s signature whistles (e.g. Tyack 1986; Janik 2000). I have not yet explored these possibilities.

Why, then, these variants? Are they biologically meaningful? The common thread of the situations when Type A calls were favoured was a need to establish or maintain contact. However, several different variables could have called for different messages or elicited different levels of arousal, such as the distress induced by the death of a calf (wild and captive) or by live-captures (wild) and veterinarian procedures (captive), alarm at the intrusion of divers in the tank, or the need for rapidly forging an acoustic bond with a newborn calf in an aquatic environment. There is evidence that species vary the rate or number of times a particular call type is emitted, its intensity and even its acoustic structure in response to different degrees of risk or predator types (e.g. alarm calls of three marmot species, *Marmota sp.*: Blumstein 1999; of suricates, *Suricata suricatta*: Manser 2001; and of white-browed scrubwren, *Sericornis frontalis*: Leavesley and Magrath 2005), stress (e.g. signature whistles of bottlenose dolphins: Esch et al. 2009a), or need (e.g. separation calls of domestic piglets, *Sus scrofa*: Weary and Fraser 1995). It is conceivable that variant types and repetition rates of beluga contact calls are associated with urgency, distress, need, or alarm. Aurora produced the same variant types after the death of her year-old calf, Nala, as after the births of the same and an earlier calf, but at double the rate. The
acoustic dimension made a critical difference in our perception of maternal distress (there was nothing particularly obvious in her non-acoustic behaviour that indicated distress).

As previously proposed (Chapter 2), the pulsed/tonal variant, A1, may play an important functional role in mother-offspring recognition. Aurora used this variant predominantly when she was alone with her calf Tuvaq from the 3rd day after his birth until the re-introduction of the rest of the whales 3 months later. Tuvaq and his half-sister Qila shared this mixed variant, and I have documented its ontogeny. Tuvaq did not emit these stereotyped contact calls at birth, but rather produced unsterotyped pulse-trains and rudimentary whistles (Chapter 2). Additional data on two calves born at the Vancouver Aquarium in 2008 and 2009 confirm that calves are not born knowing these contact calls, and must probably learn them. A1 was the first call type for which I documented full stereotypy, past a year of age.

The combination of early mobility and extended dependence of the young, invoked to explain the development of a mother-offspring recognition system in bottlenose dolphins (Tyack 2003), might generate the same strong need in belugas. In this species, the long-lasting mother-calf bond is evidenced by the prolonged lactation period (Brodie 1969; Drinnan and Sadleir 1981; Leung et al. 2010) and the maternally-directed philopatry demonstrated by genetic studies (Brennin et al. 1997; see Chapter 1 for details). In view of this, my captive findings suggest that wild beluga mothers and their offspring may share contact call variants that could serve for long-term acoustic recognition amongst large aggregations of females and young, a testable hypothesis.

A final but important consideration regarding contact call variants is the observation that one captive animal could produce up to three of these variants in the same vocal bout, or that four variants were produced in a brief vocal exchange between two animals. This greatly exacerbates the problem of assigning context to the variants, and hints at the possibility that these within-category distinctions may not be biologically meaningful. Perceptual experiments are essential to answer whether these variants are perceived categorically by the whales; playback experiments of these variants might reveal functional differences if they elicit sufficiently different responses (see Teixidor and Byrne 1999).

**Future research**

Although I successfully recorded only one temporarily restrained wild beluga, her unequivocal vocal response not only validated the captive results, but is also in agreement with a similar study on temporarily restrained whales by Van Parijs et al. (2003). Given the prevalent
use of satellite telemetry, there is much to be gained by incorporating acoustic recordings as a standard protocol of beluga tagging operations, especially those operations that take place in the same area year after year. An example of this approach is the excellent body of data on signature whistles that resulted from acoustic recordings of temporarily restrained bottlenose dolphins in Sarasota, Florida (Sayigh et al. 1990; 1995; 1999; 2007; Fripp et al. 2005; Watwood et al. 2004; 2005; Esch et al. 2009a; 2009b). For example, based on the captive data, I predicted that females and their calves may share specific contact call variants used for recognition. The sort of data needed to test this prediction could be obtained by recording temporarily restrained mother-calf pairs, or temporarily restrained mothers whose calves are swimming nearby, as is often the case during these operations (Orr, personal communication, July 2005).

Jointly, results from the captive and wild components of this study have provided reasonable certainty about the structure of beluga contact calls. Simply put: we now know what beluga contact calls sound like, and we may begin to refine the details of their function. The playback technique has been widely used to study specific aspects of contact calls, such as whether kin selectively answer each other’s contact barks in baboons, (Cheney et al. 1996), long term vocal recognition in fur seals, Callorhinus ursinus (Insley 2000), responses to family vs. strangers’ infrasonic contact calls in African elephants (McComb et al. 2000), and individual recognition in bottlenose dolphins (Sayigh et al. 1998). The fact that in captive belugas Type A calls elicit calls of the same type from conspecifics makes them ideal candidates to utilize a playback technique in order to further investigate the function and mechanisms of such calls in the wild. To my knowledge, only one playback study exists that broadcasted beluga calls to free-ranging belugas (Morgan 1979). A two-week pilot study in the summer of 2008 in the St. Lawrence Estuary helped begin to elucidate the appropriate and realistic design required for playback studies with free ranging belugas (Vergara et al. 2009, unpublished manuscript).

Lastly, propagation experiments of this signal type and an evaluation of how habitat acoustics might be related to the transmission of these signals are needed (e.g. Mercado III and Frazer 1999; Mercado III et al. 2007, for humpback whales; Lammers and Au 2003 for spinner dolphins). This species evolved in the arctic environment. In their recent review of signal propagation in the Arctic, Au and Hastings (2009) emphasized that the uniqueness of the environmental conditions in the Arctic contribute to unique sound propagation characteristics. In addition, proper testing of the propagation properties of the Type A calls would require wider band recordings, since my findings are mostly band-limited to 22 kHz (48 kHz for A1, A4, and A5, see results).
In many locations, belugas and other marine mammals are faced with increasing degrees of noise pollution from human activities. It is difficult to evaluate the consequences of the interference of noise on sound production and reception without understanding the specific function of such sounds, an understanding that has been enhanced by captive work. By continuing to shed light on the characteristics and functions of beluga calls, we may use changes in the type and rate of emission of such vocalizations to evaluate not only how the whales respond to various kinds of human disturbances, but also to assist such things as assessments of group composition. For instance, the notable predominance of the contact calls described here during all three captive births and their role in mother-calf vocal exchanges – coupled with our recent and ongoing familiarity with what beluga calves sound like, might enable us to predict the presence of young calves in a group simply through acoustic monitoring. By the same token, captive work means little if we do not validate the results with “real life” data – as I attempted to do here. In sum, it is clear that a fluid exchange between captive and wild research is crucial for a more integrated picture of beluga communication.
4. CALL USAGE LEARNING IN BELUGAS: CATEGORICAL MATCHING

4.1 Introduction

The essential role of vocal learning in the development of human speech has catalyzed a great deal of comparative research on phylogenetically diverse species. Marine mammals are conspicuous on the growing list of species known to be flexible vocal learners (see Chapter 2). Janik and Slater (1997; 2000) differentiate production learning, the modification of the structure of signals through auditory experience with those of other individuals, from contextual learning, which describes changes in the comprehension and the usage of pre-existing sounds precipitated by learning and experience. Both forms of vocal learning, especially when combined, can increase the message repertoire (Janik and Slater 1997; 2000). My longitudinal vocal development study (Chapter 2) suggests – albeit not conclusively – that both production and usage learning play a role in the development of the beluga’s natural vocal repertoire.

Vocal comprehension learning occurs when a receiver extracts the correct meaning for a signal from a range of existing possibilities, and vocal usage learning arises when animals learn to use signals already present in their repertoire in specific contexts based on experiences of other individuals’ use of such signals (Janik and Slater 1997; 2000). Numerous examples illustrate the relevance of both forms of contextual learning in nature. Learning to recognize conspecifics from their signals is comprehension learning. For instance, bottlenose dolphins learn to recognize the signature whistles of individuals with whom they share a strong social bond (Sayigh 1992). Recognizing the meaning of alarm calls designating different predators (e.g. Seyfarth et al. 1980) is also a product of comprehension learning. Animals can also learn to comprehend the meaning of some signals produced by sympatric species (e.g. vervet monkeys: Hauser 1988; Seyfarth and Cheney 1990; Diana monkeys, Cercopithecus diana: Zuberbuhler 2000), and by taxonomically distant species (e.g. yellow-casqued hornbills, Ceratogymna elata, and black-casqued hornbills, Ceratogymna atrata, can distinguish between Diana monkeys’ alarm calls, Rainey et al. 2004a,b).

Usage learning is a more complex form of contextual learning, although Schusterman (2008) hypothesized that all mammalian infants exhibit call usage learning to some extent when

---

8 A version of this chapter will be submitted for publication. Vergara, V. and Barrett-Lennard, L. Call usage learning in belugas.
they adapt their nursing calls in an interactive and synergistic fashion to the responses of their mothers. He noted, for example, that dependent Northern elephant seal pups soon learn that their own nursing calls can influence their mother’s orientation postures, facilitating access to the teat. Usage learning can thus be understood as an individual learning the social or ecological context in which a call is used (Egnor and Hauser 2004). Some of the most compelling examples of usage learning in nature by non-human animals are in primates. Hauser (1989) used playbacks of vervet monkeys’ intergroup *wrr* vocalizations to study the effect of social and auditory experience on the usage and comprehension of this vocalization by vervet infants. He found that infants in groups that had been exposed more often to intergroup encounters comprehended the meaning of the intergroup *wrr* vocalization and used adult-like exemplars of the call at an earlier age than infants in other groups (Hauser 1989). Vervet monkeys also need to learn the appropriate context for giving the correct alarm call, and are in fact given positive feedback by knowledgeable adults in the form of a second appropriate alarm call (Seyfarth and Cheney 1986).

Vocal production and usage learning are best demonstrated through conditioning experiments (Janik and Slater 2000). The vocalizations of ecologically and phylogenetically diverse animals such as marine mammals, primates, birds, cats, dogs, and rodents have been brought under control of discriminative stimuli using operant conditioning techniques with food reinforcement (reviewed in Adret 1993; Janik and Slater 1997; Shapiro et al. 2004; Schusterman 2008). Operant vocal behaviour is voluntary (rather than reflexive) and shaped by its consequences (Adret 1993; Manabe et al. 1997). Thus, to the extent that vocal behaviour is plastic, it should be susceptible to operant reinforcement techniques in captivity.

Learning to produce particular calls in response to conditioning stimuli shows usage learning (Janik and Slater 2000). Shapiro et al (2004) identified four levels of increasing complexity in the experimental demonstration of usage learning. In the first level, the animal learns to vocalize on command. In essence, the animal learns to respond with any call to a conditioning stimulus (a new context). Since it can always be argued that an animal may vocalize in an arbitrary context simply because it enters an excited motivational stage, the second level involves training an animal to vocalize or remain silent on command. Many species of birds and mammals can achieve the first two levels of usage learning under controlled conditions (reviewed in Shapiro et al. 2004).

In the third level, the animal must generalize what it learned in order to make specific vocalizations in response to different visual cues (hand signals) or acoustic cues (playbacks) – a
slightly more complicated level of usage learning. It implies that the animal can learn to associate different vocalizations of its own repertoire with different contexts. Conditioning experiments successfully demonstrating this third level of usage learning are much less common, and include common hill mynas, *Gracula religiosa* (Ginsburg 1963), budgerigars (Manabe et al. 1995), grey seals, *Halichoerus grypus* (Shapiro et al. 2004), and Pacific walrus, *Odobenus rosmarus* (Schusterman 2008).

The fourth and most complex level of usage learning, *categorical matching*, requires an animal to select and produce a vocalization from its repertoire of calls that matches the call type presented to it (for instance, produce a particular whistle type after hearing that type, a growl after a growl, etc.). This level is deemed more complex than the previous ones because a true understanding of the matching task requires not only that the animal associates a particular acoustic stimulus with a particular vocalization, but that it generalizes to any vocalization type, matching the stimulus regardless of the call type presented (Shapiro et al. 2004). Categorical vocal matching has not been conclusively demonstrated experimentally, but is common in nature and mediates important aspects of social relationships in species in which it has been proven to exist. When a vocal response to a call occurs within a few seconds of the eliciting vocalizations and consists of a vocalization of the same type (a categorical match), the process is referred to as antiphonal calling (see Chapter 3). This is especially evident in signals that facilitate contact between conspecifics over long distances or in dense environments (e.g. mangabeys, *Cercocebus albigena*: Waser 1977; Japanese macaques, *Macaca fuscata*: Sugiura and Masataka 1995; spider monkeys, *Ateles geoffroyi*: Teixidor and Byrne 1999; African elephants: McComb et al. 2003). Avian examples are conspicuous in group-living species. Loud contact calls that elicit calls of the same type have been described in nearly all the species of parrot studied, and establish vocal connections between individual birds (reviewed in Bradbury 2003). Additionally, breeding pairs of some birds engage in dueting (again, calling in an antiphonal manner) for various adaptive reasons such as maintaining reproductive synchrony and coordinating joint territorial defense (Hall 2004).

Cetacean examples of categorical matching also abound. For example, tight temporal exchanges of matching stereotyped calls between animals that are out of visual range of one another are an important aspect of intragroup calling in killer whale pods (Miller et al. 2004). Since the matching calls in these exchanges do not seem to be signature calls, Miller and colleagues hypothesized that the exchanges may allow group members to inform each other of their positions and movement trajectories almost simultaneously, although this remains to be
explored. Another plausible explanation of this instance is that by responding with a categorically-matched call in a tight temporal sequence a whale is confirming in an unambiguous, easily recognized manner that it has heard the first caller. Wild bottlenose dolphins use their learned signature whistles in matching interactions, presumably to address associates, whereby an individual responds to a whistle of a conspecific by emitting the same whistle type (Janik 2000). Sperm whales exchange matching codas (patterned sequences of clicks) seemingly to reinforce social bonds (Schulz et al. 2008). Belugas engage in antiphonal exchanges of contact calls, described in this dissertation (Chapter 3).

In chapter 1 I discuss the reputation of belugas as vocal mimics, based on anecdotal evidence of them mimicking interspecific sounds. Vocal imitation of sounds not found in the species’ typical repertoire provides compelling evidence for production learning. Given that belugas are so plastic in their vocal production, it would be surprising if they did not have the flexibility to adapt their vocal usage to novel circumstances (usage learning). Indeed, at the Vancouver Aquarium, the belugas have been trained to vocalize in air and to remain silent on command as part of their daily show routines. Furthermore, three of them can produce one type of call (a scream) in response to one hand signal (a visual discriminative cue) and a different call (a pulse-train) in response to a different hand signal. Thus, these animals are already proficient at the first three levels of usage learning. In addition, one of these signals (the scream) was not part of their natural repertoire, but was slowly shaped (see glossary) through selective reward, adding to the evidence that these animals are adept at production learning.

Abundant anecdotal evidence for production learning in belugas (Chapter 1), the indication that at least part of their large repertoire might be learned (Chapter 2), and these captive whales’ proficiency at the first three levels of usage learning, make belugas a compelling species for studying learning in the vocal domain. Their vocal response flexibility has not been formally explored experimentally.

Since antiphonal calling plays a role in beluga communication (Chapter 3), it seems reasonable to expect that they can categorize incoming calls and respond with calls of the same type. We should thus be able to demonstrate this mechanism, categorical matching - the most complex level of usage learning - under controlled conditions in captivity. The objective of this study is to examine the ability of belugas to respond to playbacks of two categories of beluga calls with matching vocalizations. The investigation of categorical matching in belugas is a step toward understanding the extent of their vocal control and flexibility, and acoustic perception
and categorization of sounds. This is the first attempt to document the training of beluga whales to control their vocalizations.

4.2 Methods

Subjects, training regime and general procedure

This study consisted of two phases. During Phase 1, three belugas were trained to vocalize after acoustic playbacks. Due to limited trainer availability and scheduling conflicts, only one of these three whales moved on to Phase 2, the categorical matching phase of the experiment.

The subjects of this study were Aurora (female) and Imaq (male), both 18 years old at the time of the study (Phase 1), and Qila, 10 year old daughter of Aurora (Phase 1 and Phase 2). All three animals were housed in an outdoor pool (18 m x 29 m; depth 6 m) connected to an adjacent 3 m deep smaller medical holding pool (3 m x 3 m; see Figure 2.2 in Chapter 2 for diagram of the pools). They shared this enclosure with a 2 year old male calf and two older, unrelated females (see Figure 3.1 and Table 3.1 in Chapter 3 for genealogy and origin of Vancouver Aquarium belugas).

Each animal was trained separately 1-2 times a day, 3 days a week, by one of the marine mammal trainers at the Aquarium. Sessions were 5-20 minutes long, depending on whether they were training sessions or experimental sessions (details below), to fit them within the institution’s busy schedule. Sessions took place either in the smaller medical holding pool, or at a training station at the corner of the larger pool (Figure 2.2). I was present at every session, and videotaped each one for later analysis.

During training procedures, the trainer identified a correct response by immediately blowing a whistle (a “bridge” or secondary reinforcer, see glossary), which was followed by a fish reward as a primary reinforcer (often accompanied by praising and petting). A Least Reinforcing Stimulus (LRS) was used for incorrect responses (see glossary). This consisted of reducing all stimuli (the trainer would stand still and not look at or interact with the animal in any way) for 5-10 seconds.
**Playback stimuli**

The belugas were trained with the two sound types that they regularly produce after hand signals during public shows: the aurally and spectrographically distinct “scream” and “pulse-train” (Figure 4.1). Screams are series of short burst pulsed sounds produced in rapid succession, each burst pulsed sound lasting $0.4 \pm 0.1$ seconds and with a mean pulse repetition rate of 513 pulses/second ($\pm 79$, $n = 12$). Most energy is below 16 kHz, with dominant frequency below 4 kHz. Pulse-trains are longer, averaging $2.02 \pm 0.7$ s and having a pulse repetition rate of $118 \pm 4$ pulses/second ($n = 12$). Most energy is below 16 kHz, with a dominant frequency below 5.5 kHz.

![Oscillograms and spectrograms of the two signal types used in training, the scream and the pulse-train](image)

**Figure 4.1:** Oscillograms and spectrograms of the two signal types used in training, the scream and the pulse-train (FFT length: 512, window: hamming, frame size: 100%, overlap: 75%, frequency resolution: 86 Hz, time resolution: 2.9 ms).
In the captive setting, these two call types function differently. Screams are produced only during shows (and during the current study), and only in air (never underwater). They are not in these belugas’ natural repertoire; they were shaped (see glossary) entirely by trainers through selective reward (suggesting production learning)\(^9\). Pulse-trains, on the other hand, are used spontaneously both underwater and in air in contexts other than shows, and were captured (see glossary) by bridging and reinforcing their production, to be used on command during shows.

The recordings for the playbacks were made from each whale’s vocal responses to the two hand signals that elicit these two sound categories, using a Sony TCD-D100 Digital Audio Tape recorder and a Logitech microphone (this system had a frequency response of 0.02 Hz – 22 KHz ± 1 dB). Playbacks were broadcasted in air using a Sony digital audio player CFD-S20CP with a frequency response of 20 Hz-20 kHz (+1/-2dB). The audio player was positioned on the pool ledge, about 1 m from the head of the whale. Playbacks were standardized in amplitude and were of similar duration: 2.5 to 3.5 seconds per playback, a length of time that could fit a total of 2 individual pulse-trains or 4 individual screams (the latter were much shorter vocalizations).

**Phase 1: Training procedure**

From October 25, 2004 until March 08, 2005 Qila, Aurora, and Imaq were trained to vocalize after the playbacks of their own calls, thus shifting the context of these two vocalizations from visual stimuli (hand signals) to acoustic stimuli (playbacks). The belugas never responded with any other call type; the only ambiguous responses occurred when they occasionally offered both signals following a playback stimulus, which was considered an incorrect response. Other than this occasional dual response, the belugas adhered strictly to pulse-trains or screams, so that assessing the correctness of the response was unproblematic.

The procedure consisted of three stages, and the number of whales participating in each training stage was reduced from three in the first stage to one (Qila) in the final stage. This was due, as mentioned earlier, to trainer availability and scheduling conflicts. One trainer was needed per whale, which was difficult given the restrictions imposed on the trainers’ schedules by public shows, husbandry duties, etc. Individuals were trained at different times so that I could be present at every training session, but they could probably hear each other’s sessions.

---

\(^9\) Unfortunately, I did not have the chance to document the process of the selective shaping of this vocalization because this occurred well before my time as a graduate student. For an excellent description of the process involved in this kind of shaping in a male harbour seal see Schusterman 2008, pg. 12.
Each session began with the animal stationed by the pool ledge, with its head out of the water, facing the trainer. The trainer ensured that he/she had the animal’s attention by bridging it for touching the trainers’ hand. The animal then kept its head vertically above the water, awaiting the next signal. I was positioned behind the trainer and off to the side, handling the playback equipment and the video camera. Training sessions were as short as 5 minutes, during busy Aquarium visitor times, and as long as 20 minutes. Thus, the number of training trials per session varied between 5 and 20 (usually approximately one trial per minute).

Stage 1: Training to match scream playbacks

Three whales, Aurora, Qila and Imaq, were taught to scream after hearing screams. Initially, I played a scream, the trainer made the corresponding hand signal, then rewarded a correct vocal response. The hand signal was gradually phased out (using them only on some trials, judged by the trainer) so that the whales would learn to attend only to the acoustic stimuli, in essence substituting one discriminative stimulus (hand cue) for another (acoustic cue). The number of training sessions that each whale required to first produce a scream after hearing a scream playback, without relying on the hand cue, was 9 for Aurora (50 trials), 12 for Qila (80 trials), and 17 for Imaq (158 trials).

Stage 2: Training to match pulse-train playbacks

Imaq’s sessions were discontinued due to his trainer’s unavailability. The pulse-train playback was introduced to Qila and Aurora and the same procedure repeated, i.e. the hand signal indicating “pulse-train” was gradually phased out as the whales learned to respond to the pulse-train playback with a pulse-train vocalization. They learned this stage more rapidly than Stage 1. The number of training sessions required in order for Aurora and Qila to first produce a pulse-train after a pulse-train playback, without relying on hand cues, were 2 (13 trials) and 8 (55 trials) sessions respectively, a notable reduction from the number of sessions required in Stage 1.

Stage 3: Training to match both playback types

Qila was the only individual who participated in the remainder of the study (again, due to shortage of trainers for this time-consuming study). This stage consisted of training her to respond to both signal types in the same session, matching the appropriate sound category. The problems I encountered throughout the course of this training stage required an ongoing rethinking and adapting of the training procedure. When both playback types were first introduced into the same session, Qila responded with only one vocalization type, regardless of
which vocal class I broadcasted. She first responded to both playback types with pulse-trains, the last vocalization that we had worked with (Stage 2). After some LRSs (see methods) from the trainer for her incorrect pulse-trains, she dropped them and began offering only screams. I thus returned to the accompanying hand signals to facilitate her understanding of the concept, then phased them out again and imposed strict reinforcement for categorical matching of both signal types.

Qila began to alternate between both vocal responses by vocalizing as soon as she heard the onset of the playback and not waiting to hear the full signal. The trainers called this guessing behaviour “anticipating” the playback. We (the trainer and I) thought Qila was beginning to understand that we wanted sometimes one vocal response and sometimes the other, and that she had to offer these responses upon hearing the playbacks, but that she still did not understand that we wanted the same vocal category that was being broadcast (for which she would have had to listen to the full playback stimulus). Consequently, we had to teach her to wait until the end of the playback before responding. This was a time consuming process. It consisted of bridging and rewarding Qila for not vocalizing too quickly upon hearing the onset of the playback - in essence, giving her the message “do not vocalize.” She then began remaining silent after playbacks, not responding with any vocalizations. After a few more sessions, however, she began to vocalize after each playback, appropriately waiting until the end of the playback to respond, no longer ‘anticipating’ the playback.

Up until this point, playback sequences were not randomized, but were selected by Qila’s trainer, depending on her performance. For instance, if Qila responded incorrectly to a scream, I was asked (after an LRS), to play another scream to give Qila the chance to match it and receive a reward. After session 65 (45 sessions after I started playing both types in the same session) the trainer considered Qila ready for randomized training trials.

**Phase 2: Categorical matching experiment**

From March 15, 2005 until May 10, 2005 I tested Qila’s categorical matching ability, adapting some of the methodology from Shapiro et al.’s (2004) study of usage learning in captive grey seals. Experimental sessions were 10-20 minutes long, to fit them within the institution’s busy schedule. This length of time allowed for 10 to 20 trials per session. My experimental protocol consisted of the following three trial types:
Training trials (15 March – 7 April, 2005)

On March 15, 2005 I began presenting random sequences of the same two vocalizations with which Qila had been trained. Sequences were generated with a random number generator before every session, allowing for no more than 3 vocalizations of the same type to be played consecutively. I used exemplars of the scream and pulse-train playback with which she had been trained (3 of each). Each track on the CD played only one vocalization exemplar, i.e. tracks 1 to 3 for scream exemplars # 1 to # 3, and tracks 4 to 6 for pulse-train exemplars #1 to #3. The number of trials varied among sessions, and was determined by trainer availability. The short duration of some training sessions permitted only 10 trials, for a total of 180 trials in 12 training sessions.

Transfer trials (14 April – 10 May, 2005)

The purpose of this phase was to assess Qila’s ability to generalize the task to novel stimuli. I tested her responses to new versions of the two call categories in a series of 10 sessions (this number of transfer trial sessions to be run was predetermined), running at least 10 trials per session, for a total of 185 trials. I made a recording of 7 novel exemplars of Qila’s scream vocalizations and 6 novel exemplars of her pulse-train vocalizations not previously used in training. I generated a random sequence for each session playing no more than 3 vocalizations of the same type consecutively.

Calibration trials (16 February – 07 March, 2006)

Prior to this set of experiments, Qila never had to repeat her vocal responses more than 2 or 3 consecutive times, because she was usually asked to produce just one vocalization of each of the two trained types during shows. I thus considered the possibility that Qila would lose motivation when the same task was asked of her 12 or more consecutive times, session after session, as in the training and transfer trials above, and that this might affect her success rate. To help distill the motivation factor from any real difficulty that she might have had in the above experiments, I ran a series of “calibration” trials testing Qila with the two types of hand signals that she responded to correctly during the shows. If her motivation was decreasing, I would expect her performance, usually perfect with just two trials per show, to drop with 12 or more consecutive trials. These calibration trials were conducted well after the playback trials were completed, so there was no chance that the calibration trials themselves would increase the chance of Qila losing her motivation to respond to the playback trials.
I conducted 10 sessions, each consisting of 12 to 16 hand signals (trials), for a total of 134 trials. As with the playback trials, I generated a random sequence for each session allowing for no more than three hand signals of the same type to be offered consecutively.

**Statistical analyses**

I used generalized linear mixed effects models (GLMM), implemented with the lmer function in the ‘lme4’ package (Bates and Sarkar 2006) of the *R System for Statistical Computing* (R Development Core Team, 2010). GLMMs provide a more flexible approach than strictly fixed effect methods (such as logistic regression or contingency tests) for analyzing binary data that involve random effects (Bolker et al. 2008). This approach allowed for different sources of variation in whether Qila succeeded or failed at matching the vocalization type. Playback type was treated as a fixed effect. Session (i.e., the day of the trial) was treated as a categorical random effect, allowing for the individual (Qila) to have good days and bad days. To allow for the possibility that Qila learned over time, which would be demonstrated by a consistent increase in success rate over the 10-12 sessions of an experiment, time was also treated as a continuous fixed effect.

### 4.3 Results

For all three experiments (training, transfer, and calibration), I accounted for the possibility that Qila learned over time. There was no consistent change over time in success rate in any of the experiments (all Z-values < 0.798, all p-values > 0.425), although there was a slight increasing trend for the training and transfer trials, and a slight decreasing trend for the calibration trials. I thus dropped this covariate from the model to increase the precision of the remaining parameter estimates and looked at whether Qila’s overall success rate, and her success rate for each call type, differed statistically from chance, accounting for random variation between sessions.

**Training trials**

Figure 4.2 illustrates Qila’s performance on the training trials. Qila’s overall success in matching all playback stimuli was 66%, not significantly better than predicted by chance ($z = 1.068$, $n = 180$, $p = 0.285$). However, this might be because she had more difficulty matching
screams (54% success) than pulse-trains (80% success). Indeed, when playback type is included as a covariate, Qila matched the pulse-trains more often than predicted by chance \((z = 3.268, n = 85, p < 0.01)\) but not the screams \((z = -1.713, n = 95, p = 0.087)\). Her success rate matching pulse-trains and screams was statistically different (i.e. in a model testing for the difference in success rate between call types, \(z = -3.639, n = 180, p < 0.001\)).

Some of the responses were classified as incorrect because Qila offered both vocalization categories (shifting from one to the other without a break; interestingly, she did this only when I played screams). She replied to 10 out of 95 scream playbacks with both vocalizations, always producing the pulse-train ahead of the scream.

![Figure 4.2](image-url)

**Figure 4.2:** Proportion of categorical matches by a female beluga, Qila, at the Vancouver Aquarium, in response to the scream (grey squares) and pulse-train (white circles) playbacks with which she was trained, and overall success rate for both playback types combined (red triangles). The numbers in brackets on the x axis indicate the number of sessions from the beginning of Qila’s training. Session # 66 was the first session where I began randomizing the playbacks to test Qila on the training trials.

It is evident from Figure 4.2 that some sessions were better than others. In the 5th and 12th (last) sessions, Qila performed similarly well matching pulse-trains and screams, for an overall success rate of 80% (5th session) and 75% (last session), but her performance oscillated over the interim sessions. After session 4, success on the two types was slightly positively correlated \((R = 0.4)\), suggesting that Qila’s ‘good’ days and ‘bad’ days were the same for both vocal types.
Transfer trials

Figure 4.3 illustrates Qila’s performance on the transfer trials. Overall, Qila responded correctly to the set of novel stimuli of both call types in 64% of the trials, a success rate that did not differ statistically from chance ($z = 0.431, n = 185, p = 0.666$). As in the training trials, this might be because she had difficulty matching screams (55% success), relative to pulse-trains (74% success). Qila matched novel pulse-trains more often than predicted by chance ($z = 2.509, n = 88, p = 0.012$), but not novel screams ($z = -1.778, n = 97, p = 0.0753$). The difference in success rate between pulse-train and scream playbacks was statistically significant ($z = -3.070, n = 185, p = 0.002$).

![Graph showing success rate]

**Figure 4.3:** Proportion of categorical matches by a female beluga, Qila, at the Vancouver Aquarium, in response to seven novel exemplars of the scream playbacks (grey squares) and six novel exemplars of the pulse-train playbacks (white circles), and overall success rate for both playback types combined (red triangles). The numbers in brackets on the x axis indicate the number of sessions from the beginning of Qila’s training.

Qila responded to 8 out of 97 screams with both vocalization types, always producing the pulse-train first, and to 1 out of 88 pulse-trains with both vocalization types, offering the scream first. These were considered incorrect responses.
Figure 4.3 indicates some clear variation among sessions. On the last two sessions of the transfer trials Qila matched screams and pulse-trains with approximately equal success, for an overall success of 75% (40 trials).

**Calibration trials**

Figure 4.4 illustrates Qila’s performance on the calibration trials. Excluding 5 non-responses (i.e. Qila remained silent following 2 of the scream hand signals and 3 of the pulse-train hand signals), Qila responded to the two visual stimuli (hand signals) with the correct vocalization type 91% of the trials, significantly better than chance ($z = 5.894, n = 130, p < 0.001$).

![Graph](image)

**Figure 4.4:** Proportion of correct vocal responses by a female beluga, Qila, at the Vancouver Aquarium, to the scream hand signals (grey squares), the pulse-train hand signals (white circles), and to both signal types combined (red triangles).

Although she performed well above chance in response to each hand signal type (98% for pulse-trains: $Z = 4.107, n = 65, p < 0.001$; 85% for screams: $Z = 3.504, n = 65, p < 0.001$), there was a statistically significant difference in performance between call types ($z = -2.192, n = 130, p = 0.028$).
4.4 Discussion

The initial stage-by-stage procedure (Phase 1) to train the whales for this experiment indicated that all three animals changed the context of the scream vocalization from a visual (hand signal) to an acoustic (playback) stimulus, and that two of them (Qila and Aurora) rapidly shifted the context of their pulse-train vocalization from a visual to an acoustic stimulus. This agrees with the idea that these animals are capable of vocal usage learning.

The results of the categorical matching task are more difficult to interpret. Overall, the task seemed difficult for Qila. She required what seemed like a large number of training sessions (although this is arguable; I will return to this later) to achieve a level of proficiency that was not statistically better than chance. This poor overall success appears to be related to her difficulty matching screams. She performed above chance for pulse-trains but not screams, both when responding to call exemplars with which she had been trained, and to novel exemplars of the same call types.

Shapiro et al. (2004) trained two grey seal pups to respond to playbacks of moans and growls with vocalizations of the same category. The seals did not generalize this response: both novel moans and novel growls tended to elicit growls. In much the same way, Qila could have favoured one call type (pulse-trains) in her responses regardless of which playback stimulus was presented, which would explain her success matching pulse-trains. However, if Qila was indeed defaulting to pulse-trains, we should have seen lower than 50% success rate matching screams, which was not the case (Qila’s performance matching scream playbacks was not statistically below chance in the training nor in the transfer trials).

It appears then that Qila’s success in matching pulse-trains, and in generalizing this response to new auditory stimuli of the same class, indicates a true categorical matching ability for this particular vocalization type. Nevertheless, for a convincing demonstration that Qila understood the premise of the categorical matching task, she would have had to perform above chance at matching each type. Further, the next step would require presenting Qila with an altogether different vocal class – a third category – to determine if she would transfer her knowledge of matching screams and pulse-trains by producing the same category of call as the new stimulus.

In light of the above, the results do not warrant a firm conclusion that Qila can match, and discriminate between, playback stimuli in trained tasks – at least not these particular
vocalizations, as we shall see below. Moreover, Qila’s above chance performance matching pulse-trains, a natural vocal category, compared to her poor performance matching screams, a shaped vocal category, is intriguing. We will explore this and other possible explanations for Qila’s poor performance in the remainder of this discussion.

**Perception of vocal categories**

The question of whether categories of sounds distinct to researchers are distinct to animals arises repeatedly in the literature (Snowdon and Pola 1978; Sinnott 1987; Nelson and Marler 1989; Owren 1990; Fischer 1998; Baugh et al. 2008). A handful of studies have found that humans differ from the studied species in the partitioning of the species-typical vocal repertoire (e.g. Snowdon and Pola 1978, for trill vocalizations of pigmy marmosets; Cheney and Seyfarth 1982, for vervet monkey grunts; Weary 1989, for songs of great tits). For example, Cheney and Syfarth (1982) found, through playbacks, that the grunt vocalization of vervet monkeys, which humans perceive as one call type, is perceived by the monkeys as four different call types that elicit different responses.

We currently lack an adequate understanding of what components of the vocalizations are perceptually important to belugas. Differences in perceptual boundaries between cetaceans and humans could partly explain Qila’s difficulty. The two pulsed signals I used are clearly distinct to humans, based on their pulse repetition rate. Above a pulse repetition rate of around 300 pulses per second, humans can no longer discriminate the pulses as separate from one another (Au and Hastings 2009), and begin to hear continuous tonal-like sounds, which sound subjectively like screams, squeals, or moans (Caldwell and Caldwell 1967; Murray et al. 1998). Screams in this study were well above this threshold, whereas pulse-trains were well below it (see methods), which explains why we perceive them as distinct. Cetaceans, on the other hand, have a much finer temporal resolution of sounds. Belugas in particular have extremely refined bioacoustic abilities suited for the Arctic environment (Turl et al. 1987; Turl and Penner 1989; Turl et al. 1991, see Chapter 1). Qila may have had problems with the categorical perception task if these particular signals represent, to a beluga, points on a continuum rather than discrete categories. Pulse repetition rate may not be a perceptually salient feature of these two vocalizations.

To determine the perceptual validity of the two vocal classes in my study, we would need to disentangle the categorical perception task from the vocal matching task. If Qila had difficulty understanding the vocal matching task, but had no trouble distinguishing the categories of sounds
presented to her, she might achieve proficiency in the categorical perception task if she was asked to indicate her understanding with a response other than a matching vocalization. Training her to touch different objects with her melon upon hearing the different vocal categories might work. An experiment of this sort was conducted by Le Prell et al. (2002), who trained rhesus monkeys to either maintain contact with, or to release, a metal cylinder during presentation of two rhesus scream categories, a tonal scream vs. an arched scream. The monkeys learned to classify initial exemplars correctly, but did not generalize the response to novel exemplars of the two categories, indicating that these classes were not perceptually discrete. The variable responses to novel tonal and arched screams suggested that the monkeys did not perceive these scream classes defined by the researchers as categorically distinct.

Qila’s poor performance in the categorical matching experiment might also be explained if this species’ capacity to perceive sounds broadcasted in air is reduced. Auditory systems of cetaceans are highly specialized for underwater hearing, whereas their hearing capacity in air is not well documented (Liebschner et al. 2005). To my knowledge, hearing sensitivity in-air was investigated in only three cetacean species, bottlenose dolphins (Babushina 1979), harbour porpoises (Kastelein et al. 1997), and tucuxis (Sotalia fluviatilis guianensis; Liebschner et al. 2005). All three species had better underwater than aerial hearing sensitivity, for all frequencies tested, when using sound intensity. However, when aerial and underwater sound detection thresholds were compared via sound pressure instead of intensity, the results were less clear-cut, with two of the species (Tursiops and Sotalia) showing more sensitivity to airborne sounds at particular frequencies (summarized in Liebschner et al 2005).

For belugas, the frequency bandwidth of best underwater hearing is from 11 to 105 kHz, with an upper frequency limit of 120 kHz (Au and Hastings 2009). This species has, along with bottlenose dolphins, the widest auditory bandwidth underwater (reviewed in Au and Hastings 2009). However, the extent of their hearing sensitivity to airborne sounds is unknown, so the question of whether her performance was affected by the fact that the playbacks were broadcasted in-air must, for the moment, remain unanswered.  

---

10 To determine if Qila’s performance improved markedly underwater, I initiated an experiment to test Qila’s ability to match underwater playbacks, broadcasting the playbacks via an LL916 speaker (Lubell Labs) deployed at a depth of approximately 1 m. Qila did not perform better than chance during the initial training sessions, but the experiment could not be carried to completion.
Categorical perception in relation to function

There is mounting evidence that animals categorize signals based on their meaning under natural circumstances (e.g. alarm calls of vervet monkeys: Seyfarth et al. 1980; lactating calls of house mice, *Mus musculus*: Ehret and Haack 1981; vervet monkey grunts: Cheney and Seyfarth 1982; rhesus monkey screams: Gouzoules et al. 1984; food calls of rhesus monkeys: Hauser 1998; mating calls of túngara frogs, *Physalaemus pustulosus*: Baugh et al. 2008). Baugh et al.’s (1998) study is a good example. Closely related species of túngara frogs have mating calls named “whines” that elicit the approach of females. Baugh et al. tested female *Physalaemus pustulosus* responses to a series of synthetic calls that varied continuously in several dimensions between the conspecific whine and the call of an allopatric heterospecific. Females discriminated the calls categorically into either conspecific or not conspecific. When they were presented with pairs of calls that differed by the same magnitude, females discriminated between those in different categories (a conspecific and a heterospecific call) but not between those that belonged to the same category (both conspecific calls).

Hauser’s (1998) study on rhesus monkeys’ food calls provides another revealing example. Rhesus monkeys produce five acoustically distinctive calls in the presence of food. Hauser conducted a habituation-dishabituation playback experiment to determine how the monkeys classify such acoustic variation. He showed that a call’s functional referent carries more weight during perceptual classification than its acoustic features. For instance, ‘warbles’ and ‘harmonic arches’ are acoustically distinctive (at least to humans) but rhesus perceive them as one functionally meaningful category based on their referent (high quality food).

Quite relevant to this discussion is the growing body of literature on phoneme (see glossary) recognition in humans, which stresses that knowledge of word meaning facilitates the perceptual discrimination of speech sounds. Research on cross-language speech perception is particularly illuminating. Kazanina et al. (2006) examined the categorical perception of [d] and [t] sounds by Russian and Korean speakers. Only in Russian are these two sounds used to encode word meaning. Predictably, Russians could distinguish [d] from [t] easily, whereas discriminating the two sound categories was difficult for Koreans, as indicated both by behavioural tasks and electrophysiological measures of brain responses. Similarly, Davidson et al. (2007) designed a series of experiments that demonstrated that associating meaning (pictures)

---

11 It has been brought to my attention after writing this paragraph that, very unfortunately, Hauser was recently found guilty of scientific fraud. Whether we can thus believe any of his studies is questionable, but I have chosen to give this study the benefit of the doubt.
with words containing non-native phonotactics (sequential combination of phonemes that are not allowed by the native language of the participants) improves the ability of listeners to distinguish the fine phonetic differences between non-native sequences and native ones.

Along the same lines, although with an innovative angle, Pfordresher and Brown (2009) compared musical pitch perception in tone language speakers (Mandarin, Cantonese, and Vietnamese) to that of native English speakers. Individuals whose native language is a tone language, in which pitch conveys information about word meaning, are better able to imitate (through singing) and perceive pitch differences in music than speakers of English (which does not employ pitch to convey word meaning).

Collectively, the evidence reviewed in this section sheds light on Qila’s difficulty in matching screams. This call type is not used by belugas for communication; at least not by this group of captive whales. It is a vocalization that was shaped (see glossary) by trainers in the older animals (Aurora and Imaq) through selectively rewarding successive approximations to the final, desired call type (a scream). Qila learned screams by copying her mother during training sessions; once she began volunteering scream-like sounds, their production was appropriately reinforced and further shaped by the trainers. Before the current study, Qila produced this call type only in air (never underwater) and only during shows, strictly upon presentation of the discriminative stimulus (the hand signal representing this call type). She was never heard to produce this airborne sound at any other time. The pulse-train, on the other hand, is used spontaneously underwater by all the individuals in contexts other than shows. It is not surprising, then, that Qila was successful only at matching pulse-trains – part of her natural repertoire – but not screams, in both the training and the transfer trials. It is possible that Qila had difficulty perceiving screams categorically because they play no functional role in this species’ natural communication system. Interestingly, the difference between her performance matching pulse-trains and matching screams was supported statistically in all three experiments. Even during the successful calibration trials, in which Qila responded to both hand signal types better than chance, she performed better for pulse-trains than for screams. These results point to the notion that Qila could not deal efficiently with a call type that is foreign to her natural vocal repertoire in tasks that involve the categorical perception and matching of such sound.

**Procedural limitations**

It is possible that Qila’s poor performance stems from procedural limitations rather than from a true difficulty performing the task. The training method - operant conditioning with food
reinforcement - might not be conducive to stimulating vocal learning in these animals. Several authors have proposed that food may not be the best type of reinforcer to control vocal output when vocalizations in nature are largely under the control of social reinforcers (Pierce 1985; Adret 1993; Manabe et al. 1997). Belugas have developed an elaborate system of vocal signals to mediate their social interactions. This begs the consideration that perhaps social reinforcers should be expected to influence their vocal behaviour to a larger extent than food.

The idea that animal vocalizations are highly susceptible to training protocols that emphasize interaction and social rewards has gathered a great deal of evidence from studies on grey parrots (*Psittacus erithacus*) (e.g. Todt 1975; Pepperberg 1985; Pepperberg and Neapolitan 1988; Pepperberg 1993; 1999; Pepperberg et al. 1999). Pepperberg (1985, 1993) used human social modeling theory as a framework to guide her avian vocal learning studies, and showed that exposing parrots to the presence of interacting human “tutors” clearly accelerates the learning and appropriate usage of new vocal labels and concepts.

Pepperberg’s experimental paradigm has parallels in cetacean research. Reiss and McCowan (1993b) manipulated the acoustic environment of subjects sufficiently to observe spontaneous (untrained) mimicry without altering their social environment. They designed a large underwater keyboard that stimulated the dolphin’s natural curiosity, so that the animals spontaneously pressed visual forms on the keyboard, which were linked to different synthetic whistles. Each whistle was in turn linked to a specific object (e.g. ball, ring) or activity (e.g. rub). The rapidity with which the dolphins mimicked the model sounds and used them in appropriate contexts contrasted greatly with the results reported by previous studies (Richards et al. 1984; Sigurdson 1993), which used traditional food reinforcement methods and required many trials to train initial vocal mimicry. This suggests that a social and interactive environment, which also allows for spontaneity, whereby the animal may initiate the interaction, might be more conducive to the learning of new signals, categorical matching, and mimicry than training protocols divorced of social factors.

Herman (2002), recounted the three decades of prolific studies on dolphin cognitive, sensory and linguistic competencies, noting that *teaching* models were often more efficient and effective than *training* models based on traditional instrumental conditioning techniques. He pointed out that these teaching models employed techniques that are commonly used with young children, such as showing and demonstrating.
Motivation

Another idea worth considering is the importance of an animal’s motivation on the success rate of a task. Pepperberg (2002), when training parrots in vocal learning tasks, noted that concurrent work on several tasks was crucial because the birds would become restless in single-topic sessions, ceasing to work, preening, or interrupting with requests such as "want cork". Testing would thus be impeded and test scores might decline “for reasons unrelated to competence” (Pepperberg, 2002, pg 56). Similarly, in an experiment to train a female killer whale to label visual objects with acoustic symbols, she would refuse to cooperate – and food made no difference – when bored (Dudok van Heel et al. 1982).

Since prior to this set of experiments Qila had rarely been asked to repeat her vocal responses more than 2 or 3 consecutive times, the number of repeated trials in my experiments might have caused her to lose motivation to perform well. I examined this possibility by asking Qila to perform repeatedly a task she knew well, producing a scream or a pulse-train in response to two different hand signals. Qila achieved a high level of success during these calibration trials, suggesting that her difficulty during the playback experiments was not a result of lack of motivation arising from the recurrent nature of the task. It could be argued, of course, that Qila simply disliked the vocal matching task compared to the calibration trials. However, she had ample opportunity to leave the training station. Had she disliked or become bored with the task, she could have simply swum away and refused to cooperate, but she did not. She stayed with us and was quite attentive throughout most sessions, even while performing poorly.

Learning to learn

There is evidence that animals use prior learning experiences to facilitate subsequent learning – the “learning-to-learn” phenomenon, or deuto-learning. Visser (2003) traced the development of the concept of deuto-learning to Gregory Bateson. In his early essay on social planning and the concept of deuto-learning (1942, reprinted in Bateson, 1972, pp. 159-176) Bateson distinguished two types of learning: simple learning, denominated “proto-learning”, the adaptation of behaviour in response to reinforcement contingencies, and “gestalt learning”, defined as deuto-learning, which refers to changes in the rate of proto-learning as a result of “insight” into the contexts of problem solving. “We might say that the subject is learning to orient himself to certain types of contexts”, noted Bateson (p. 166).
Schusterman (2008) described the process of “learning-to-learn” in one of his earlier experiments with a California sea lion. The author was testing whether the sea lion could transfer its knowledge of when to vocalize and when to remain silent to many pairs of visual stimuli of different shapes and sizes. Every 160 trials a new problem was presented, for a total of 45 such problems. Interestingly, during the first five problems the subject averaged 59% correct responses, but in the following 10 problems correct responses jumped to 75%, reaching 88% correct responses on the last five problems. In other words, the sea lion solved each problem more rapidly than the earlier one (but note that 6400 trials were required for his best performance; many more than the number presented to Qila in this study). Similarly, the “creative porpoise” (Pryor et al. 1969), a classic experiment in training for novelty and creativity, illustrated deuteero-learning. A rough-toothed dolphin (Steno bredanensis) was rewarded only when it showed a new behaviour. This took many sessions and seemed difficult for the animal until it understood the principle, after which new tricks came easy. In much the same way, training sniffer dogs to discriminate odors takes a great deal of time initially, but once trained, learning to discriminate new odors becomes progressively less time consuming (McGreevy and Boakes 2007). A good example of this phenomenon in humans is foreign language learning: a person learning a new language not only learns the language but also learns how to learn a language, becoming metalinguistically aware and thus acquiring subsequent languages more easily (Klein 1995).

The concept of deuteero-learning is, I believe, key in helping us understand Qila’s performance. Qila was naive to research training. This was the first time that she was exposed to this kind of learning and had previously only been trained to perform a series of fairly simple behaviours for public shows. Previous experience participating in inherently more complex scientific research might have provided the necessary conceptual scaffolding to perform better in this experiment. Furthermore, in light of the enormous number of trials that some species require in order to make the cognitive switch that will suddenly accelerate their learning and render increasingly correct responses (e.g. sea lion experiment above), I probably did not run too many trials for Qila, but not enough. The fact that by the last two sessions of the transfer experiments she performed equally well matching both call types suggests that she might have gotten better with more sessions - unless this was simply an ephemeral, random peak (given the oscillation between sessions).

At Kewalo Basin Marine Mammal Laboratory, six dolphins participated, for over 30 years, in a prodigious amount of research that produced groundbreaking discoveries on vocal and
motor mimicry, language comprehension, memory, cross modal matching, sensory abilities, conceptual processes, and self awareness, amongst others (e.g. Beach and Herman 1972; Thompson and Herman 1977; Herman et al. 1984; Shyan and Herman 1987; Pack and Herman 1995; Ralston and Herman 1995; Herman et al. 1998; Pack and Herman 2004; Herman et al. 2008). These animals were embedded in an environment in which this kind of “education” was the norm; their entire lives were about learning in this manner. It is not surprising that Herman (2002) made the following observation when describing his own work:

It seems clear that many of the studies we undertook would not have been possible, and many of the dolphin capabilities described would have gone unrevealed, without the implementation of the initial guiding educational philosophy. Immersion in a long-term program of intensive special education results in the accumulation of knowledge, concepts, rules, strategies and a general level of intellectual sophistication that allows for the understanding and solution of a broad range of increasingly complex problems or tasks. (p.9)

Animals consistently trained or taught in this manner simply have more experience interpreting the world according to the scientists’ demands. Qila had no such prior experience.

Conclusion

The ability of the belugas in this study to emit different vocalizations and to remain silent on command, and the rapidity with which they switched the context of these vocalizations from visual to acoustic stimuli during the training process, point to a flexibility in the vocal domain and to a mastering of the first three levels of usage learning described by Shapiro et al. (2004). Qila’s performance in the categorical matching experiment strengthens the notion that belugas are, at the very least, proficient in the first three levels – since she had to learn to produce elements of her own vocal repertoire in this very specific playback context. This study, however, is not conclusive about whether belugas are capable of the fourth level of usage learning in trained tasks.

The problem inherent in this sort of experiment is that a high above-chance response provides an unequivocal answer that a particular capability is present, but a poor or near-chance performance does not provide an equally unequivocal answer that the capability is absent, because we are unsure about the extent to which various procedural factors limited success. McGreevy and Boakes (2007) question the tacit belief that an individual animal that is very trainable in a particular task has a better understanding of this task than another individual that was difficult to train. Instead, they argue, variability in individuals’ responses to training
probably reflects various factors, including the trainer’s skills. A combination of the issues explored here might explain Qila’s difficulty. For example, performance might have improved if I had used more biologically appropriate vocal categories.

Compelling evidence of vocal production learning in adult belugas (Chapter 1), the insinuation that at least part of their extensive repertoire must be learned (Chapter 2), and the groundwork on vocal usage learning and categorical matching provided by the present study make belugas a compelling species for continuing studies of vocal learning at facilities harbouring these animals.
5. GENERAL DISCUSSION

In this dissertation I have sought to fill in some of the conspicuous gaps in our knowledge of the communication system of belugas, a highly vocal cetacean species. I proposed that the socio-ecological characteristics of belugas (long lives, extended maturation periods, long-term individual relationships, and cooperation, in an aquatic environment) would have favoured a system of vocal learning and plasticity. I explored this premise by examining production and contextual learning - the core components of vocal learning - both during repertoire development and in adult whales. I also began an inquiry into the function of beluga calls by investigating vocalizations associated with social bonding, i.e. contact calls. I focused my efforts on a captive beluga group at the Vancouver Aquarium, and used some of my findings to generate predictions about the vocal behaviour of their wild counterparts. I then tested some of these predictions opportunistically in the wild environment. In this chapter I will summarize the main findings of this dissertation, discuss some important ideas that have emerged from it, and outline future work.

5.1 Summary

In chapter 2 I documented vocal development in a captive beluga calf, Tuvaq, providing the first account of the sequence and timing of vocal acquisition in a beluga whale. The first sounds that Tuvaq produced after birth were pulse-trains, which continued to be the only sounds in his repertoire before he acquired rudimentary whistles at two weeks of age. His whistle production increased substantially in the second month of life. These whistles lacked a regular frequency modulation, increased in dominant frequency with age, and, at least in his first year, whistles did not attain full stereotypy. Tuvaq started using mixed pulsed-tonal calls consistently at 4 months. Two additional calves, born after the main period of data collection for this dissertation, also produced only pulse-trains after birth, and increased their pulse repetition rate over time. These findings indicate that belugas can produce burst pulse sounds almost immediately after birth, but that the mechanisms of whistle and mixed call production may require further maturation of the sound production system.

In terms of the roles of vocal learning in call acquisition, two findings are revealing. Firstly, Tuvaq’s mixed calls progressively lost variability and increasingly resembled the mixed stereotyped contact call of his mother and half sister. Such ontogenetic changes in the acoustic structure of calls are not conclusive evidence for vocal production learning, because maturational
processes could explain those modifications. Nevertheless, such changes are certainly a prerequisite for vocal learning, because when vocalizations are adult-like from infancy, in the absence of auditory experience, vocal production learning is unlikely (Egnor and Hauser 2004). Secondly, Tuvaq’s vocal repertoire lacked his father’s distinct calls before their first acoustic contact, but he incorporated at least one of his father’s call types after contact. Again, maturational changes and the close genetic relationship between the two animals disqualify this as conclusive evidence of production learning. Moreover, Tuvaq may have simply reached an appropriate age to use a specific genetically programmed call type. However, his renditions of this call type were less stereotyped than his father’s, suggesting that Tuvaq had not yet perfected its production. Further, Tuvaq did not produce any vocalization resembling this call type at all before exposure to his father’s usage of this signal. If we concede that learning can be genetically guided, it is possible that some call types have a genetic basis but that their first production is triggered by hearing another individual’s use of it. Tuvaq may have had to learn to use the vocalization – and perfect its production - by hearing another individual’s (his father in this case) use of the call (sensu Janik and Slater, 2000). This is, of course, speculative. Further data on more calves exposed to different acoustic environments might shed light on the question of the role of learning in vocal development.

In Chapter 3 I showed that captive belugas at the Vancouver Aquarium use a broad band pulsed call type, labeled “Type A”, for establishing contact and maintaining group cohesion. Of 28 classified types, this particular call type was favoured in situations that seemed to require maintaining contact and group cohesion: during isolation, husbandry and pool-cleaning procedures, re-union of animals after separation, and immediately after births and deaths. In contrast, the Type A call made up a very small proportion of the repertoire during regular sessions where none of these contact-eliciting events were taking place. Further, this call type was used in mother-calf vocal exchanges, and in matching vocal exchanges between adults (antiphonal calling). I subjectively and quantitatively identified variants of this call type, some of which were shared only by one mother and her two offspring. The five variants do not represent individual signatures, although the possibility remains that there may be more subtle individually distinct parameters within each variant.

Taken together, the findings of Chapters 2 and 3 suggest that wild belugas utilize broadband pulsed vocalizations as contact calls, that females and their offspring share variants that can serve for long-term recognition, and that calves slowly develop such variants. These hypotheses developed from captive research serve as a springboard for research in the wild. I
verified the existence of this call type in the repertoire of groups of females and young and adult males in the St. Lawrence River estuary, and documented its usage by two wild individual belugas from different populations, Hudson Bay and St. Lawrence Estuary, in contexts that strengthened the inference that it functions as a contact call. Jointly, my captive and wild observations have provided reasonable certainty about how beluga contact calls sound, so that we may next begin to refine the details of their function (more about this in Future Studies).

In Chapter 4 I investigated vocal learning in adult belugas using operant conditioning. I tested the ability of a female beluga, Qila, for categorical matching - to respond to playbacks of two call types with vocalizations of the same type. This represents the most complex of the levels of usage learning in trained tasks identified by Shapiro et al. (2004). During the training stage of the experiment, Qila and two other whales, Aurora and Imaq, demonstrated call usage learning by quickly shifting the context of their vocalizations from hand cues to the new acoustic stimuli. However, the study is not conclusive about the extent to which belugas can be trained to vocally match sound categories, although an interesting and unexpected finding emerged: Qila successfully matched only one of the two call types, pulse-trains, precisely the class that is part of this species’ natural repertoire. She had difficulty matching screams, an artificial vocal class that is not used by these captive belugas for communication but was shaped by training. Given mounting evidence that animals categorize signals based on their meaning, the screams’ lack of function in the natural repertoire of belugas hints at the possibility that Qila had difficulty perceiving this sound categorically. This might explain, at least partly, her difficulty matching screams. Indeed, Schusterman (2008) noted that some difficulties inherent in vocal conditioning studies tend to remain unpublished, including observations that only certain call types, combined with appropriate modes of reinforcement, may render successful results.

Interpreting Qila’s limited success in the categorical matching task is problematic. As noted by Tyack (1993a p.125), “if one does not find strong evidence for the existence of a cognitive capability, one cannot conclude that this is strong evidence for the absence of the capability.” A combination of procedural factors might explain Qila’s difficulty, including the call categories selected for the experiment, the relatively short training sessions and overall time period allocated to this study (given Qila’s lack of prior experience participating in this type of study), and the training protocol.
5.2 A sound-centered species in today’s noisy world

I could not end a dissertation that focuses exclusively on the communication system of belugas without both addressing the problems such sound-centered animals face in their increasingly noisy environment, and discussing how my data can help us think about these problems. Anthropogenic noise from a number of sources (e.g. hydrocarbon and seismic exploration, ocean dredging, military activities, commercial shipping, fishing vessels, passenger ferry traffic and recreational boats), has increased significantly in most marine areas over the past decades (Richardson et al. 1995; Erbe and Farmer 2000; Tyack 2008a). In the Arctic, noise sources are largely from oil and mineral exploration (Erbe 1997), although as Arctic ice cover continues to decline as a result of climate change, northern passages will become more navigable (IUCN 2010), which will increase the exposure of belugas and other marine mammals to shipping noise. In the busy St. Lawrence Estuary, a mainstream route for commercial shipping and home to a thriving whale-watching industry (Scheifele et al. 2005), noise pollution can exacerbate the strain that chemical pollution has already imposed on this threatened beluga population.

Noise can disrupt echolocation and mask environmental sounds that the animals use for navigation and the sounds of predators and prey. Further, it can trigger avoidance reactions and stress responses, with their associated physiological and energy costs, and lost opportunities such as mating, nursing, or feeding (Erbe and Farmer 2000; Tyack 2008a). In extreme cases, it can cause temporary or permanent hearing loss. Of most relevance to this thesis is the potential for anthropogenic noise to mask biologically important signals, interfering with acoustic communication (Erbe and Farmer 1998; Miksis-Olds and Tyack 2009).

The consequences of the interference of noise with communication become more obvious in the case of critical vocalizations such as those that maintain contact between mothers and calves. Based on my data on beluga calves’ vocalizations (Chapter 2), and on contact calls (Chapter 3), my colleagues and I (McKillop et al. 2010) looked at the potential for shipping noise to mask calls used by beluga mothers and offspring to maintain or regain contact. In Chapter 2 I described the vocalizations of newborn beluga calves, and noted the low peak frequencies (2-3.4 kHz) and narrow energy distributions as compared to the broad band adult contact calls. Analysis of the peak frequency of one calf’s vocalizations showed an increase by 4-5 months (see Figure 2.7), a time when calves in the wild would begin to encounter ice and extreme weather conditions, both of which elevate broadband sound fields. McKillop et al.
(2010) noted that shipping in the Arctic is effectively limited to the ice-free summer months and therefore exposes calves to broadband sound at a point in their vocal development when their low energy, undeveloped pulse-trains are easily masked, reducing the ability of mothers to hear them and thus increasing the risk of mother-calf separation. This is one example of how learning to identify the beluga call types that have important survival value puts us in a better position to evaluate the impacts of human-related noise on their communication signals.

5.3 Future work

The research in this dissertation raises several important questions. A number of possible experiments could follow from the work presented here. I alluded to some in the discussion section of each chapter, and I will propose others here. Some seem like such logical, elegant steps that I cannot help but finding it somewhat disillusioning to not have undertaken them myself (more on this in ‘Final Thoughts’). But I will have to content myself with giving them some consideration in these final pages.

The common question of what constitutes a functionally meaningful vocal category to the studied species did not elude this thesis. We learned here that loud, broad band, stereotyped, pulsed calls generally emitted in bouts function to establish or maintain contact, and I have named this signal class “Type A”. But are the five Type A sub-categories, or variants, identified here biologically meaningful categories? Playback experiments are the best available tool to learn not only how animals perceive call categories identified by humans, but also what these calls may mean to the animals. Results of playback experiments with several species of non-human primates indicate that their vocal repertoire, when assessed by the primates themselves, is larger than classification by humans would suggest (Cheney and Seyfarth 1982). We need similar investigations into whether playbacks of different beluga signals reliably elicit different responses. In the wild, playback experiments with belugas are extremely difficult and costly (I will come back to this later). In captivity, however, the possibilities for playback experiments are endless.

A useful playback technique to uncover perceptual categorization of acoustic stimuli is the habituation-dishabituation paradigm. This method is used with non-human animals and pre-linguistic human infants to determine if an individual treats two stimuli as the same or different (Hauser 1996). Experiments consist of repeating the presentation of a stimulus until the subject becomes habituated to it, and then introducing a stimulus of a different class. If both stimulus
classes are perceived as the same, the subject will show no dishabituation, defined as a significant change in response; if, on the other hand, the new stimulus is perceived as different from the first one, the subject will show some evidence of dishabituation. Measurable responses widely utilized in playback studies of primates are the direction of the gaze and the duration of looking (Hauser 1996). Habituation is shown, for instance, by a failure to look towards the speaker or by greatly reduced looking time (e.g. alarm calls of vervet monkeys, Cheney and Seyfarth 1988; 1990). Belugas differ from other cetaceans in that they do not have fused necks and the direction of their gaze is easily observable, affording the rare opportunity to use this response measure. Presentation of pairs of Type A variants within a habituation-dishabituation paradigm might reveal if the different variants are meaningful to the belugas themselves. For example, we could present the A1 variant during the habituation phase, and introduce the A2 variant during the test phase. If a significant change in response occurs, the two variants are probably also categorized as different by the whales (although this test would tell us nothing about what the variants actually mean). Furthermore, taking advantage of the fact that, at the Vancouver Aquarium, Imaq and Kavna are often kept in a different pool acoustically and visually isolated from the rest of the social group, we could present to them Qila’s rendition of the A1 variant during a habituation phase, followed by Aurora’s A1 during the test phase. A dishabituation reaction in this case would inform us that the whales attend to caller identity, and that there might be individual signature information within each variant type.

Related to this is the question of which acoustic dimensions are salient to belugas when they categorize their signals. Classification by humans on the basis of particular parameters may be statistically justified but may lack biological relevance unless the belugas are using the same parameters to categorize their vocalizations. The Type A variants identified in this study (Chapter 3) are quantitatively justified on the basis of two main parameters, pulse repetition rate and peak energy, but we have yet to determine if belugas use the same features to categorize these calls. Similarly, the pulse-trains and screams used for playbacks in Chapter 4 are distinct to humans based primarily on their markedly different pulse repetition rates. But do the whales use the same acoustic parameter to categorize these two call types? An important follow-up to both the contact call study and the categorical matching experiment might address the question of which acoustic dimensions of these signals are perceptually important to the belugas. What features do they use to differentiate between scream and pulse-train playbacks? And between Type A variants? Again, we could use a habituation-dishabituation paradigm and design different sets of playback pairs of calls differing in a particular parameter.
Another important consideration that emerged from this thesis is that of categorical perception in relation to function. In Chapter 4 I attempted to demonstrate categorical matching in trained tasks utilizing at least one vocalization type that has no particular function other than obtaining a food reward during shows. Given the inconclusive results of the study, the experiment could be repeated with vocal categories known to be functional, as these may be more easily categorized by the whales. In retrospect I feel that I might have approached this study with the wrong question: rather than “can captive belugas categorically match playbacks of their calls?” asking “which call types are captive belugas most likely to match?” might have been more rewarding. Vocal matching in nature occurs in a wide variety of species and serves different functions (see Chapter 4, Introduction). In belugas, the demonstration of natural vocal matching exchanges of Type A calls (Chapter 3) shows that categorical matching plays a role in their communication. It follows, then, that a more productive line of research into categorical matching could use the Type A calls described in Chapter 3. Given their function as contact calls, would belugas match Type A calls in trained tasks more readily than other call categories? One caveat is that animals may be reluctant to use signals with a strong emotional content out of context.

Despite its challenges, the categorical matching experiment did corroborate the idea that belugas can learn to shift the context of their calls, indicating usage learning. In Chapter 1 I reviewed reports of spontaneous imitation by captive belugas, intriguing observations that suggest that production learning plays an adaptive role in this species. The next level of complexity following demonstration of usage and production learning might be to train the whales to associate particular objects with specific sounds (i.e. referential labeling), so that we may infer something about this species’ ability for referential signaling. This step is not only a clear extension of usage learning (an object provides a context for a particular sound), but also of production learning, since learning to produce a sound to match a particular object can only proceed if the animals can imitate that sound. As put by Andrew (1962, pg. 586): “A particular meaning for a particular pattern of vocalizations could hardly be transmitted from one individual to another until the pattern itself could be readily imitated [italics mine].” There is now ample evidence, both from the wild and captivity, of species that can produce and understand signals that have external referents (i.e. that correspond to particular stimulus categories), including primates (e.g. vervet monkeys: Seyfarth et al. 1980), birds (e.g. African grey parrots: Pepperberg 1991; 1999), and cetaceans (e.g. killer whales: Dudok van Heel et al. 1982; bottlenose dolphins:
Richards et al. 1984; Reiss and McCowan 1993; see Evans 1997 for a thorough and critical evaluation of referential signaling). It is valuable to keep adding to this body of data, not only for comparative purposes, but also to provide an empirical basis for discriminating between those aspects of language that are unique to humans and those that are more widely shared (Evans 1997; Hauser et al. 2002).

However, such a dedicated set of future experiments at the Aquarium would necessitate somewhat of a shift in its operational priorities. Studying vocal learning experimentally within the framework of an aquarium presented more challenges than studying vocal development and call function. The latter required only that I adapt my observation and recording schedule to the institution’s routines, simply incorporating all the various contexts, such as feedings, shows, pool cleaning, husbandry, and ‘quiet’ times. Studying vocal learning, however, required the active involvement of the trainers and the specific allocation of the belugas to research procedures as often as possible. Despite the good will of the trainers, such requirements often conflicted with other institutional priorities. Clearly, at aquaria this kind of research can be constrained by the competing uses of the animals for demonstrations or display, although this situation may improve as the benefits and educational value of research gain appreciation by managers (Herman 2002) and the public.

Additionally, to study vocal learning experimentally I had to adhere to the training paradigm of the institution: operant conditioning with food reinforcement. Yet a wealth of evidence (see Chapter 4) suggests that methodologies that deviate from the standard operant conditioning techniques prevailing at aquaria and zoos should be considered when working with highly social, intelligent animals. Herman (2002), writing about the philosophy behind the long-term research program into dolphin cognition and behaviour at the Kewalo Basin Marine Mammal Laboratory, noted that

...the intellectual potential of a long lived, presumptively intelligent species, like the dolphin, is best revealed within a long term program of intensive, special education within a culture that values education. As these conditions surely favor the emergence of the full flower of human intellect, can comparable conditions also reveal the intellectual potential of other targeted species? (pg.1).

Indeed, given the long lives and high sociability of belugas, protocols that emphasize interaction, demonstration, and social rewards - used successfully with other long-lived social species (dolphins, grey parrots) - might uncover a wealth of capabilities.

In closing, Chapter 3 discusses a number of research possibilities in the belugas’ natural environment - ideas that have sprung from the captive research. These ideas include the
following: (a) Obtain acoustic recordings as often as possible during beluga field studies that involve capture or restraint. Ideally, acoustic recordings would become a standard protocol of such operations. Since belugas produce contact calls when restrained, and calves are known to swim near the restrained adults, systematic recordings could be used to identify individually distinct calls, and to determine if females and their calves share contact call variants. (b) Perform signal propagation experiments to elucidate the advantages of the kind of broadband contact call structure described in this dissertation, for both the pulsed variants (A2, A4 and A5), and the mixed variants (the tonal-pulsed A1, and the double-pulsed A3). (c) Conduct playback studies with free ranging belugas. The tendency of the Type A calls to be produced in antiphonal call matching exchanges in captivity make them ideal candidates to use playbacks to verify that this call type is used similarly in the wild environment. If these vocalizations function in nature, as in captivity, to maintain or establish contact, we would expect an increase in the rate at which these calls are produced in a beluga group after a playback sequence of such calls. Assuming that the calls contain cues to individual identity, the playback sequence would have to originate from animals known to the group to which it was being broadcasted (i.e. belugas may only respond to contact calls of specific, known individuals). To begin to explore this prediction, my colleagues and I conducted a two-week pilot study in August, 2008, at the confluence of the Saguenay River and St. Lawrence Estuary, and identified some of the methodological considerations and logistic challenges inherent to playback studies with free-ranging belugas. A summary report of the pilot study includes recommendations for the realistic design of future playback studies with this species (Vergara et al. 2009, unpublished manuscript).

5.4 Final thoughts

A king summons the blind men of the capital to the palace, where they are brought to stand by an elephant. ‘Well, blind man, have you seen the elephant? Tell me, what sort of thing is an elephant?’ asks the King. One, feeling the leg, declared the elephant to be like a tree trunk; the second blind man touched the elephant’s stomach and said that the elephant was a wall; the third, touching only the elephant’s ear, said that it was a fan; a fourth, grasping the tail, said that the elephant was like a rope; the fifth blind man felt the tusks and described the elephant as a spear; and the sixth blind man rubbed the elephant’s snout and got very scared because he thought it was a snake.

Popular Buddhist fable in the Avadānas

Through the rollercoaster years as a doctoral student I have learned how not to do science – and, as a consequence, how to do things well, how to become a better scientist. This is, I
understand, not uncommon – after all, that is precisely the purpose of a PhD. Unfortunately, this understanding often comes much too late in one’s project, when there is no going back and fixing things. There was a time not long ago when I had trouble considering my thesis finished without attempting at least some of the further studies outlined earlier. I felt that I was just beginning to uncover the questions most worth asking, and, most importantly, the questions that can be answered with the captive beluga group that was accessible to me. Such is the nature of science. Each discovery leads to more questions, and the more immersed one is in a subject the better one becomes at identifying worthwhile research. What a pity to have to write and submit just when I could do a little more research and produce a better thesis! But then a wise friend reminded me that “you never finish your PhD project; you merely abandon it.” It is thus with equal parts relief and regret that I put this thesis to rest – but not without a final word. Early in our trajectories to become good biologists we are taught to mind the ‘so what’ question. Why is this study interesting? Why did I bother? How has it enriched our understanding? I hope I did a decent job of addressing these questions throughout this dissertation. But I must confess the real, unpainted truth: ultimately, what compelled me to embark on this colossal journey is a deep sense of curiosity and wonder. The fundamental question that drove me has been all along: what is it like to be a beluga? What is it like to be an intelligent, long-lived, social creature with a predominantly acoustic perception of the world - an aquatic, three-dimensional world at that? Can we ever know? Perhaps not, as the philosopher Thomas Nagel (1974) concludes in his famous essay “What is it like to be a bat?” But I, like others, agree that Nagel was perhaps “too pessimistic and declared impossible what is merely difficult – and fascinating” (Cheney and Seyfarth 1990, p. 2). Studying communicative signals certainly allows us to make substantial progress, as they might just be the best “windows” into what animals think and feel (Griffin 2001). Like “the blind men feeling a great elephant”, the Buddhist proverb which alludes to the celebrated fable, my thesis may have illuminated only one fragment of what a beluga is, but I hope that the questions that it helped answer, and the questions that it gave rise to, helped us get a little closer to seeing the whole animal.
REFERENCES


Baptista LF, Schuchmann KL (1990) Song learning in the Anna hummingbird (Calypte anna). Ethology 84:15-26


Beach FAI, Herman LM (1972) Preliminary studies of auditory problem solving and intertask transfer by the bottlenose dolphin. The Psychological Record 22:49-62

Belikov RA, Bel'kovich VM (2003) Underwater vocalizations of the beluga whales (Delphinapterus leucas) in a reproductive gathering in various behavioural situations. Oceanology 43:112-120


Caldwell MC, Caldwell DK (1972) Vocal mimicry in the whistle mode by an Atlantic bottlenosed dolphin. Cetology 9:1-8


Fish M, Mowbray W (1962) Production of underwater sounds by the white whale or beluga, *Delphinapterus leucas*. Journal of Marine Research 20:149-162


Hauser MD (1989) Ontogenetic changes in the comprehension and production of vervet monkey (Cercopithecus aethiops) vocalizations. Journal of Comparative Psychology 103:149-158, doi: 10.1037/0735-7036.103.2.149


Kroodsma DE (1984) Songs of the alder flycatcher (Empidonax alnorum) and willow flycatcher (Empidonax traillii) are innate. Auk 101:13-24


Lilly JC (1965) Vocal mimicry in Tursiops: Ability to match numbers and durations of human vocal bursts. Science 147:300-301


Nagel T (1974) What is it like to be a bat? The Philosophical Review 83:435-450


Schevill WE, Lawrence B (1949) Underwater listening to the white porpoise (*Delphinapterus leucas*). Science 109:143-144


Suthers RA (1990) Contributions to birdsong from the left and right sides of the intact syrinx. Nature 347:473-477, doi: 10.1038/347473a0


Turl CW, Skaar DJ, Au WW (1991) The echolocation ability of the beluga (Delphinapterus leucas) to detect targets in clutter. Journal of the Acoustical Society of America. 89:896-901


Tyack PL (1991) If you need me, whistle. Natural History 8:60-62


Sample spectrograms of each of 28 call types classified from recordings of the Vancouver Aquarium belugas from 2002 to 2008 (FFT length: 1024, frame size: 100%, window: hamming, overlap: 75%, frequency resolution 43 Hz, time resolution: 5.8 ms). All spectrograms are drawn to scale, the X axis mark (time) reading “2 s”, and the Y axis (frequency) ranging to 24 kHz. Between brackets are the names used to refer to some of the vocalizations, due to their aural quality. This figure illustrates the distinctiveness of the Type A call in relation to the overall repertoire. Several of these call types have variants that aurally fall into the same general call category but that show some spectrographical differences in duration, energy distribution, or pulse repetition rate. In this paper (Chapter 3) I considered only the five variants of the Type A call (A1 – A5), quantitatively verified. Note that three vocal types that might look similar in these small scale spectrograms (A, C, and P) are indeed aurally and spectrographically quite different. For instance, the scratch cannot be broken into harmonic intervals nor resolved into clear pulses, it is a noisy call that appeared enough times in the repertoire to be warranted a category.