Using Neural Networks to Categorize Bottlenose Dolphin (*Tursiops truncatus*) Whistles: Age and Sex Differences

Brittany McIntosh

M.S. Research Project

Evolution, Ecology, and Behavior Program

University at Buffalo, SUNY

Table of Contents

Abstract

Bottlenose dolphins (*Tursiops truncatus*) have a complex communication system composed of a variety of sounds including narrow-band frequency modulated whistles. These have typically been studied in the context of a signature whistle framework, and reported whistle variation has been largely ignored in the literature. In order to investigate the information content of whistles from a broader perspective, a modified source-filter theory was applied to understand how the whistle producing mechanism in dolphins (phonic lips) may directly affect acoustic parameters of the whistle. In other mammals, the source-filter theory states that acoustic parameters of vocalizations potentially contain information about the physical characteristics of the caller. To test the possibility of this in bottlenose dolphins, a study was conducted that investigated whether there were perceivable acoustic differences in dolphin whistles according to age and sex by testing their potential for categorization by a neural network (NN). Three types of neural networks were created: gender classifier (SEX-NET), calf/adult classifier (CALF-NET), and age class classifier (AGE-NET). Thirteen acoustic parameters measuring frequency, energy (spectral density), and duration were selected as inputs for the networks. Results showed that all NNs were able to learn to categorize whistles according to their respective category with 100% performance accuracy, but only the SEX-NET could generalize these categories to unfamiliar whistles with above chance performance. Fundamental frequency and energy parameters were shown to be most critical to the success of all three network classifiers. These results indicate that information content regarding age and sex of whistlers are available to dolphins, but if they perceive these cues remains unknown.

I. Introduction

The bottlenose dolphin (*Tursiops truncatus*) is one of the most cognitively advanced and socially complex mammal species in the world (Marino et al. 2007; Reiss et al. 1997). These dolphins have remarkably large brains, with an encephalization quotient closer to human range than even that of chimpanzees (*Pan troglodytes*) and other great apes, our closest genetic relatives (Marino 1998). Decades of experimental behavioral studies in captive dolphins show complex capacities consistent with the sophisticated structure of the delphinid brain. Results from these studies have documented that dolphins excel at learning and memory tasks (spatial

and vocal), vocal and motor mimicry, novel problem solving, symbolic/representational concepts and tasks, and have even shown some evidence of self-awareness (for review see, Herman 2002).

Socially, dolphins live in fluid and dynamic 'fission-fusion societies,' similar to that of chimpanzees (Smolker et al. 1993). Individual dolphins will encounter both familiar and unfamiliar dolphins over their lifetime, forming both long-term and short-term relationships. They are even known to form long-lasting male coalitions comprised of two to three males who work together to gain access to mates, and in some cases will form super alliances, in which two coalitions work together to defend against other coalitions (Connor et al. 2001). The maintenance of these intricate social relationships seems crucial for dolphins to contend with the problems they face within their aquatic environment, such as: finding food, avoiding predators, securing mates, and caring for young (Connor 2007; Reiss et al. 1997; Tyack & Miller 2002).

As mammals living in a marine environment, evolution has led to certain anatomical and physiological specializations in dolphins, favoring auditory and vocal adaptations over visual ones (Ketten 1997). Marine environments are conducive to the use of acoustics as the main method of long-range communication, because sound propagates much faster in water than in air, and because visual signals under water are unreliable and limited. Dolphins show some of the most striking and significant adaptations for acoustic communication, as they have evolved to echolocate (Tyack & Miller 2002).

Dolphin Communication- Whistles

Some researchers suggest that the complex social demands of dolphin societies in combination with their acoustic and vocal adaptations to the sea may have shaped an extremely advanced and sophisticated communication system in order to maintain individual and group relationships (e.g., Reiss et al. 1997). This communication system consists of some non-vocal sounds and visual displays such as aerial displays, bubbling, body posture, gestural displays, and tactile interactions (Pryor 1973), and vocal signals. Vocalizations are generally grouped into three categories; broad-band clicks (echolocations), narrow band frequency-modulated whistles, and broad band burst-pulsed sounds (Lilly 1962; Reiss et al. 1997). How exactly dolphins, and other odontecete cetaceans produce these vocalizations continues to be explored and there is much we do not know about their sound-production mechanisms. The study of whistles is a valid approach to better understanding dolphin communication since whistles are often used in social

interactions (Caldwell et al. 1990; Ferrier-i-Cancho and McCowan 2009; Herzing 2000; McCowan and Reiss 2001)

The sound production mechanism responsible for whistles and clicks in odontocete was initially thought to be the larynx, as is typically found in terrestrial mammals, but after Norris *et al*. (1961) first proposed nasal passages as the possible source of vocalizations most scientists began to study cetacean sound production in this context (Tyack & Miller 2002). Lilly (1962) was the first to propose that whistles, specifically, are produced within the nasal passages, in which air sac volume change alters resonance frequency and produces a frequency modulated whistle. Later, Mackay & Liaw (1981) proposed that whistles and clicks were produced by tissue vibrations without the use of the larynx. Recent work by Madsen *et al.* (2012) confirmed these speculations after the conduct of a heliox breathing study, which found that dolphins actually produce whistles via pneumatically induced tissue vibrations, rather than air resonance changes, making the term "whistle" a functional misnomer. The authors proposed that the phonic lips, two pairs of internal nasal lips (Cranford et al. 1996), were a "good candidate" for being the source of those vibrations, in which mass and tension of the lips would determine fundamental frequency, and would only be slightly affected by air sac changes (Madsen et al. 2012). A follow up study by Madsen et al. (2013) determined that phonic lips pairs are responsible for both clicking and whistling in dolphins, with the right lip pair mostly used for clicking and the left for whistling. This helps to explain how dolphins are able to echolocate and whistle simultaneously.

The "meaning" or functionality of dolphin whistles as a mode of communication is an area of study that is dominated by the "signature whistle hypothesis" (for review see, Harley 2008 and Janik & Sayigh 2013). Melba and David Caldwell first proposed the existence of a "signature whistle" in 1965, when they found that in a specific group of dolphins, each individual animal had a tendency to emit the same whistle repeatedly. They reported that though the whistle could vary slightly, the basic contour remained the same, and this whistle comprised 90% of their whistle repertoire. Since then, signature whistles have been recorded from 300 captive and wild bottlenose dolphins reported by several research teams in a range of contexts. It is widely accepted that these whistles function as individual recognition signals, which broadcast a dolphin's identity to its conspecifics (Esch et al. 2009; Harley 2008; Sayigh et al. 2007). Janik & Slater (1998) found that captive dolphins, when isolated, would repeatedly vocalize their signature whistle, but when allowed to swim together signature whistles were rarely used,

functioning as cohesion calls. Later work by Janik et al. 2006, reported that signature whistle shape, or contour, does indeed convey identity information to conspecifics, which they determined by synthesizing signature whistles and playing them back to isolated dolphins during capture-release events. Proponents of the signature whistle hypothesis admit that signature whistle acoustic parameters fluctuate, but believe these variations may convey state or location information to conspecifics rather than identity information (Caldwell 1990; Janik et al. 1994; Janik & Sayigh 2013; Harley 2008). Non-signature whistles have been reported during signature whistle studies (Sayigh et al. 1990; Janik et al. 1994) and have been found to make up about 48% of all whistles recorded in a study examining whistle variation in free-ranging bottlenose dolphins (Cook et al. 2004), but they have been relatively unstudied and largely ignored in the literature.

Though most studies have examined whistles in the context of understanding how signature whistles are used, some have reported variation in the whistle repertoire of the dolphin (Dreher 1966; Lang and Smith 1965; McCowan & Reiss 1995; Reiss et al. 1997). McCowan & Reiss (1995b, 2001) suggest that rather than dolphins having an individualized whistle type, dolphins across varying social groups produce a predominant whistle type, which can show acoustic variability in some parameters. They claim that this may also help explain the existence of regional dialects in different geographic dolphin populations (May-Collado & Wartzok 2008). There has been much discord between advocates of the signature whistle hypothesis, and those, like McCowan and Reiss, who question its validity. Researchers have debated methodologies for assessing whistle similarity and each has pointed out the limitations of each party's theories and experiments (Janik 1999; McCowan & Reiss 2001; Sayigh et al. 2007). Regardless of these two opposing views on the 'meaning' and information content of whistles, biologists on both sides of the spectrum report overall variation in the whistle repertoire of the bottlenose dolphin, as well as variation in the acoustic parameters of an individual dolphin's signature whistle contour (Caldwell 1990; Janik et al. 1994; Janik et al. 2013; McCowan & Reiss 2001; Sayigh et al. 1998). Researchers who have reported these variations have pointed out the importance and value of future studies to investigate these variations (Harley 2008; Janik et al. 1994; Janik et al. 2013; Kershenbaum et al. 2013; McCowan & Reiss 1997).

In order to examine whistle variation in a broader context than that of signature whistles, it is necessary to consider what purpose that variation might serve in dolphin communication.

Typically, variation in communication signals increases the information being transmitted between individuals (Tyack & Miller 2002). Brownlee and Norris (1994) postulated that the variation of whistles could carry information about a variety of social contexts and behavioral states. Modulation of whistle parameters ("intensity, frequency, and frequency pattern through time") can "…produce a complex system of great potential information-carrying capacity" (Brownlee & Norris 1994, pg.180). In comparison to most other animals that use individual acoustic variations of a shared call type to communicate information, dolphins are atypical in that they can be identified by an individual stereotypic whistle (Janik et al. 2006; Reiss et al. 1997). This may in fact be true, but it is still very likely that dolphin whistles have the potential to provide basic contextual cues about the whistlers. The idea that vocalizations can contain information about the signaler has been extensively studied with regard to mammal (and avian) vocal production and so it is useful to explore some of the existing research as it relates to dolphin vocalizations.

Source-Filter Theory as a Model for Dolphin Whistles

The field of mammal vocal communication has been greatly advanced by the application of the source-filter theory to non-human vocal production, by promoting a better understanding of vocal production mechanisms and resulting acoustic variations (Fitch & Hauser 2003; Taylor & Reby 2009). Source-filter theory, originally developed to study human speech (Fant 1960), links a mode of vocal production to acoustic characteristics which have the ability to reveal information about a caller's identity or phenotype (Briefer & McElligott 2011; Fitch & Hauser 2003; Taylor & Reby 2009). For humans and most non-human mammals, source-filter theory relies on the fact that vocal signals depend on the independent contribution of different parts of the vocal anatomy; namely, that vocal signals are generated by the airflow from the lungs which sets the vocal folds of the larynx into vibration ("source") which is then filtered in the vocal tract ("filter") (Briefer & McElligott 2011; Charlton et al. 2009; Fitch 2006; Fitch & Hauser 2003; Taylor & Reby 2009). Since individual differences would be expected in the morphology of the source and/or the filter, acoustic call parameters have the potential to contain indexical information, such as, body size, weight, age, hormonal state and sex (Briefer & McElligott 2011; Charlton et al. 2009; Fitch 2006; Fitch & Hauser 2003; Taylor & Reby 2009). Evidence that vocal signals contain indices of a caller's physical attributes has been documented in mammal

species such as, the red deer (*Cervus elaphus)* (Reby & McComb 2003), giant panda (*Ailuropoda melanoleuca*) (Charlton et al. 2009a, 2009b), hamadryas baboon (*Papio hamadryas*) (Pfefferle & Fischer 2006), yellow bellied-marmot (*Marmota flaviventris*) (Blumstein & Munos 2005), goat (*Capra aegagrus hircus*) (Briefer & McElligott 2011), and southern elephant seal *(Mirounga leonina*) (Sanvito et al. 2006) (for further review see, Taylor & Reby 2009). Similar results have been found in avian species such as, the banded wren (*Pheugopedius pleurostictus*) (Vehrencamp et al. 2013), Bengalese finch (*Lonchurastriata domestica*) (Cooper et al. 2012), European starlings (*Sturnus vulgaris*) (Pavlova et al. 2010), and the black-capped chickadee (*Poecile atricapillus*) (Hahn et al. 2013). As vocal communication is used to facilitate interactions between individuals of a species, such as competition, territorial disputes, acquisition of mates, and parent/ young recognition, the ability for receivers to differentiate among callers (size, age, sex, etc.) would seem beneficially adaptive (Charlton et al. 2009a; Owings & Morton 1998) For example, male yellow baboons (*Papio cynocephalus cynocephalus*), were able to discriminate copulation calls of their consort female and other females (Semple 2001), and Australian fur seals (*Arctocephalus pusillus doriferus)* may be able to recognize rival male calls and avoid unnecessary conflict (Tripovich et al. 2008). The ability to differentiate among callers in bottlenose dolphins would seem highly adaptive as they are an extremely social species that frequently encounter familiar and unfamiliar individuals. In this way the source-filter theory provides a tentative model for thinking about the possible sources of information content encoded within dolphin whistles, but their novel vocal production may require some modifications to the theory.

Unlike humans and most other non-human mammal vocalizations, dolphin whistle production mechanisms are much less understood and not perfectly analogous to the source-filter theory (Reidenberg & Laitman 2010; Tyack & Miller 2002). In terrestrial mammals airflow triggers vibration of the vocal folds and acts as the source of vocal production, while in dolphins the phonic lips are pneumatically driven to vibrations via adjacent air sacs in the nasal passage which act as the corresponding "source;" if applied to source-filter theory (Huggenberger et al. 2009; Madsen et al. 2012, 2013; Reidenberg & Laitman 2010). It is important to note that the phonic lips are not homologous to the vocal folds found in terrestrial mammals, but are more of a unique adaptation to an underwater environment, comparable to the avian syrinx which evolved to produce sound via pneumatic vibrations in birds (Madsen et al. 2012; Taylor & Reby 2009).

As of now, no filter has yet been determined for the vocal production of whistles in cetaceans, though some have suggested the melon to act as a filter (Reidenberg & Laitman 2010). If the source-filter theory is generalized to a model simply linking vocal production anatomy to acoustic call parameters it can be a useful method for predicting what information might be contained within dolphin vocalizations. If this simplified model is applied to bottlenose dolphins, the phonic lips would act as the major source (without a known filter) potentially producing perceivable differences in call parameters according to physical attributes of individuals.

To investigate this possibility, a study was conducted analyzing acoustic parameters of whistles as they relate to physical attributes of the whistler (age and sex). To do this, frequency parameters of whistles were measured, as the vibrating mass and tension of the phonic lips directly determines the fundamental frequency of dolphin whistles (Madsen et al. 2012). The vibrating mechanism in other mammals (vocal folds) also determines the fundamental frequency. (Taylor & Reby 2009). In humans, longer and heavier folds vibrate at a slower rate than smaller vocal folds, which explains why adult males have pitches about an octave lower than adult females, since vibrating segments of male vocal folds are 60% longer than those in females, and are also much larger (Titze 1989). The sexual dimorphism found in human vocal folds is originally caused by increased testosterone levels that males experience at the onset of puberty which stimulates fold growth (Beckford et al. 1985). Pitch differences related to vocal fold length and mass also allows human listeners to differentiate between children and adult voices, as children will have higher pitched voices than adult women and men (Smith & Patterson 2005). Evidence of fundamental frequency differences in the age and sex of individuals related to "source" morphology has been found in other non-human mammals, such as the giant panda (Charlton et al. 2009b), baboon (*Papio spp.*) (Rendell et al. 2004), and goat (Briefer & McElligott 2011). In a study examining sex differences in the vocalizations of the syrinx of collared doves (*Streptopelia decaocto*), the authors found evidence that vocal morphology may contribute to sexual dimorphic vocalizations (Ballintijn & Cate 1997). This study is particularly relevant to the possibility that acoustic cues in bottlenose dolphins may encode indexical information of body characteristics, because the avian syrinx is so similar in function to the phonic lip mechanism. Though bottlenose dolphins are moderately dimorphic (Tolley et al. 1995), there are no studies which have measured variations of size and function of the phonic lip, so it is impossible to correlate phonic lip dimorphisms to acoustic parameter differences in

whistles. This study seeks to investigate the possibility that there may be perceivable acoustic parameter differences in dolphin whistles produced by individuals of varying age and sex, as have been found in many other mammal species.

Study Objective

Sex and age related differences in non-signature whistles have not been investigated as they relate to acoustic parameters, and general frequency measures have been ignored in favor of frequency contour measures. For example, Sayigh *et al.* (1995), examined sex differences in signature whistle production of free-ranging dolphins, and reported that female calves produced whistles that were different from those of their mothers, whereas male calves produced similar ones, but the authors ignored non-signature whistles, and did not examine any variation beyond the main frequency contour. Studies regarding whistle ontogeny have revealed that infant dolphins produce sounds that are tremulous and quavery, with small frequency modulations; most similar to whistles that are produced by adult dolphins when under stress. In young dolphins this may be indicative of a lack of muscular development and control (Caldwell & Caldwell 1979; Killebrew et al. 2001). Sex related differences in vocalizations have been reported in another cetacean species, the killer whale (*Orcinus orca*). Miller *et al.* (2007) found that adult sex influenced the relative energy of the first two harmonics of the low frequency component of killer whale calls. The authors predicted that these sex differences should be detectable by killer whales. This would suggest that body size differences due to age would cause similar differences, though this would have to be investigated further. Results such as these, suggest the possibility that dolphin whistles may encode similar information.

The purpose of this study is to explore the communicative value of dolphin whistles from a previously unexplored perspective. This study aims to examine whistle variation, and provide insight into the types of whistle sound features that may contain indexical information that dolphins are potentially transmitting and receiving between individuals. Past studies have measured and categorized whistles in a variety of qualitative and quantitative ways, including but not limited to, the human observer method, McCowan's (1995a) computer frequency analysis method, Janik's (1999) two cross-correlation techniques, Janik's (2012) SIGID method, Buck and Tyack's (2003) dynamic time-warping technique, and Deecke and Janik's (2006)

ARTWARP neural network technique (for review see, Harley 2008). Neural network haves been used to classify many different types of animals vocalizations, such as, deer (*Dama dama*). vocalizations (Reby et al. 1997), black-capped chickadees (*Poecile atricapillu*) calls (Nickerson et al. 2006), bat (*Rhinolophus spp.*) echolocations (Parsons & Jones 2000), and tungara frog (*Physalaemus pustulosus*) mate recognition calls (Phelps & Ryan 1998). They have also been used in several studies regarding categorization of marine mammal calls; sperm whale (*Physeter macrocephalus*) clicks (Schaar et al., 2007), false killer whale (*Pseudorca crassidens*) vocalizations (Murray et al. 1998), dolphin echolocations (Au et al. 1995), and humpback whale (*Megaptera novaeangliae*) songs (Mercado et a. 2008). Essentially, neural networks are computer programs that simulate neural information processing in learning to classify various inputs through pattern recognition (Ripley 1996; Haykin 1994). They are typically composed of three or more units (neurons), an input layer, an output layer, and at least one hidden layer. These layers are connected by a set of processing units between them, which are modified by vector weights (Reby et al. 1997; Dawson 2004). Neural networks are valuable as a classification method because they do not necessitate human qualitative analysis, and have the benefit of being able to process large variable data sets quickly and without bias; therefore, they have been suggested as a reliable and effective method to categorize animal vocalizations (Reby et al. 1997; Murray et al. 1998).

This study investigates if acoustic differences in dolphin whistles differentiate individuals by age and sex by testing whether a neural network could learn to categorize individuals using whistle features. It examines the information content of dolphin whistles, regardless of "signature and "non-signature" categories, by considering a broader range of acoustic features than just frequency contour, which has been the focus of most studies investigating dolphin whistles (Janik et al. 2006; for review see, Harley 2008). Rather, in this study measurements pertaining to spectral frequency were the focus, as many of the mammal vocalization studies previously cited in this paper, have examined the fundamental frequency parameters and found them to serve as reliable indicators of physical attributes of the caller (Ballintijn & Ten Cate 1997; Briefer & McElligott 2011; Hahn et al. 2013; Pfefferle & Fischer). Though it is likely that the fundamental frequency of whistles strongly affected frequency measures in this study, there were some instances of whistles overlap with trains of echolocation clicks and minor ambient noise, which may have also contributed to the measured values. Therefore, the claim cannot be

made that frequency measures collected in this study are exclusively of the fundamental frequency, and so they will simply be referred to as frequency measures. Measures of duration and energy (spectral density measures) were also included in the study to ensure acoustic parameters represented a wide range of variables that could contain information content in whistles. The objective of the study is not to contradict, nor refute existing theories of whistle functions, but to examine the use and potential communicative value of whistles from a new perspective. To further explore the value of whistles as a mode of dolphin communication, they need to be evaluated with a form of quantitative analysis, which addresses whistle variation in a reliable and objective manner. Using neural networks to classify bottlenose dolphin whistles according to relevant social cues could potentially provide insight as to the information value of whistles.

II. Methods

Study Site and Subjects

Data was collected in 2009 and 2010 at the Roatan Institute for Marine Sciences (RIMS) at Anthony's Key Resort in Roatan, Honduras which is home to a captive population of bottlenose dolphins (*Tursiops truncatus*). The group is housed in a 300m^2 fenced enclosure within a natural lagoon covered by native coral, sand and sea-grass beds. The enclosure occupies depths ranging from the shallows of the shoreline to about 8 m (Dudzinski et al. 2010). Group size ranges from 20-24 dolphins (due to new births), and includes both wild caught and captive born individuals with ages ranging from neonate to 30+ years (Dudzinski et al. 2010; K. Dudzinski, personal communication, February 13, 2014). Socially, this captive group is comparable to wild bottlenose dolphins populations as their group composition (age and sex) is very similar to that observed in the wild (Connor et al. 2006; Dudzinkski et al. 2010; Kogi et al. 2004).

Data Collection

Dolphin whistle audio recordings were isolated from archived video/audio footage collected by the Dolphin Communication Project (DCP) (2009-2010) at RIMS. A customized underwater mobile video/acoustic system (see Dudzinski et al. 1995), composed of a digital

Eduardo Mercado 2/24/14 8:31 AM
Deleted:

video camera (Sony HDR HC5/HC7 in HD format, set to stereo audio, 12 bit amplitude resolution, 32 kHz sampling rate, compressed HDV audio) secured in an underwater housing with two attached omni-directional hydrophones (normal receiving sensitivity of -155dB re1v/uPa) was used to capture synchronized video/audio recordings of RIMS dolphins. Video/audio footage was converted into digital video files (.dv) with two audio channels using a Sony Video Cassette Recorder (model GV-D1000 NTSC) processed through iMovie software. Subjects were filmed according to a focal follow sampling method, in which one individual within field of view was randomly chosen and opportunistically observed (and recorded) until they traveled out of frame (K. Dudzinski, personal communication, February 13, 2014).

Digital video files were subjectively examined aurally and through visual analyses of spectrograms to identify high quality whistles; those with minimal echolocation overlap, and those in which the whistling dolphin could be accurately identified. The whistling dolphin was identified via aural detection from the stereo-audio recordings, as the hydrophone spacing on the video/acoustic system was specifically designed to accommodate the human interarual distance based upon the speed of sound in water (approximately 5 times faster in water) (Dudzinski et al. 1995). The ability to detect directionality (left, right, and center) of sound sources on the stereo-audio recordings and associate those sounds with visual distribution of subjects on screen allows for the identification of a whistling dolphin when in field of view (Dudzinski et al 1995). Whistling dolphins could only be identified using this method when the vocalizing dolphin was fairly isolated (no more than three dolphins on screen), clearly visible on screen, only a short distance from the camera (~5m or less), and whistling was loud with little interference from other sound sources. Additionally, identification of the whistling dolphin was facilitated by behavioral and contextual information, such as emission of bubbling streams from the blowhole and/or unusual swim patterns and movements. Dolphin subjects were identified in video footage by DCP trained researchers via naturally occurring marks and scars, and were previously archived in video logs that were used for this study.

During video analysis, every whistle that met the criteria mentioned previously was used for this study, with no preferential selection of whistling from individuals of a certain age or sex. Whistles used in this study were not equally representative of all twenty dolphins identified as whistlers, since certain dolphins were more likely to swim and vocalize close to the camera. Since the amount of video footage is biased towards certain dolphins that are more apt to come

close to the camera, whistle selection was maximized to obtain as many whistles as possible, in the hopes that it would also maximize the amount of individuals that were represented in the sample.

Four hundred whistles were analyzed using the bioacoustics analysis software, RavenPro 1.5, wherein whistle spectrograms were examined and selected whistles saved as AIFF files (32 kHz sampling rate, 16-bit amplitude resolution), spectrogram FFT was 512 pts. The audio channel with less echolocation noise was selected as the whistling spectrogram to be used during analysis. Thirteen acoustic parameters representing a wide array of measures were chosen to increase the information available to the NNs (Table 1). These included: bandwidth 90%, average entropy, IQR bandwidth, center frequency, frequency 5%, frequency 95%, peak/max frequency, average power, duration 90%, energy, IQR duration, mean frequency, aggregate entropy, begin time, and end time. Total whistling duration was calculated manually. Descriptive statistics (min, max, mean, standard deviation) were calculated for all acoustic parameters for each category (as well as overall). Additional information recorded for each whistle included: whistle screen time occurrence, vocalizer ID and age class, instances of echolocation overlap, instances of bubbling during whistling, and any notable behavior observed during whistling (e.g. parallel swimming, sexual behavior, object play, etc.).

Neural Network Analysis

Artificial neural network software from the University of Alberta (available for download http://www.bcp.psych.ualberta.ca/~mike/Software/index.html) was used for all sound classifications in this study. Specifically, Rumelhart software, a multi-layer perceptron neural network (NN) program used to explore pattern classification was employed to differentiate dolphin whistles according to categories of age and sex. Three separate NN classifiers were created for this study. The first was trained to discriminate sex differences between males and females (2 outputs) based on whistling characteristics, and will be referred to as the SEX-NET (see Figure 1 for example NN structure). The remaining two NN classifiers were trained to discriminate age differences between whistlers. In order for the NNs to discriminate between certain age groups, age classes were assigned to all study subjects according to four broadly used categories: calf (3 wks-4 yrs.), juvenile (4-7 yrs.), sub adult (7-10 yrs.), and adult (11+ yrs.) (K. Dudzinski, personal communication, February 13, 2014; Kogi et al. 2004; Melillo et al. 2009).

The first age related NN classifier was trained to discriminate between calves and adults (two outputs) and will be referred to as the CALF-NET, while the other NN classifier was trained to discriminate between the four age classes (four outputs), and will be referred to as the AGE-NET. All NN classifiers consisted of 13 input neurons representing the 13 measured acoustic whistle parameters noted earlier. All of these input values were standardized before being used in the NNs. Simulations with varying amounts of hidden units were tested for each NN classifier, and the one that yielded the best network generalization was selected. Ten hidden units were utilized for the SEX-NET, five hidden units for the CALF-NET, and ten hidden units for the AGE-NET.

In each NN simulation connections between units were initialized with random weights, and the learning rate was fixed at 0.05. Minimum squared error defining a hit was set at 0.01. Each of the three NNs was put through a training session, in order to test their ability to learn the age and sex categories, and a generalization test, in order to test their ability to generalize what they learned during training to novel whistles. Approximately half of the whistles collected from each respective category of whistlers (e.g. male, female, calf, juvenile, and sub-adult, adult) were used during training and the other half during testing. All whistlers were included in both training and testing periods. Each NN was trained for 10,000 trials, or until all outputs were identified 100% accurately. Generalization tests were run after training, in which the novel, unfamiliar test set whistles were run through the NN using the hidden unit weight value transformations that were achieved during testing. No training occurred during the generalization test period. The training and testing process was repeated ten times for each NN, so that performance accuracy measurements could be averaged over multiple NN configurations.

The percent of whistles correctly classified by each of the NNs during training and testing periods was used as a measure of performance accuracy. For the SEX-NET and CALF-NET, output unit weights were assessed to identify which input parameters were driving hidden units to suppress or bolster respective outputs. This assessment revealed which input parameters were specifically associated with correctly identifying particular outputs. Negative output unit values suppress the output, which pushes the output value towards zero (incorrect), and positive values bolster the output, which pushes the output towards one (correct). Once the association between a hidden unit and the output that activated it was established, absolute value means were calculated for those hidden units which related to each output, thereby determining which

acoustic parameters were critical to categorizing each respective output. The smaller the absolute value of the weighted connection the less affected the hidden unit is by that input, whereas large absolute values signify that hidden unity activity is strongly affected by that input. For the SEX-NET the three most significant hidden units associated with each output were selected for the analysis. To do this, the weight vectors connecting each whistle to each hidden unit were assessed. These vectors essentially represent the contribution each hidden unit had in correctly categorizing the outputs. The higher the mean vector weight value the more significant the hidden unit was in contributing to the success of the network. For the CALF-NET there were only five functioning hidden units. Rather than selecting the three most significant units for analysis, all high functioning units bolstering calf and adult classification, respectively, were selected. In most cases there was one hidden unit bolstering calf whistle categorization and the remaining four hidden units bolstering adult whistles. Hidden units associated with adult whistles were generally all included for the analysis, unless mean vector weights were significantly lower than the others.

T-tests were performed on those parameters which were shown to contribute most to the network, in order to determine whether there were significant differences in the raw data measures between competing categories (male/female, calf/adult). Binomial probability tests were performed to determine statistical significance of generalization testing performance for the SEX-NET and CALF-NET.

Output units could not be assessed for the AGE-NET, since hidden units could not be associated with a single output as there are four, rather than two, output categories. To analyze the significance of input parameters for this NN, absolute value of weighted connections between input units and hidden units were examined to determine how hidden unit activity relied on specific inputs. Mean hidden unit weight values were averaged for all hidden units from all ten NN configurations. Higher mean weight values determined which input parameters contributed more to the success of the network.
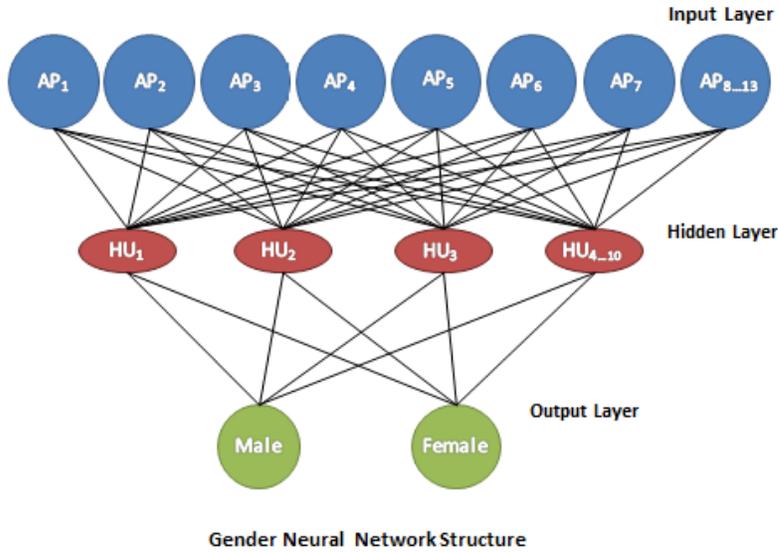
Figure 1. Example of the neural network structure for gender classification. Input layer, AP= acoustic parameter. Hidden layer, HU= hidden unit.

Table 1. Definitions for acoustic parameters used in whistle analysis. Definitions adapted from Raven Pro 1.5 Manual (Charif et al. 2010).

Acoustic Parameter	Definition
Energy (spectral density)	
Aggregate Entropy	The overall "disorder" in a sound, measured by analyzing the energy distribution within a selection. Higher entropy values correspond to greater disorder in the sound (i.e., larger differences in spectral peaks across frequencies).
Average Entropy	The amount of disorder for a typical spectrum within the selection, calculated by finding the entropy for each short-duration frame in the selection and then taking the average of these values.
Average Power	The value of the spectrogram's power spectral density, as it appears in each pixel of the spectrogram, averaged over the selection. Units: dB
Energy	The total energy within the selection bounds. Units: dB
Temporal	
Duration	Total time duration of the sound. Calculated by subtracting start time from end time (measured manually). Units: seconds.
Duration 90%	The difference between the 5% and 95% times. These are the points in time that divide the selection into two time intervals containing 95% and 5% of the energy in the selection. Estimates the duration across which most energy is distributed. Units: seconds.

IQR Duration	The difference between the 1st and 3rd Quartile Times. These are the points in time that divide the selection into two time intervals containing 25% and 75% of the energy in the selection and vice versa. Estimates the duration across which the highest energy elements are distributed. Units: seconds
Spectral (Frequency)	
Bandwidth 90%	The difference between the 5% and 95% frequencies. Units: Hz.
Peak Frequency	The frequency at which max peak power occurs within the selection.
Center Frequency	The frequency that divides the selection into two frequency intervals containing equal energy. Units: Hz
Frequency 5%	The frequency that divides into two frequency intervals containing 5% and 95% of the energy selection. Similar to center frequency except summed energy has to exceed 5% of total energy, instead of 50%. Units: Hz
Frequency 95%	The frequency that divides the selection into two frequency intervals containing 95% and 5% of the energy in the selection. Similar to center frequency except summed energy has to exceed 95% of total energy, instead of 50%. Units: Hz
IQR Bandwidth	Inter-quartile range. The difference between the 1st and 3rd Quartile Frequencies. These are the frequencies that divide the selection into two frequency intervals containing 25% and 75% of the energy in the selection (and vice versa). Units: Hz

IV. Results

Video Analysis

A total of thirteen hours of video (approximately half each from 2009 and 2010 video), was analyzed in order to collect a large sample size (398 total whistles) to obtain a wide variation of whistles from as many dolphins as possible. Whistles used in this study were recorded from twenty dolphins (11 males/ 9 females), 16 individuals recorded in 2009 and 18 in 2010 (Table 2). Three of the twenty dolphins are not included as whistlers for both years, as they were only observed whistling during one year's data collection. Also, three calves were birthed in 2010 that are only included in the 2010 data set of whistles.

During video analysis supplemental information was recorded in addition to acoustic parameters of whistles. Based upon inspection of whistle spectrograms, 36% of recorded whistles co-occurred with minor overlap with echolocation clicks. Observation of bubble emissions during whistling occurred for 63% of recorded whistles. Behaviors observed during whistling included, sexual play (with few erections), parallel swimming with another dolphin, object play (with fence, seagrass, fins of swimmer), and jumping out of water (interspersed with

whistling).   Descriptive statistics (min, max, mean, and standard deviation) for all acoustic parameters for each category are listed in Table 3.

Table 2. Total number of whistlers and whistles recorded during video analyses according to categories of age class and sex. Two adult dolphins were only recorded as whistlers in 2009 and one adult was only recorded as a whistler in 2010. Three calves were birthed in 2010 which were not included in the analyses of 2009 data. All other dolphins were present in both years.

Data Collection Year	Age				Sex		Totals
	Calf	Juvenile	Sub-adult	Adult	Male	Female	
Number of Whistlers- 2009	n= 2	n= 5	n= 2	n= 7	n= 8	n= 8	$n_{2009}=$ **16**
Number of Whistlers- 2010	n= 3	n= 6	n= 3	n= 6	n= 11	n= 7	$n_{2010}=$ **18**
Number of Whistles	43	204	75	75	251	147	**398***

* One whistle could only be identified according to gender; therefore there are 398 total whistles according to the sex category, and 397 total whistles according to age class category.

SEX-NET

　　　　The NN which was trained to classify whistlers according to gender was able to correctly categorize individual whistles (128 male whistles/ 75 female whistles) as coming from either a male or female dolphin with an average performance accuracy of 99% after an average of 8,977 training trials (Fig. 2). The network configuration that performed best yielded an accuracy of 100% after 4,592 trials (chance performance was 50% correct).
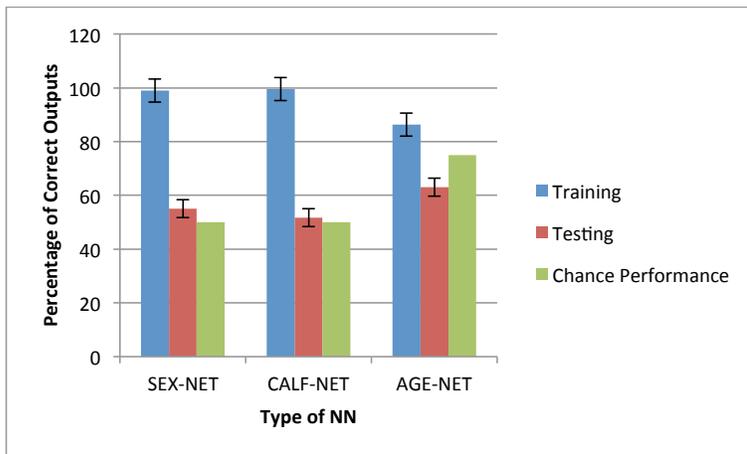

Figure 2. Performance accuracy (%) for all three types of NN classifiers, during training and testing. Chance performance % bars are included as a reference.

Summary statistics from the network were calculated for hidden unit weight vectors that connected the fourteen acoustic parameter inputs to the ten hidden units. Output unit weights were also evaluated for each of the five best performing network configurations. For these five networks, weight vectors from the three most significant hidden units for each output category were analyzed and used to determine which acoustic features were most associated with identifying male and female whistles respectively. The highest hidden unit weight means (absolute value) for the male-driven hidden units showed that frequency bandwidth, whistle duration, and entropy measures were most relevant for correctly identifying male whistles (Fig. 3, Table 4). Measures of hidden weight values for the female-driven hidden units showed that frequency measures, entropy measures, and whistle duration were most relevant for correctly identifying female whistles (Fig. 3, Table 4). The input parameters with the largest mean weight differences between male-driven units and female-driven units were center frequency and frequency 5% (Fig.3, Table 4). T-tests performed were performed for these input parameters using the raw measurements obtained during data collection. These tests determined that frequency (5%), bandwidth (90%), and entropy (average) were the acoustic parameters with significant differences between males and females (Figure 3).
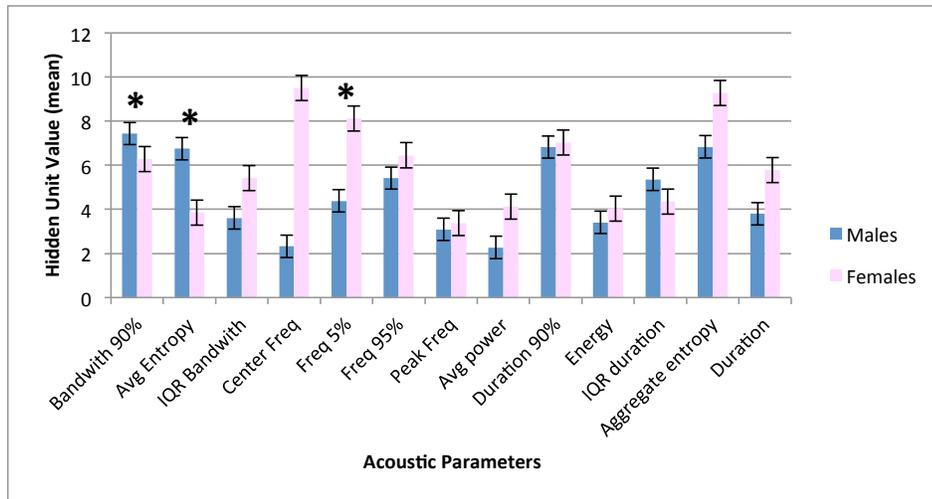

Figure 3. Mean hidden unit values calculated from the five highest performing gender classification networks during training. Means were calculated using the three hidden units that contributed the most to the success of the network. Higher hidden unit values (means) indicate significance to the NN. Bars are standard errors of the means. Asterisk indicates statistical significance for that acoustic parameter, $p < 0.05$.