Proceedings of Meetings on Acoustics

Volume 29 http://acousticalsociety.org/

172nd Meeting of the Acoustical Society of America

Honolulu, Hawaii

28 November – 2 December 2016

Animal Bioacoustics: Paper 4pABb3

Human listening experiments provide insight into cetacean auditory perception

Caroline M. DeLong

Department of Psychology, College of Liberal Arts, Rochester Institute of Technology, Rochester, NY; cmdgsh@rit.edu

Odontocete cetaceans produce an array of sounds including echolocation clicks that can be used for object recognition and whistles used primarily for communication. Echoes and whistles contain a multitude of acoustic features, and the salient features used by animals when identifying objects or in social contexts are often difficult to isolate. One method for detecting salient acoustic features is to compare the performance of cetaceans and humans on the same auditory perception tasks. Human listeners can be presented with echoes produced with simulated cetacean clicks or whistle-like stimuli and provide verbal feedback on discriminatory cues. An early human listening study was performed by Au and Martin (1989). Seven studies performed over the past 15 years show that humans perform as well or better than cetacean subjects in a variety of tasks: echoic discrimination of objects varying in size, shape, material, texture, or wall thickness; echoic recognition of objects from various aspect angles; echoic object discrimination using clicks from different cetacean species; and recognition of frequency-modulated whistle-like sounds. Analyzing the error patterns of humans and cetaceans alongside the reported cues reveals processing mechanisms and decision strategies cetaceans may use. This comparative approach sheds light on how cetaceans perceive and represent sounds.

Published by the Acoustical Society of America

1. INTRODUCTION

Odontocete cetaceans produce an array of sounds including echolocation clicks and whistles (Au, 1993). There are many acoustic features that can be measured in echoes and whistles (e.g., duration, target strength, peak frequency, spectrum shape, bandwidth). However, the salient acoustic features used by animals utilizing echoes and whistles in certain tasks can be difficult to identify. One approach to identifying the acoustic features in echoes that dolphins may use to discriminate among objects is to present a dolphin with an echoic discrimination task, then measure the object echoes and statistically analyze acoustic differences in the objects. These between-object differences can be examined in conjunction with the dolphin's errors to identify the acoustic features the dolphin may have used in the task (DeLong *et al*., 2006). Another approach is record echoes from objects used in a discrimination task with a dolphin subject and use artificial neural networks to identify echoic cues that enable objects to be recognized (DeLong *et al*., 2014).

A third method for detecting salient acoustic features is to compare the performance of cetaceans and humans on the same auditory perception tasks. In a typical study, a dolphin and a group of human participants are asked to discriminate among the same set of objects. The human listeners are presented with slowed-down echoes produced with simulated cetacean clicks, or whistle-like stimuli. An advantage to using human listeners is that they can verbally report discriminatory cues. Echolocating dolphins may use complex nonlinear combinations of acoustic features, but it may be difficult to identify them using statistical analyses of the dolphins' performance (DeLong *et al*., 2006). It may be easier to identify the use of complex combinations of features in the reports of human listeners.

There have been reports for over fifty years of both blind and sighted humans using echolocation to navigate and judge the distance, shape, or size of distal objects (e.g., Kellogg, 1962; Rice, 1967; Thaler, 2015). In those studies, the humans investigated objects by projecting self-generated broadband signals (e.g., tongue clicks, snaps, or hisses) and listening to the returning echoes. When using self-generated signals, dolphins are better able to discriminate smaller differences among objects than humans. However, when presented with echoes that were generated using broadband dolphin signals instead of their own signals, humans were capable of discriminating among objects with approximately the same accuracy as dolphins (Au and Martin, 1989; Fish *et al*., 1976; Helweg *et al*., 1995). The chief goal of these studies was to begin gaining insight into the performance of cetaceans by using humans as models. These early studies showed that humans were capable of discriminating between broadband echoes and reporting relevant echo features, but the researchers did not always attempt to determine whether the echo features or strategies used by the humans were the ones likely to have been used by the dolphins.

I have performed seven human listening studies over the past fifteen years with the goal of identifying acoustic features, processing mechanisms, and strategies used by humans that may also be used by cetaceans. Where possible, I have closely analyzed the errors made by both humans and dolphins to discover whether both species are using similar acoustic features. I have also examined how humans and dolphins identify objects from various orientations, and under what conditions learning is required for successful object recognition. I have investigated object recognition using both artificial objects popular in tests with captive animals and naturalistic objects likely to be encountered by dolphins (e.g., fish). I have tested recognition using echoes and whistles. This paper reviews both published and unpublished work I have conducted using the human listening study methodology.

2. GENERAL METHODOLOGY

The methodology described below pertains to studies I have conducted. Other studies on echolocation in humans have utilized different signals (e.g., mouth clicks) or procedures (Thaler, 2015).

A. PARTICIPANTS

In any given experiment, 14 – 65 female and male participants with normal vision and hearing were tested ranging in age from 18 – 55 years (M = 23 years). Participants were given a hearing test that spanned at least 250 – 8,000 Hz at a professional audiology clinic or using a commercially available hearing test (Digital Recordings or Home Audiometer). Only those participants that showed no signs of hearing loss and reported no hearing difficulties completed the experiments. Participants were often asked to fill out a questionnaire about musical experience and ability level, and personal and family history of hearing difficulties. Participants typically received $10 or course credit in a psychology class for completing a study. All participants provided written informed consent before beginning a study.

B. MATERIALS

i. Objects

The objects used in these studies fell into two categories: 3D artificial objects (human-made), and naturalistic objects (live fish). Examples of artificial objects include hollow aluminum cylinders, solid spheres (e.g., steel, aluminum, brass, glass, nylon), ceramic objects (in a cross, line, square, or stacked formation), foam cones in different formations, and figure 8's made of different materials (copper, rope, plastic tubing). The fish prey included Atlantic cod (*Gadus morhua*), grey mullet (*Chelon labrosus*), pollock (*Pollachius pollachius*), and sea bass (*Dicentrarchus labrax)*. The live fish were anesthetized and secured in a monofilament net when the echoes were recorded (see Au *et al*., 2007 for a description of recording fish echoes).

ii. Acoustic stimuli

Echoes from the objects were usually recorded using a typical bottlenose dolphin click (70 µs with a peak frequency of 120 kHz and a 60-Hz bandwidth; see Au, 1993). Echo recordings were made from multiple aspect angles of each object (e.g., every 1.3º) so that there is a sequence of echoes ("echo train") for each object. Detailed descriptions of typical echo recording methods and procedures can be found in DeLong *et al.* (2006). To shift the echoes into the human hearing range, they were time-stretched by a factor of 50 or more. For example, echoes digitized at 1 MHz with center frequencies around 125 kHz became audible for the human listeners at 2.5 kHz.

Acoustic features of echoes were typically measured in a number of ways. Four acoustic features were measured for each individual echo: (a) target strength (the ratio in decibels of the echo intensity measured 1 m from the target to the intensity of the incident signal at the location of the target), (b) echo length, (c) peak frequency (frequency of the signal at which the spectrum has its maximum value), and (d) center frequency (the frequency that divides the power spectrum into two equal energy parts). Additional acoustic features were measured for echo trains (e.g., 23 echoes collected from –15° to +15°) instead of individual echoes (e.g., target strength bumpiness: the number of slope direction changes in the plot of target strength shown as a function of object orientation).

In one study, instead of echoes we used frequency-modulated tones created to loosely mimic dolphin whistles (Branstetter *et al.,* 2016). The three FM sounds were identical (duration, bandwidth, center frequency) except for the frequency contour. The center frequency of the sounds presented to the dolphin was 10 kHz. They were down-sampled (44.1 kHz) and frequency-transposed to 1000 Hz for human listeners.

C. PROCEDURE

Participants were tested individually by a single experimenter in an anechoic chamber or quiet room. Both the participant and experimenter listened to the stimuli via on-ear headphones. The procedure typically included a hearing test, instructions, training sessions, test sessions, and one or more interviews in which participants were asked to describe the auditory features of the sounds they used to discriminate among objects. After passing the hearing test, participants were typically given a vocabulary sheet with definitions that they could use to describe differences among the sounds (e.g., loudness, pitch, timbre, duration) in later interviews. They also listened to examples of sounds differing in loudness, pitch, or timbre to ensure they understood the terms.

i. Training

Training could be completed in two ways. In some studies, participants listened to the echo train stimulus for each object as many times as they liked before beginning the test (typically 3-6 times). The experimenter would say "here are the echoes for object A (play echoes), here are the echoes for object B (play echoes), etc." In other studies, training consisted of multiple sessions containing 10-50 trials per session (total of 100-200 trials). In each training trial, the experimenter played a stimulus, the participant gave a verbal response (name of the object), and then the experimenter provided feedback (correct, or incorrect and name of correct object).

ii. Testing

Test sessions were completed immediately after training. There were one or more test sessions containing a total of approximately 20 – 200 trials, depending on the study. Test trials consisted of the presentation of a test stimulus, the verbal response of the participant, then feedback from the experimenter (correct, or incorrect and name of correct object). In some studies, there were test sessions in which no feedback was given after each trial to prevent learning from occurring during the session (prior to the test sessions containing feedback).

iii. Interview

An interview was conducted after each test session or once at the end of the experiment. Participants were asked to refer to the vocabulary sheet provided to them to describe differences in the sounds for each of the objects. They were often asked to identify the most important sound feature (or two most important features) they used to discriminate among the objects. They were sometimes asked to identify the pair of objects with the most similar sounds and to characterize how their strategies for identifying objects using auditory features shifted throughout the sessions.

3. DISCRIMINATION OF OBJECTS VARYING IN SIZE, SHAPE, MATERIAL, AND/OR STRUCTURE

A. OBJECTS VARYING IN MULTIPLE DIMENSIONS

In this study, an adult male bottlenose dolphin and fourteen human listeners were asked to discriminate among objects in six three-object sets that varied in size, shape, material, and surface texture (DeLong *et al*., 2007A). The objects can be seen in Figure 1 of DeLong *et al.* (2007A). The dolphin and the human listeners were successful at discriminating among the objects within the sets (M_{humans} = 88%). The dolphin's errors (confusions between objects) were evenly distributed in three of the sets, but in the other three sets the dolphin confused one of the three pairs of objects more frequently than the other two pairs (we called this the "predominant error type"). The humans had a predominant error type for all six sets. Comparing the human and dolphin predominant error types alongside the reported cues revealed that the dolphin did not seem to rely on overall amplitude differences to discriminate among these objects, but may have used pitch, timbre, and the pattern of changes in these features as the object was ensonified from various orientations. There was one object set in which the predominant error made by the humans and dolphin strongly matched: the foam cones – which varied primarily in shape. In this set, all the human listeners reported using the pattern of change in the echo train in loudness (or pitch) as the object rotated. This was a very salient and effective cue, as evidenced by their high choice accuracy. It is very likely that dolphins use this dynamic cue when scanning objects across a range of aspect angles.

To test how critical it was to have access to echoic information from a range of aspect angles, we ran another experiment in which there were two echo conditions: multiple aspects (echoes from multiple object orientations) and single aspect (echoes from one object orientation). A different group of human listeners were asked to perform the same matching task for four object sets from the previous experiment in both echo conditions. If multiple object orientations are necessary for discriminating among the objects, then the participants should perform significantly better in the multiple aspect condition than in the single aspect echo condition. If multiple object orientations are not necessary, then there should be no difference in performance between the two echo conditions. We found that having information from multiple object orientations was advantageous for the object set that varied primarily in shape (foam cones) but not necessary for the sets that varied primarily in material (figure 8s), shape and size (stone shapes), and size (stone squares). The participants did not need information from multiple aspects to discriminate between different-sized objects (they used overall loudness differences) or to discrimination between objects made of different material (they used overall pitch and/or timbre). Shape information in aspect-dependent shapes that did not vary in other ways (same material, same surface area) was conveyed by listening to how the echoes changed as the objects rotated. Learning this led to a series of studies examining how human listeners recognize objects using echoes from novel aspect angles (see section 4 below).

B. CYLINDERS VARYING IN WALL THICKNESS, SPHERES VARYING IN MATERIAL

Human listeners in this study were presented with echoes from aspect-independent cylinders and spheres that had been used in previous studies with dolphin subjects (DeLong *et al*., 2007B). The hollow cylinders had varying wall thicknesses: 6.35 mm ±0.2, 0.3, 0.4, and 0.8 mm, and solid spheres 7.62 cm in diameter made of different materials (steel, aluminum, brass, nylon, and glass). The cylinders and spheres were tested in separate studies for the dolphin and in separate trials for the humans (i.e., participants had to compare the cylinders only to other cylinders and compare the spheres only to other spheres). Both the humans and the dolphin had to identify the "standard" object (e.g., cylinder with 6.35 mm wall thickness or steel sphere). In addition, the humans were asked to identify which of the comparison objects was presented on any given trial after identifying which echoes were from the standard object.

The human listeners performed as well or better than the dolphins at the task of discriminating between the standard object and the comparison objects on both the cylinders (humans = 97.1%; dolphin = 82.3%) and the sphere (humans = 86.6%; dolphin = 88.7%). The human listeners actually performed better than the dolphin when comparing the standard cylinder to the comparison cylinders with ±0.2 or ±0.3 mm wall thickness. However, the humans listened to a noise-free set number of echoes from both the standard and a comparison cylinder on each trial, whereas the dolphin typically chose to emit a variable number of clicks towards only one cylinder on each trial (making it a memory test for the dolphin to compare the current echoes to a memory of the standard echoes). When both the dolphin and humans relied on memory (i.e., when the humans were asked to identify the specific comparison cylinder presented on a given trial), both species showed graded performance in which better performance was shown on the cylinders most different than the standard (0.4, and 0.8 mm). The human listeners reported using primarily pitch and duration to discriminate among the cylinders, and pitch and timbre to discriminate among the spheres. When listening to cylinder echoes, the main discriminatory cue was pitch, and the pitch they heard could be the time-separation pitch (TSP) created by the two main highlights in the cylinder echoes, since this type of signal would produce TSP in the human auditory system (Thurlow, 1957). The dolphin may also have used TSP cues. However, there is no direct evidence that a dolphin can perceive TSP. This study suggests that further research is warranted to determine whether dolphins perceive TSP cues.

4. OBJECT RECOGNITION ACROSS MULTIPLE ASPECT ANGLES

The ability to recognize an object from any viewpoint is called object constancy, and it is one of the fundamental and essential properties of visual perception (for a review see Jolicoeur and Humphrey, 1998; Graf, 2006). Past studies on visual object constancy have led to competing theories of how people represent objects. One theory is that recognition performance is view-invariant, and a single underlying representation is constructed of the object (e.g., Biederman and Gerhardstein, 1993; Marr, 1982). In this view, object constancy is achieved because the same object representation is activated for any orientation of the object. Another theory is that recognition performance is view-dependent, and different object representations are formed with each different view (e.g., Tarr and Pinker, 1989). Differing experimental conditions and changing the properties of the objects being compared can cause performance to shift from view-dependent to view-invariant, or vice versa.

Most studies of object constancy have addressed the ability of humans and other animals to visually recognize objects from different orientations (Biederman and Bar, 1999; Kirkpatrick, 2001; Logothetis *et al*., 1995; Shepard and Metzler, 1972; Zoccolan *et al.,* 2009). Object constancy is rarely investigated in humans and other animals in modalities other than vision (bats: DeLong *et al.,* 2008). I have done a group of studies in which I have investigated object constancy in the auditory modality using dolphin echoes. Dolphins use echolocation to recognize objects, and they must do so underwater where they ensonify the object from many viewpoints (e.g., they can swim above, beneath, or sideways relative to fish). Most objects dolphins encounter underwater are aspect-dependent, and the echoes from these objects vary considerably depending on aspect angle (DeLong *et al*., 2006). For example, Figure 1 shows that echoes from different aspect angles of a fish vary across many acoustic features (Au *et al*., 2009). Thus, dolphins need to correctly identify objects despite changes in the specific echo acoustic features that result from variations in viewpoint.

A. OBJECTS VARYING IN MATERIAL

In this study, we tested the ability of human listeners to identify objects varying along several dimensions (material, shape, size) from different aspect angles (DeLong *et al*., 2014). Participants listened to pre-recorded echoes from a copper figure 8, a ceramic cross, and a wood rectangle (Figure 2A). Each echo recording was an echo train consisting of 45 echoes ranging from -30º to +30º. The echo trains were divided into five aspects. Aspect 1 contained echoes from approximately -19º to -30º, aspect 2 contained echoes from approximately -7º to -18º , aspect 3 contained echoes from approximately -6º to +6º (including 0º , the front face of the object), aspect 4 contained echoes from approximately +7º to +18º , and aspect 5 contained echoes from approximately +19º to +30º. In Experiment 1, twenty-six participants were allowed to listen only to echoes from aspect 3 during the three training sessions (108 total trials). Performance accuracy was 80% or better on the last 27 training trials. Participants then completed two test sessions (72 trials) in which they had to discriminate among the three objects using echoes from all five aspects (four novel transfer aspects and one training aspect). The participants' performance was significantly above chance for all objects in all five aspects (M = 77%). There was no significant difference in performance between the training and transfer aspects (except performance on aspect 3 and 4 of copper figure 8 was significantly higher than on aspect 5).

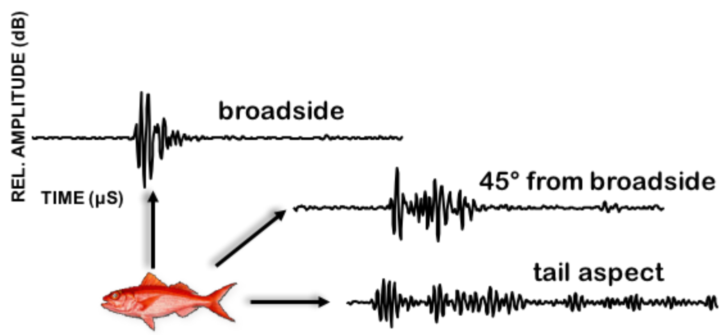

Figure 1. The acoustic features of echoes from different aspect angles of a fish. Echoes typically have the highest amplitude and simplest structure at the broadside angle (perpendicular to the longitudinal axis of the fish), and increase in complexity as you move away from broadside.

In Experiment 2, thirty-five participants were randomly assigned to one of five groups, where each group listened to echoes from a different aspect during training (e.g., group 1 only heard echoes from aspect 1, group 2 only heard echoes form aspect 2, etc.). Again, performance accuracy was 80% of better on the last 27 training trials. When averaged across the three objects, performance was significantly above chance for all groups in all aspects (M = 79%). There were some differences between groups. Groups 2 and 3 showed no differences in performance between training and transfer objects. Group 4 had better performance on the training aspect vs. transfer aspect 5 (but only for copper figure 8). Groups 1 and 5 had better performance on the training aspect vs. transfer aspect 3 (but only for the ceramic cross). The overall pattern of performance suggests that the participants had mostly view-invariant object recognition in this study– they were able to identify the objects from aspects other than the ones they had been trained to recognize. Listeners reported a wide variety of cues they used to achieve success; including overall loudness, pitch, timbre, and duration as well as the pattern of change in features produced by the succession of echoes in the train. Participants appeared to use a wider variety of cues during the test sessions that they did in the training sessions.

B. OBJECTS MADE OF THE SAME MATERIAL

The goal of this study was to compare and contrast the performance of human listeners presented with objects varying in material (see section 4A above) vs. objects made of the same material. The similarity of objects within a visual task is known to change the performance of humans and non-human animals. When objects within a set are more similar, it can make it more difficult to recognize them from novel aspect angles. In this study, twenty participants had to discriminate among three objects made of the same material: ceramic cross, line, and stack (see Figure 2B). Like the previous study (DeLong *et al*., 2014), each object's echo recording was an echo train consisting of 45 echoes and the echo trains were divided into five aspects. The participants were randomly assigned to one of five groups, where each group listened to echoes from a different aspect during training (e.g., group 1 only heard echoes from aspect 1, group 2 only heard echoes form aspect 2, etc.). We found that 108 training trials was insufficient to achieve the same criterion as in DeLong *et al*., 2014, so we increased it to 252 training trials. Performance was consistently above chance (33%) on the last 90 trials and averaged 65% or better on the last 27 trials.

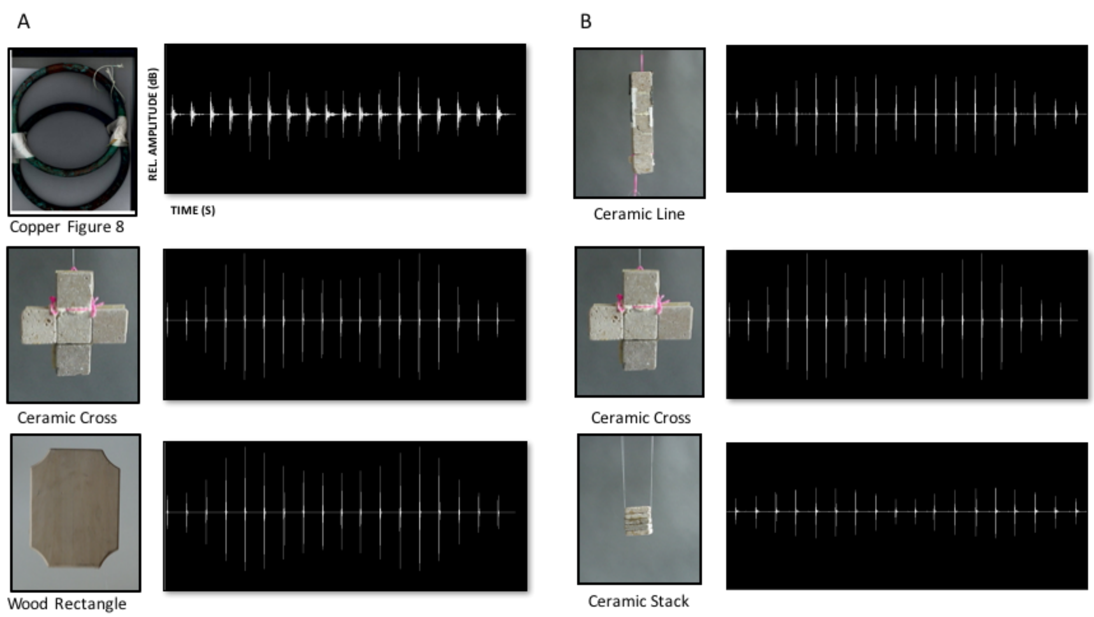

Figure 2. Objects and echoes used in two auditory object constancy studies with artificial objects. The displayed echoes are from aspect 3 only (contains broadside echoes). The 18-echo trains include two sweeps of nine echoes to mimic a dolphin sweeping its head back and forth. (A) Three objects varying in material, size and shape used in Study 1. (B) Three objects made of the same ceramic material, but varying in size and shape that were used in Study 2.

Test sessions contained echoes from the one training aspect and the four novel transfer aspects. We had two types of test sessions: a no feedback session and a feedback session. The no feedback session had 30 trials (no feedback was provided after each trial). Participants were not able to discriminate among the objects echoes at novel aspects without feedback. Their performance was not significantly different from chance for all objects in all aspects except the stack at aspect 1 (M = 37%). There were 60 trials in the feedback test session (they were told whether they were correct/incorrect and identity of correct answer after each trial). Performance improved when they were given feedback (M = 47%). However, they only performed significantly better than chance on three of the five aspects for ceramic cross and ceramic stack and only two aspects for ceramic line. Participants' performance in this study would be classified as view-dependent, not view-invariant. Participants report that the echoes from three objects made from the same material sound very similar (in pitch and timbre), so it is difficult to locate sufficient discriminatory cues to identify them at different aspects. The availability of feedback was necessary for participants to succeed at all – they needed learning opportunities when the objects were made of the same material.

The results of the human listening studies in section 4A and 4B suggest that the characteristics of the objects being discriminated plays a role in whether performance is viewpoint-invariant or viewpoint-dependent. When objects vary in material, both human listeners and dolphins show viewpoint-invariant performance (Au and Turl, 1991; DeLong *et al*., 2014). When objects are the same material, both human listeners and dolphins show viewpoint-dependent performance (Nachtigall *et al.,* 1980; current study). These findings parallel studies on visual object constancy in humans and other non-human animals. Efforts are underway to present the same object echoes shown in Figure 2 to a dolphin so that human and dolphin performance can be compared and discriminatory cues used by dolphins can be identified.

C. NATURAL OBJECTS: FISH ECHOES

Most echolocation studies with captive cetaceans use artificial (human-made) objects selected to vary precisely along certain dimensions (e.g., spheres of the same size that vary only in material) or selected for convenience. These objects are unlikely to ever be encountered by cetaceans in their natural habitat and studies using those objects lack ecological validity. In this study, we used echoes from four fish species that would be encountered by bottlenose dolphins: Atlantic cod (*Gadus morhua*), grey mullet (*Chelon labrosus*), pollock (*Pollachius pollachius*), and sea bass (*Dicentrarchus labrax;* Figure 3A). The live fish were anesthetized, enclosed in a monofilament net bag, and then the net bag was mounted to a monofilament net suspended from a rotor. Echoes were collected from all lateral aspect angles of each fish using dolphin signals (Figure 3B; Au *et al.,* 2007). Echoes were collected from the fish every 2.5º as the fish rotated through 360 degrees. The echo train was divided into 8 aspects, each containing 10 echoes (covering 25 degrees).

In Experiment 1, sixty-four participants were randomly assigned to one of eight groups, where each group listened to echoes from only one aspect during training (group 1 listened to echoes from aspect 1, group 2 listened to echoes from aspect 2, etc.). There were 200 training trials. Participants' achieved 66% correct over the last 32 trials, and were significantly above chance on all but the first 16 trials (chance was 25%). Performance varied as a function of fish species during training – their performance accuracy was typically 10-25% higher with pollock and mullet echoes compared to cod and sea bass echoes.

Test sessions contained echoes from the training aspect and the seven novel transfer aspects. There were two types of test sessions: a no feedback session (32 trials) and a feedback session (64 trials). In the no feedback session, participants were able to successfully identify the pollock and mullet at the novel aspect angles (except mullet aspects 6, 7, and 8, see Table 1). However, they were not able to identify the sea bass or cod at any novel aspect angles (except cod aspect 5). In the feedback test session, performance improved across the board. Participants again successfully identified the pollock across all aspect angles, but now they could also identify the other three fish at most aspect angles (Table 1). This experiment suggested that human listeners could not identify two fish species at novel aspects without learning opportunities (i.e., feedback on novel aspect trials).

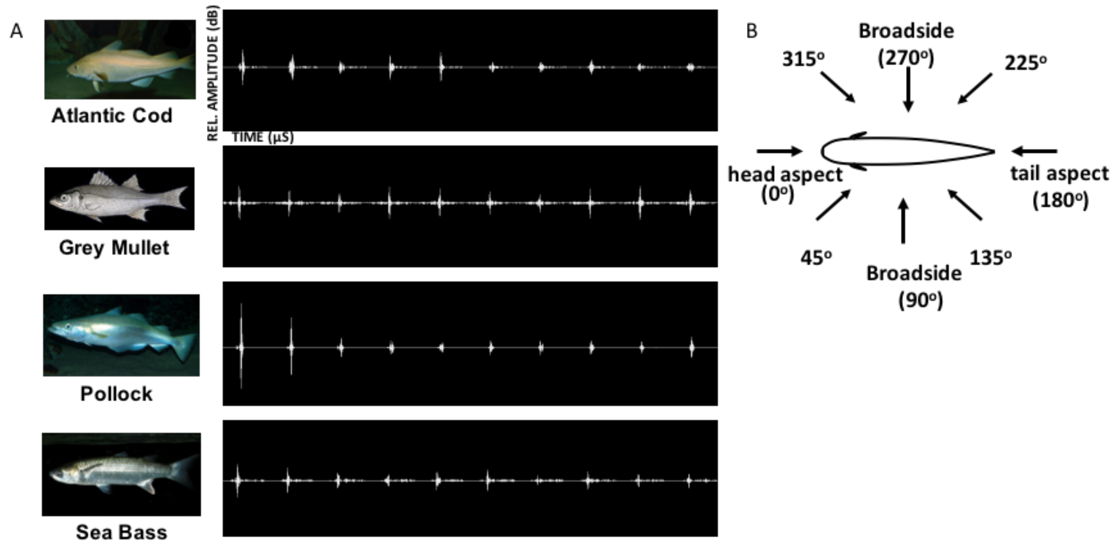

Figure 3. Four fish species and their echoes from the broadside aspect angle (90º). (A) Ten echoes from each of the fish. (B) Diagram showing how echoes were recorded from all aspect angles of the fish.

We conducted Experiment 2 to explore the possibility that an extended training phase would improve performance on the test phase. Twenty-five participants received 280 training trials (as opposed to 200 trials) with echoes from only aspect 2. Their performance was significantly above chance on all except the first 8 trials. Their performance on the last 32 trials was 72%, which exceeded performance in Experiment 1 (M = 66%). In Experiment 2, the no feedback test session and the feedback test session both contained 64 trials. Throughout both test sessions, participants continued to perform above chance on the trained aspect angle (M = 62%). In the no feedback test session, participants successfully identified the mullet and pollock at most novel aspect angles, but failed to identify the cod and sea bass (except for one or two aspects; Table 1). In the feedback test session, participants again successfully identified the mullet and pollock at most or all novel aspect angles, failed to identify the cod, and were moderately successful with the sea bass.

Table 1. Test results – accuracy for each test session averaged across aspects (percent correct).

	Experiment 1		**Experiment 2**	
Fish species	**No Feedback**	**Feedback**	**No Feedback**	**Feedback**
Atlantic cod	31	39*	31	32
Grey mullet	42*	51*	49*	52*
Pollock	48**	58**	56*	64**
Sea bass	29	37*	39	$40^{\#}$

Note. Participants received the no feedback test session first. ** denotes performance that was significantly above chance (25%) on all 8 aspects. *denotes performance that was above chance on all but 1, 2, or 3 of the 8 aspects. #denotes performance that is not different from chance on 4 aspects. No asterisk denotes performance that is not different from chance on 6, 7, or 8 aspects.

In this study, human listeners did not show immediate view-invariant auditory object recognition of all fish species. They were able to identify the mullet and pollock at novel aspect angles without any feedback, but learning was required to generalize the discrimination to novel aspect angles for cod and sea bass. Even with learning opportunities, recognizing these fish using echoes from different viewpoints was challenging. Participants reported using multiple cues to discriminate among the fish – loudness, pitch, timbre, duration, and the pattern of changes in these cues across the echo train. The mullet and pollock were more distinctive (e.g., in loudness and timbre) compared to cod and sea bass. Human listeners confused the cod and sea bass based on similarity in echoic auditory features. In addition, they often needed to utilize different cues in the test sessions (with the novel aspect angles) than they used in the training session. Identifying the fish from a new aspect angle was not a trivial problem, and even their best performance (about 60% correct) was far from perfect.

Although mature dolphins have much more experience identifying objects using echoes than the human listeners in this study, it is still likely that dolphins require learning experiences to identify aspect-dependent objects using echoes from various orientations. For example, they may ensonify an object from various orientations whilst gathering information via other sensory systems (e.g., vision, touch, taste) to verify the identity of the object. Dolphins also have a variety of other cues to assist them other than the fish echoes in identifying fish species while foraging– the swimming pattern of the fish, water depth, nearby structures, and other animals present in the scene. It has been suggested that there are sufficient acoustic cues available in echoes for dolphins to discriminate between the four species of fish in this study, regardless of aspect angle (Au *et al*., 2009). However, aspect-dependent discrimination of fish echoes has not yet been demonstrated with dolphin subjects in a controlled psychophysical study. Further studies on auditory object constancy are needed to shed light on the mechanisms underlying performance in cetaceans and humans.

5. CONCLUSIONS

In these types of studies, humans listening to prerecorded echoes typically perform as well or better than dolphin subjects and report discriminatory cues. If both dolphins and humans perform the same task using the same objects, analyzing the error patterns of dolphins and humans alongside the reported cues can reveal echo auditory features that are likely to be salient for dolphins. Human listening studies can also reveal processing mechanisms and decision strategies that may be used by dolphins, and indicate whether task performance is likely to be improved with learning opportunities.

Of course, similarities and differences between human and dolphin auditory perception should be considered when evaluating the results of human listening studies. Both dolphins and humans can discriminate sounds that vary in intensity by about 1 dB (Evans, 1973; Green, 1993) and the frequency discrimination abilities of dolphins and humans for tonal stimuli are comparable in the range of best hearing for each species (Herman and Arbeit, 1972; Thompson and Herman, 1975; Weir *et al.,* 1977). However, dolphins have sharper frequency tuning curves than humans (Supin and Popov, 1995) so they may be better able to interpret frequency information in echo stimuli than the human listeners. Dolphins have better temporal resolution ability than humans, which means the dolphins' hearing is sensitive to extremely short gap durations as compared to other animals and humans (Popov and Supin, 1997). However, it is important to note that when the echoes are slowed down to bring them into the human hearing range, the duration of the echoes is lengthened relative to the echoes the dolphin hears.

Human listening studies are one way of investigating how cetaceans use information in echoes or whistles to classify objects or agents. Although caution must be used in interpreting the results, since there are some ways in which dolphin and human hearing are different, it is still a useful method because of the considerable challenges involved with dolphin research – extensive training time, few research animals, and expensive animal care and maintenance. Human listeners can be tested quickly and identify salient echo features that permit object discrimination. The knowledge we gain can be used to generate hypotheses that can be tested using dolphins as subjects. The most productive approach to studying auditory perception in dolphins involves combining multiple methods – directly testing dolphins, comparative studies with human listeners, and modeling performance with artificial neural networks and other tools.

ACKNOWLEDGMENTS

I would like to thank the colleagues and coauthors who worked with me on these studies, especially Whitlow W.L. Au, Brian K. Branstetter, Brandon Dziedzic, Heidi E. Harley, Amanda Heberle, Irene Kannyo, Eduardo Mercado III, and Matthew Wisniewski. This work was supported by many grants and awards, including research grants from the Rochester Institute of Technology, the American Psychological Association, the American Association of University Women, Soroptimist International, and SEASPACE (please see DeLong *et al*., 2007A; DeLong *et al.,* 2007B; DeLong *et al*., 2013; DeLong *et al*., 2014; and Branstetter *et al*., 2016 for a complete listing of funding support).

This paper was presented in an Animal Bioacoustics Session in Honor of Whitlow Au. Whit was my mentor at the Marine Mammal Research Program in the Hawaii Institute of Marine Biology from 1997-2003. He served on my dissertation committee at the University of Hawaii. After Whit introduced me to his human listening study (Au and Martin, 1989), I decided to use that methodology in one study of my dissertation research. Whit was an important factor in my success in graduate school, and continued to mentor me as a postdoc and a young professor after I left his lab. Whit has always been both professionally and personally supportive of me during my career, a rare and valuable trait in a graduate mentor. I will always be grateful to him for his contributions to my own career and to the field of animal bioacoustics.

REFERENCES

Au, W. W. L. (**1993**). *The sonar of dolphins*. (Springer, New York).

Au, W.W.L., Benoit-Bird, K.J., and Kastelein, R.A. (**2007**). "Modeling the detection range of fish by echolocating bottlenose dolphins and harbor porpoises," J. Acoust. Soc. Am. **121**, 3954-3962.

Au, W.W.L., Branstetter, B.K., Benoit-Bird, K.J., and Kastelein, R.A. (**2009**). "Acoustic basis for fish prey discrimination by echolocating dolphins and porpoises," J. Acoust. Soc. Am. **126**, 460-467.

Au, W. W. L., and Martin, D. W. (**1989**). "Insights into dolphin sonar discrimination capabilities from human listening experiments," J. Acoust. Soc. Am. **86**, 1662–1670.

Au, W. W. L., and Turl, C. W. (**1991**). "Material composition discrimination of cylinders at different aspect angles by an echolocating dolphin," J. Acoust. Soc. Am. **89**, 2448-2451.

Biederman, I., and Bar, M. (**1999**). "One-show viewpoint invariance in matching novel objects," Vis. Res. 3**9**, 2885–2899.

Biederman, I., and Gerhardstein, P.C. (**1993**). "Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance," J. Exp. Psychol. Hum. Percept. Perform. **19**, 1162–1182.

Branstetter, B., DeLong, C.M., Dziedzic, B., Black, A., and Bakhtiari, K. (**2016**). "Recognition of frequency-modulated whistle-like sounds by a bottlenose dolphin (*Tursiops truncatus*) and humans with transformations in amplitude, duration, and frequency," *PLoS ONE* **11**(2): e0147512. doi: 10.1371/journal.pone.0147512

DeLong, C.M., Au, W.W.L., Harley, H.E., Roitblat, H.L., and Pytka, L. (**2007A**). "Human listeners provide insights into echo features used by dolphins to discriminate among objects," J. Comp. Psych. **121**, 306-319.

DeLong, C.M., Au, W.W.L., Lemonds, D.W., Harley, H.E., and Roitblat, H.L. (**2006**). "Acoustic features of objects matched by an echolocating bottlenose dolphin," J. Acoust. Soc. Am. **119**, 1867-1879.

DeLong, C.M., Au, W.W.L, and Stamper, S.A. (**2007B**). "Echo features used by human listeners to discriminate among objects that vary in material or structure: Implications for echolocating dolphins," J. Acoust. Soc. Am. **121**, 605-617.

DeLong, C.M., Bragg, R., and Simmons, J.A. (**2008**). "Evidence for spatial representation of object shape by echolocating bats (*Eptesicus fuscus*)," J. Acoust. Soc. Am. **123**, 4582–4598.

DeLong, C.M., Heberle, A.L., Mata, K., Harley, H.E., and Au, W.W.L. (**2013**). "Recognizing objects from multiple orientations using dolphin echoes," POMA, **19** (010034), 1-9. doi: 10.1121/1.4799409

DeLong, C.M., Heberle, A.L., Wisniewski, M.G., and Mercado, E., III. (**2014**). "The ability to recognize objects from dolphin echoes generalizes across multiple orientations in humans and neural networks," Anim. Cogn. **17**, 543-557. doi: 10.1007/s10071-013-0685-0

Evans, W. E. (**1973**). "Echolocation by marine delphinids and one species of fresh-water dolphin," J. Acoust. Soc. Am. **54**, 191–199.

Fish, J. F., Johnson, C. S., and Ljungblad, D. K. (**1976**). "Sonar target discrimination by instrumented human divers," J. Acoust. Soc. Am. **59**, 602–606.

Graf, M. (**2006**). "Coordinate transformations in object recognition," Psych. Bull. **132**, 920–945.

Green, D. M. (**1993**). "Auditory intensity discrimination," in *Human psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 13–55.

Helweg, D. A., Roitblat, H. L., Nachtigall, P. E., Au, W. W. L., and Irwin, R. J. (**1995**). "Discrimination of echoes from aspect-dependent targets by a bottlenose dolphin and human listeners," in *Sensory systems of aquatic mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil Publishers, Woerden, The Netherlands), pp. 129–136.

Herman, L. M., and Arbeit, W. R. (**1972**). "Frequency discrimination limens in the bottlenose dolphin: 1–70 Ks/c," J. Aud Res. **2**, 109–120.

Jolicoeur, P., and Humphrey, G.K. (**1998**). "Perception of rotated two-dimensional and three-dimensional objects and visual shapes," in *Visual constancies: why things look as they do,* edited by V. Walsh and J. Kulikowski (Cambridge University Press, Cambridge), pp 69–123.

Kellogg, W. N. (**1962**). "Sonar system of the blind," Science **137**, 399–404.

Kirkpatrick, K. (**2001**). "Object recognition," In *Avian visual cognition* [Online], edited by R.G. Cook, Retrieved from www.pigeon.psy.tufts.edu/avc/kirkpatrick/

Logothetis, N.K., Pauls, J., Poggio, T. (**1995**). "Shape representation in the inferior temporal cortex of monkeys," Curr. Biol. **5**, 552–563.

Marr, D. (**1982**). *Vision*. (Freeman, San Francisco).

Nachtigall, P.E., Murchison, A.E., and Au, W.W.L. (**1980**). "Cylinder and cube discrimination by an echolocating blindfolded bottlenose dolphin," in *Animal sonar systems*, edited by R.G. Busnel and J.F. Fish (Plenum Press, New York), pp. 945-947.

Popov, V. V., and Supin, A. Ya. (**1997**). "Detection of temporal gaps in noise in dolphins: Evoked potential study," J. Acoust. Soc. Am. **102**(2), 1169– 1176.

Rice, C. E. (**1967**). "Human echo perception," Science **155**, 656–664.

Shepard, R.N., and Metzler, J. (**1972**). "Mental rotation of three-dimensional objects," Science **171**, 701–703.

Supin, A. Ya., and Popov, V. V. (**1995**). "Frequency tuning and temporal resolution in dolphins," in *Sensory systems of aquatic mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil Publishers, Woerden, The Netherlands), pp. 95–110.

Tarr, M.J., and Pinker, S. (**1989**). "Mental rotation and orientation-dependence in shape recognition," Cog. Psychol. **21**, 233–282.

Thaler, L. (2015). "Using sound to get around: Discoveries in human echolocation," Observer, **28**(10), 24-27.

Thompson, R. K. R., and Herman, L. M. (**1975**). "Underwater frequency discrimination in the bottlenose dolphin (1–140 kHz) and the human (1–8 kHz)," J. Acoust. Soc. Am. **57**, 943–948.

Thurlow, W. R. (**1957**). "Further observations on pitch associated with a time difference between two pulse trains," J. Acoust. Soc. Am. **29**, 1310–1311.

Wier, C., Jesteadt, W., and Green, D. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.

Zoccolan, D., Oertelt, N., DiCarlo, J.J., and Cox, D.D. (**2009**). "A rodent model for the study of invariant visual object recognition," PNAS **106**, 8748–8753.