Anim Cogn
DOI 10.1007/s10071-013-0685-0

ORIGINAL PAPER

The ability to recognize objects from bottlenose dolphin (*Tursiops truncatus*) echoes generalizes across multiple orientations in humans and neural networks

Caroline M. DeLong · Amanda L. Heberle · Matthew G. Wisniewski · Eduardo Mercado III

Received: 19 June 2013 / Revised: 5 September 2013 / Accepted: 16 September 2013

Abstract Object constancy, the ability to recognize objects despite changes in orientation, has not been well studied in the auditory modality. Dolphins use echolocation for object recognition, and objects ensonified by dolphins produce echoes that can vary significantly as a function of orientation. In this experiment, human listeners had to classify echoes from objects varying in material, shape, and size that were ensonified with dolphin signals. Participants were trained to discriminate among the objects using an 18-echo stimulus from a 10° range of aspect angles, then tested with novel aspect angles across a 60° range. Participants were typically successful recognizing the objects at all angles ($M = 78\ \%$). Artificial neural networks were trained and tested with the same stimuli with the purpose of identifying acoustic cues that enable object recognition. A multilayer perceptron performed similarly to the humans and revealed that recognition was enabled by both the amplitude and frequency of echoes, as well as the temporal dynamics of these features over the course of echo trains. These results provide insight into representational processes underlying echoic recognition in dolphins and suggest that object constancy perceived through the auditory modality is likely to parallel what has been found in the visual domain in studies with both humans and animals.

C. M. DeLong (✉) · A. L. Heberle
Department of Psychology, College of Liberal Arts, Rochester Institute of Technology, 18 Lomb Memorial Drive, Rochester, NY 14623, USA
e-mail: cmdgsh@rit.edu

M. G. Wisniewski · E. Mercado III
Department of Psychology, College of Arts and Sciences, University at Buffalo, The State University of New York, Park Hall Room 204, Buffalo, NY 14260, USA

Keywords Object constancy · Bottlenose dolphin · Human listening · Echolocation · Object discrimination · Artificial neural networks

Introduction

Every day, organisms quickly and accurately recognize familiar objects despite changes in object orientation, size, or distance. This 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). Although humans typically recognize objects visually without apparent effort, no artificial computational system is yet able to successfully recognize objects over a wide range of orientations and contexts (Graf 2006). The inability of computational vision researchers to match natural object recognition abilities suggests that the representational and transformational processes underlying object constancy remain unclear (DiCarlo and Cox 2007).

Past research on visual object constancy has focused on recognition of objects that vary in shape. These studies 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). In this framework, object constancy is achieved by a transformation process that involves multiple stored representations. Object features

Published online: 27 September 2013

and experimental conditions can cause performance to shift from view-dependent to view-invariant, or vice versa.

Although many studies 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. 1994, 1995; Shepard and Metzler 1972; Zoccolan et al. 2009), very few have investigated object constancy in humans and other animals in other modalities (DeLong et al. 2008). The current study examined whether auditory representations of echoes coming from objects show object constancy effects comparable to those seen in visual object recognition.

Studies of object constancy can benefit from examining multiple species that solve the same object recognition problem, but in different ways. For example, whereas humans rely heavily on vision to identify objects, dolphins often use echolocation (i.e., biological sonar). Dolphins echolocate by emitting a series of ultrasonic clicks and listening to the returning echoes (Au 1993). Dolphins use echolocation to navigate, avoid predators, and track moving prey. Most of the objects dolphins encounter are aspect-dependent, meaning that the size and shape of the surfaces of the object will change as they are viewed from different orientations. The echoes from these types of objects can vary considerably depending on the angle from which they are inspected by the dolphin (Helweg et al. 1996a, b). In fact, echoes from different orientations of a single object can vary more from each other than do echoes from different objects (DeLong et al. 2006). Thus, the problem echolocating dolphins face in recognizing objects is very similar to the one humans face visually—the object must be correctly identified despite large changes in the specific sensory information (the visual image or the auditory event) that result from changes in the orientation and location of the object.

Human listening studies can inform our understanding of how both humans and dolphins extract information about object identity from echoes (Au and Martin 1989; DeLong et al. 2007a, b; Gorman and Sawatari 1985; Gorman and Sejnowski 1988; Helweg et al. 1995). This general approach involves recording echoes from objects using a dolphin's echolocation signal, slowing down the echoes in time to lower the frequency into the human hearing range, and then presenting those echoes to human participants in an object discrimination task. In the task presented to humans, the echo stimulus is played via headphones, after which the participant must choose the correct object from among multiple objects. Human participants have the advantage that they can verbally report auditory features of the echoes they used to discriminate objects. Whether salient auditory features reported by humans overlap with those used by dolphins can be assessed, in part, by comparing the errors made by humans and dolphins on the object discrimination task. Matching error patterns would suggest use of similar auditory features, whereas mismatching error patterns would imply the use of different features.

This general approach is a reasonable way to investigate auditory processing in both species because auditory perception in humans and dolphins appears to be similar in certain ways. For example, both humans and dolphins can discriminate between sounds that differ in intensity by 1 dB (Green 1993; Au 1993), and the frequency discrimination abilities of humans and dolphins 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. 1976).

Several prior studies have shown the usefulness of comparing human and dolphin performance on object discrimination tasks. In one study, human listeners were presented with echoes from different angles of aspect-dependent objects that had been used in a dolphin discrimination task (DeLong et al. 2007a). Aspect-dependent objects appear different based on their orientation, and echoes from different angles differ substantially. The objects varied in size, shape, material, and/or texture. In two experiments, the human listeners performed as well or better than the dolphin at discriminating objects and reported the salient acoustic cues. The error patterns (object confusions) of the humans and the dolphin were compared to determine which acoustic features reported by the humans were likely to have been used by the dolphin. The results indicated that the dolphin did not appear to use overall echo amplitude but that it attended to the pattern of changes in the echoes across different object orientations. Having information in multiple echoes gleaned from multiple object orientations appeared to be particularly advantageous when discriminating among objects that varied in shape. These results were consistent with another study in which the human listeners had to identify three aspect-dependent objects (rectangle, pyramid, and cube) that had been used in a dolphin discrimination task (Helweg et al. 1995). The human listeners used the pattern of changes in amplitude across successive echoes to identify the differently shaped objects.

In another study, researchers ensonified aspect-independent objects with dolphin-like signals and then recorded the echoes produced by each object (DeLong et al. 2007b). The objects, which had been used in discrimination studies with dolphins, varied in specific ways: hollow aluminum cylinders varied only in wall thickness (a standard had a 6.35 mm wall thickness and eight others varied by ±0.2, ±0.3, ±0.4, ±0.8 mm) and solid 7.62-cm-diameter spheres varied only in material (a standard was stainless steel, and the four others were aluminum, brass, glass, and nylon). Like the dolphins, the human subjects discriminated between echoes from the standard objects and the comparisons within each set, but the humans also identified which comparison object produced the echoes. The human and dolphin subjects performed similarly in most object discriminations. The humans reported using pitch (potentially time separation pitch) and duration to identify the cylinders and using pitch and timbre to identify the spheres.

There are differences between humans and dolphins that could result in differences in the perception of echo stimuli (e.g., sharp frequency tuning and short auditory integration time in dolphins; Supin and Popov 1995), and transforming the echoes to fall within the human hearing range will produce changes in the timing and spectra of the echoes. Thus, human listening studies have to be supplemented with other approaches, such as computational modeling and direct investigation of specific auditory features that could be used by dolphins. Although humans probably do not apprehend objects echoically in the same way that dolphins do, they can perform as well or better than dolphins on object recognition tasks using echo stimuli (DeLong et al. 2007a, b), once those echoes have been transformed to fall within the range of human hearing. Furthermore, the objects that humans tend to confuse when classifying echoes are often the same objects that dolphins performing echolocation tasks confuse, indicating that at times humans and dolphins may use the same echoic features (DeLong et al. 2007a). These findings suggest that despite differences in how humans and dolphins may perceive the echoes and represent the sounds, there may be underlying similarities in how both acquire and generalize auditory information relevant to recognizing objects.

Computational models of object recognition provide a second useful tool for exploring the representational processes underlying object constancy. For example, early attempts to classify sonar targets with artificial neural networks revealed that relatively simple networks with multiple layers of processing discovered features within sonar returns that were similar to the features that human listeners learned to use (Gorman and Sejnowski 1988). Studies of neural networks trained to recognize objects using either single echoes or multiple successive echoes generated from a dolphin's sonar signal showed that networks that integrate information from successive echoes are better able to mimic the recognition performance of dolphins (Moore et al. 1991) and that such networks can recognize objects at above chance levels even when the echoes were generated by objects that varied freely in orientation (Helweg et al. 1996a, b). As with human listening studies, neural networks almost certainly do not replicate the perceptual processes employed by echolocating dolphins. Nevertheless, the ability of such computational models to recognize objects at levels comparable to dolphins and humans indicates that they may be extracting similar information. Major advantages of neural networks are that they can be used to rapidly test various hypotheses about which features of objects are critical for object recognition and generalization and can be used to explore how different networks transform stimulus features in ways that facilitate recognition of novel exemplars (Guillette et al. 2010; Wisniewski et al. 2012) as well as recognition of familiar exemplars presented in novel orientations (Gorman and Sejnowski 1988).

Examining object constancy in modalities other than vision may yield a deeper understanding of the cognitive processes that support this fundamental property of perception. The purpose of the current study was to explore how changes in viewpoint affect participants' ability to identify an object using auditory instead of visual stimuli. Only two studies to date have assessed how well a dolphin can identify objects rotated away from familiar orientations. In Nachtigall et al.'s (1980) study, a dolphin learned to discriminate between an upright foam cylinder and a foam cube with its flat face forward with high accuracy. In the test phase, there were probe trials in which the cube and cylinder were presented in different orientations (e.g., cube edge forward and cylinder top face forward). The dolphin showed poor performance when both objects were presented flat face forward ($M = 57$ %). This suggests a lack of object constancy on this shape discrimination task when training consisted of limited views of the objects.

In Au and Turl's (1991) study, a dolphin learned to discriminate among cylinders made of different materials (e.g., aluminum and coral rock) at three orientations (0°, 45°, and 90°). When later tested at novel orientations (15°, 30°, 60°, and 75°), the dolphin continued to discriminate among the stimuli with high choice accuracy (96–100 %). The high choice accuracy on novel orientations probably reflects the fact that they were similar to the object orientations used during training (within 15°) and that the object discrimination involved a material discrimination instead of a shape discrimination.

The current study expands upon Nachtigall et al.'s (1980) study and Au and Turl's (1991) study by presenting echoes from objects that vary in shape and material, as well as presenting novel aspect angles that are more than 15° away from the aspect angles used during training and by examining whether human participants were able to identify sequences of echoes as corresponding to particular objects when the object orientations differed from those that they were originally trained to recognize. Neural networks were also trained to classify these same objects using the spectral and temporal features of the echo sequences learned by participants and tested on their ability to accurately classify representations of novel echo sequences corresponding to different aspects. Neural network performance was then compared to human performance.

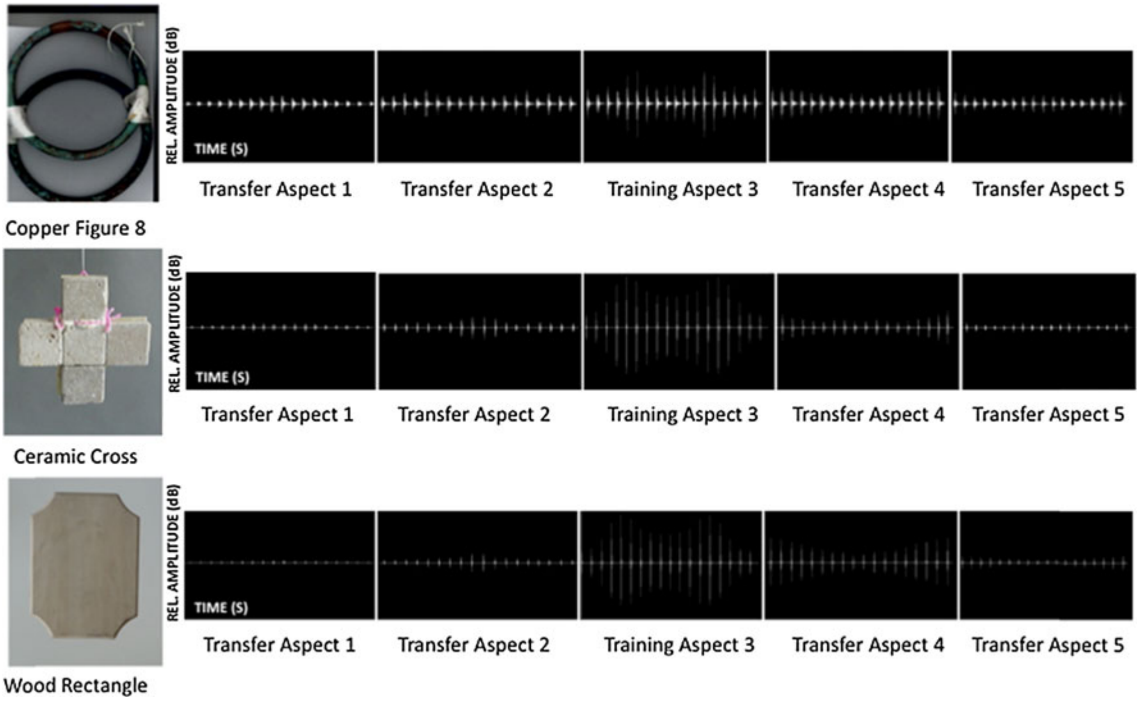

Fig. 1 The three objects and the stimulus echo trains for each of the five aspects, including the training aspect (3) and the four transfer aspects (1, 2, 4, and 5)

Human listening experiment

Methods

Participants

Twenty-six participants (14 females and 12 males) volunteered to be tested. Participants ranged in age from 19 to 49 years ($M = 22.6$ years). All participants were students or staff at Rochester Institute of Technology. All participants were tested for normal hearing with the Home Audiometer 2.0 hearing test that spanned 125–8 kHz (Esser Audio 2009), and all had normal sensitivity in the frequency range of the echo stimuli. Participants received either extra credit for a psychology course or were paid $10.

Stimuli

The stimuli used in this experiment were pre-recorded echoes from three objects: a ceramic cross, a copper fig. 8, and a wood rectangle (see Fig. 1). These objects were used in a previous echolocation study with a bottlenose dolphin (DeLong et al. 2006). Echoes from the three objects were recorded using a typical bottlenose dolphin click recorded from a male dolphin that has been used in numerous studies (70 µs long with a peak frequency of about 120 and a 60-kHz bandwidth; see Au 1993). Echo recordings were made from multiple orientations of each object in which one echo was captured for each aspect (1.3° apart) between −30° and +30°, where 0° is the front face of the object. Nine echo train measurements were collected for each object. A detailed description of the echo recording apparati and procedure can be found in DeLong et al. (2006).

The stimuli were slowed down to shift the spectra of the echoes into the human hearing range using Avisoft-SAS Lab Pro version 5.0.11 (Avisoft Acoustics, 2010). The original echoes were digitized at 1 MHz and had center frequencies around 125 kHz. The echoes were all time stretched by a factor of 50 by converting the echoes from digital to analog at 20 kHz. The time-stretched echoes had center frequencies around 2.5 kHz. Other studies in which dolphin echoes were presented to human listeners used the same factor of 50 (Au and Martin 1989; Helweg et al. 1995).

Each echo recording was an echo train consisting of 45 echoes ranging from −30° to +30°. The echo trains were divided into five aspects. Aspect 1 contained echoes from approximately −19° to −30°, aspect 2 contained echoes from approximately −7° to −18°, aspect 3 contained echoes from approximately −6° to +6° (including 0°, the front of the object), aspect 4 contained echoes from approximately +7° to +18°, and aspect 5 contained echoes from approximately +19° to +30°. Each aspect contained nine echoes, but dolphins rarely use so few echoes to discriminate among objects in a laboratory setting (Au 1993). For example, one dolphin typically used 20–40 echoes to detect a target (Au and Turl 1983). To increase the ecological validity of the task, the echo stimuli were formatted to mimic a dolphin getting multiple "looks" at each object so that there was a total of 18 echoes in each sequence of echoes. The first nine echoes correspond to a short left-to-right sweep of the dolphin's head, and the last nine echoes correspond to a short right-to-left sweep of the dolphin's head. Echoes were presented consecutively in each stimulus echo train with no delays introduced between echoes (although individual echoes were audible in the stimuli). The total duration of each stimulus echo train was approximately 1 s. Since there were nine echo recordings for each object, nine exemplars of stimulus echo trains were produced for each of the five object aspects (three objects × nine exemplars × five aspects = 135 total echo train stimuli).

Procedure

The participants were tested with an iMac desktop computer (Apple Inc. 2009) that channeled the echo stimuli via two Bose on-ear headphones (Bose Corp. 2009) to the participant and experimenter (the headphones were not noise canceling). Participants sat with their back to the experimenter and were not able to see the computer screen or the experimenter. The echo stimuli were played using Quicktime version 7.6.6 (Apple Inc. 2010).

Participants were tested individually by a single experimenter in a quiet room. After participants passed the hearing test, they heard a set of instructions and were given a sound vocabulary sheet with terms to describe the echo stimuli. The participants were instructed to use these terms and their definitions and were allowed to reference the sound vocabulary sheet throughout testing. Five terms and their operational definitions were given: "loudness" (how loud or quiet the sound is, related to the sound's amplitude or intensity), "pitch" (how high or low the sound is, related to the sound's frequency). "Length" (how long each echo is in duration), "timbre/sound quality" (the property in musical tones that makes it possible to distinguish one instrument from another when the pitch and loudness of the tones are the same), and "pattern of change in echoes across the echo train" (the change in the echoes within the echo train may form a different pattern for different objects). This ensured that all participants had the same set of descriptive tools to describe differences between echo stimuli in the interviews that followed each training and test session. Participants were told they could also use other terms that were not listed on the vocabulary sheet.

Participants were then played three pure tone sounds with the same pitch and timbre demonstrating a loudness difference (the tones rose in volume), two pure tone sounds demonstrating a difference in pitch (880 vs. 220 Hz tone played at the same volume by a piano), three tones demonstrating a differences in timbre (middle C played by a French horn, trumpet, and soprano sax at the same volume). Participants viewed photographs of the ceramic cross (6.3 cm × 8.1 cm), copper fig. 8 (7 cm x 8.2 cm), and wood rectangle (6.7 cm x 8.2 cm) throughout the study. All participants listened to the echo stimuli at the same overall volume, with the exception of one participant who preferred to listen to all the echoes at two-thirds volume (for all participants, the original amplitude relations between echoes in an echo train and between objects were retained).

First, participants received three training sessions. In each training trial, the experimenter played a stimulus echo train, the participant gave a verbal response (the name of the object), and the experimenter provided feedback (correct/incorrect). If the participant was correct, she/he would move onto the next trial. If the participant was incorrect, she/he was allowed to give a second answer and would again receive feedback (correct, on incorrect and the correct choice). In each trial, the participant was allowed to listen to the stimulus echo train as many times as they liked before they made their first answer (but they could not hear it again before their second answer). Participants usually requested to hear each stimulus echo train once or twice. The average number of times in a trial the participant requested to hear each stimulus across three training sessions is as follows: copper fig 8. = 1.03, wood rectangle = 1.03, and cross = 1.07.

Only aspect 3 for each object was used in the three training sessions. Each training session consisted of four blocks, with nine trials per block (three trials for each of the three objects) for a total of 36 trials per session. Stimulus echo train exemplars 1, 2, and 3 were used in the first two training sessions. The first 18 trials of training session three contained stimulus echo train exemplars 4, 5, and 6 and the last 18 trials contained exemplars 7, 8, and 9. The trial order within each training block was randomized separately for each participant. There was no pause between the trials.

Participants completed two test sessions after the training was complete. Each test session consisted of 36 trials (four blocks with nine trials per block). Each block contained the training aspect (aspect 3) and two transfer aspects: a near aspect (2 or 4) and a far aspect (1 or 5) for each of the three objects (only exemplar 1 was used in the test sessions). Test session 1 contained the following transfer aspects: cross aspects 1 and 2, copper aspects 4 and 5, and wood aspects 1 and 4. Test session 2 contained the following transfer aspects: cross aspects 4 and 5, copper aspects 1 and 2, and wood aspects 2 and 5. Half of the participants were randomly assigned to receive test session 1 first, and the other half received test session 2 first. The

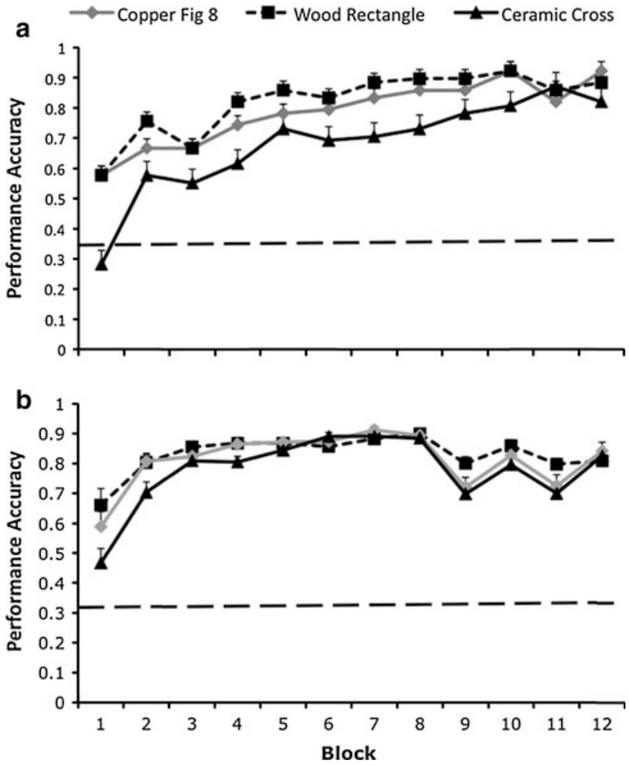

Fig. 2 Discrimination performance in the training sessions for humans (**a**) and artificial neural networks (**b**). Training session 1 included blocks 1–4, training session 2 included blocks 5–8, and training session 3 included blocks 9–12 (all training trials included only aspect 3). *Error bars* show standard error. The dotted line shows chance performance (33 %). For the human listeners in (**a**), performance was significantly above chance ($p < 0.001$) for all three objects in all blocks with the exception of the ceramic cross in blocks 1, 2, and 3

trial order within each test block was randomized separately for each participant. Participants were provided with feedback in each trial and allowed to select a second answer in the same manner as in the training sessions. Participants could listen to the echo train stimulus more than once before the first answer. The average number of times in a trial the participant requested to hear each stimulus across two test sessions was as follows: copper = 1.02, wood = 1.04, and cross = 1.05.

Participants completed an interview immediately after each training and testing session. Participants were asked to describe how the auditory features of the echo stimuli (e.g., loudness, pitch, and timbre) differed between the three objects, to report the feature that was most important to them when discriminating among the objects, and to describe whether their performance changed as the session progressed. After participants completed the entire experiment, participants were asked to specify auditory features for each object that helped them identify objects throughout the study. A SONY IC-Recorder ICD-PX720 (2009) was used to record interviews. Each training or test session followed by the interview took approximately 10–15 min. All participants completed the entire experiment in approximately 60–80 min.

Table 1 *T* tests for the training sessions with human participants

Block	Object		
	Copper fig. 8	Ceramic cross	Wood rectangle
1	4.11	0.94	4.11
2	5.25	3.39	7.46
3	5.49	3.00	4.70
4	7.33	4.17	10.63
5	10.03	8.18	14.00
6	8.34	6.56	11.88
7	10.11	6.31	17.49
8	12.58	7.65	15.81
9	16.06	7.75	15.81
10	21.11	11.35	17.63
11	10.63	15.51	11.51
12	17.64	10.63	15.11

Each cell shows the t value for a t test that was performed separately for each object in each block using one score for each participant (average performance on the block) and comparing the participants' scores to a value of 0.33. The degrees of freedom equal 25 for each t test and $p < 0.001$ for each cell except for the ceramic cross in blocks 1, 2, and 3

Results

Training sessions

Figure 2a shows the discrimination performance in the training sessions. Chance choice accuracy is 33 %, because the participants could choose from among three alternatives. To determine whether the participants' performance was above chance in the training session, a t test was performed separately for each of the three objects in each of the 12 blocks using one score for each participant (average performance on the block) and comparing the participants' scores to a value of 0.33. Since there were 36 t tests performed, a Bonferroni adjustment was made and the alpha level was set at $p = 0.001$. The participants' performance was significantly above chance for all objects in all blocks except for the ceramic cross in blocks 1, 2, and 3 (see Table 1). Performance accuracy was 80 % or higher on the last three training blocks for all objects.

Test sessions

The participants successfully transferred their discrimination to the four novel object aspects. Figure 3a shows the discrimination performance in the test sessions for the aspect participants were trained on (3) and the four transfer aspects (1, 2, 4, and 5). To determine whether the

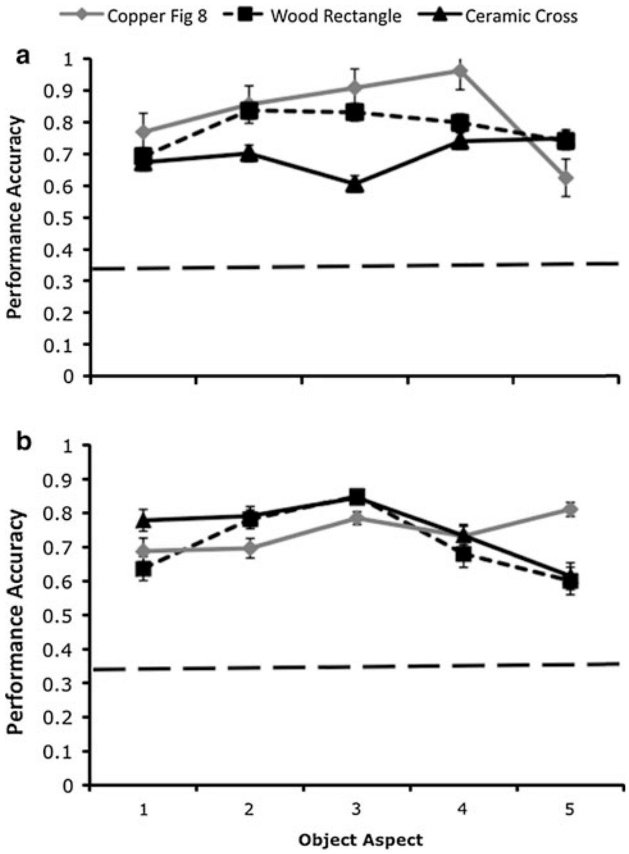

Fig. 3 Discrimination performance in the test sessions for humans (**a**) and artificial neural networks (**b**). Aspect 3 was the training aspect, and aspects 1, 2, 4, and 5 were the transfer aspects. *Error bars* show standard error. The *dotted line* shows chance performance (33 %). For the human listeners in (**a**), performance was significantly above chance ($p < 0.001$) for all three objects in all five aspects [Cross—Aspect 1, $t(25) = 6.04$; Cross—Aspect 2, $t(25) = 6.50$; Cross—Aspect 3, $t(25) = 5.82$; Cross—Aspect 4 $t(25) = 8.34$; Cross—Aspect 5, $t(25) = 7.57$; Copper—Aspect 1, $t(25) = 8.49$; Copper—Aspect 2, $t(25) = 12.52$; Copper—Aspect 3, $t(25) = 24.28$; Copper—Aspect 4, $t(25) = 27.75$; Copper—Aspect 5, $t(25) = 4.51$; Wood—Aspect 1, $t(25) = 5.95$; Wood—Aspect 2, $t(25) = 12.21$; Wood—Aspect 3, $t(25) = 16.34$; Wood—Aspect 4, $t(25) = 9.01$; Wood—Aspect 5, $t(25) = 9.61$]

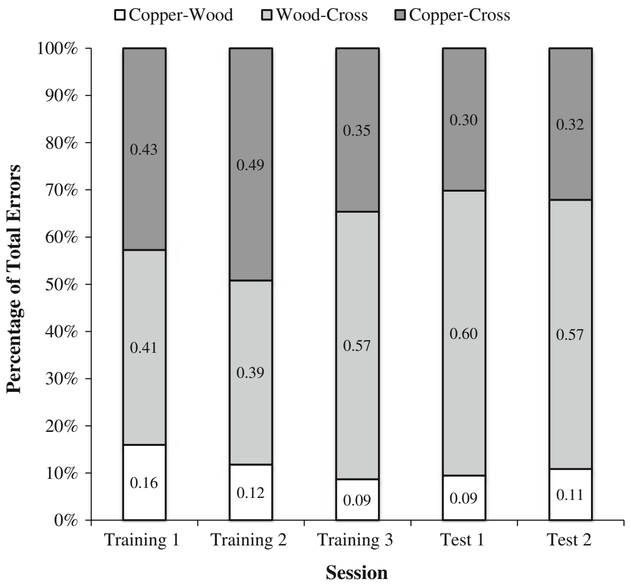

Fig. 4 The proportion of errors in each error type shown for all sessions of the human listening experiment. Participants could make three types of errors: confusing the copper figure 8 and wood rectangle, confusing the wood rectangle and ceramic cross, or confusing the copper figure 8 and the ceramic cross. The predominant error type in the test sessions was a wood rectangle–ceramic cross error

participants' performance was above chance in the test session, a t test was performed separately for each of the three objects in each of the five aspects using one score for each participant (average performance on the block) and comparing the participants' scores to a value of 0.33. Since there were 15 t tests performed, a Bonferroni adjustment was made and the alpha level was set at $p = 0.01$. The participants' performance was significantly above chance for all objects in all five aspects (see Fig. 3 for t values).

A 2 (order: order 1 is test session 1 first, order 2 is test session 2 first) × 2 (test session) × 3 (object: ceramic cross, copper fig. 8, and wood rectangle) × 3 (aspect: training aspect, transfer near aspect, and transfer far aspect) analysis of variance (ANOVA) with the last three factors as repeated measures was conducted on the proportion of correct answers made by the participants. There was a significant effect of object, $F(2, 48) = 9.42$, $p < 0.001$, a significant effect of aspect, $F(2, 48) = 7.99$, $p < 0.01$, and an interaction between object and aspect, $F(4, 96) = 5.73$, $p < 0.001$. Post hoc analyses revealed that performance for the copper fig. 8 on the transfer far aspect (70 %) was significantly worse than on the transfer near aspect (91 %) and the training aspect (91 %). Choice accuracy did not differ significantly between the training aspect, transfer near aspect, and transfer far aspect for the ceramic cross (61, 72, and 71) and the wood rectangle (83, 82, and 72 %; Newman–Keuls tests, $p < 0.05$).

There was also a significant effect of test session, $F(1, 24) = 6.34$, $p < 0.05$, and a three-way interaction between test session, object, and aspect, $F(4, 96) = 2.72$, $p < 0.05$. Post hoc analyses show that performance for the copper fig. 8 on transfer aspect 5 (62 %) was significantly worse than on training aspect 3 (94 %) and transfer aspect 4 (96 %) in session 1. In session 2, there was no significant difference in performance for the copper fig. 8 on training aspect 3 (87 %), transfer aspect 1 (77 %), and transfer aspect 2 (86 %). Choice accuracy did not differ significantly between the training aspect, transfer near aspect, and transfer far aspect for the ceramic cross and the wood rectangle in test session 1 or test session 2 (Newman–Keuls tests, $p < 0.05$).

Errors

Figure 4 shows the types of errors participants made in each session. Errors were divided into three types by pairing the objects: copper figure 8-wood rectangle confusions, wood rectangle-ceramic cross confusions, and copper figure 8-ceramic cross confusions. For example, when participants heard echo stimuli from the copper figure 8 but reported the wood rectangle or vice versa, those errors were classified as copper-wood confusions. The predominant error type in the test sessions was a wood-cross confusion. In the training sessions, wood-cross and copper-cross confusions were approximately equal in likelihood. There were very few copper-wood confusions in the experiment.

Second answers

All the results given above are calculated based on the first answer provided by the participant on each trial. Participants were allowed to select a second answer when their first answer was incorrect in both the training and testing sessions to provide them with further learning experiences. Because there were three objects, once a participant provided a first answer that was incorrect, they had to select one of the two remaining objects for their second answer. When taking into account both the first and second answers, participants' performance was nearly perfect on all but the first training session [training session 1 (89 %), training session 2 (97 %), training session 3 (99 %), test session 1 (96 %), and test session 2 (97 %)].

Reported auditory features

Table 2 shows the auditory features reported by human listeners after each session was completed. The participants reported using between one and five features to discriminate among the objects in the sessions. Participants reported using more features in the test sessions ($M = 3.83$) compared to the training sessions ($M = 3.21$). The features reported most often in the training sessions was pitch and timbre, whereas in the test session the features reported most often were pitch, timbre, and loudness. A majority of the participants (60 %) reported pitch to be the most important feature in the training session (e.g., some participants reported ranking the objects in terms of pitch; they said copper had the highest pitch, ceramic was intermediate, and wood had the lowest pitch). There was no consensus on the most important feature in the test sessions; pitch (36 %), timbre (33 %), loudness (19 %), and pattern (8 %) were all reported by some participants. The majority of participants (88 %) reported the echo stimuli from the novel aspects sounded different than the training aspect and many noted using a different strategy in the testing sessions as compared to the training sessions. Many participants reported using more auditory features (e.g., the addition of pattern and loudness) in the test sessions than in the training sessions (e.g., pitch and timbre).

Table 2 Echoic cues reported by human participants during the interview phase after each session

Echoic cues	Training sessions			Test sessions	
	1	2	3	1	2
Loudness	12	13	14	23	25
Pitch	26	26	25	25	25
Length	8	8	10	13	10
Timbre	20	20	24	23	23
Pattern of change in echo train	14	12	14	15	16
Average number of cues reported	3.12	3.12	3.38	3.81	3.85

Each cell contains the number of participants ($N = 26$) who reported using each cue for each session. Participants could report multiple cues for each session

Table 3 Echoic cues reported by human participants during the end-of-the-study interview

Echoic cues	Object		
	Cross	Copper	Wood
Loudness	9	9	15
Pitch	20	19	20
Length	4	1	5
Timbre	14	15	15
Pattern of change in echo train	7	7	8
Average number of cues reported	2.08	1.96	2.42

Each cell contains the number of participants ($N = 26$) who reported using each cue for each object after all sessions was completed. Participants could report multiple cues for each object

Table 3 shows the echoic cues reported by human listeners during the end-of-the-study interview. The participants reported using multiple features to identify each object in the study ($M = 2.15$). A majority of participants reported using pitch (76 %) and timbre (56 %) for all objects, but loudness was reported by 58 % of the participants to identify the wood rectangle (e.g., several participants said copper was the loudest, ceramic cross was intermediate, and wood was the quietest).

Discussion

The participants were able to learn to discriminate among the three objects using echoes from aspect three, and they successfully generalized that discrimination to four novel transfer aspects. In the test sessions, their performance on the training aspect was not significantly different than their performance on the transfer aspects (except with copper aspect 5). Their performance on the transfer aspects does not look like a generalization gradient in which performance on the transfer aspects near to the training aspect exceeds performance on transfer aspects far from the training aspect (c.f. visual object discrimination by pigeons; Kirkpatrick 2001). This may be because the "far" aspects are only 13°–24° away from the training aspect. A steeper generalization gradient would be expected if the transfer far aspects were 90° or 180° away from the training aspect.

The auditory features reported by participants suggest that they used some combination of amplitude and frequency over the course of echo trains to identify objects. However, whether these reported features are actually sufficient for identifying objects or how these features might be used to identify objects remains unclear. The acoustic features that identify an object might vary across objects or across aspects, such that amplitude cues are useful for identifying a particular object at a particular aspect, but not other objects at that same aspect. Similarly, participants that reported using patterns of frequency and amplitude over time could have been using those patterns for some echoes, but not for others. They might also have been assessing variations in frequency and amplitude over the course of the entire echo train or for only portions of the train. To more closely evaluate the acoustic information relevant for object identification within echo trains, we trained artificial neural networks (ANNs) to classify objects based on their echoes and then examined how the networks accomplished this task.

Neural network modeling

Methods

Network architecture

The ANNs used were multilayer perceptrons with a layer of 40 input units, a layer of 30 hidden units, and a layer of 3 output units (Fig. 5). Most input units reflected individual echoes within trains (18 echoes) and features extracted from those echoes (peak frequency and amplitude). Four additional units corresponded to summary features (outlined below). The appropriate number of hidden units was determined by pilot testing. No attempt was made to optimize or minimize network performance with different architectures. Several networks were piloted and the architecture that best matched human performance was used in the reported simulations. Output units corresponded to object categories. Each hidden unit's net input from connections with the input layer was converted to an activation level using a sigmoid activation function. In Fig. 5, the dotted line within hidden units shows the point at which the sum of the weighted input was converted to an activity level of zero. In the output layer, weighted inputs from the hidden units were converted to an activation level using a linear activation function. Twenty networks initialized with random weights were trained using a standard backpropagation learning algorithm (Rumelhart et al. 1986) with a learning rate of 0.05. All simulations were implemented using the Neural Network Toolbox running in MATLAB (R2010a).

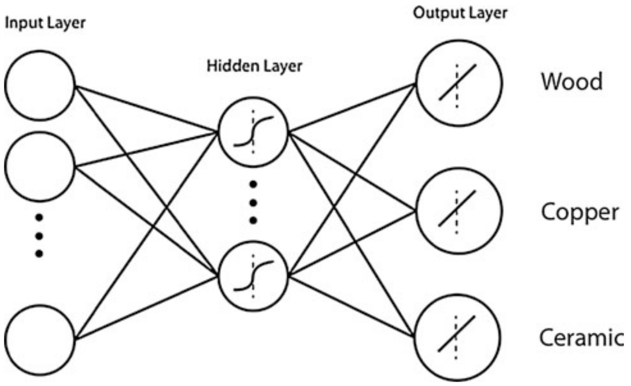

Fig. 5 A depiction of the network architecture used in simulation 2. Each input layer unit was associated with a feature extracted from echo trains. Input units had weighted connections to each hidden unit, and each hidden unit had a weighted connection with each output unit. Hidden units transformed the sum of the weighted input, plus a bias parameter, to an activity value with a sigmoid activation function. Output units operated similarly, except that they employed linear activation functions

Echo representations

Each echo recording was automatically analyzed using a customized MATLAB (R2010b) script to extract measurements of amplitude and frequency from echoes. Eighteen separate amplitude and peak frequency values were measured from each echo train (i.e., there was a peak frequency and amplitude measure for each echo). Additionally, these measures were used to compute means and standard deviations of amplitude and frequency across all echoes within a train. These measures were combined into a vector, normalized to fall between ±1, and then used as inputs for the ANNs.

Presentation of stimuli to networks

Artificial neural networks were given the same number of training blocks, test blocks, and stimuli as participants in the human listening experiment. The number of trials per block was increased from 9 to 48, because pilot studies revealed that ANNs required more training to reach comparable performance. Noise consisting of randomly generated values from a uniform distribution between ±0.2

was added to inputs before presentation to ANNs on each trial. This was necessary for ANNs to show variations in performance similar to those observed in humans. As with humans, networks were given feedback on all trials; weights were updated when the response generated in the output units differed from the target response. Target responses for each echo train were always set at one for the output unit associated with the object that generated the echoes and at zero for the remaining two output units. The output unit with the highest activity was considered to be the object identity endorsed by a network on a given trial. The percent of correct identifications during each block was used as a measure of performance accuracy.

Analyzing network and echo train structure

The absolute value of a weighted connection between an input unit and a hidden unit within the ANNs indicates how dependent hidden unit activity is on a particular input feature. Values close to zero indicate that hidden units are little affected by differences in the value of an input feature, whereas absolute values above zero indicate that the activation of hidden units is more strongly modulated by changes to that feature. The mean absolute values of weights between input features and hidden units were analyzed to determine which of the 40 acoustic features most strongly determined how echoes were classified.

In the human listening experiment, participants reported using frequency and loudness cues, but we do not know from those interviews whether these cues varied in importance at different points in the echo trains. To get a sense of how different parts of the echo train and the features that described entire echo trains were weighted, mean absolute weight values were also calculated for the beginning, middle, and end of the train. To further evaluate the relative importance of frequency versus amplitude cues, as well as possible shifts in their usefulness across training sessions, we calculated the difference in frequency and amplitude weights after training and after testing.

The ANNs created internal representations of echo trains via their hidden units in order to identify objects. It was the weights on these representations that ultimately determined the output values for any given echo train. Analyses of input weights only reveal the relative importance of input values. They do not reveal how ANNs use inputs to identify each object. To explore this issue, the most heavily weighted hidden units for each output unit from a single representative network were examined.

The ability of humans or ANNs to distinguish and categorize echo trains depends on the similarities between trains. To get a better sense of how frequency and amplitude features varied between echo trains recorded from different objects, we analyzed ANN inputs using a self-organizing map (SOM). The SOM learned to spatially organize echoes based on feature similarity, making it possible to visualize differences between echo trains varying along several dimensions (for further details, see Kohonen 2001). SOMs cluster inputs without information about object identity, providing a way to show how echoes from different objects differ acoustically. A 3 × 3 SOM was implemented, allowing echoes to be sorted in terms of their similarity to nine prototypes automatically constructed from all echo trains.

Results

Training sessions

Figure 2b shows the accuracy for the ANNs in the training sessions. As with human participants, the ANNs learned to identify the ceramic cross more slowly than the other two objects. Networks showed larger decreases in performance accuracy during the blocks where new echo exemplars were introduced than did humans (>80 % in blocks 9 and 11). Nevertheless, on the last block of network training, performance accuracy was 80 % or higher for all objects.

Test sessions

The ANNs successfully transferred their discrimination to the four novel object aspects. Figure 3b shows the discrimination performance in the test sessions for the aspect ANNs were trained on (3) and the four transfer aspects (1, 2, 4, and 5). ANN performance matched human performance in that the networks successfully identified all three objects at all five aspects. However, the simulations did not replicate the finding from the human listening experiment that performance on aspect 5 for the copper figure 8 was worse than aspects 3 and 4.

Errors

As in the human listening experiment, errors were divided into three types by pairing the objects: copper-wood confusions, wood-cross confusions, and copper-cross confusions. The mean percentages of confusions were: wood-cross = 35 %, copper-cross = 35 %, wood-copper = 30 %. Each confusion type contributed to total error similarly across blocks of training and test. The ANNs therefore replicated the finding from the human listening experiment that wood-copper confusions occurred less than wood-cross and copper-cross confusions. However, the percentage of wood-copper confusions made by ANNs was not as low as in the human listening experiment.

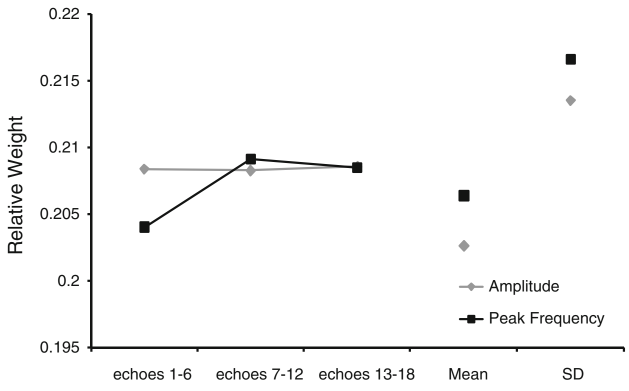

Fig. 6 Analysis of the relative value of weights between input units and hidden layer units. Relative weights are shown for each third of the echo train along with mean and standard deviation features derived from the entire echo train

Auditory features

Figure 6 shows the relative importance given to auditory features. Amplitude features were weighted similarly throughout echo trains. Peak frequency information from echoes, however, was weighted higher for the later portions of echo trains. Figure 6 shows that weights for the frequency of echoes 7–18 are higher than for echoes 1–6. The figure also shows that mean peak frequency was weighted more heavily than mean amplitude and that variation in frequency was weighted more heavily than variation in amplitude. Weighing of cues by networks was consistent with the cues reported by humans (i.e., more humans reported using frequency than amplitude; see Tables 2 and 3). However, difference scores (frequency |weight|—amplitude |weight|) for individual echoes showed that whether ANNs weighted frequency or amplitude more heavily was dependent on the echo. Comparisons of cue weighting before and after testing showed that the ANNs learned to weight amplitude features more heavily during testing than was the case in training.

Hidden unit patterns

Table 4 lists the most positively and most negatively weighted connections for each output unit. The connections that maximally activated or maximally inhibited output units overlapped across objects. For example, the hidden unit that most strongly activated the output unit corresponding to the wooden object, simultaneously strongly inhibited the output unit corresponding to the copper object (designated Wood +/Copper). This particular hidden unit responded to inputs with high amplitudes, high peak frequencies, and large variations in amplitude. Consequently, the ANN used these features as cues to both reject a train as being from a copper object and to accept the train as coming from a wooden object when the inputs were high. When the inputs were low, the reverse classification occurred. The hidden unit that most strongly inhibited activation of the wooden object unit was sensitive to these same cues, inhibiting this output unit most strongly when an echo train included low amplitude echoes, low peak frequencies, and little variation in amplitude. Other hidden units had similarly symmetric effects on the activation of output units indicating a copper or ceramic object (e.g., strong activation of the ceramic unit coupled with strong inhibition of the copper unit). The copper object output unit was triggered by echo trains with high peak frequencies and highly variable peak frequencies, whereas the ceramic output unit was activated by echo trains with low peak frequencies that varied little across echoes.

Echo similarity map

Figure 7 shows results from SOM analyses of echo trains. Each unit within the SOM corresponds to one of the nine prototypical sets of input values identified by the SOM. Pie charts associated with each unit show how many inputs of each object type are associated with each prototype. Some units within the SOM were activated by echoes from only one type of object, showing that the features given to the SOM were sufficient for identifying a subset of objects. For instance, nine echo trains from the wooden object activated one unit at the bottom of the map, and nine echo trains from the ceramic object activated an adjacent unit. Thus, activation of either of these two units identifies the object that was ensonified. Echo trains from the copper figure 8 activated unique units on both the left and right sides of the SOM, effectively identifying that object for at least a subset of echo trains. The three SOM units along the top, however, each responded to echo trains from multiple objects. The overlap in these units suggests that several echo trains have amplitude and frequency properties that are not systematically associated with particular objects.

Discussion

We trained and tested ANNs using procedures comparable to those used in the human listening experiment. ANNs were able to identify objects from echo trains at levels comparable to human participants and generalize the identification of objects to novel aspects in similar ways. One benefit of ANNs is that the mechanisms that they use to differentiate patterns can be analyzed in detail and can then be used to generate novel hypotheses about the kinds of cues or higher-level features that humans or dolphins use to perform this task.

The way in which ANNs learned to identify objects was in some manners consistent with human self-reports of the

Table 4 The most positively and negatively weighted hidden units for each output unit

Input feature	Hidden unit				
	Wood+/copper−	Wood−	Copper+	Ceramic+	Ceramic−
Mean amplitude	0.3838	−0.067	−0.1113	0.0391	−0.2189
Mean peak frequency	0.4182	−0.427	0.3296	−0.2885	0.0419
SD amplitude	0.3579	−0.3635	−0.0594	−0.0752	−0.2555
SD peak frequency	0.0706	0.1192	0.4054	−0.2793	0.3501

Hidden units are labeled with respect to how output units weighed their activities. The object name indicates the hidden unit to output unit connection. Plus and minus symbols indicate whether that connection was positively or negatively weighted. For instance, the wood unit is a unit whose connection to wood was inhibitory (high activities inhibited wood identification). Input features determining hidden unit activations are also shown

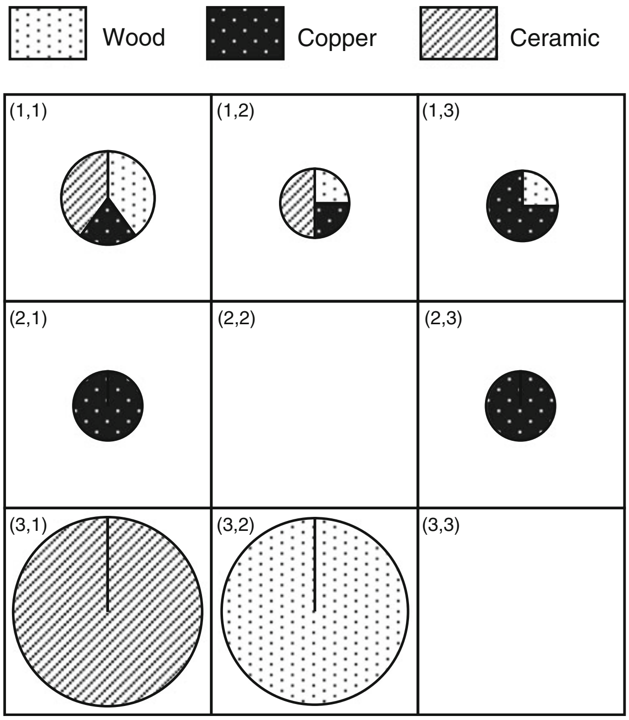

Fig. 7 Results from a 3 × 3 SOM grid trained with all echo trains used in the human listening experiment. A depiction of units that represent prototypical echo trains from each object activated in the trained map. *Pie chart* size represents the number of echoes activating that unit and shading indicates the proportion of those echoes that were from each object

features they used. For instance, more humans reported using amplitude after testing and the networks similarly weighted amplitude more after testing than after training. Also, more humans reported using frequency than amplitude and networks weighed the mean frequency more than mean amplitude. However, analysis of ANN input weights did not show a bias for making greater use of frequency information over amplitude information for individual echoes. For some echoes, frequency information was more important, but for others, amplitude was more important. The simulations suggest that differentially weighing the importance of features from each echo facilitates object identification and may account for participants' self-reports of using a combination of frequency and amplitude features. Tracking the variations in features across echoes may also be useful for identification as ANNs weighed the standard deviation of frequency and amplitude higher than most other cues. Half to two-thirds of the human listeners reported using the pattern of change over time to identify objects in the experiment.

Self-organizing map analyses showed that frequency, amplitude, and combinations of those features within an echo train can distinguish some objects. However, one feature may not be sufficient to distinguish an object from all other objects. In particular, the SOM grouped some echo trains from different objects as all being similar. Separating these similar echoes may require the extraction of higher-level features involving a combination of frequency and amplitude cues that can only be identified when specific information about object identities is available (e.g., in the form of feedback). It may also be the case that more fine-grained analyses of acoustic properties are necessary to reliably distinguish a subset of echo trains.

Given the structure of the SOM, participants in the human listening experiment may have picked up on the fact that the amplitudes of echoes coming from the wood rectangle and ceramic cross were higher than echoes coming from the copper figure 8 or were varying less in peak frequency across echoes. Distinguishing wood rectangle examples from ceramic cross examples might then have been frequency-dependent, because peak frequency for several echoes and mean peak frequency was higher for the SOM unit that ceramic examples activated. We cannot know for sure whether humans used the same acoustic features that the SOM and ANNs used, but the simulations clearly show that when ANNs differentially weigh acoustic cues in the ways shown in these simulations, then this leads to the levels of performance, generalization profiles, and errors that are observed in humans.

General discussion

Both humans and neural networks were able to learn to identify objects using echo trains from those objects and to recognize the objects based on echoes reflected from novel aspects. These findings provide evidence that representations of acoustic information from these objects can support object constancy. Different theories propose that visual object constancy is achieved through the formation of a single representation based on the distal stimulus (e.g., Marr 1982) or the formation of different representations constructed from different views of the object (e.g., Tarr and Pinker 1989). These theoretical explanations about the representations underlying object constancy have yet to be thoroughly explored for the auditory modality. The results from the current study with echo stimuli suggest that generalizations to novel aspects can occur without assuming a distal representation of the object.

The objects used in the current study varied in material, shape, and size. In another study, human listeners presented with echoes from objects made from the same material that vary only in shape and size and trained on a limited set of object aspects show poor performance when discriminating between objects at some novel aspects (DeLong et al. 2013). Studies of visual object recognition also show that object characteristics and the specific features of the objects play a role in determining performance on novel aspect angles (e.g., Tarr et al. 1997). The finding that human listeners were able to generalize across novel aspects when the objects varied in material in this study, but were not always able to do so when objects did not vary in material (DeLong et al. 2013) matches the performance of dolphins in two previous studies (Au and Turl 1991; Nachtigall et al. 1980).

The more interesting predictions of the current study relate to which objects and which aspects are most likely to prove challenging for a dolphin and more specifically which acoustic features may be particularly important for performing such tasks. Self-reports from human participants and detailed analyses of ANN classification mechanisms both suggest that learning to recognize objects from echoes involves using combinations of cues when making a classification decision. Attending to either absolute frequency differences or absolute amplitude differences is not sufficient to correctly classify all echo trains. In some cases, selective attention to acoustic features during segments of an echo train or to variations in amplitude or frequency across echoes will lead to more successful identification. Discovering such idiosyncratic feature patterns likely requires incremental perceptual learning mechanisms. In principle, the kinds of auditory patterns identified by humans, dolphins, and neural networks could differ significantly. However, the fact that humans and ANNs discovered similar acoustic features when learning to distinguish man-made objects suggests that some features may be intrinsically more useful for making such distinctions. In particular, analyses of hidden unit responses within ANNs indicate that opponent process mechanisms (in which features that distinguish objects lie on opposite ends of a continuum) may be particularly useful in identifying complex objects.

In the current experiment, the relevant dimensions for classification did not correspond to the simple dimensions of frequency, amplitude, or time usually emphasized in auditory analyses, but instead to more complex spaces in which dimensions might correspond to correlated changes in frequency, amplitude, and echo stability. A better understanding of how dolphins organize auditory space (and echoically identify targets) might thus be gained by monitoring the echoes received by freely moving dolphins that are echoically interrogating objects of various types from various angles and then training ANNs to associate trains of echoes from these recordings with objects of particular types. To the extent that ANNs can succeed at this task, they are likely to do so using complex object-related dimensions that do a good job of differentiating objects, as was observed in the current study. It seems likely that the neural systems underlying dolphin echolocation would, through either evolution or experience, converge on similarly useful dimensions for differentiating objects.

There were some differences in the performance of ANNs in comparison with humans during both training and transfer tests. Several factors can potentially account for these differences. Human participants came to the task with extensive auditory experience, biases about what sounds are like, and biases about how to interpret them. In contrast, ANNs started from a *tabula rasa* when learning the identification task. Additionally, the auditory features that humans were presented with differed from the information made available to the neural networks. In particular, ANNs were presented with precise absolute measures of amplitude and individual frequencies that were retained with perfect accuracy, whereas human participants heard physiologically transformed variations in loudness, pitch, and timbre. Unlike the inputs used with ANNs, the auditory features experienced by human participants had a resolution that varied as a function of frequency and amplitude, as well as across trials due to habituation. Some of these differences in acoustic signal processing could potentially be overcome by transforming measured acoustic features in ways that mimic processing in human auditory pathways (Patterson et al. 1995). Similarly, the performance of both ANNs and humans could potentially be made more similar to that of echolocating dolphins by transforming signals in ways that account for differences in the temporal and spectral resolving capacities available to dolphins (Branstetter et al. 2007). Even if these more sophisticated approaches succeed in homogenizing performance across species and machines, this still leaves open the possibility that a dolphin's auditory perception of objects differs radically from that of humans. After all, no one would suggest that the ANNs in the current study are perceiving echo trains in the same ways as human participants (or at all) despite many similarities in their learning and generalization profiles.

One limitation of the current study is that although both ANNs and human participants acquired at least somewhat aspect-independent auditory representations of objects, it remains unclear what properties of objects made them recognizable across aspects. Ensonified objects varied in shape, size, and material. Certainly, differences in material between objects could potentially generate differences in reflected echoes that would enable them to be distinguished. In this case, one might expect that the same material-generated cues would be available to distinguish objects regardless of aspect (since the material does not change with aspect). However, the SOM analyses revealed that there were no acoustic cues that distinguished all three objects across all aspects. This may be because the reflectivity of different materials can differentially impact echoes depending on an object's shape. For example, a glass sphere will not return the same echoes as a glass disk of equal radius. It seems likely that some combination of material-related cues and shape-related cues combine to enable identification of objects from multiple aspects.

One way to experimentally assess whether the acoustic features used by humans and ANNs to identify objects from their echoes are also used by dolphins is through phantom echo studies. Phantom echo experiments involve projecting acoustic signals that mimic the echo returns that would have occurred if an object had actually been present. The advantage of this approach is that specific acoustic cues can be selectively filtered from (or added to) the phantom echoes. So, for example, the echo train recorded from a ceramic object might be normalized so that all amplitude information is removed or so that all frequencies are equalized in terms of energy. Furthermore, the acoustic features of some objects can be gradually superimposed on or morphed into echo trains from other objects so that the threshold at which categorical perception switches from one object to another can be identified (as has been done in speech perception studies). A long-term study is just beginning in which echoes from the same three objects used in the current study will be presented to a dolphin subject. The dolphin will be exposed to the same training stimuli and novel test stimuli heard by the human listeners. In subsequent tests, the dolphin may be presented with modified echoes based on the salient auditory features identified by the human listeners and the ANNs.

Experiments with dolphins and other echolocating species generally require extensive resources and time. Consequently, techniques for identifying key hypotheses that can be tested through such experiments are particularly important. A combination of experimental results from humans, computer simulations, and naturally echolocating species is likely to provide greater understanding of the representational processes underlying echoic identification of objects and can also clarify the extent to which findings from visual object recognition research are applicable to perception in other modalities.

Acknowledgments Portions of this work were presented at the 165th meeting of the Acoustical Society of America in Montreal, Quebec, Canada, June 2013. This project was supported by a grant to CMD from the Rochester Institute of Technology Office of the Vice President for Research, a grant to CMD from the Rochester Institute of Technology College of Liberal Arts, and by NSF Grant #SMA-1041755 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center. The experiment described here was reviewed by the Rochester Institute of Technology Institutional Review Board and comply with the current laws of the country (USA) in which they were performed.

Conflict of interest The authors declare they have no conflict of interest.

References

Au WWL (1993) The sonar of dolphins. Springer, New York

Au WWL, Martin DW (1989) Insights into dolphin sonar discrimination capabilities from human listening experiments. J Acoust Soc Am 86:1662–1670

Au WWL, Turl CW (1983) Target detection in reverberation by an echolocating Atlantic bottlenose dolphin (*Tursiops truncatus*). J Acoust Soc Am 73:1676–1681

Au WWL, Turl CW (1991) Material composition discrimination of cylinders at different aspect angles by an echolocating dolphin. J Acoust Soc Am 89:2448–2451

Biederman I, Bar M (1999) One-show viewpoint invariance in matching novel objects. Vis Res 39:2885–2899

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

Branstetter BK, Mercado E III, Au WWL (2007) Representing multiple discrimination cues in a computational model of the bottlenose dolphin auditory system. J Acoust Soc Am 122: 2459–2468

DeLong CM, Au WWL, Lemonds DW, Harley HE, Roitblat HL (2006) Acoustic features of objects matched by an echolocating bottlenose dolphin. J Acoust Soc Am 119:1867–1879

DeLong CM, Au WWL, Stamper SA (2007a) 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 CM, Au WWL, Harley HE, Roitblat HL, Pytka L (2007b) Human listeners provide insights into echo features used by dolphins to discriminate among objects. J Comp Psychol 121:306–319

DeLong CM, Bragg R, Simmons JA (2008) Evidence for spatial representation of object shape by echolocating bats (*Eptesicus fuscus*). J Acoust Soc Am 123:4582–4598

DeLong CM, Heberle AL, Mata K, Harley HE, Au WWL (2013) Recognizing objects from multiple orientations using dolphin echoes. POMA. doi:10.1121/1.4799409

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cog Sci 11:333–341

Gorman RP, Sawatari T (1985) The use of multidimensional perceptual models in the selection of sonar echo features. J Acoust Soc Am 77:1178–1184

Gorman RP, Sejnowski TJ (1988) Analysis of hidden units in a layered network trained to classify sonar targets. Neural Net 1:75–89

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

Green DM (1993) Auditory intensity discrimination. In: Yost WA, Popper AN, Fay RR (eds) Human psychophysics. Springer, New York, pp 13–55

Guillette LM, Farrell TM, Hoeschele M, Nickerson CM, Dawson MRW, Sturdy CB (2010) Mechanisms of call note-type perception in black-capped chickadees (*Poecile atricapillus*): peak shift in a note-type continuum. J Comp Psychol 124:109–115

Helweg DA, Roitblat HL, Nachtigall PE, Au WWL, Irwin RJ (1995) Discrimination of echoes from aspect-dependent targets by a bottlenose dolphin and human listeners. In: Kastelein RA, Thomas JA, Nachtigall PE (eds) Sensory systems of aquatic mammals. De Spil Publishers, Woerden, pp 129–136

Helweg DA, Au WWL, Roitblat HL, Nachtigall PE (1996a) Acoustic basis for recognition of aspect-dependent three-dimensional targets by an echolocating bottlenose dolphin. J Acoust Soc Am 99:2409–2420

Helweg DA, Roitblat HL, Nachtigall PE, Hautus MJ (1996b) Recognition of aspect-dependent three-dimensional objects by an echolocating Atlantic bottlenose dolphin. J Exp Psychol Anim Behav Process 22:19–31

Herman LM, Arbeit WR (1972) Frequency discrimination limens in the bottlenose dolphin: 1–70Ks/c. J Aud Res 2:109–120

Jolicoeur P, Humphrey GK (1998) Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In: Walsh V, Kulikowski J (eds) Visual constancies: why things look as they do. Cambridge University Press, Cambridge, pp 69–123

Kirkpatrick K (2001) Object recognition. In: Cook RG (ed) Avian visual cognition [Online]. Retrieved from www.pigeon.psy.tufts.edu/avc/kirkpatrick/

Kohonen T (2001) Self-organizing maps. Springer, Berlin

Logothetis NK, Pauls J, Bülthoff HH, Poggio T (1994) View-dependent object recognition by monkeys. Curr Biol 4:401–414

Logothetis NK, 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

Moore PWB, Roitblat HL, Penner RH, Nachtigall PE (1991) Recognizing successive dolphin echoes with an integrator gateway network. Neural Net 4:701–709

Nachtigall PE, Murchison AE, Au WWL (1980) Cylinder and cube discrimination by an echolocating blindfolded bottlenose dolphin. In: Busnel RG, Fish JF (eds) Animal sonar systems. Plenum Press, New York, pp 945–947

Patterson RD, Allerhand MH, Giguere C (1995) Time-domain modeling of peripheral auditory processing—a modular architecture and a software platform. J Acoust Soc Am 98:1890–1894

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal errors by error propagation. In: Rumelhart D, McClelland J (eds) Parallel distributed processing: explorations in the microstructure of cognition, vol 1., FoundationsMIT Press, Cambridge, pp 318–362

Shepard RN, Metzler J (1972) Mental rotation of three-dimensional objects. Science 171:701–703

Supin AY, Popov VV (1995) Frequency tuning and temporal resolution in dolphins. In: Kastelein RA, Thomas JA, Nachtigall PE (eds) Sensory systems of aquatic mammals. De Spil Publishers, Woerden, pp 95–110

Tarr MJ, Pinker S (1989) Mental rotation and orientation-dependence in shape recognition. Cog Psychol 21:233–282

Tarr MJ, Bulthoff HH, Zabinski M, Blanz V (1997) To what extent do unique parts influence recognition across changes in viewpoint? Psychol Sci 8:282–289

Thompson RKR, Herman LM (1975) Underwater frequency discrimination in the bottlenose dolphin (1–140 kHz) and the human (1–8 kHz). J Acoust Soc Am 57:943–948

Weir C, Jesteadt W, Green D (1976) Frequency discrimination as a function of frequency and sensation level. J Acoust Soc Am 61:178–184

Wisniewski MG, Radell ML, Guillette LM, Sturdy CB, Mercado E III (2012) Predicting shifts in generalization gradients with perceptrons. Learn Beh 40:128–144

Zoccolan D, Oertelt N, DiCarlo JJ, Cox DD (2009) A rodent model for the study of invariant visual object recognition. PNAS 106:8748–8753