Imitation of Computer-Generated Sounds by Wild Atlantic Spotted Dolphins (*Stenella frontalis*)

Denise L. Herzing [1,2,*], Adam A. Pack [3,4], Fabienne Delfour [5], Thad Starner[6], Celeste Mason[6], Scott Gilliland[6], Charles Ramey[6], and Daniel Kohlsdorf [6]

[1] Wild Dolphin Project, North Palm Beach, USA
[2] Department of Biological Sciences, Florida Atlantic University, USA
[3] Departments of Psychology and Biology, University of Hawai'i at Hilo, USA
[4] The Dolphin Institute, Hilo, Hawai'i, USA
[5] Ecole Nationale Veterinaire de Toulouse, Toulouse, France
[6] Animal Computer Interaction Lab, Georgia Institute of Technology, USA

*Corresponding author (Email: dlherzing@wilddolphinproject.org)

Abstract – Since 1985 a community of wild Atlantic spotted dolphin (*Stenella frontalis*) have been observed underwater in the Bahamas. A human-worn, acoustic underwater two-way communication interface was developed and deployed from 2013-2016. Dolphins were exposed to an acoustic referentially based wearable underwater computer/interface. A model/rival system was used with dolphins and human participants during in-water sessions. Artificial and natural objects were labeled with computer generated sounds. Female juvenile spotted dolphins dominated the activity. Group size averaged seven dolphins for an average duration of 37 minute over 58 sessions. Of 243 video audio imitations and 56 Cetacean Hearing Augmentation Telemetry (CHAT) audio imitations, six potential response types were documented and measured. *Stand-alone* vocal contour mimics and Frequency *Modulated Contours* were the most common imitations. Within 5 sec of a computer-generated sound playing, of the 191 non-stand-alone vocal responses that were produced, 114 of them (59.7%) were judged as partial accurate matches, 3 of them (1.57%) were judged as non-matching partial imitations of a computer-generated sound, 67 of them (35.08%) were signature whistles, and seven of them (3.67%) were either non-signature whistle vocalizations or a mimic of the start or end tones. Thus, the majority of vocalizations produced by the dolphins within five seconds of a computer-generated sound were partial accurate imitations for the computer-generated sound played. Dolphins demonstrated both immediate and delayed vocal imitation and flexible attempts at imitation but did not show signs of a functional understanding of object labels. Atlantic spotted dolphins showed vocal flexibility in reaction to humans broadcasting computer generated sounds.

Keywords – Dolphin, Communication, Cognition, Interspecies, Imitation, Language

_____

Imitation, although historically considered a simplistic intellectual feat, is now regarded as a complex cognitive activity whose mechanisms and relationships to sociality, communication, and learning are still being investigated (see reviews in Dautenhan & Nehaniv, 2002; Huber et al., 2009; Subiaul, 2016; Whiten et al., 2004). A capacity for "true" imitation (distinguished from simpler forms of copying such as emotional contagion, observational conditioning, stimulus or local enhancement, or goal emulation, see discussion in Dautenhan & Nehaniv, 2002), may enhance social learning, social bonding, and cultural transmission of information. A key question in studies of imitative ability is whether a species is limited to

"contextual imitation" (i.e., imitation of pre-existing behaviors within its repertoire), or whether it can also engage in "production imitation," copying behaviors that are beyond its natural behavioral repertoire including those that are novel (Byrne, 2002; Byrne & Russon, 1998). A secondary question is whether the imitative act is a result of a response-facilitation mechanism that automatically activates an imitation program (Heyes, 2011) or whether imitation is more conceptual allowing for control and flexibility in the imitative process (e.g., Jaakkola et al., 2013). Answers to these questions can be facilitated by the task chosen to investigate imitation. For example, motor imitation has been demonstrated using the so-called two-action test in Japanese quail (*Coturnix japonica*) (Akins & Zentall, 1996), rats (genus *Rattus*) (Heyes & Dawson, 1990), and chimpanzees (*Pan troglodytes*) (Whiten, 2002; Whiten et al., 1996). In such tests a subject observes a model perform on a single device either one of two actions or a series of different single actions each with a binary choice to obtain a reward, and then is given access to that device to examine if they copy the witnessed action(s) more often than the alternative(s). However, this type of test is arguably limited in its ability to distinguish between contextual imitation and production imitation as well as to determine whether an individual can form a flexible imitation concept (Herman, 2002).

A task better equipped to handle these questions is the "do-as-I-do" task. Here, an observer witnesses a model perform one of a variety of different behaviors or behavioral sequences, including those that are novel, and must copy faithfully what it witnesses for reward. Using the do-as-I-do task, motor imitation on a conceptual level (i.e., the subject has demonstrated a concept of imitation by imitating novel acts beyond those familiar in its repertoire), has been demonstrated to various extents in chimpanzees and bonobos (*Pan paniscus*), (e.g., Custance et al., 1995; Myowa-Yamokoshi & Matsuzawa, 1999; Tomasello et al., 1993), orangutans (*Pongo pygmaeus*), (Call, 2001), parrots (order *Psittaciformes*) (Moore, 1992); bottlenose dolphins (*Tursiops truncatus*) (Herman, 2002; Jaakkola et al., 2010, 2013), and dogs (*Canis familiaris*) (Topal et al., 2006).

Production imitation, concept formation, and flexibility of the imitative process have also been demonstrated in non-human animals in the vocal domain. For example, harbor seals (*Phoca vitulina*) (Ralls, Fiorelli & Gish, 1985), Asian elephants (*Elephas maximus*) (Stoeger et al., 2012), beluga whales (*Delphinapterus leucas*) (Eaton, 1979; Murayama et al., 2014; Ridgway et al., 2012), and the superb lyrebird (*Menura novaehollandiae*) (Dalziell & Magrath, 2012) are all adept at imitating the vocalizations of other species. One of the strongest examples of vocal production imitation, as well as the ability to form an "imitate" concept that is flexible, is in bottlenose dolphins. Individual bottlenose dolphins naturally produce signature whistles that contain frequency-modulated contours that are unique to individuals and are developed through a combination of vocal learning and individualized changes (Janik & Sayigh, 2013). In the wild, as well as in captivity, signature whistles may be imitated by close companions (e.g., during visual separations) (Janik, 2000; Tyack, 1986). In the wild, King and Janik (2013) demonstrated that bottlenose dolphins produced their own version of their own developmentally learned, individually unique signature whistle when hearing a broadcasted facsimile of that whistle.

In an "imitate-what-I-play" variant of the do-as-I-do paradigm, Richards et al. (1984) trained a bottlenose dolphin to produce vocal facsimiles of computer-generated sounds following an "imitate" sound cue that was broadcasted underwater. Initially, in response to the sound cue followed by a "model sound," the dolphin was shaped to produce a whistle that first approximated the duration of the model, and then matched its base frequency and modulation. Eventually, the dolphin was able to imitate a variety of non-natural sounds with both precision and fidelity including novel sounds on first trials in response to the imitate sound cue. In other words, it had learned a concept of vocal mimicry that later in a separate study was extended to the motor behavior domain (Herman, 2002; Xitco, 1988; see also Jaakola et al., 2013 for a demonstration of motor and vocal imitation in the same dolphin in the same paradigm). Conceptual flexibility was demonstrated by the dolphin spontaneously shifting its production of some vocal facsimiles by an octave of the model.

In a second study, Reiss and McCowan (1993) showed that two other bottlenose dolphins could spontaneously vocally imitate computer-generated sounds that were broadcasted underwater without any preliminary operant shaping (as had been used in the Richards et al., 1984 study). Instead, a human demonstrated that the pressing of each of three unique symbols on an underwater keyboard resulted in a

unique sound being played as well as a unique physical consequence being produced (either one of two toy objects for the dolphin to play with or a hand rub from the human). As the dolphins observed the human actions and their consequences as well as explored on their own the contingencies of the keyboard (e.g., by pressing a key, hearing the sound associated with that key and receiving the associated toy or the rub), they began spontaneously imitating the sounds after hearing them play and also began producing facsimiles to the sounds just prior to pressing the key associated with that sound as well as while playing with the associated object or receiving the rub.

An important feature of vocal imitation captured by the Reiss and McCowan (1993) paradigm was the connection to social learning that involved function and consequence. That is, the dolphins observed a human interact with the keyboard which resulted in the consequences of a sound and an appealing object or behavior. In a follow up study, Hooper et al. (2006) compared the rates and quality of vocal imitation of three computer-generated sounds unpaired with any objects, and three other sounds paired with three objects the dolphins found appealing. For the latter sounds, after a sound was played the associated toy object was tossed into the habitat. Although the dolphins in this study vocally imitated spontaneously in both conditions without explicit training (partial imitations occurred for all six sounds, as they had been in Reiss and McCown, 1993, and improved in quality over time), the contextual pairing with toy objects appeared to enhance saliency which resulted in imitations occurring more frequently, after fewer exposures and of higher quality than when sounds were unpaired with toy objects.

The Hooper et al. (2006) study shows that vocal imitation in bottlenose dolphins can be enhanced in several ways by increasing the saliency of the sounds to be imitated. Salience of sounds to be imitated can also be enhanced by presenting them as object labels within a model/rival paradigm as was successfully done with an African grey parrot (*Psittacus erithacus*) by Pepperberg (1981) (see also Todt, 1975). In this paradigm, the African grey parrot observed as a human trainer interacted with another human who served as both model for the parrot and rival for the human trainer's attention. On a trial, the trainer requested in English that the model identify a presented object using its verbal label. If the human acting as model/rival correctly named the object, it was given the object and social approval through verbal praise. If it did not, the object was removed from sight and the model/rival received social disapproval by being scolded by the trainer. The roles of trainer and model/rival were then reversed for the next trial. During or between these trials, if the parrot appeared to attempt the vocal label through a partial or full imitation, it received attention from either the model/rival or the trainer in the form of a question about whether the parrot wanted the object and asking for its name or requesting that the parrot produce a better pronunciation. Successful verbal labels as determined by the humans resulted in the parrot receiving the object and verbal praise, as he had seen occur with the model/rival. A follow on "productive" procedure involved a single human acting as trainer presenting an object in front of the parrot while labeling it using different declarative sentences and then withholding the object until the parrot produced the correct label (in which case it received the object and the verbal label). After a little over two years with these two procedures, the African grey parrot was able to label as well as request different objects. Thus, for the parrot vocal imitation of the model/rival played a key role in the acquisition of vocal labels for objects although the degree to which it understood the referring function of these labels at this stage is debatable (see Savage-Rumbaugh et al., 1980a, 1980b for a discussion of the essentials of referential communication).

A common feature of the studies of Pepperberg (1981), Reiss and McCowan (1993), and Hooper et al. (2006) is that saliency of sounds appears to facilitate imitation of those sounds, and that vocal imitation, like occurs in humans first acquiring language (e.g., Bloom, 1970; Bloom et al., 1974; Kuczaj, 1982, 1987), often begins with partial imitations. However, all three studies were confined to captive situations. Here, we explore whether known individuals from a resident community of wild Atlantic spotted dolphins (*Stenella frontalis*) with long-term experience interacting with humans (Herzing 1997, 2011, Herzing et al., 2012), will imitate computer-generate sounds that are made salient through a procedure similar to the model/rival paradigm of Pepperberg (1981).

Methods

Ethics Statement

All research met animal standards and was approved under the Bahamas Department of Natural Resources and their permit process for wild marine mammal research (DNR - MAMR/FIS/12A).

Field Site and Research Vessel

The study took place in two separate field locations in the Bahamas, Little Bahama Bank (LBB) and Great Bahama Bank (GBB) (Figure 1). These areas form home ranges for over 300 wild Atlantic spotted dolphins. Since 1985, researchers of the Wild Dolphin Project have identified individuals in these communities through photographs and video footage of their dorsal fins, tail flukes, and constellations of spots on their bodies, have sexed them through direct observation of the genital area, categorized them into age classes by their degree of spotting and color phase (T = calf, S = Juvenile, M = young adult, F = Old adult), and have traced their individual life histories for over 35 years (e.g., Herzing, 1997). Given that much of this work has involved close underwater observations, photography and videography, the dolphins have largely become habituated to the presence of humans and sometimes engage with them in synchronous swimming and interactions involving the exchange of natural and non-natural objects (Herzing et al., 2012). The research vessel used to access the field site and from which researchers entered the water for this study, R/V Stenella, is a 20 m live-aboard power catamaran.

Figure 1

Map of the Study Area
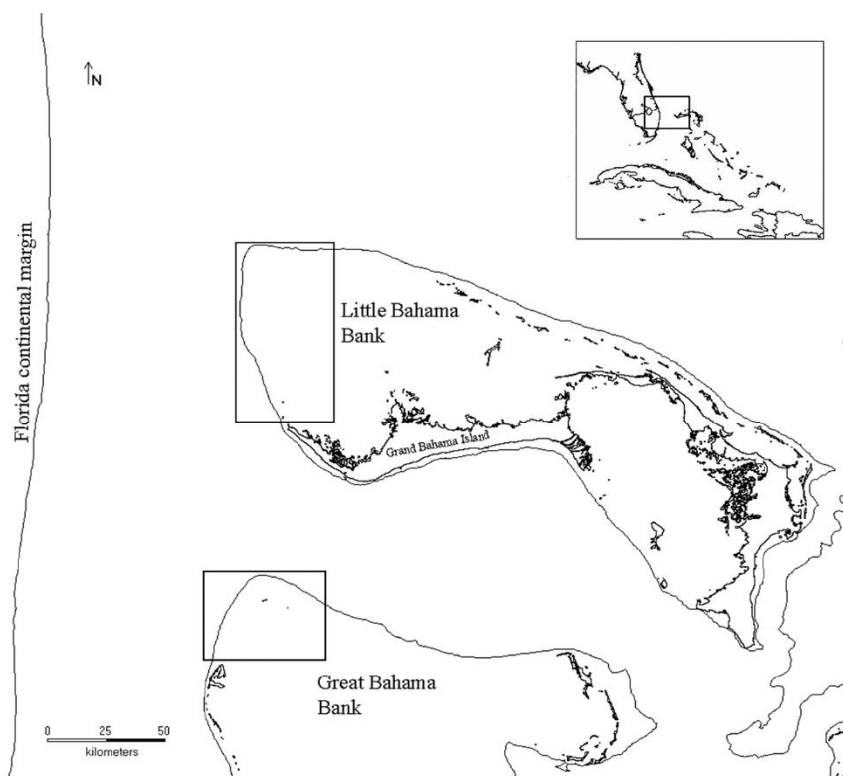
Note. This figure shows Little Bahama Bank (LBB) and Great Bahama Bank (GBB) study areas indicated by the boxes.

Subjects

We worked with a subset of the LBB Atlantic spotted dolphin community, and occasionally with a subset of the GBB Atlantic spotted dolphins. All the LBB dolphins had known life histories (Herzing, 1997). The dolphin society was already habituated and well exposed to the presence of humans in the water and an initial group of LBB dolphins had been exposed to humans pointing to natural as well as non-natural objects or pointing at or touching visual symbols associated with these objects either silently or in conjunction with an associated sound (Herzing et al., 2012). However, this early technology did not permit the examination of whether these dolphins ever imitated the sounds associated with objects. This paper describes the use of a second underwater interface utilizing only acoustics and real-time sound recognition and pattern recognition software with a different subset of Atlantic spotted dolphins that had grown up in the same community and were also habituated to human presence.

Technology

In 2013 a prototype wearable underwater computer Cetacean Hearing Augmentation Telemetry (CHAT) (Figure 2) was designed to play artificially created frequency modulated (FM) whistles in Audacity 2.4.2 (an open-source software for recording and editing sounds) and recognize these same whistles in real time (Herzing et al., 2012). Detailed technical descriptions of the system can be found in Kohlsdorf et al. (2013). Briefly, each artificially created sound could be initiated by depressing one of the 12 keys on the keypad; four whistles for objects on the top row, four signature whistles for individual humans in the middle row (created to be outside the repertoire of the dolphins), and four of six signature whistles of existing dolphin candidates on the bottom row, which were rotated depending on the dolphins' presence in a given year. Each human participant wore a bone-conducting headset that allowed preloaded English words to be heard, in a male voice, for each signal triggered from an operator, and a female English voice for any external incoming sound matches.

Whistle sounds used in 2013 were designed using the computer program Audacity 2.4.2 and covered 3.5-9 kHz range and were played at a 44.1 kHz sample rate. The CHAT system and underwater video cameras recorded audio at 44.1 kHz rate, allowing detection of sounds up to 22 kHz. In 2014, the CHAT system was improved and after initial testing in the field, the system played back and sampled at 192 kHz and had pattern recognition abilities. This same system was also used during 2015 and 2016 field seasons. Other experimental projects have used a variety of programs to synthesize whistles for a variety of uses (Janik et al., 2006).

Three identical CHAT boxes were used during the field seasons. Boxes 1 and 2 were used to produce sounds in the water with the dolphins while humans interacted with them, and Box 3 was used for additional recording of the sessions in the water, or as a substitute for Box 1 or 2 if either box was unavailable. All CHAT computers were time synchronized with each other and any other underwater equipment (videos, cameras, stand-alone acoustic recorders). The computer logs registered all activity on the computer including keypad hits, sound matches, and other activities. Computer logs were downloaded wirelessly after each session, once the systems were out of the water.

Figure 2

The Cetacean Hearing Augmentation Telemetry (CHAT) Device
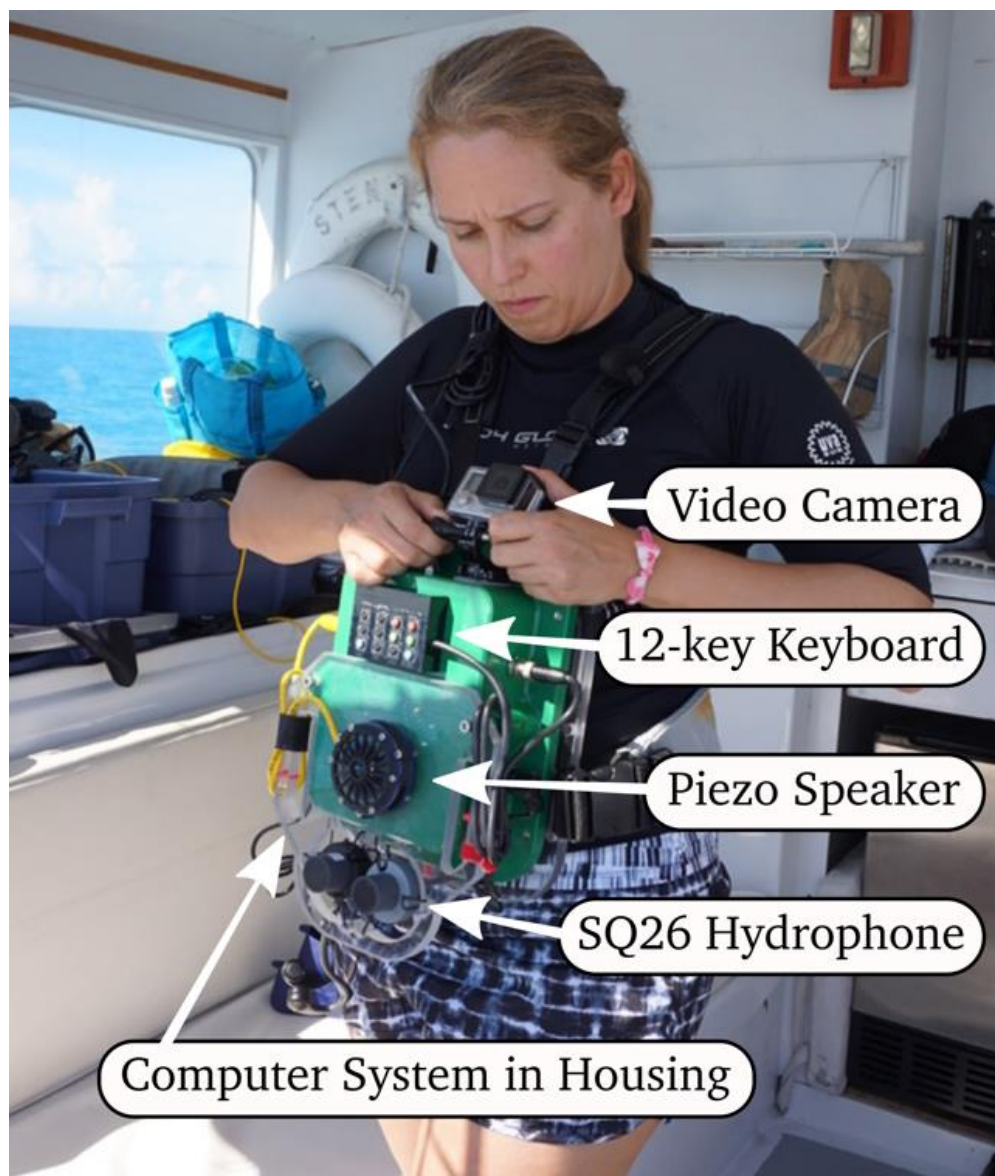
Note. This photo shows the researcher with a CHAT box on and component parts.

Signal Design and Labels

Eight simulated computer-generated sounds (CGS), all FM whistles, were created and associated with different objects and humans. CGS whistles ranged from 1-2 sec in duration. CGS whistles were designed to be outside the dolphin's natural repertoire. To verify this, we tested four years of data (pre-CHAT years) to examine if the CHAT computer generated signals that we artificially created were found in the dolphin's normal repertoire. We used a machine learning tool that identifies sound matches of any queried sound files (Kohlsdorf et al., 2016). This tool allows the user to upload an audio file, then it presents the first best match (if any), in the dataset, of the component parts of a signal. A second-best match (if any) follows after a first match. Of the eight computer generated sounds (Scarf, Sargassum, Rope, Grass, Adam

Pack Signature Whistle, Denise Herzing Signature Whistle, Bethany Augliere Signature Whistle, and Fabienne Delfour Signature Whistle) inputted into this program, none showed a match out of 4 years of pre-CHAT data. Thus, the CGS sounds represented novel signals that were not within the dolphins' vocal repertoire prior to this study. In addition, we found no evidence that they had been produced in response to the earlier study sound exposure (Herzing et al., 2012).

Each whistle was preceded by a 250 ms "start" pure *Tone* and a 50-85 ms silent interval and followed by a 50-85 ms silent interval and a 250 ms "stop" pure tone. This pure tone was also not in the normal repertoire of the dolphins. Four CGS FM whistles used in earlier work (Herzing et al., 2012) were associated with two natural objects found in the dolphins' wild habitat SA (*Sargassum*) and GR (*Seagrass*), and two non-natural objects SC (*Scarf*), and RP (*Rope*) (Figure 3). In addition, in 2013 and 2014 an 8 kHz "start" pure tone and 7 kHz "stop pure tone" were used (Figure 3). In 2015 the "start" and "stop" pure tones were changed to a 10kHz signal (Figure 4). Again, none of these pure tones were in the normal repertoire of the dolphins. The *Seagrass* signal GR was previously labeled as *Bowride*, (as described in Herzing et al., 2012), but this CGS was unused in the system, so it was repurposed in 2014 as *Seagrass*.

Each of the four trained CHAT operators (Adam Pack, Denise Herzing, Bethany Augliere, and Fabienne Delfour) were associated with a unique synthesized signature whistle programmed into the computer symbolized by their initials followed by SW for signature whistle: APSW, DHSW, BASW, and FDSW. Each synthesized whistle had varying durations and similar start and stop tones as the object-associated whistles (Figure 5). A protocol was adopted in which each CHAT operator used his or her own signature whistle when "greeting" a dolphin in the water, or to refer to another human prior to requesting an object from them in the presence of the dolphins. Previously recorded signature whistles of the six most highly interactive dolphin candidates (Meridian, LPrawn, Bijyo, Palette, Brat, Neridae) were also put into the system, over the years, to acknowledge their presence during CHAT sessions (Figure 6).
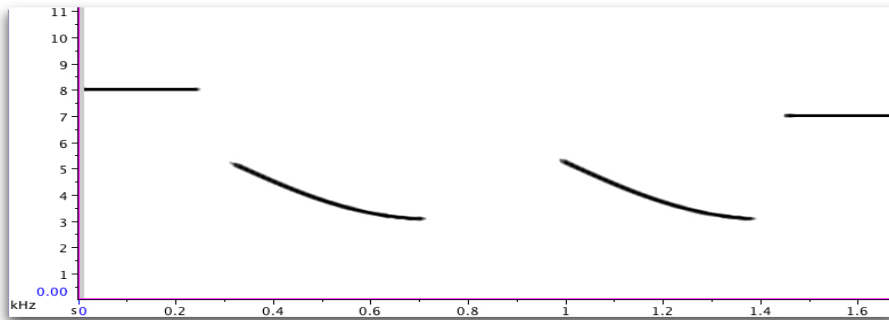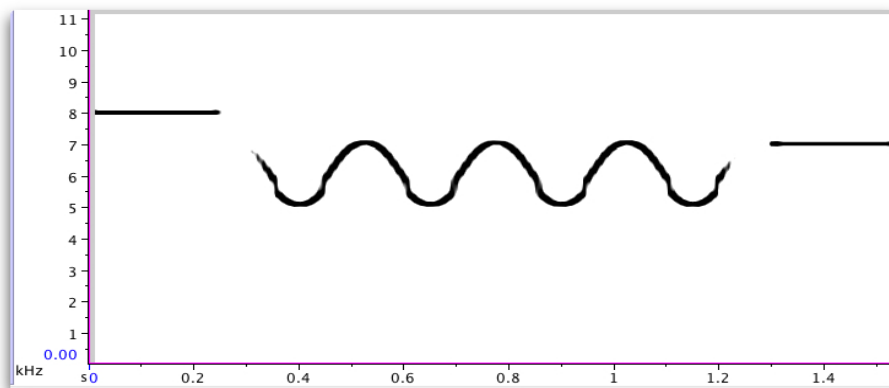
Figure 3

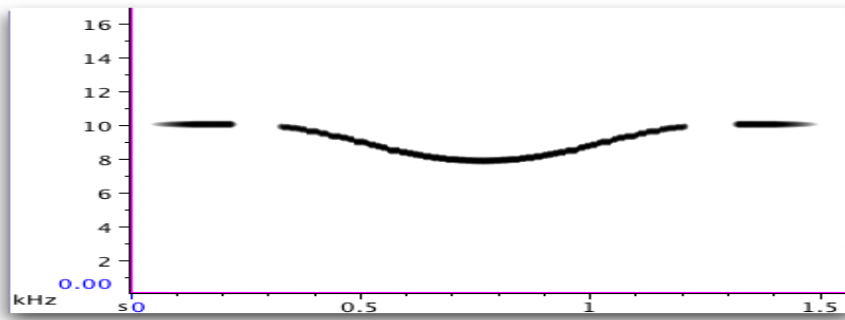*CGS Object Label Spectrograms*
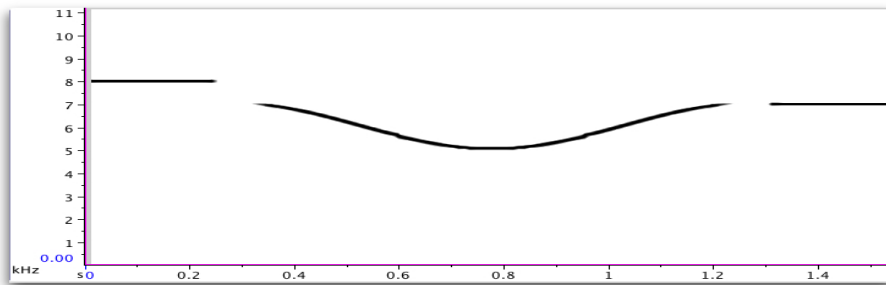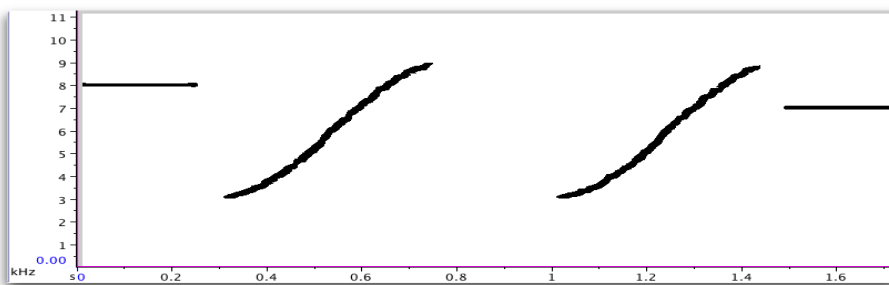
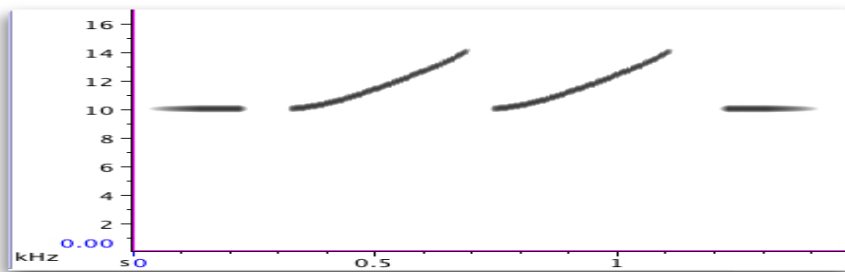SCARF

SARGASSUM

ROPE

SEAGRASS

Note. This figure shows CGS object label spectrograms with "start" and "stop" pure tones at 7 and 8 kHz.
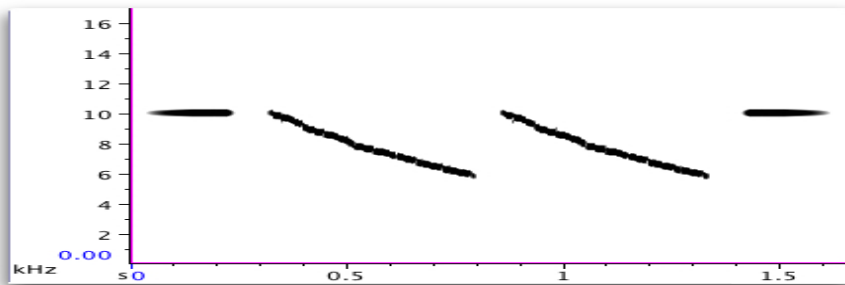
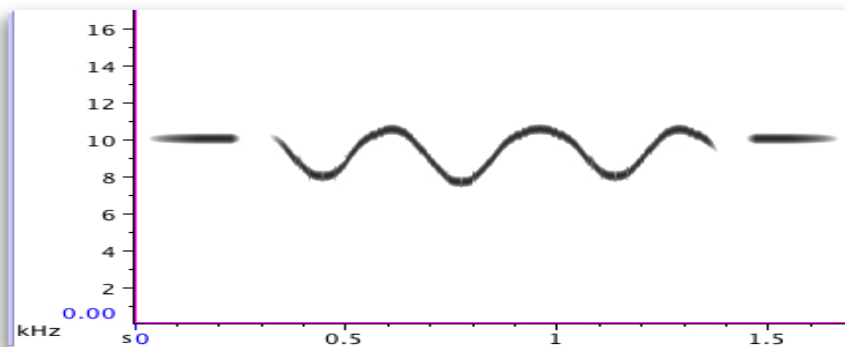**Figure 4**

CGS Object Label Spectrograms
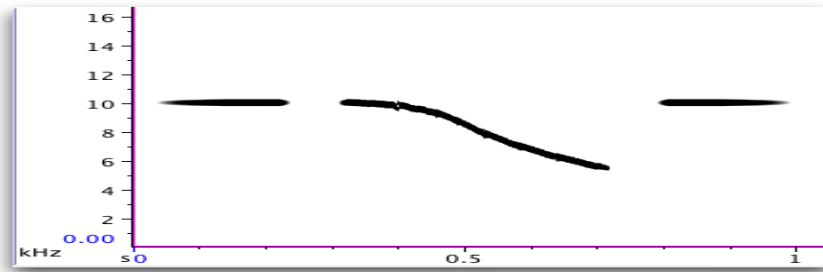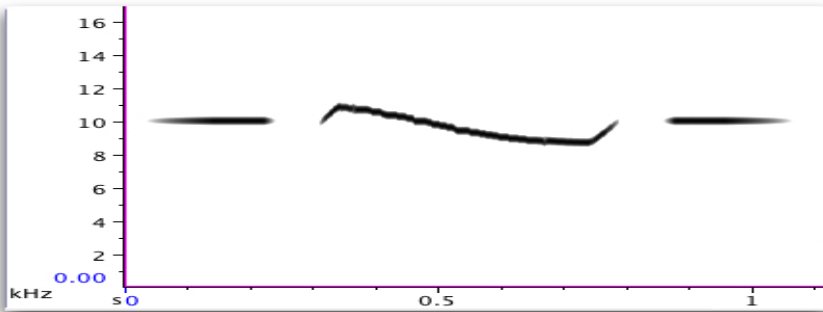

SCARF

SARGASSUM

ROPE

SEAGRASS

Note. This figure shows CGS object label spectrograms with "start" and "stop" pure tones at 10 kHz.
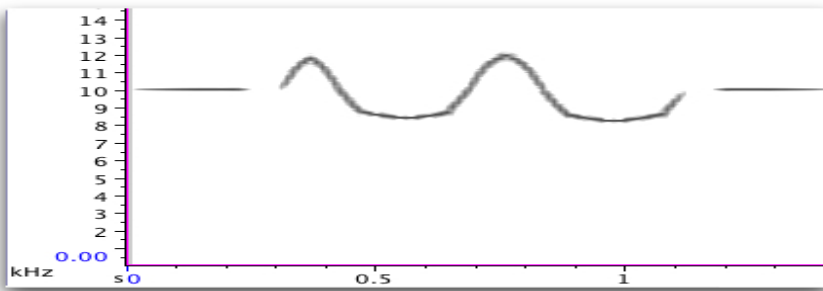
Figure 5
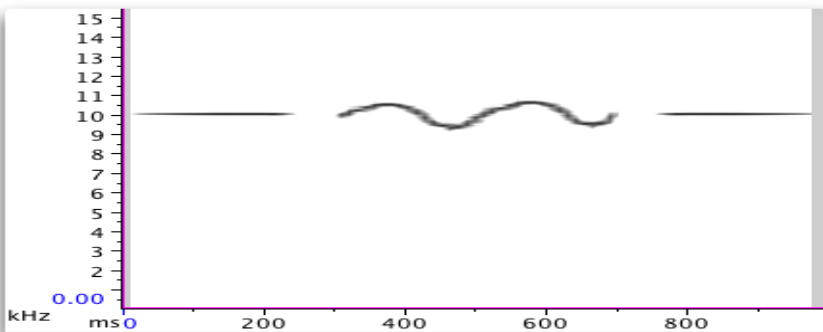
*Signature Whistles Created for Researchers*

DHSW

APSW

BASW

FDSW

Note. This figure shows CGS Signature Whistle Labels created for Researchers including a "start" and "stop" pure tones at 10 kHz. DHSW is Denise Herzing Signature Whistle, APSW is Adam Pack Signature Whistle, BASW is Bethany Augliere Signature Whistle, and FDSW is Fabienne Delfour Signature Whistle.

Figure 6

Signature Whistles of Dolphins Used in CHAT Work
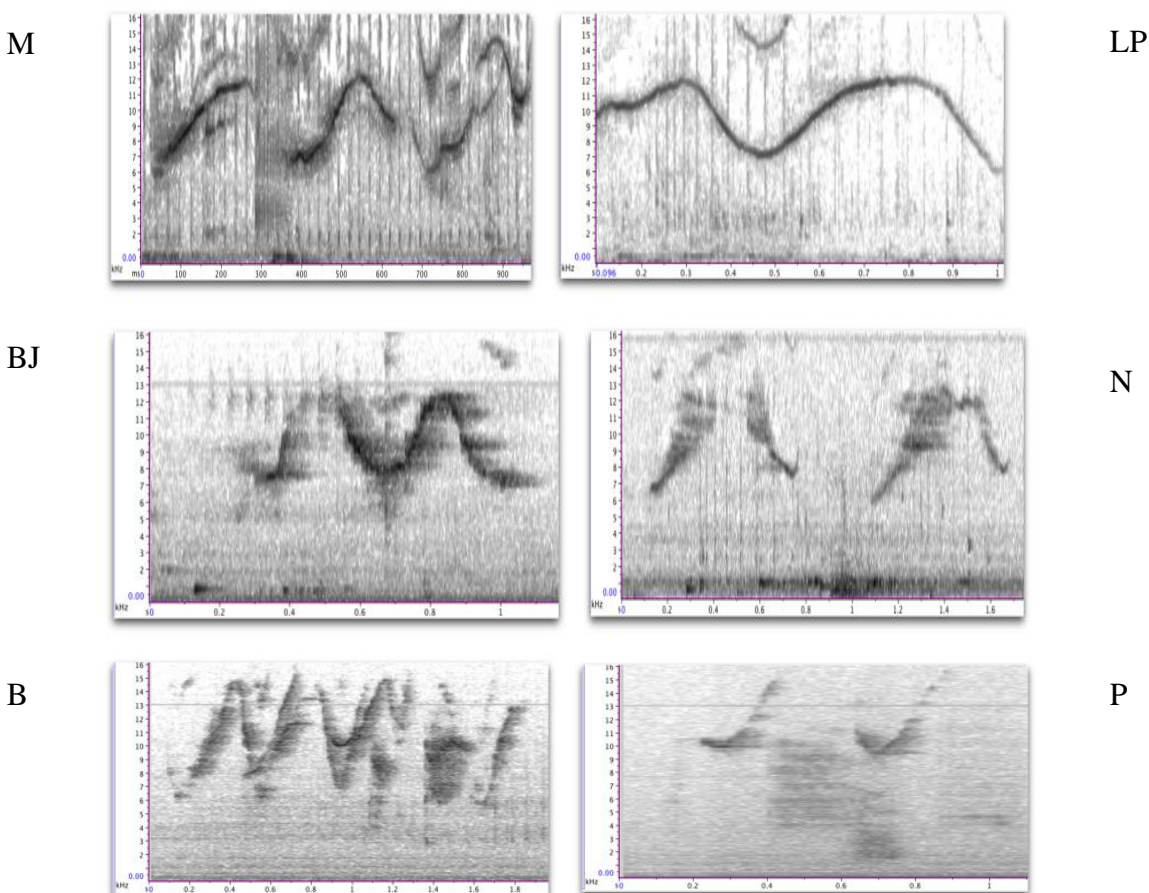
Note. This figure shows recorded signature whistles of individual juvenile dolphins incorporated into the CHAT computer for playback including Meridian (M), LPrawn (LP), Biyjo (BJ), Neridae (N), Brat (B), Palette (P).

Field Deployment, Testing, and Protocols

Before using CHAT in the presence of dolphins, researchers tested the accuracy of the boxes in the water systematically at a variety of distances, specifically for box-played sounds. Two researchers wearing the CHAT boxes on their chests faced each other while holding a safety line aft of the research vessel. One researcher played each preprogrammed sound at 3m, 7m, 10m, or more away from the second researcher. The same sets of sounds were then played while one researcher was oriented at different angles up to 90 degrees, and spaced at the same distances as above. After researchers tested all sounds multiple times, at multiple angles, and switched locations on the safety line, and after boxes reached over 95% accuracy on the recognition of the preprogrammed sounds (some up to 30m) the system was deemed ready to use with the dolphins. CHAT was designed to be quickly and easily put on by a human so that they could be deployed expeditiously in the presence of dolphins. Researchers practiced using the system, with each other, on multiple occasions, to familiarize themselves with operating the units while swimming and to coordinate communication protocols with other humans using CHAT in various contexts.

Using CHAT with the Dolphins

When dolphins were encountered, CHAT operators entered the water and used the system when the following conditions occurred: 1) the dolphins approached and showed an interest in the humans and were not engaging in specific intraspecific behaviors, and 2) sea surface and current conditions were deemed safe for humans. When entering the water, humans carried one or more of the objects associated with the different sounds in order to engage the dolphins in play if appropriate (i.e., if they were not engaged in intraspecific behaviors and also showed interest in human interaction). Before every encounter, the vessel's sonar system and operation of any non-essential equipment which produced noise that might interfere with CHAT system operation was suspended.

During interactive encounters, one to three humans wore CHAT boxes. Although on occasion a single CHAT operator worked in the water without a second modeling human, more typically two CHAT operators were interactive with the dolphins. If a third CHAT operator entered the water, he or she passively recorded sounds on the CHAT box but did not interact. Other human team members were designated either videographer or photographer so that all sessions were recorded visually and acoustically for the purposes of verifying the identity of the dolphins present and provided a wide-angle view of the events and context of the CHAT sessions.

As noted earlier, CHAT operators entered the water with one or two of the objects either in their hands or partially tucked into their bathing suit to be visible to another human and to a dolphin. Once a dolphin showed interest in an object, for example by orienting its rostrum to that object or echolocating on it, only the CHAT operators engaged the dolphins in activities. Humans modeling the use of CHAT CGS to refer to the object in the presence of one or more dolphins occurred in three scenarios:

1. Two CHAT boxes and two humans: CHAT Operator A requested a particular object from CHAT Operator B (who had one or two objects visible on him/her), either by acoustically playing the label associated with that object or by pointing at it with their arm and body in alignment towards the object. In the former case, CHAT Operator B responded by giving the acoustically labeled object to CHAT Operator A. In the latter case, CHAT Operator B responded by first "commenting" by relabeling the object that had been pointed at, by playing the sound associated with it, and then giving the object to CHAT Operator A.
2. One CHAT box and two humans: A human without a CHAT device requested an object from the CHAT box operator by pointing at it. The CHAT box operator's response was the same as when pointing was employed in the two CHAT box scenario.
3. One CHAT box and one human: A human with a CHAT box labeled objects as dolphins explored them. The human operator would occasionally change to a different toy during a session if desired.

In the two human types of scenarios, once a human received the object that he/she had requested from another human, they engaged in various forms of play with that object, were available to model an additional object exchange with a human, and were available to any dolphin requesting an object either by pointing at it with its rostrum and aligned body and/or through the production of a facsimile of the sound associated with that object.

Additional CHAT Scenarios

Beyond human-human modeling of requests for objects, the CHAT system was used in a variety of additional ways as listed below:

1. Signature Whistle Greeting: A CHAT Operator played his or her human SW when encountering a dolphin in the water, usually after the dolphin vocalized his or her own signature whistle. If the dolphin was recognized as one whose signature whistle was in the CHAT computer program, the CHAT Operator could also play that dolphins' whistle in greeting to that specific dolphin.
2. Human Requesting an Object from a Dolphin: If a dolphin possessed an object that a CHAT Operator desired, the CHAT Operator played the sound for that object with or without simultaneous

pointing to request that object. This could also be interpreted as a human labeling an object in a dolphin's possession. If directly after the human signal, the dolphin dropped the object, the human provided positive feedback in the form of excited motions such as hand-clapping underwater and either labeled the object again, retrieved the object, or pointed to it with and without a label directing their attention either to a dolphin or to another human as a signal to retrieve the object.

3. <u>Human Declarative Pointing and Acoustic Labeling:</u> A human dropped an object in the water column close to another human, pointed to that object while simultaneously labeling it acoustically, then provided positive feedback in the form of underwater clapping once the human retrieved the indicated object. This type of feedback was also provided if a dolphin retrieved the object and was sometimes accompanied by the human replaying the acoustic label for that object.

After a session ended (either because the dolphins exited the area, the dolphins began to engage in intraspecific behavior unrelated to the objects, or the weather or sea conditions became poor), and the humans returned to the research vessel, data were logged including date, time, video and audio equipment in the water, location, humans with CHAT devices on, dolphin identification and sex, age class, level of engagement with humans and the system, names of specific dolphins that were present, the level of attentiveness and engagement with the system, objects available, sounds played in the water and general notes about what activities occurred. CHAT automatically recorded all keypad input and output activities in a time log as well as audio data. These exposure data were also extracted for every CHAT session from computer logs, or manually from the audio files when logs were not available.

Processing of Audio Data with Vocal Responses to Computerized Sounds

Audio files (using Audacity 2.4.2) from the video recordings (sampling rate 44.1 kHz) and from every CHAT box (sampling rate 192 kHz) in the water during each encounter were extracted to determine what responses occurred during a CHAT session. Audio files were snipped when a match was noted. A total of 1319 min of video audio recordings (2013 - 213 min, 2014 - 560 min, 2015 - 350 min, 2016 - 196 min) were reviewed from spectrograms (max freq 22 kHz, window 1024, Hann). Each file was displayed as a spectrogram and scanned visually for the presence of any vocal whistle responses after playback of any CGS sounds played. A total of 468 GB of recordings were reviewed from spectrograms (max freq 96 kHz, window 2048, Hann) manually (2014 - 45 GB, 2015 - 404.4 GB, 2016 - 18.6 GB) to check for the occurrence of vocal responses to any CGS played.

Dolphin vocalizations that occurred close in time (within 5 s) after the CGS played were considered immediate responses to the CGS and were extracted as a short duration audio segment by the primary rater. If their contour (regardless of absolute frequency) visually resembled, either in whole or in part, one of the CGS played, they were extracted and analyzed. Contours that visually resembled a CGS within the audio file, outside the five-second criterion, were extracted and considered potential delayed imitation or independent productions and were also extracted. When the CGS overlapped a dolphin vocalization, these sounds were not measured, but were extracted in some cases where the contour directly overlapped the CGS.

All vocal responses were tallied by types. Signature whistle responses (recognized as a complete signature whistle, not as separate upsweep or downsweep) and stand-alone contour matches were considered responses, not immediate imitation, and were analyzed separately from contour imitations within the five-second window. Imitations within the five-second windows were labeled a Vocal Contour Match (*VCM*) and were put into three categories: Frequency variant match (*VCM1*), Time variant match (*VCM2*), and modulated click match (*VCM3*). *VCM*s from each category, both from the video audio and CHAT recordings, were then measured based on the CGS they visually resembled in contour. Because of the angular slope distortion of CHAT data (which changes with increased frequency), we attempted to measure the harmonic of the *VCM* in the same frequency band as the CGS to ascertain the degree of similarity of the *VCM* to the generated CGS.

Reliability of Primary Rater's Categorization of Dolphin Vocal Responses

In order to determine the reliability of the primary rater's categorization of dolphin vocal responses recorded by CHAT, a random sampling of 14 dolphin response vocalizations to CHAT sounds that were identified as facsimiles of CGS sounds by the primary rater (DH), were reviewed for facsimile identification by a second rater (CV), who was "blind" to the CHAT signals preceding the facsimiles (and thus not biased by that knowledge). The second rater was also presented with a random sampling of 16 stand-alone dolphin vocalizations (i.e., those that occurred outside the five-second window following a CHAT sound) for facsimile identification.

Effort, Duration, Group size, and Age Classes

Since different dolphins were exposed across years, and locations, we felt it worthy to look at the spread of exposure. During a four-year period, the CHAT system was used during encounters a total of 58 times (30 on LBB, 28 on GBB). In 58 encounters in four years, when interacting with the underwater interface, dolphins remained within approximately 10m of the CHAT Operators for an average of 37 min (M ± SD duration: 37.32 ± 32.29, n = 58) and were in groups of approximately seven dolphins (M ± SD dolphin: 6.79 ± 5.36, n = 58). On LBB the majority of dolphins were in one mixed age group (Adult, juvenile, calf) or solely juveniles, and on GBB the age class groups were more varied. In total, dolphins were exposed to the CHAT system for respectively 36.1 hours on LBB over four seasons and for 16.8 hr on GBB over three seasons.

Results

Exposure: Types of Sounds Played During CHAT Interactions

Dolphins across years were exposed to a total of 2,534 CGS signals through CHAT. Different CHAT signals had different frequencies of exposures across CHAT sessions (Figure 7). Of the object associated signals, Scarf and Sargassum were played most often, and of the human researcher associated signals, DHSW was played most often.

Despite no knowledge of the specific CGS preceding a dolphin vocal response, the second rater nonetheless identified the correct facsimile type in 100% of the cases. High inter-rater reliability (between the primary and secondary rater) (85%) also occurred when the second rater labeled a sampling of 16 stand-alone facsimile vocalizations. When the results for the two facsimile types were combined (N = 30), inter-rater reliability was 90%.

Reliability of Judgements of Matches Between Dolphin Vocalizations and CHAT CGSs
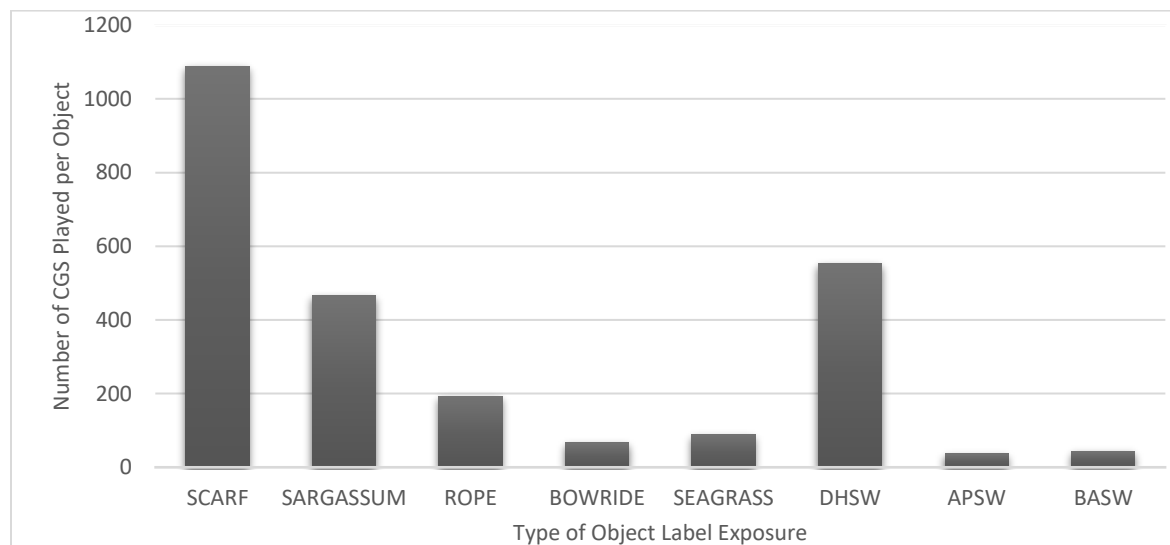
Types of Dolphin Vocal Responses to CGS

Dolphin vocal responses to CGS were of five types (Figure 8). Three of these were considered Vocal Contour Matches (*VCM*) which were CGS imitations extracted within the five-second window. None were complete imitations that replicated all acoustic parameters of a CGS. Instead, *VCM*s were partial imitations of a CGS. *VCMs* fell into three categories:

- **VCM TYPE 1** - *Flexible Frequency – VCM* whistle contour matches CGS in contour but varies in frequency (kHz). *VCM* may be anywhere in frequency space but is separated in time from CGS less than five seconds. *VCM* is usually a tonal whistle.
- **VCM TYPE 2** - *Temporal Match*. *VCM* whistle contour matches CGS in contour. *VCM* may start at a different frequency but is simultaneous, attached to CGS, or starts when CGS ends with no temporal delay after CGS.

- **VCM TYPE 3 -** *Frequency Modulated Contours*. *VCM* whistle contour matches are found within the frequency modulations of click trains. Banding (white banding spectrum) is caused by the production of the dolphin clicks by phasing out certain frequency areas of clicks, resulting in a contour.
- The remaining two types of responses to CGS were noted but not scored as CGS contour imitations due to their other nature and/or existence outside the five-second response time criterion. These two remaining types included:
  - *Stand-alone contours* that resembled a CGS were scored separately as they occurred anywhere in the sequence separated from CGS itself more than five seconds. Stand-alone contours were either an upsweep or a down-sweep. Because they occurred outside of the five-second window, following a CGS, they were considered a form of delayed imitation.
  - *Dolphin Signature Whistles* often occurred immediately after, or overlapped with a CGS, and were not considered an imitation. They were scored simply as a potential response to a CGS.

\
Figure 7

Dolphins' Exposure to Computer-Generated Sounds

Note. This figure shows the dolphins' exposure to sounds played through the CHAT system over four years.

Figure 8

Dolphin Vocal Response Type Examples

VCM TYPE 1

VCM TYPE 2

VCM TYPE 3

STAND ALONE

SIGNATURE WHISTLE
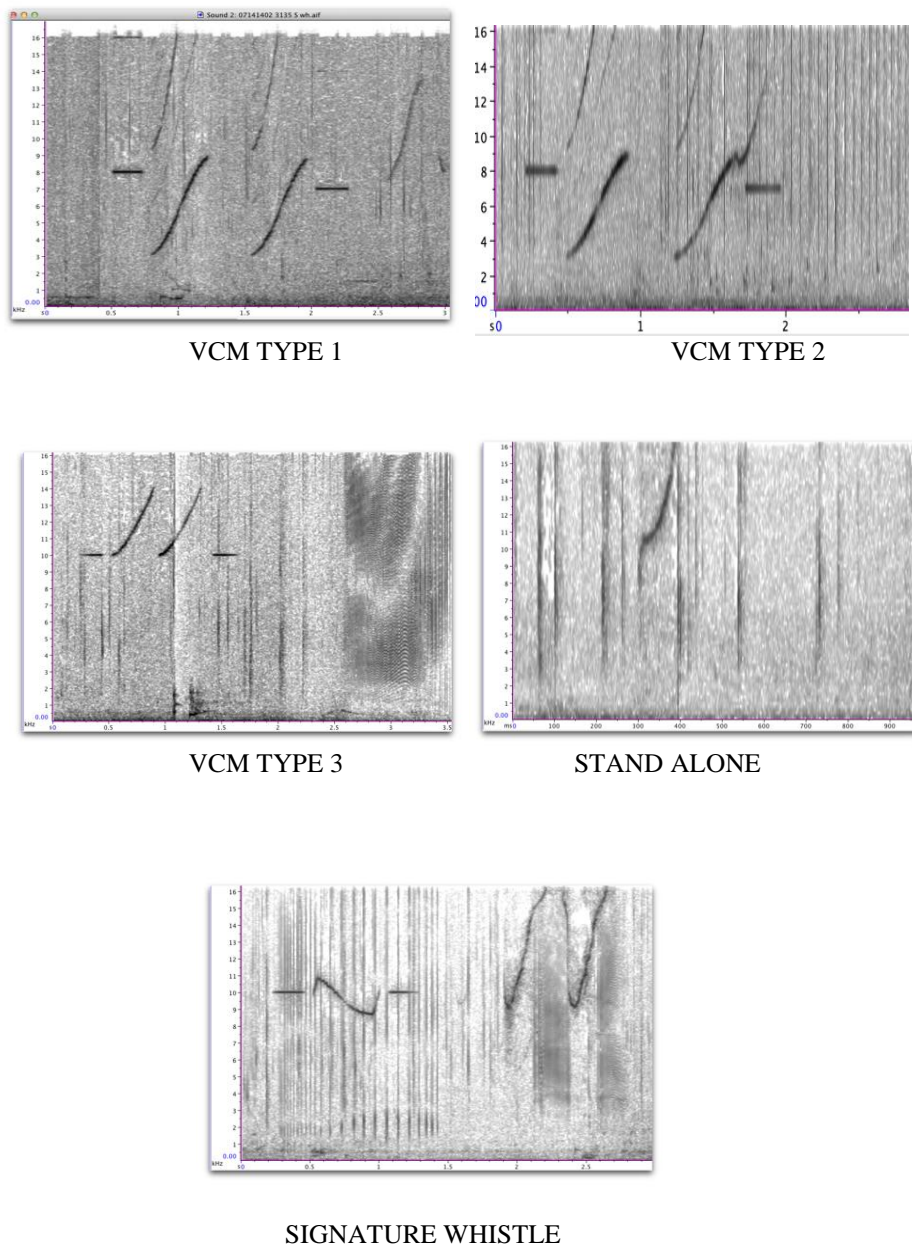
Note. This figure shows spectrograms of five different types of responses made by dolphins after a CGS. Boxes represent the CGS played. The arrows point to the dolphin response. The stand-alone example resembles a CGS sargassum upsweep.

A Detailed Description of the VCM3

Although banding patterns, as an artifact of high repetition rates of click trains, is well- known (Watkins, 1968) and has been observed frequently in sound measurements of this population, we observed another type of banding pattern, the *VCM3*. The *VCM3* type sometimes appeared during our CHAT

sessions. This type of banding pattern can be an artifact, observed in shallow water when dolphin echolocation clicks bounce off the surface or the bottom during multipath projection (Aubauer et al., 2000). This unusual contour banding pattern can also potentially be created by a dolphin producing and manipulating its two separate sound generators for clicks (Cranford et al., 2011).

To test whether the *VCM3* was an artifact of the habitat or equipment reflection vs. a dolphin-created contour within a click train, we examined the waveforms from the CHAT data signals when such banding occurred in two ways. First, *VCM3* waveforms were visually reviewed to determine the presence of pulses. If an inverted pulse occurred after the initial dolphin click pulse (potentially a surface reflection and the effect of multipath reflections (Aubauer et al., 2000), we then attenuated the inversion throughout the wav file whenever it occurred. After attenuation was complete, we reproduced an FFT to see if banding still occurred. If banding was still present, the signal was determined to be a dolphin-generated signal, not an artifact in the environment.

Second, the spacing between two closely spaced clicks, collected simultaneously from the two different hydrophones on the CHAT box, was measured to determine if there was a consistent spacing between these clicks, suggesting that the clicks were produced by a dolphin with two sound generators separated by a consistent distance in the head, or if these clicks varied in spacing, suggesting that the signal changed with geometry or angle to the surface, which would confirm that the effect was from multipath reflection. After these known artifacts and confounds were removed, the *VCM3* signals were considered to be produced by a dolphin. Given the contextual nature of these *VCM3* signals, and their appearance in close proximity (< 5 s) with the CGS playback, we suggest another possibility; that the dolphins are able to manipulate their clicks to create a reduced amplitude banding pattern in their click trains, in this case resembling specific contours.

It should be noted that these two processes do not need to be mutually exclusive; the shallow water habitat could also produce contour banding via multipath propagation and the dolphin may also be able to control aspects of their two pulse generators. Therefore, we only extracted *VCM3*'s that fit the criterion of being produced within 5 s of a CGS, and matched the contour of the CGS. These *VCM3*s were extracted and scored, since we considered these to be real signals produced by the dolphin in response to a CGS.

Dolphin Vocal Imitation Attempts

A breakdown of all potential instances of vocal responses, classified by five types, in video audio recordings within the five-second window can be seen in Figure 9 and for CHAT data recordings in Figure 10.

Recordings from Video Cameras

A total of 243 vocal responses were extracted and measured from 44 encounters (from a total of 58 encounters) using CHAT. Signature whistles and stand-alone responses accounted for 70% of the responses (Figure 9). Of the vocal response types, *VCM3* and *VCM1* made up 28% of the responses.

Recordings (up to 96kHz)- from CHAT, 2014-2016

A total of 56 responses were extracted and measured CHAT recordings from 15 encounters (from a total of 17 encounters with CHAT data available) using CHAT. Of the vocal response types, *VCM3* made up over 70% of the responses (Figure 10).

Figure 9

*Vocal Responses by Dolphins Found in Video Audio Recordings Following CGS*
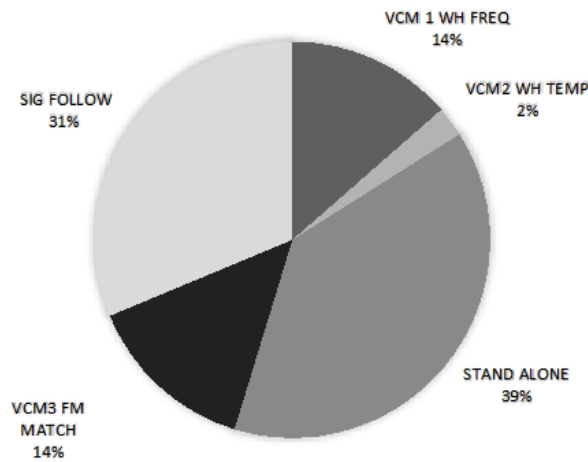
Note. Sound types following a CGS are as follows: SIG follow represents a signature whistle following CGS, VCM 1 Whistle Frequency represent a Type 1 Frequency match, VCM 2 Whistle Temporal represents a Type 2 Whistle match in time, Stand Alone represents a singular upsweep or downsweep, and VCM 3 FM Match represents a click-modulated contour pattern.

Figure 10

Vocal Responses by Dolphins Found in the CHAT Recordings Following CGS
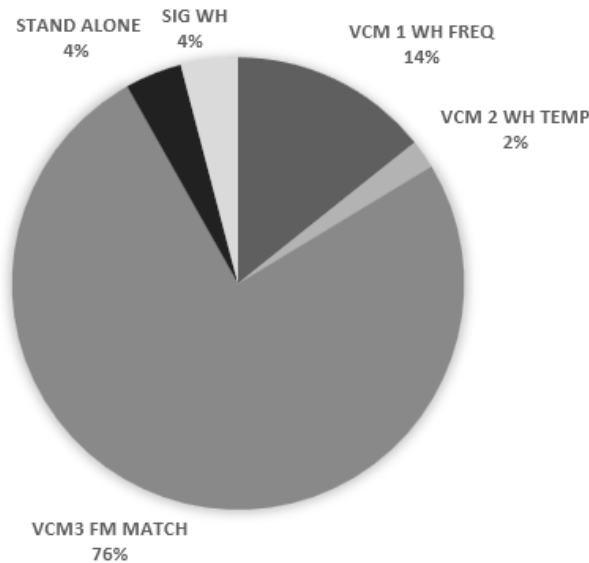
*Note.* Sound types following a CGS are as follows: SIG WH represents a signature whistle following, VCM 1 WH FREQ represent a Type 1 Frequency match, VCM 2 WH TEMP represents a Type 2 Whistle match in time, VCM 3 FM Match represents a click-modulated contour pattern, Stand Alone represents a singular upsweep or downsweep match.

Types of Responses – Dolphin Vocalizations Following

After eliminating signature whistle (SW) responses, and Stand-Alone (SA) responses, dolphin imitations were labeled as Vocal Contour Matches (VCM) and fell into the three main categories by type. This breakdown of Vocal Contour Matches (VCMs) can be seen in Figures 11 and 12. Responses were made up mostly from Whistle Frequency types (VCM1) and Frequency-Modulated types (VCM3) as shown in Figure 11. HF responses were VCM3 by a very high percentage (Figure 12).

Figure 11

Number and Type of VCM Responses to CGS from Video Audio Recordings
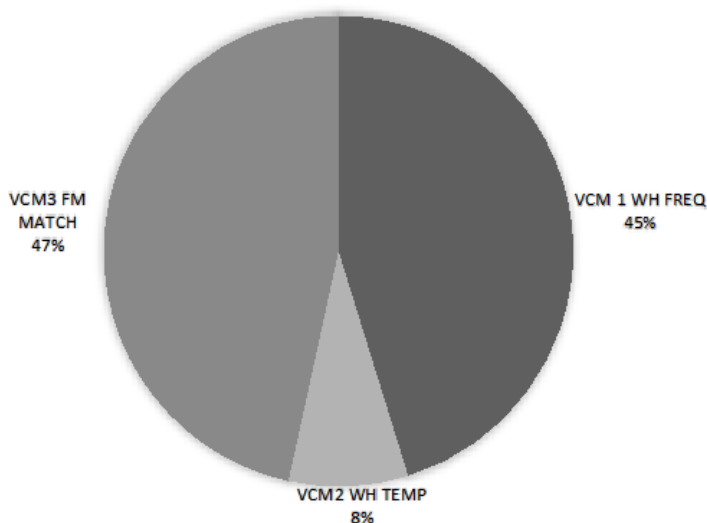
Note. Sound types following a CGS are as follows: VCM 1 WH FREQ represent a Type 1 Frequency match, VCM 2 WH TEMP represents a Type 2 Whistle match in time, VCM 3 FM Match represents a click-modulated contour pattern.

Figure 12

Number and Type of VCM Responses to CGS from CHAT Recordings
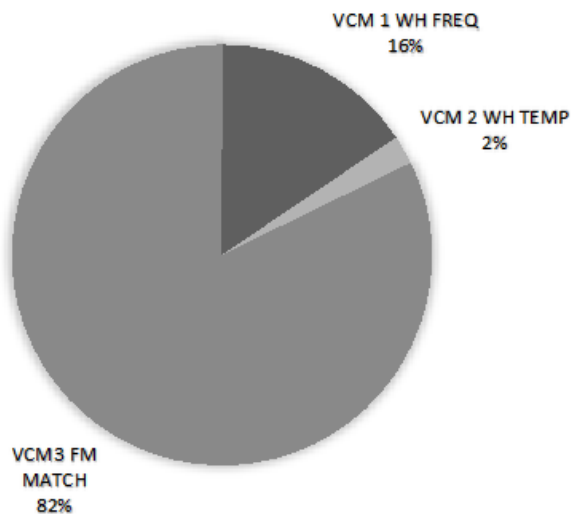
Note. Sound types following a CGS are as follows: VCM 1 WH FREQ represent a Type 1 Frequency match, VCM 2 WH TEMP represents a Type 2 Whistle match in time, VCM 3 FM Match represents a click-modulated contour pattern.

Signal Exposure and Signals Most Often Imitated

The degree of exposure to a CGS and VCM responses are seen in Figure 13. Although *Scarf* was played in the water more than any other CGS, *Sargassum* was responded to more frequently. DHSW was also the second most frequently played CGS but had a relatively low response rate. Responses to objects varied between the video audio and CHAT data responses (Figure 14 and 15).
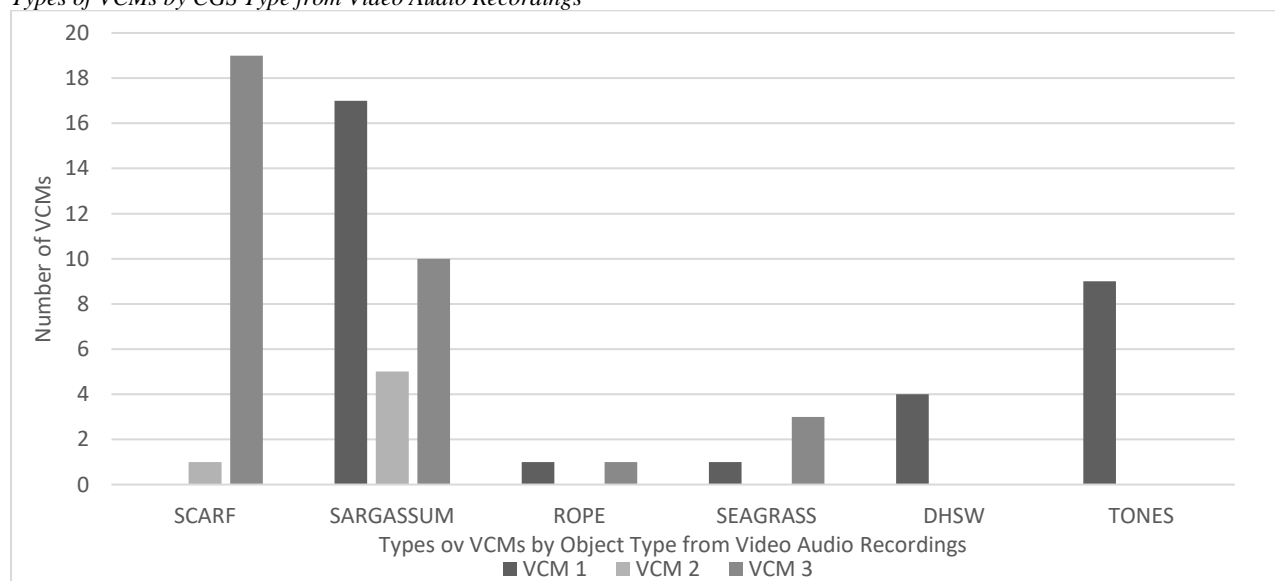
Figure 13

Total Exposure to CGS and VCMs Produced by Object Type

Object Type CGS and VCMs produced
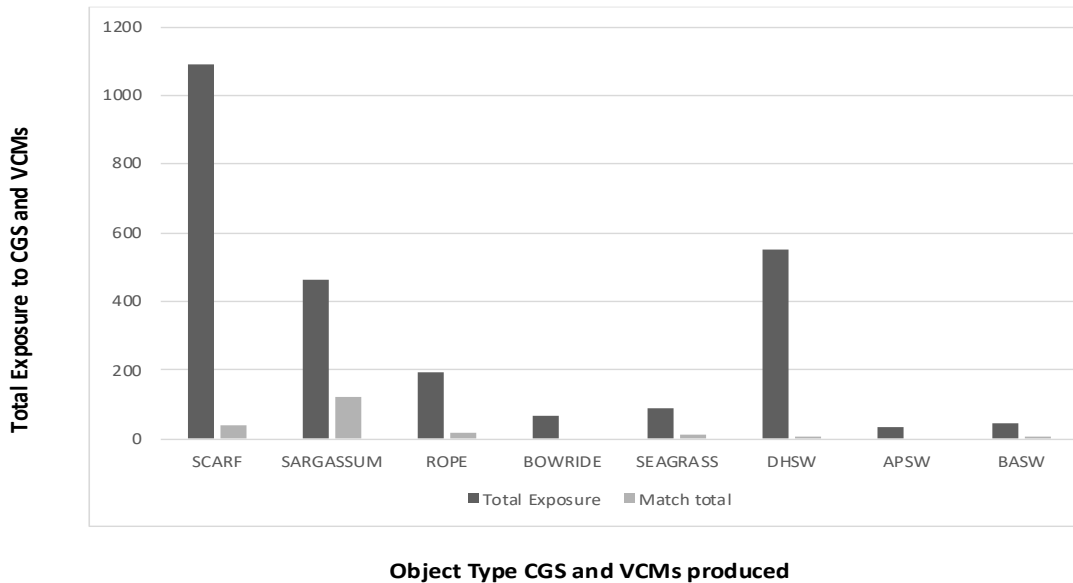
Note. This figure represents a comparative graph of rates of response relative to exposure rates. (# of times imitated verses # of exposures to CGS type).

Figure 14

Types of VCMs by CGS Type from Video Audio Recordings

Note. Response types, Type 1 and 3, were frequently produced for *Sargassum.*

Figure 15

Types of VCMs by Object Type from CHAT recordings.
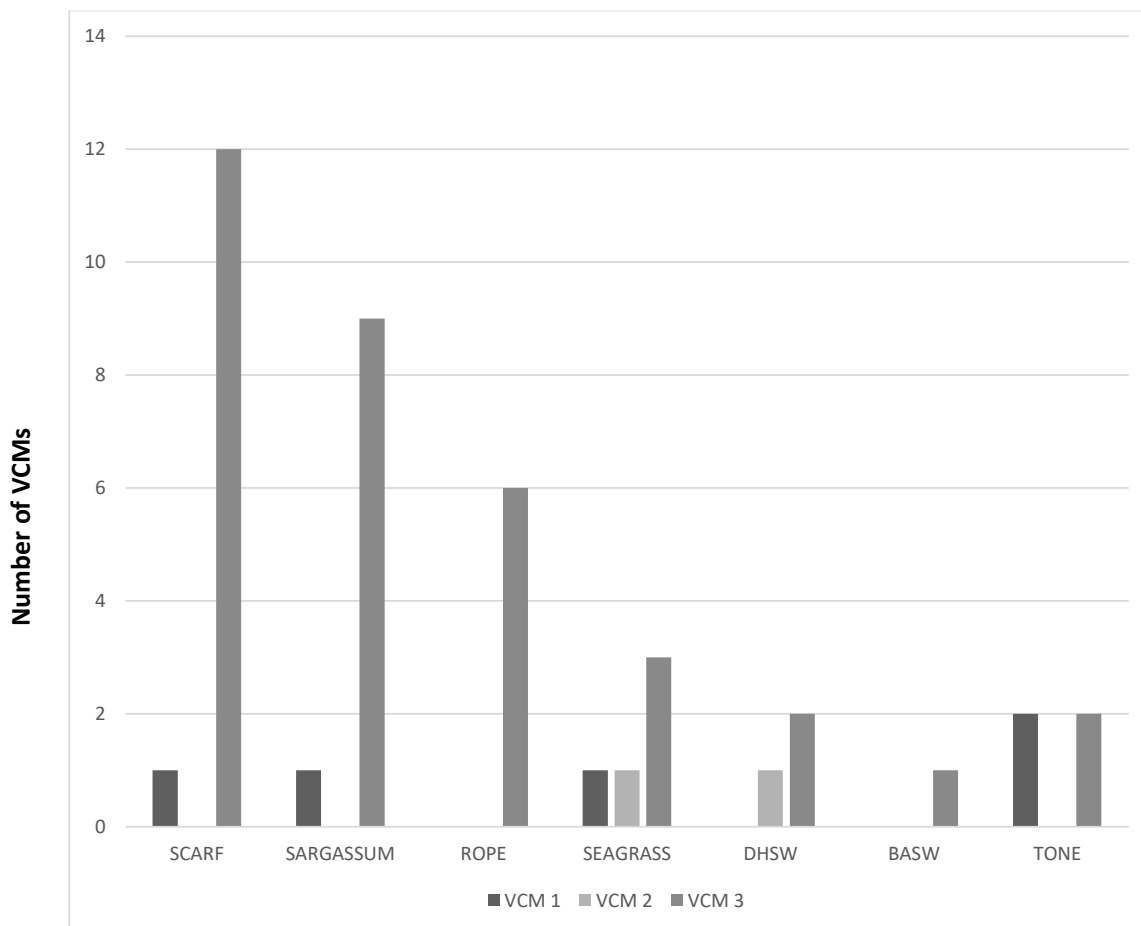
Types of VCMs by Object Type from CHAT recordings

Note. VCM3 Frequency Modulated Contour responses were most frequently produced for all objects including Scarf, Sargassum, Seagrass and Rope and for DHSW.

There were a variety of ways imitations could match a CGS. Figure 16 shows examples of various types of responses, some of which were considered matches.

Figure 16

Imitation Types

CGS SCARF

CGS SCARF/DOLPHIN MIMIC OVERLAP

CGS SARGASSUM

CGS SARGASSUM   DOLPHIN MIMIC

CGS ROPE

CGS ROPE   DOLPHIN MIMIC

CGS SEAGRASS

CGS SEAGRASS   DOLPHIN MIMIC

CGS DHSW

CGS DHSW   DOLPHIN MIMIC
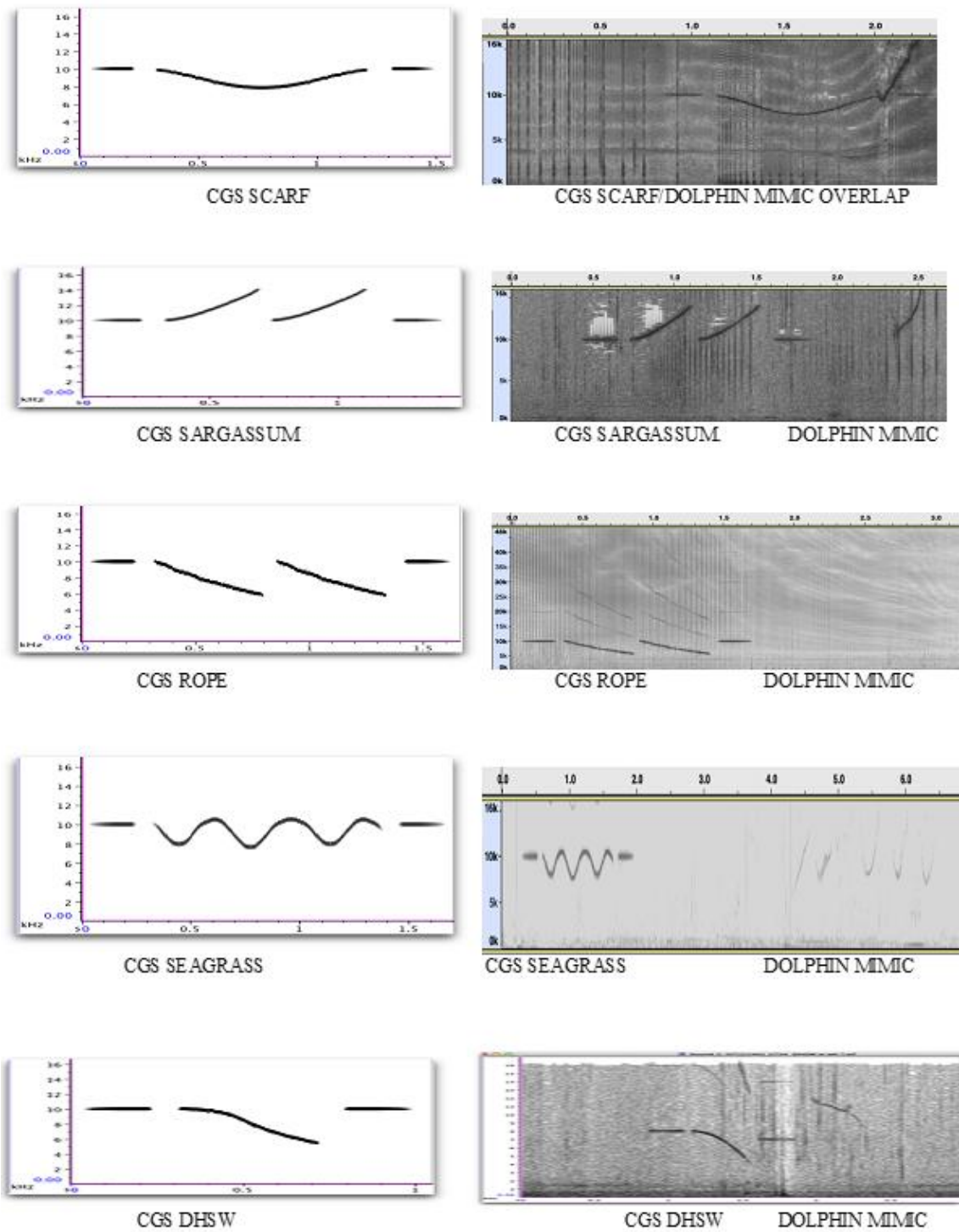
Note. Figure 16 shows examples of some of the responses scored by type. Boxes represent the CGS. Arrows point to the matches. The Scarf example has an overlapping response (technically not scored) but shows the VCM3 type signal. The white banding (within the box) over the CGS scarf is the mimic itself, not the check mark which is not a mimic. The Sargassum example shows an upswept contour within 5 sec. The Rope example shows the VCM3 white banding type after the CGS is produced. The seagrass example shows a multi-loop response. The DHSW shows a mimic of the CGS contour without the start or stop tones.

Accuracy of Imitations

Six of the seven CGS sounds to which the dolphins were exposed evidenced the production of one or more dolphin "response" vocalizations within five seconds of the CGS being broadcast (APSW was never responded to with a dolphin vocal response within five seconds). Table 1 is a response (error) matrix showing the compilation of these dolphin vocal responses. Each row represents all the dolphin responses to a different CGS sound with the shaded cell representing matches of that sound and unshaded cells representing responses that were judged as non-matches and fell into other categories. After taking out stand-alones which were outside the five-second window (94 from 243 video audio responses, and 15 stand alones from the CHAT audio) a total of 191 imitations were noted. In total, of the 191 presentations of CGS sounds representing objects (SA, SC, RP and GR) and agents (DHSW and BASW), a correct dolphin match to that CGS sound was produced 114 times (59.69%) and a vocal response other than a match to that sound was produced 77 times. The latter responses included seven matches to the start/end tones accompanying the production of each CGS, 67 non-match productions of signature whistles, and three non-matched productions of other CGS sounds.

If one conservatively considers only two options for a dolphin when hearing a CGS (i.e., produce a match versus produce another non-matched vocalization) (i.e., chance = 0.5), the overall results indicate that in response to a CGS, matched and non-matched vocalizations were not produced at equivalent frequencies. That is, dolphins were significantly more likely to produce a matched sound (of either the specific CGS representing an object or agent or start/end tone) (N = 121) than another non-matched vocalization (N = 70), in response to a CGS (x^2 (1) = 13.62, p = .00022). Furthermore, when producing a match, they were more likely to produce a match of the CGS sound representing an object or agent (N = 70) rather than the start/end tone accompanying the CGS sound (N = 7) (x^2 (1) = 51.55, p < .0001). In addition to these overall results, we compared the dolphins' responses to a CGS representing an object versus an agent. Again, conservatively considering only two options for a dolphin (match the CGS being presented versus produce another vocalization including either a non-match or a match of the start/end tone), the results show that whether a dolphin imitated or did not imitate a CGS was dependent on whether the CGS represented an object or an agent (x^2 (1) = 11.51, p = .000691) with imitation of the CGS occurring in 65% of the instances of the former compared to 33% of the latter. Finally, Table 1 shows that the dolphins produced imitations of each of the six types of CGS sounds. Also, in response to a CGS, they produced signature whistles and imitations of the computer-generated tone sounds accompanying each CGS. Thus, there were in fact eight possible vocal production options that a dolphin could make in response to the playing of CGS. Taking these options into account (i.e., chance = 1/8 or p = .125), the cumulative binomial test reveals that each CGS was imitated at above chance levels (p < .05 or smaller). Aside from CGS imitation, the dolphins also imitated the pure tones placed prior to and following a CGS tone, revealing the flexibility of imitation abilities beyond frequency contours.

Table 1

Response Matrix Showing the Dolphin Vocal Responses Within Five Seconds of Each CGS

		Dolphin Vocal Responses with 5 s of CGS						Other: Sig. Whistle	Other: Non-Sig. Whistle Tone mimics	Total
		SA	SC	RP	GR	DHSW	BASW			
CGS	SA	43	-	-	-	-	-	12	1	56
	SC	2	37	-	1	-	-	31	3	74
	RP	-	-	13	-	-	-	3	-	16
	GRR	-	-	-	10	-	-	1	1	12
	DHSW	-	-	-	-	8	-	20	2	30
	BASW	-	-	-	-	-	3	-	-	3
	Total	45	37	13	11	8	3	67	7	191

Note. This figure shows a response matrix showing the vocal responses from the dolphins within 5 seconds after each CGS. SA=sargassum, SC=scarf, RP=rope, GR=seagrass, DHSW = Denise signature whistle, BASW= Bethany signature whistle. Shaded cells indicate a correct match of a CGS. With 8 potential responses, chance = 1/8 or p = 0.125. This graph includes both video audio recording matches and CHAT recording matches, not including standalones.

Dolphin Vocal Production Prior to a CGS

In five cases, the dolphins showed a vocal response that preceded the playing of a CGS. In two of these cases, the response was a signature whistle that preceded the CGS for sargassum and for DHSW. The other three responses included a sargassum facsimile preceding the CGS for sargassum, a DHSW facsimile preceding the CGS for DHSW, and a scarf facsimile preceding the CGS for scarf.

How Many Dolphins Imitated a CGS?

Although the technology did not allow us to determine which specific dolphins in CHAT sessions were producing imitations to a CGS, by documenting the individual identifications of all dolphins in each CHAT session, through the process of elimination we are able to determine that the vocal imitation ability was demonstrated by at least seven Atlantic spotted dolphins. On LBB, of 16 CHAT sessions in which imitation of one or more CGS was documented, 13 involved the same 4 dolphins and two of these were also found one each in the remaining two sessions. On GBB, of the six CHAT sessions in which imitation of one or more CGS was documented, three involved different dolphins, and the three remaining sessions contained one of the three dolphins. Thus, conservatively between the two areas, at least seven dolphins showed this capability.

Discussion

The findings of this study provide the first preliminary evidence that human-habituated wild Atlantic spotted dolphins that are exposed to computer-generated underwater sounds in the presence of humans will vocally imitate these sounds. Six of seven different frequency modulated CGS sounds to which the dolphins were exposed were partially vocally imitated three or more times within five seconds of the CGS sound being produced and at levels greater than what would be expected by chance (Table 1). These included CGS sounds associated by humans with four different objects that the dolphins showed interest in playing with and CGS sounds associated with two human agents. The dolphins also imitated "start" and "stop" computer-generated pure tones that directly preceded and followed CGS sounds, thus demonstrating a flexibility of vocal imitation beyond frequency modulated computer-generated sounds. This finding of

imitation of a computer-generated sound beyond those specifically intended for imitation was similar to what Richards et al. (1984) reported in their study of vocal imitation in the bottlenose dolphin Akeakamai: "Of particular interest for the analysis of the extent and flexibility of mimicry capability is the following serendipitous occurrence. Owing to the way the sounds are produced, broad-band clicks are occasionally produced at the onset and offset of a model… [and] in addition to the pure- tone quality of the sound, Akeakamai also occasionally mimicked the onset click and swept her frequency down to the frequency of the model. This mimicry was not detected until later, off-line spectrum analysis, and it is further confirmation of the generality of the mimicry concept in the absence of specific training" (p. 17).

The context within which Atlantic spotted dolphin vocal imitations occurred largely involved a model/rival paradigm in which two humans demonstrated in the presence of one or more dolphins that each of the four objects could be referred to with a unique CGS as well as through occasional accompanying human pointing (at the object). The goal of this paradigm (modeled loosely after Pepperberg's studies of vocal imitation and labeling with an African grey parrot, e.g., Pepperberg, 1981), was to show the dolphins socially that these CGSs could have a referring function extending from commenting on an object to requesting that object. Aside from the association with toy objects that the dolphins historically showed interest in playing with (Herzing et al., 2012), making the CGSs more salient to the dolphins as was shown in captivity with bottlenose dolphins by Reiss and McCowan (1993) and Hooper et al. (2006), if the dolphins were able to appreciate the referring function of the CGS sounds as demonstrated by the humans, it was hypothesized that they would spontaneously begin to produce facsimiles to those sounds to gain access to the objects. Although in the present study, it is unclear whether the Atlantic spotted dolphins ultimately understood this referring function (cf. Herman et al., 1993; Savage-Rumbaugh, 1993), they not only vocally imitated CGSs within five seconds of their production, but also produced these signals outside of this time frame including prior to any CGS production, when the associated objects were simply visibly in the possession of a human.

This was a first attempt to examine vocal imitation of CGS using sounds that we believed might be within the dolphins' capability. Whether the dolphins can also vocally imitate more complex signals like some of those used in Richards et al. (1984) is the subject of a future study.

Degree of Fidelity of Atlantic Spotted Dolphin Vocal Imitation

A common feature of the studies of Pepperberg (1981), Reiss and McCowan (1993), and Hooper et al. (2006) is that vocal imitation, like occurs in humans first acquiring language (e.g., Bloom, 1970; Bloom et al., 1974; Kuczaj, 1982, 1987) often begins with partial imitations.

Our finding that all of the mimics of CGS sounds by the dolphins were partial imitations, is similar to several previous studies investigating vocal imitation, vocal labeling and stimulus matching by captive and wild bottlenose dolphins. With dolphins in captivity, Reiss and McCowan (1993) described a gradual progression of immediate vocal imitation of computer-generated whistles associated with various objects from partial to complete vocal facsimile reproductions. For example, partial vocal productions began after 19 exposures to the model sound for a toy ball with initial attempts by the dolphin focusing on reproducing the end of the fundamental frequency of the whistle followed by attempts at reproducing the beginning of the fundamental frequency of the whistle, then focusing on the harmonic structure of the whistle, and finally experiencing success at matching the whistle model fundamental in its entirety including its frequency modulation. Early partial reproductions also showed variation in matching various acoustic and temporal characteristics of the model sound. Early partial reproductions of the end of the model sound revealed success in matching temporal parameters, absolute frequency and frequency modulation. Dolphin attempts to reproduce the beginning of the model sound replicated the temporal parameter, but only the relative rather than the absolute frequency modulation (i.e., the frequency was shifted downwards). A similar initial focus on reproducing the end portion of a sound was witnessed with the "rub" model sound. This was followed by a full, but compressed, reproduction of the model, which was then followed by the dolphin focusing on the beginning of the sound. These examples appear to reveal the flexibility of the dolphin in focusing its attention on different portions of a sound to be vocally imitated as well as the sound in its

entirety, with perhaps initial attention of the ending portion of the sound reflecting a recency effect in memory (i.e., the most recently experienced part of the sound is remembered best, see Thompson and Herman, 1977). Overall, Reiss and McCowan (1993) showed that their bottlenose dolphins were able to spontaneously produce facsimiles to model whistle sounds that typically maintained their duration, relative frequency, and frequency modulation but often differed in fundamental frequency.

Similarly, in a vocal matching study with bottlenose dolphins both in captivity and in the wild, King et al. (2013) showed that dolphins who copied the signature whistles from either their mother or calf, close associate, or alliance member in male-male coalition closely matched the frequency modulation pattern of the model whistle but added variability that differed from the model in one or more other parameters. Parameters of the model signature whistle that were varied by the copier included: mean frequency, maximum frequency, starting frequency, ending frequency, frequency range, and number of loops. The copies also differed significantly in some parameters from the copier's own signature whistle.

Because of its functional limitations, the CHAT system did not allow us to localize the individual Atlantic spotted dolphin initiating a vocal imitation attempt or stand-alone production of a CGS**,** thus preventing us from tracing the development of an individual dolphin's vocal imitation ability. Because the CHAT boxes never reported a dolphin imitation or stand-alone production of a CHAT sound in real time, the dolphin producing that sound was unlikely to receive human feedback (unless it also pointed at the object, in which case it would have received it from the human). Thus, dolphin partial imitations could have been a case of vocal contagion, or productions because of an intrinsically motivating curiosity associated with the perception of the CHAT sounds in a highly social context.

Nonetheless, our findings demonstrate a variety of different types of partial vocal matching attempts that maintained with varying fidelity various acoustic and temporal parameters of the model sounds. The replication of various sound characteristics was not equivalent within or across the different types of sounds. Results indicated that dolphins did not preferentially vocally imitate the most frequently played computer sounds and rates of vocal imitation for most object signals were approximately 10% of the signal exposure time.

Unanticipated Reactions to the System – VCM3 and What It Means

Although the dolphins attempted partial imitations in a variety of ways, the *VCM3* type was unusual and represents an interesting finding of the work. Although the *VCM3* signal type can be a repercussion of multi-path projections of a click train, we tested *VCM3 s*ignals found in close proximity to CGS for this feature. It appears that the dolphins were creating a whistle contour utilizing click interactions that demonstrates their understanding of the equivalence of different types of signals and flexibility in vocal imitation, and potentially demonstrates their preferred range and production methods for imitation (see Richards et al., 1984 and Ralston & Herman, 1995 for evidence of the bottlenose dolphin's understanding of the equivalencies of octave generalization of vocally imitated sounds, and relative frequency changes of sounds, respectively). These phenomena need not be mutually exclusive. That is, the dolphins could be intentionally manipulating their clicks to create *VCM3*s and *VCM3* could also be a result of multi-path propagation. Lammers and Castellote (2009) demonstrated that beluga whales can use both their pulse generators to create one pulse, modulate the amplitude of each pulse while possibly manipulating the spectral content of the click, allowing contribution of energy at different frequencies. Contour shapes matching a CGS played within seconds of these *VCM3*s, suggests the dolphins were actively engaged in its production in an attempt to imitate.

Past studies report on bottlenose dolphins adjusting their vocal imitations up an octave from a CGS model as a strategy for mimicking, and also recognizing as equivalent the same sequences of sounds shifted relative to each other in frequencies (Richards et al., 1984; Ralston and Herman, 1995). Here, it appears that they also move up in frequency, although the mimicked sound contour is created utilizing click train amplitude adjustments instead of a tonal mimic.

Overall, the Atlantic spotted dolphins' vocal imitations and productions of CGS sounds occurred spontaneously without humans employing explicit operant shaping through direct reinforcement of vocal

approximations to the CGS sounds. Indeed, because the CHAT system virtually never triggered a dolphin vocal imitation to a CGS in the water real time (due to the challenges of real-time vocal matching in the water as noted earlier), the CHAT operators were blind to vocal mimics until the spectrograms were visually reviewed after in-water sessions.

Dolphin vocal reproductions that occurred immediately before CGSs

Occasionally the dolphins produced a *VCM* (usually VCM3s) ahead of the researcher playing the CGS, although these events were in a larger context of CGS use over minutes of interaction. The potential anticipation of the dolphins of computer-generated sounds associated with consequences (i.e., objects or actions) has been noted in other interface studies. Reiss and McCowan (1993) defined a dolphin vocal facsimile of their computer-generated sounds as a *production* if it was not immediately produced after a computer-generated sound. They also reported occasional productions that preceded key hits, before the dolphins approached their keyboard apparatus, suggesting that the dolphins may have been anticipating various sounds being played, as has been described in a few other species (Spruijt et al., 2001). In our study, it is unclear if the dolphins were perhaps anticipating a human operator playing a CGS sound, or viewed an available object in the human's possession, and either labeled that object or produced the facsimile to request the object.

The relation of immediate and delayed vocal mimicry

Human infants, parrots, and bottlenose dolphins, as cited earlier, have each demonstrated an aptitude for both immediate vocal imitation (i.e., producing a facsimile to a model word or non-word sound within a few seconds of hearing that sound) and delayed vocal imitation (i.e., producing a facsimile to a model word or non-word sound long after experiencing the model sound). Our findings indicate that Atlantic spotted dolphins are also capable of both immediate mimicry and delayed (i.e., what we termed "stand alone") vocal imitation. While immediate vocal imitation was likely to occur using either video audio or CHAT audio partial matches, delayed vocal imitation was biased towards audio from the video partial matches. Stand-alone facsimiles were found outside the 5 sec CGS response time criterion and were prevalent in the video audio data and may represent instances of delayed imitation as has been shown in captive dolphin studies (Reiss & McCowan, 1993, Richards et al., 1984, Kremers et al., 2011). For other species, delayed imitation has been proposed to have functional benefits in the wild for example in the case where potential prey bird species copy the alarm calls of benign species to deceive predators (Igic et al., 2015).

The relation of the present study to vocal production and contextual learning

Vocal production learning and contextual learning, both of which influence vocal communication, form a cornerstone of this study. Whereas vocal production learning influences the generation of calls (i.e., an individual learns to modify a sound they produce including in novel or modified ways as a consequence of their experience with others producing similar sounds), contextual vocal learning influences the usage and comprehension of sounds (i.e., an individual learns appropriate contexts to produce specific calls, and learns to recognize specific calls and how to react to them, Janik, 2014). Vocal production learning is relatively rare compared with contextual vocal learning across animal species (Janik & Slater, 1997, 2000; Janik, 2014). In cetaceans, contextual vocal learning has been demonstrated in bottlenose dolphins (Hooper et al., 2006; Lilly, 1965; Richards et al., 1984; Reiss & McCowan, 1993) and beluga whales (Murayama et al., 2012). Vocal production learning has been demonstrated in several species of songbird (suborder *Passeri*) parrots, hummingbirds (family *Trochilidae*), and in non-human mammals including bats (order *Chiroptera*), pinnipeds (suborder *Pinnipedia*), elephants (order *Proboscidea*), and cetaceans (order *Cetacea*) and includes in some cases not only vocal learning influenced by conspecifics but non-conspecifics including electronically-generated non-natural sounds (see reviews in Briefer & McElligott, 2012; Janik and Slater, 1997, 2000; Tyack, 2008). As noted earlier, multiple studies have provided strong

evidence of vocal production learning in the bottlenose dolphin (reviewed in Janik and Sayigh, 2013). Evidence also indicates vocal production learning in humpback whales *Megaptera novaeangliae*, and perhaps in other Mysticete species (Janik, 2014), as well as in several delphinid species beyond bottlenose dolphins such as killer whales (*Orcinus orca*) (Crance et al., 2014; Foote et al., 2006), Risso's dolphins (*Grampus griseus*) (Favaro et al., 2016), and Beluga whales (Panova & Agafonov, 2017). Our study provides evidence of contextual vocal learning in at least seven wild Atlantic spotted dolphins. It also indicates that Atlantic spotted dolphins can engage in "production imitation," (i.e. copying behaviors – in this case CGSs including pure-tone sounds - that are beyond its natural behavioral repertoire including those that are novel (Byrne, 2002; Byrne & Russon, 1998). The present work suggests that imitation by Atlantic spotted dolphins has some conceptual aspects (in contrast to simply being a result of a response-facilitation mechanism that automatically activates an imitation program, see Heyes, 2011), allowing for control and flexibility in the imitative process (see also Jaakkola et al., 2013). And because the start and stop tones were imitated several times, this supports the case for imitation of sounds beyond upsweeps and downsweeps, as the tones were pure tone frequency, not modulated.

Finally, the flexibility demonstrated in vocal imitation by Atlantic spotted dolphins is likely an extension of a more generalized ability for copying observed motor behaviors and learned signature whistles of family members and close associates, as has been demonstrated in spotted dolphins (Bebus & Herzing, 2015) as well as bottlenose dolphins (Tyack, 2008). The pressures selecting for such broad copying capabilities are likely associated with the extensive long-term social networks and cooperative behaviors that have evolved in both species.

Future Directions

Lessons learned from the current study and our previous pilot work (Herzing et al., 2012), suggest that wild Atlantic spotted dolphins show interest and flexible vocal imitation strategies when presented with artificial acoustic signals in an interspecific play situation. As noted earlier, future studies should expand on the variety of novel non-natural sounds dolphins are exposed to, such as those used by Richards et al. (1984) to investigate the extent of generalization of a concept of imitation. One limitation of the present CHAT system is its apparent limitation in recognizing and reporting in real-time dolphin attempts at producing vocal facsimiles to CGSs as well as identifying the location of the vocalizing dolphin. This in turn limits the ability of human CHAT operators to respond in real time to dolphin vocal productions in ways that can scaffold the dolphin in understanding the referring function of CGS symbols in social exchanges (cf. Rumbaugh et al., 1986, 1993). Additionally, extensive and repeated individual exposure may be critical to the dolphin's understanding the referring function use of a two-way communication system, both of which are difficult to obtain in open ocean situations. Improved hardware and software may assist in solving the former challenge, and if attractive to the dolphins, may help ameliorate the latter challenge.

Conflict of Interest: We have no known conflict of interest to disclose.

Data Availability: Data can be made available upon special request to corresponding author.

References

Igic, B., McLachlan, J., Lehtinen, I., & Magrath, R. D. (2015). Crying wolf to a predator: deceptive vocal mimicry by a bird protecting young. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1809), 20150798.

Jaakkola, K., Guarino, E., & Rodriguez, M. (2010). Blindfolded imitation in a bottlenose dolphin (*Tursiops truncatus*). *International Journal of Comparative Psychology*, *23*, 671-688.

Jaakkola, K., Guarino, E., Rodriguez, M., & Hecksher, J. (2013). Switching strategies: a dolphin's use of passive and active acoustics to imitate motor actions. *Animal Cognition, 16*, 701-709.

Janik, V., & Slater, P. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, *26*, 59-99.

Janik, V., & Slater, P. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*, 1-11.

Janik, V. M. (2014). Vocal learning and communication. *Current Opinion in Neurobiology*, *28*, 60-65.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, *289*, 1355-1357.

Janik, V. M., & Sayigh, L. S., (2013). Communication in bottlenose dolphins: 50 years of signature whistle research. *Journal of Comparative Physiology A, 199*, 479-489.

Janik, V. M., & Sayigh, L. S., & Well, R. S. (2006). Signature whistle shape conveys identity information to bottlenose dolphins. *Proceedings of the National Academy of Sciences, 103(21),* 8293-8297.

King, S. L., & Janik, V. M. (2013). Bottlenose dolphins can use learned vocal labels to address each other. *Proceedings of the National Academy of Sciences, 32*, 13216-13221.

King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W., & Janik, V. M. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B, 280*, 20130053.

Kohlsdorf, D., Gilliland, S., Presti, P., Starner, T., & Herzing, D. L. (2013). An underwater wearable computer for two-way human-dolphin communication experimentation. *Proceedings of the 2013 International Symposium on Wearable Computers* (pp. 147-148).

Kohlsdorf, D., Herzing, D., & Starner, T. (2016). Feature Learning and Automatic Segmentation for Dolphin Communication Analysis. In *INTERSPEECH* (pp. 2621-2625).

Kremers, D., Jaramillo, M. B., Böye, M., Lemasson, A., & Hausberger, M. (2011). Do dolphins rehearse show-stimuli when at rest? Delayed matching of auditory memory. *Frontiers in Psychology*, *2*, 386.

Kuczaj, S. A. II. (1982). Language play and language acquisition. In H. W. Reese (Ed.), *Advances in child development and behavior, 17* (pp. 197-232). New York: Academic Press.

Kuczaj, S. A. II. (1987). Deferred imitation and the acquisition of novel lexical items. *First Language, 7,* 177-182.

Lammers, M. O., & Castellote, M. (2009). The beluga whale produces two pulses to form its sonar signal. *Biology Letters, 5*, 297-301.

Lilly, J. C. (1965). Vocal mimicry in Tursiops: ability to match numbers and durations of human vocal bursts. *Science, 147*, 300-301.

Moore, B. R. (1992). Avian movement imitation and a new form of mimicry: Tracing the evolution of a complex form of learning. *Behaviour*, 122, 231-263.

Murayama, T., Iijima, S., Katsumata, H., & Arai, K. (2014). Vocal imitation of human speech, synthetic sounds and beluga sounds, by a beluga (*Delphinapterus leucas*). *International Journal of Comparative Psychology*, 27, 369-384.

Murayama, T., Fujii, Y., Hashimoto, T., Shimoda, A., Iijima, S., Hayasaka, K., … Arai, K. (2012). Preliminary study of object labeling using sound production in a beluga. *International Journal of Comparative Psychology*, 25, 195-207.

Myowa-Yamokoshi, M., & Matsuzawa, T. (1999). Factors influencing imitation of manipulatory actions in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *113*, 128-136.

Panova, E. M., & Agafonov, A. V. (2017). A beluga whale socialized with bottlenose dolphins imitates their whistles. *Animal Cognition,* 20, 11533-1160.

Pepperberg, I. M. (1981). Functional vocalizations by an African Grey Parrot (*Psittacus erithacus*). *Zeitschrift für Tierpsychologie, 55*, 139-160.

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals (*Phoca vitulina).* *Canadian Journal of Zoology, 63*, 1050–1056.

Ralston, J. V., & Herman, L. M. (1995). Perception and generalization of frequency contours by a bottlenose dolphin (*Tursiops truncatus*). *Journal of Comparative Psychology*, *109*(3), 268-277.

Reiss, D., & McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenosed dolphins (*Tursiops truncatus*): Evidence for vocal learning. *Journal of Comparative Psychology*, *107*, 301-312.

Richards, D. G., Wolz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus. Journal of Comparative Psychology, 98*, 10-28.

Ridgway, S., Carder, D., Jeffries, M., & Todd, M. (2012). Spontaneous human speech mimicry by a cetacean. *Current Biology*, *22*, 860-861.

Savage-Rumbaugh, E. S. (1993). Language learnability in man, ape and dolphin. In H. R. Roitblat, L. M. Herman, & P. Nachtigall (Eds.), *Language and communication: Comparative perspectives*. (pp. 457-484). Hillsdale, NJ: Erlbaum.

Savage-Rumbaugh, E. S., McDonald, E., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General, 112*, 211-235.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., Smith, S. T., & Lawson, J. (1980a). Reference: The linguistic essential. *Science, 210*, 922-925.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1980b). Do apes use language? *American Scientist*, 68, 49-61.

Spruijt, B. M., van den Bos, R., & Pijlman, F.T.A. (2001). A concept of welfare based on reward evaluating mechanisms in the brain: Anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behavioural Science*, *72*, 145–171.

Stoeger, A. S., Mietchen, D., Sukhun, O., Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, *22*, 2144-2148.

Subiaul, F. (2016). What's special about human imitation? A comparison with enculturated apes. *Behavioral Sciences*, *6(3)*, 13.

Thompson, R. K., & Herman, L. M. (1977). Memory for lists of sounds by the bottle-nosed dolphin: Convergence of memory processes with humans? *Science*, *195(4277)*, 501-503.

Todt, D. (1975). Social learning of vocal patterns and modes of their application in gray parrots (*Psittacus erithacus)*. *Zeitschrift für Tierpsychologie*, *39*, 178 -188.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development*, *64*, 1688-1705.

Topal, J., Byrne, R. W., Miklosi, A., & Csanyi, V. (2006). Reproducing human actions and action sequences: "Do as I Do!" in a dog. *Animal Cognition*, *9*, 355-367.

Tyack, P. L. (1986). Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: mimicry of signature whistles? *Behavioral Ecology and Sociobiology*, *18*, 251-257.

Tyack, P. L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology*, *122*, 319-331.

Watkins, W. A. (1968). The harmonic interval: fact or artifact in spectral analysis of pulse trains. In W. N. Tavaolga (Ed.), Marine Bioacoustics Vol. 2 (pp. 15-43). New York: Pergamon Press,

Whiten, A. (2002). Imitation of sequential and hierarchical structure in action: Experimental studies with children and chimpanzees. In C. Nehaniv, & K. Dautenhahn (Eds.), *Imitation in animals and artifacts* (pp. 191-209). Cambridge, MA: MIT Press.

Whiten, A., Custance, D. M., Gomez, J-C, Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children *(Homo sapiens)* and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *110*, 3-14.

Whiten A., Horner V., Lichtfield C. A., & Marschall-Pescini S. (2004). How do apes ape? *Learning & Behavior*, *32*, 36–52.

Xitco, M. J., Jr. (1988). *Mimicry of modeled behaviors by bottlenose dolphins* (Unpublished master's thesis). University of Hawaii at Manoa, Honolulu, HI.