

Interactive Bioacoustic Playback as a Tool for Detecting and Exploring Nonhuman Intelligence: “Conversing” with an Alaskan Humpback Whale

Brenda McCowan¹, *Josephine Hubbard², *Lisa Walker³, *Fred Sharpe⁴, Jodi Frediani⁵, and Laurance Doyle⁶

¹SVM: Population Health and Reproduction, University of California, Davis

²Animal Behavior Graduate Group, University of California, Davis

³Grooved Whale Project

⁴Alaska Whale Foundation

⁵Jodi Frediani Photography

⁶SETI Institute

Corresponding author:

Brenda McCowan¹

Email address: bjmccowan@ucdavis.edu

ABSTRACT

Here we report on a rare and opportunistic acoustic “conversation” with an adult female humpback whale, known as Twain, in Southeast Alaska. Post hoc acoustic and statistical analyses of a 20-minute acoustic exchange between the broadcast of a recorded contact call, known as a ‘whup/throp’, with call responses by Twain revealed an intentional human-whale acoustic (and behavioral) interaction. Our results show that Twain participated both physically and acoustically in three phases of interaction (Phase 1: Engagement, Phase 2: Agitation, Phase 3: Disengagement), independently determined by blind observers reporting on surface behavior and respiratory activity of the interacting whale. A close examination of both changes to the latency between Twain’s calls and the temporal matching to the latency of the exemplar across phases indicated that Twain was actively engaged in the exchange during Phase 1 (Engagement), less so during Phase 2 (Agitation), and disengaged during Phase 3 (Disengagement). These results, while preliminary, point to several key considerations for effective playback design, namely the importance of salient, dynamic and adaptive playbacks, that should be utilized in experimentation with whales and other intelligent nonhuman species.

Keywords: communication, bioacoustic playback, turn-taking, vocal matching, humpback whale

INTRODUCTION

The study of nonhuman animal signals challenges and informs our search for nonhuman intelligence with the ultimate hope of deepening human-animal intercommunication. The diversity of nonhuman intelligence that exists on Earth has been revealed in a multitude of observational and experimental studies over the past several decades (Vonk, 2021). This diversity is observed from octopus to ravens to elephants to whales, thus manifesting itself in a variety of both terrestrial and non-terrestrial environments. Indeed, one need look no further for a non-terrestrial form of intelligence than in the remarkable acoustic and behavioral versatility of the humpback whale (*Megaptera novaeangliae*) (Payne and McVay, 1971; Fournet and Szabo, 2013; Fournet et al., 2015).

Whales are ancestrally remote to humans and resumed life in the oceans 60 million years ago. Here humpback and other whales found themselves in an ocean soundscape capable of distant and widespread transmission of both low and high frequency sounds (Payne and Webb, 1971). The brain of the humpback whale possesses a large auditory cortex and clusters of spindle neurons that are presumably linked to their altruistic behavior (Hof and Van der Gucht, 2007; Pitman et al., 2017). Their global recovery from whaling hints at remarkable survival intelligence. The humpback’s superlative acoustic prowess is likely underwritten by impressive mental abilities. The whales acoustic achievements include songs that are lengthy, rhythmic and constantly evolving (Payne and McVay, 1971; Garland et al., 2011). Structural parallels between human speech and humpback whale song include the use of a similar frequency band and are comprised of both tonal (voiced) and broadband (unvoiced) elements, and vocalizations of variable duration punctuated by silence (Pace et al., 2010). Their vocal

repertoire also includes a panoply of less-studied social sounds (non-song sounds), including over 40 unique social calls (Thompson PO et al., 1986; Cerchio and Dahlheim, 2001; Dunlop et al., 2007; Fournet and Szabo, 2013; Fournet et al., 2015, 2018). These calls are produced in bouts whose variability increases with social complexity and larger group sizes (Tyack, 1983; Silber, 1986; Cusano et al., 2020). A subclass of these calls exhibits stability (across generations and ocean basins), which is often considered a prerequisite for the evolution of complex communication ((Bradbury and Vehrencamp, 2011; Rekdahl et al., 2013; Wild and Gabriele, 2014)). At least one of these calls is thought to serve as a contact call known as the “whup” or “throp” call (Wild and Gabriele, 2014; Fournet and Szabo, 2013; Fournet et al., 2015, 2018). This diversity of signals represent a unique opportunity to characterize and interact with a substantive nonhuman intelligence.

Communication is a fundamental behavior of all living things that facilitates the exchange of information among signalers and receivers and is dynamically comprised of both reciprocal and emergent processes (Wiley, 1983). Researchers of communication, and specifically in nonhuman animals, have utilized a series of tools to decipher both the structure and function of these communication systems to gain greater insight into the diversity, complexity and meanings of signals animals use to navigate their social environments (Wiley, 1983; Kershenbaum et al., 2016; Fischer et al., 2013; Reiss et al., 1997; McCowan et al., 1999, 2002). While most communication systems are multimodal in nature, a wealth of research has specifically focused on the acoustic communication of a multitude of anuran, avian and mammalian species (Wiley, 1983; Kershenbaum et al., 2016) as a means toward a greater understanding of the evolution of vocal communication (including human language) as well as a tool for enhancing welfare and conservation efforts. The tools involved in this pursuit include the bioacoustic recording of signals or calls, noting their behavioral contexts, and analyzing their acoustic structure (comprised of spectral and temporal components) to decipher call meaning (Kershenbaum et al., 2016; Fischer et al., 2013). A serious challenge to these traditional types of observational approaches however is that context does not always predict the signal to be used nor conversely does the signal precisely predict its context; that is one signal does not always equal one context. Rather it depends on the signal’s function; for example, alarm calls generally tend to be stereotypical in acoustic structure and are emitted conservatively in very specific contexts (e.g., presence of a predator), while other less constrained calls, such as a general nondescript category known as “contact calls”, are more variable in acoustic structure and emitted more ubiquitously in multiple contexts (Kershenbaum et al., 2016; Fischer et al., 2013; McCowan et al., 1999; Ferrer-i Cancho and McCowan, 2009). Additionally, calls can occur in long and dynamically changing sequences such as in bird and whale song, further complicating the deciphering of call meaning not only in different social but also acoustic contexts (like in human language) (Payne and McVay, 1971; Rekdahl et al., 2015; Kershenbaum et al., 2016; Miksis-Olds et al., 2008; McCowan et al., 1999, 2002).

In addition to variation in context, many forms of communication involve turn-taking, which prevents interruptions and the inefficiency of multitasking (Pika et al., 2018). Turn-taking is defined as the alternation of information exchange resulting in observable temporal regularities between two or more interactants (Pika et al., 2018). Turn-taking aligns along a spectrum from simple to complex and its sophistication provides insight into a given species communicative prowess and perhaps intelligence. Simple forms of turn taking include insect vibration, percussion, stridulation, and bioluminescence (grasshoppers, cicadas, and fireflies), while more complex examples are observed in primates with gaze, gesticulation, and antiphonal calling. In humans, spoken language underwrites turn taking cadence and has been proposed as a potentially cognitively demanding process. This is due to the rapid turn response rate (≈ 200 ms) compared to average language processing speeds (≈ 600 ms), potentially allowing receivers to begin crafting their response before fully receiving a signal (Levinson and Torreira, 2015). As a result, speakers construct their turns out of units whose structure allows the next speaker to anticipate their completion (Pika et al., 2018). In nonhuman animals, the turn response rate is much slower (ex. marmosets: 3-5 s), likely stemming from differences in ‘units of perception’ where animals of different sizes experience the world at different rates (Healy et al., 2013). In animals, the turn response rate can also be tied to several behavioral or social factors such as territoriality (manakins: Maynard et al. 2012; blue tits: Poesel and Dabelsteen 2005, competition (barn owl siblings: Ducouret et al. 2018), or mating behavior (Kipper et al., 2015). Vocal turn-taking requires a certain level of coordination between partners to avoid overlapping (and thus inefficient) signals. As a result, turn-taking has been shown to coincide with social organization (Pougnault et al., 2022) and has been thought to influence social coordination (Starlings: Henry et al. 2015) or enhance social bonds through cooperation (meerkats: Demartsev et al. 2018). For example, in species that are highly monogamous and develop strong bonds between mating pairs it has been shown that the longer the duration of the pair bond, the more similar their calling behavior will become (i.e., their signals converge) (titi monkeys: Clink et al. 2019; zebra finches: D’Amelio et al. 2017). In fact, aspects of vocal synchrony during vocal exchanges (such as matching the type or timing of a signal or shortening inter-call intervals between signals) have been shown to correlate with acute changes in arousal, valence, or potential engagement between participants (Briefer, 2012). Synchrony, in its purest form, is a perfect (or close) matching of temporal coordination (Ravignani, 2015) and one subclass, antisynchrony consists of perfect (or close) alternation, the latter of which is more readily found in the turn-taking observed in human and nonhuman communicative interactions (Pika et al., 2018; Mondémé, 2022; Levinson and Torreira, 2015). Thus, the temporal dimension

of time matching can be an essential element to a conversation's structure. For example, inter-call intervals have been found to be significantly shorter during vocal exchanges than between solo calls (dolphins: Nakahara and Miyazaki 2010; Japanese macaques: Katsu et al. 2019). Metrical structure has also been found in the songs of humpback whales suggesting that temporal organization and coordination is not just salient but important to this species (Handel et al., 2009).

While many studies on turn-taking in humans or animals focus on intraspecies communication, very few focus on interspecies partners. However, these forms of communication have great potential to provide insights to the evolution of signaling behavior and the role of flexibility in producing acoustic signals. Intraspecies communication has likely evolved in response to direct survival needs and fitness benefits of sharing information with conspecifics. Species who can extend those communication techniques to interspecies interactions may show particular promise in responding to foreign signals in their environment, as a result of a rapidly changing world climate and other human-related disturbances. When making predictions regarding interspecies communication patterns it is reasonable to expect that organisms with closer phylogenetic relationships may share mutual entrainment processes and intelligibility. However, it is also possible that with the appropriate selection pressures that these communicative capacities may also have evolved independently in more remote taxa. Indeed, studies on interspecies communication can allow for direct inquiries into shared communicative structure between species and has the potential to reveal “unplumbed” sophistication or intelligent behavior from nonhuman communicative partners. Studies on human-animal communication are particularly promising due to their potential to elicit a unique response from the subject outside of their species-specific, contextualized behavior patterns (Mondémé, 2022).

Finally, as one approach that includes a type of interspecies communication, researchers have long explored the function and meaning of nonhuman animal signals using acoustic playback experiments (McGregor et al., 1992). These studies are designed to actively probe nonhuman communication systems beyond traditional observations to gain further understanding of signal meaning. Historically, the design and interpretation of these studies have been hotly debated, including concerns of accuracy and salience of acoustic stimuli, use of appropriate controls, pseudoreplication of playback sets, and reproducibility of methods and results (Fischer et al., 2013; King, 2015; King and McGregor, 2016; Butkowski et al., 2010; Powell and Rosenthal, 2017; Pika et al., 2018; McGregor, 2000; De Rosa et al., 2022; Deecke, 2007). One area of less focused concern however is the overall utility of acoustic playback in general, and especially passive acoustic playback, the primary mode in which acoustic playback has been conducted. Passive acoustic playback experiments take fixed stimuli of interest and broadcast them at randomized, predetermined intervals within the animal's natural range to measure their behavioral and/or vocal response. While traditional passive acoustic playback has significantly contributed to our understanding of animal communication (Fischer et al., 2013; King, 2015; King and McGregor, 2016; Butkowski et al., 2010; Powell and Rosenthal, 2017; Pika et al., 2018; McGregor, 2000; De Rosa et al., 2022), it also has significant limitations. For example, in such experiments, concern arises over the composition and size of the audience targeted as well as the types of predictions one makes about appropriate responses, whether physical or vocal, to such playbacks. Indeed, if the response is predicted to be of a vocal nature, one needs to be careful about concluding that a simple increase in calling rate at the population level during playbacks to a large group of subjects is indicative of a contingent response since, in this context, the causal relationship between playback and response is unclear at best. Similarly, physical approaches to playback stimuli could be due to its function (as a contact call for example) or merely because the stimulus appears novel (or just novel enough) to the subject(s). In addition, such passive experiments are considered non-interactive because the playback stimuli are not adaptively modified to match or probe the varying vocal or physical responses by the subjects (King, 2015; King and McGregor, 2016). Yet, substantial evidence indicates that the information conveyed in a signal may only be salient (or more salient) when presented interactively or dynamically to (thus engaging) the subjects under study (King, 2015; King and McGregor, 2016).

Given this perspective, the goal of this report is to (1) describe a novel interaction, including a matched acoustic exchange (“conversation”) via bioacoustic playback, with an individually identified humpback whale (known as Twain) in Southeast Alaska and (2) discuss how this particular interaction, and its successes and limitations, might inform future approaches in designing experimental playback studies that can more directly engage subjects and thus aid in the detection and exploration of intelligence in nonhuman species.

METHODS

On August 18, 2021, as part of collecting a diverse pool of call exemplars (Deecke, 2007), we recorded a high quality whup call from a dispersed aggregation of nine humpback whales that were sited in the general vicinity of our research vessel, *Glacier Seal*, which was drifting with engines off near Five Fingers in Frederick Sound, Alaska. This high-quality exemplar was used in a playback trial on the following day, August 19, 2021 approximately three nautical miles from the previous day's location in which we conducted a controlled experiment with 20 min. pre-, during and post- playback periods (see Figure 1) with a whale, subsequently identified via Happywhale (www.happywhale.com) as a 35-yo adult female known as Twain, who remained within 100 m of our research vessel during the entire playback period (see Figure

2). The upper deck of the research vessel had four blind observers with a 360° view continuously collecting photos, video and notes on Twain's behavior during the exchange, all of whom were blind to the playback paradigm. An additional still photographer (also blind to the paradigm) collected images on the lower deck. During the pre- and post-playback periods, four additional whales were observed via binoculars and telephoto lens between 200-1000 m of the vessel and another eight whales were sighted beyond 1000 m from the vessel. Other than Twain, no other whales were present within 200 m of the vessel during the 20-min playback period. The playback apparatus, deployed from the lower deck of the *Glacier Seal*, consisted of a single underwater Lubell speaker (Lubell LL916) emitting the exemplar at approximately 6 m in depth, deployed on the port side of the bow, and two hydrophones (Cetacean Research Technology SQ26-01), deployed on the port and starboard side of the bow, approximately 3 m apart horizontally and at depths of approximately 3 and 6 m respectively. These hydrophones recorded the entire playback trial and exchange event on a MacBook Air computer using Audacity software (<http://audacityteam.org/>) and Zoom H1n Portable Recorder (see schematic in Figure 2 and Supplementary Materials for full specifications of the equipment used and a detailed description of the encounter). A single exemplar recorded the previous day was played back at spontaneous derived intervals during the playback period from a MacBook Air computer. These playbacks were interspersed with the calls of the whale surfacing near our vessel at approximately 100 m of the playback apparatus and the two experimenters conducting the playback were blind to the whale's surface behavior. Most of the recorded whup calls were of high amplitude (ranging from -21 to 43 dB). The overall site tenacity of this individual during the 20-min exchange as evidenced by seven blows/surfacings (which allowed us to mark her position/distance/direction from the boat - see Figure 2), followed by the calls dropping significantly in amplitude as she departed (GLM: $\beta = -9.28$, $p \ll 0.0001$), indicates with almost 100% confidence that Twain was the origin of these calls. As Figure 1 shows, no vocalizations were recorded during the pre- or post-playback periods, only during the playback period. Three exemplars were broadcast prior to Twain's first response and we ceased playback of exemplars prior to Twain's 33rd call; thus, the final three calls by Twain were not in exchange with the playback exemplar after which time she ceased calling (see Figure 3A). In all, there was a total of 38 whup exemplars broadcast and 36 Twain whup calls recorded during the 20-min playback period, during which time Twain performed seven blows/surfacings and observers captured three positive fluke IDs (see Supplementary Materials).

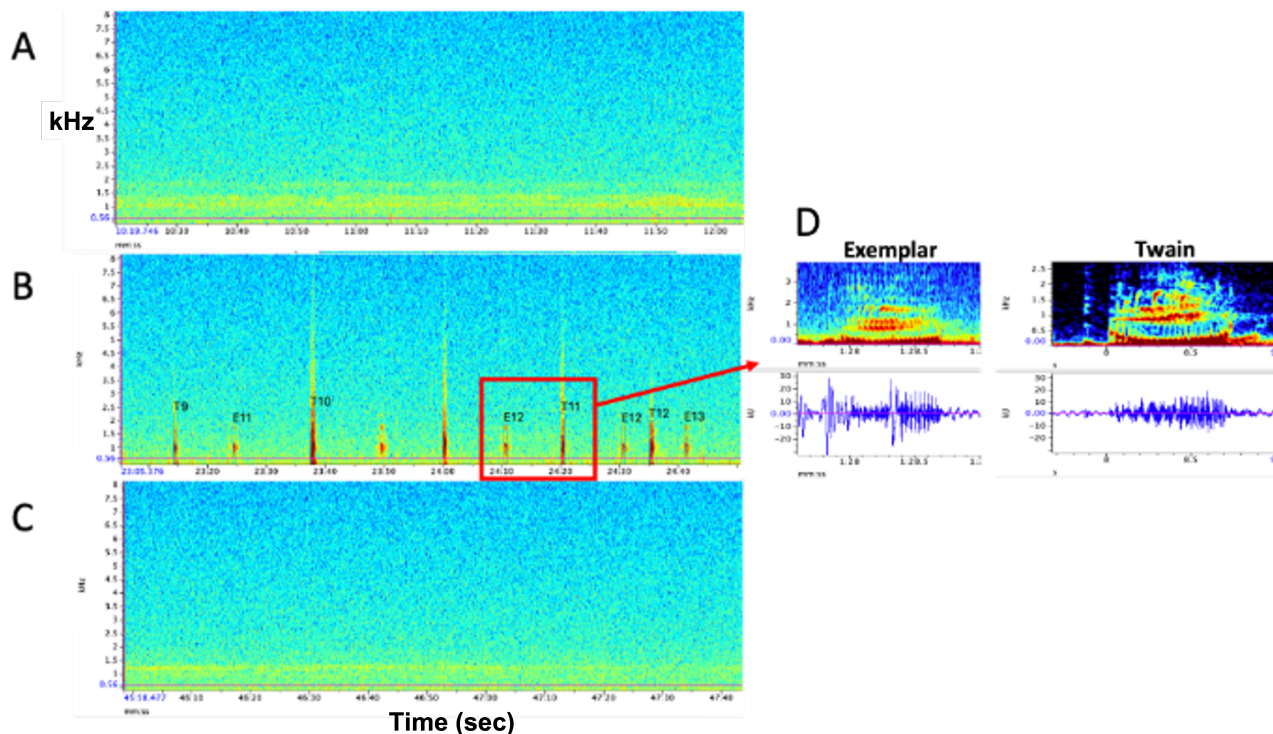


Figure 1. Representative spectrograms of **A.** Baseline control (pre) period, **B.** Experimental playback (during) period, and **C.** Follow-up control (post) period. No whup calls were recorded during pre- and post- periods. **D.** Example spectrograms and waveforms of the whup exemplar and Twain's whup call. Note exchange between E (exemplar) and T (Twain) in **B.**

To examine whether the acoustic sequence recorded during the broadcast of the exemplar and the whup vocalizations of

Twain could be considered a causal human-animal interaction or exchange, we conducted a number of post hoc analyses on aspects of call structure of Twain's whup calls recorded during this event. We considered Twain's calls either to be counter calls or not, with counter calls defined as calls that occurred following an exemplar without any other intervening behaviors (surfacing/blows) while those that were not as Twain's calls or other intervening behaviors produced in succession. We also categorized the entire interaction into three separate phases based on Twain's behavior recorded by the blind observers. The independent behaviors used to define phases were surface blows and orientation to/distance from the research vessel. Phase 1 was considered "Engagement" as this was the onset of the exchange after three exemplars were broadcast (interspersed with two cryptic or neutral blows). Phase 2 was considered "Agitation" as its onset and progression included three wheezy or reverse forced surface blows. These types of blows are considered to indicate excitement/agitation in humpback whales (Thompson PO et al., 1986; Watkins, 1977). Phase 3 was considered "Disengagement" as Twain was oriented away while increasing her distance from the research vessel (with two neutral blows). For an additional control and comparison to calls produced outside the playback trial context, we also included the latencies between whup calls from a short sequence of calls (N = 6) passively recorded from a small group of whales on August 18, 2021, the day before the playback trial (termed "Day 1"), the group membership of which included Twain and eight other whales of which Twain was the only individual common to both days (with the second day, the playback day, termed "Day 2").

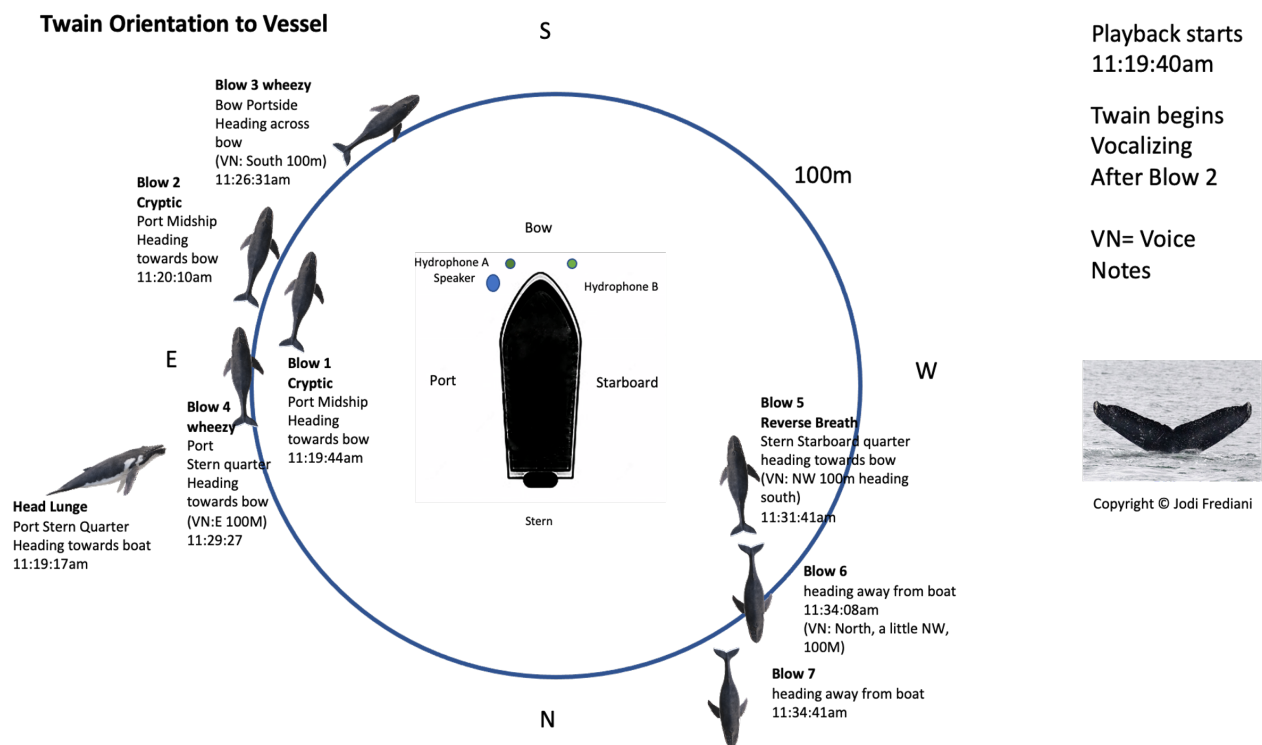


Figure 2. Twain's surface tracking and behavior during the playback phase of the trial (see Figure 1B). Whale image with text indicates position and behavior of Twain during playback trial. Boat image represents the R/V *Glacier Seal*. Hydrophones and speaker placement indicated. Fluke ID shot of Twain during the interaction.

Twain's whup calls on Day 2 and the whup calls recorded on Day 1 were acoustically analyzed for both spectral and temporal features using Raven Pro 1.6.4 (<https://ravensoundsoftware.com>); after considerable investigation we limited our analysis to one variable of interest: the inter-call interval (or latency between calls as measured by the time difference between the preceding call's offset and the subsequent call's onset). This was primarily due to the limitation in the use of a single whup exemplar that could only be manipulated via timing of broadcast. The rationale for this metric was that it potentially represents two fundamental aspects of emotional content of calls: arousal and valence (Briefer, 2012), which could be used to evaluate the motivational state of Twain and thus the causal nature of this communicative exchange.

In an additional analysis, we also examined whether the matching of Twain's calls latencies to that of the exemplars occurred and how matching differed across the phases as defined by the behavioral context. Phase 3 had only one pair to match and thus was excluded from this analysis. To standardize across the absolute duration of the latency, ratios for each Exemplar-Twain (E-T) pair were calculated (N = 27 pairs). These ratios were used to develop two metrics of discrepancy.

One was a categorical metric of “match or not”, where a perfect match would be a 50:50 ratio, in latency between Twain and the previous exemplar. Our criterion for a designating a “match” was any ratio that occurred between 41:59 and 50:50 (see Supplementary Materials). To avoid any bias in defining thresholds for categorizing a match, the other measure was the absolute difference in the ratio between Twain and the previous exemplar for each E-T pair as a second metric of discrepancy in latency matching. These metrics were included because such matching represents a type of behavioral synchronization, which is known to be an important feature in human and nonhuman animal communication (Wood et al., 2021; Bowling et al., 2013; Oesch, 2019; Imel et al., 2014; Lord et al., 2015; Xiao et al., 2013; Ravignani, 2018; Greenfield et al., 2021; Patel et al., 2009; King and McGregor, 2016; Hausberger et al., 2020; Demartsev et al., 2022; Herzing, 2015).

All metrics were subjected to GLM regression analysis where count outcomes were analyzed using negative binomial regression and binary outcomes using logistic regression in Stata 15 (StataCorp., 2017) or R Studio 7.2022 (RStudio Team, 2020). Predictors included counter calling or not, phase of interaction (Phase 1, Phase 2, Phase 3) and day of recording (Day 1 [control], Day 2 [playback]) (as defined above). Regressions on randomized data (N = 1000) were conducted for a subset of these analyzes (e.g., the counter calling analysis) to determine that these results were not spurious due to the clear limitation in sample size given an event of 1 (also see Supplementary Materials).

RESULTS AND DISCUSSION

Results shown in Figure 3 indicate that substantial variation was found among the latencies in Twain’s calling behavior both across the playback phases (Figure 3A and B) on Day 2 and between calls produced on Day 1 (control) versus Day 2 (playback) (Figure 3C). For the latter, latencies between Twain’s calls were overall shorter on Day 2 than the calls recorded on Day 1, and all phases on Day 2 showed significantly shorter latencies to those on Day 1 (Figure 3B). Notably for the former on Day 2, latencies between Twain’s calls were significantly shorter for Phase 1 during “Engagement” than either Phase 2 during “Agitation” or Phase 3 during “Disengagement”. Phase 2 latencies were also shorter than Phase 3 latencies albeit marginally so ($\beta = -0.30, p=0.05$). Day 1 latencies were significantly higher than all three phases on Day 2 (see Figure 3B; Day 1-Phase 3: $\beta = -0.30, p\ll 0.001$). Finally, latencies between Twain’s calls were shorter when the call was in direct exchange with the exemplar than when it was not (across the entire time series regardless of phase; Figure 3D). These results indicate that Twain was modifying her vocalization rate based upon the broadcast of the whup exemplar. Her motivational state in responding suggests either arousal, valence or some combination of both (Briefer, 2012).

Results shown in Figure 4 indicate that, in addition to shorter latencies between Twain’s call during interactive playback, matching between Twain’s vocalization rate and that of the exemplar was higher during Phase 1 than Phase 2 of the playback period. For matching that was categorized as a binary match or not (using the criterion indicated above), latencies between the exemplar and Twain’s calls produced in Phase 1 during “Engagement” were 9.9 times more likely to match the latencies of the previous exemplar than in Phase 2 during “Agitation” (Figure 4A). Because only one E-T pair was recorded in Phase 3 as we had stopped broadcasting exemplars early on during this phase, we omitted this phase from the analyses. For the analysis in which we compared the absolute discrepancies in latency ratios during Phase 1 and Phase 2, we found a significantly lower degree of discrepancy in Phase 1 than Phase 2 (Figure 4B). In Figure 4C, we provide a pie-chart representation of the ratio discrepancy as a time series for visualization. These results indicate that a greater amount of vocal synchrony was occurring during Phase 1 than during Phase 2. Vocal and behavioral synchrony are associated with pair bonding, group cohesion, and even empathy in natural communicative exchanges between humans (Wood et al., 2021; Lord et al., 2015; Imel et al., 2014; Xiao et al., 2013; Bowling et al., 2013) and between nonhuman animals (Ravignani, 2018; Greenfield et al., 2021; Patel et al., 2009; King and McGregor, 2016; Hausberger et al., 2020; Demartsev et al., 2022; Herzing, 2015), again suggesting that Twain was actively engaged in the turn-taking exchange with the broadcasted exemplar.

In sum then, the data on latencies between Twain’s calls and the latency matching between Twain’s calls and the exemplars shown in Figures 3 and 4, in relationship to the independent data collected on Twain’s surface and respiratory behavior, strongly indicate that Twain was actively engaged in a type of vocal synchrony, and specifically as a “coupled oscillator” with our playback system (Demartsev et al., 2018) (especially during Phase 1 of the exchange). This finding suggests that Twain was aware of our exemplar’s temporal dimension and that she had the capacity to coordinate in an antisynchronous counter calling framework. Her concurrent behavioral data suggest that she was also exhibiting changes to both arousal and valence during the encounter: certainly excitement and possibly the onset of agitation during Phase 2. However, she was not sufficiently aroused that she initiated calling in bouts or produced additional call types, which been demonstrated elsewhere (Dunlop, 2017). Interestingly, one reason that might explain the establishment of Phase 1 engagement during the playback with Twain is that the whup call used as the exemplar was serendipitously and opportunistically recorded on the previous day during a control period (Day 1). This call was recorded from one of the group of animals present during that recording, which included Twain and eight other group mates (see Supplementary Materials for details). We can only speculate that this initial strong response to our whup call exemplar was due to either a

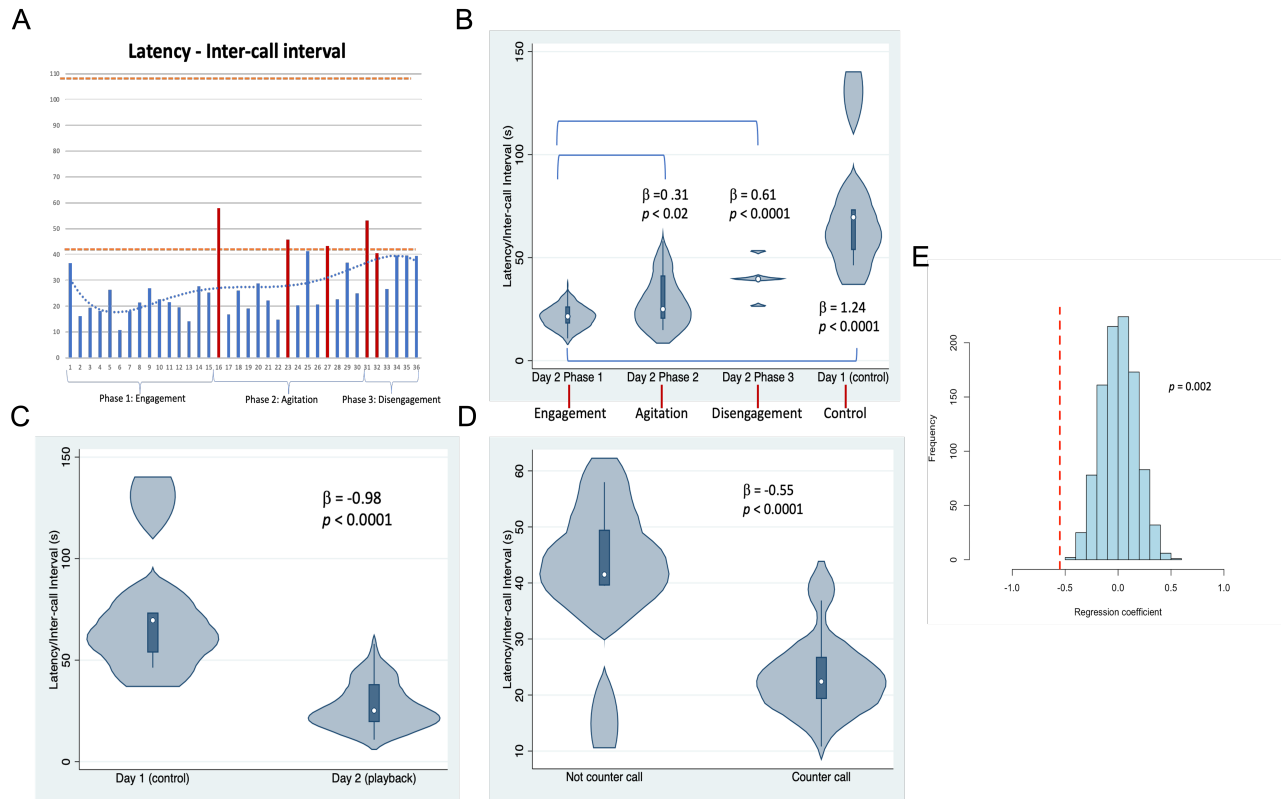


Figure 3. Latencies between Twain’s calls. **A.** As a time series showing in red the latencies where Twain had surfaced/blew prior to vocalizing on Day 2 (playback) (blows 1 and 2 occurred just before the exchange began – see Figure 2). Orange dotted lines indicate the range of latencies between whup calls on Day 1 (control). **B.** Negative binomial regression on latencies between Twain’s calls during each of Day 2’s phases and the calls recorded on Day 1. **C.** Latencies between the calls recorded on Day 1 (control) in comparison to Twain’s calls on Day 2 (playback). **D.** Latencies between Twain’s calls when a counter call (direct exchange) versus not a counter call across the entire time series. **E.** Randomizations ($N = 1000$) of regression conducted in (D) indicating that the significant difference found between the latencies of Twain’s calls in counter calling vs. non-counter calling was not spurious.

mirroring effect of playing back Twain’s own call to herself or increased interest due to the fact that the call was identifiable to one of her group members. Although perhaps unlikely, the former would suggest a highly evolved cognitive capacity of self-recognition in humpback whales that could be further explored using a type of interactive adaptive design (see “Lessons Learned: Directions for Future Research” below). If the latter, the initial and then eventual waning of engagement found across the playback exchange provides us with deeper insights into how we can improve future playback designs for unraveling vocal structure and meaning in animal communication. For example, one promising area that this study points to is discerning the salient turn-taking rhythm of each species’ communication system (Healy et al., 2013). We describe further insights in the final section below.

LESSONS LEARNED: DIRECTIONS FOR FUTURE RESEARCH

In all, from this rare opportunistic event and the accompanying acoustic analyses, we can suggest several important insights for future studies attempting to solicit prolonged acoustic interactions between humans and nonhuman species, such as humpback whales.

First, a library of salient well-processed stimuli should be prepared and include stimuli recorded both outside of and within the current field season (Deecke, 2007). For example, our capture of the exemplar on Day 1 from a group of animals that included Twain, while serendipitous, suggests that calls recorded from animals known to the interacting individual are more salient than randomly selected calls (McCowan and Newman, 2000; King, 2015; King and McGregor, 2016). In our study, playback trials conducted on multiple previous and subsequent days using exemplars from archived recordings did not solicit physical or acoustic interaction. Furthermore, the salience of playback stimuli may depend on the rate of change

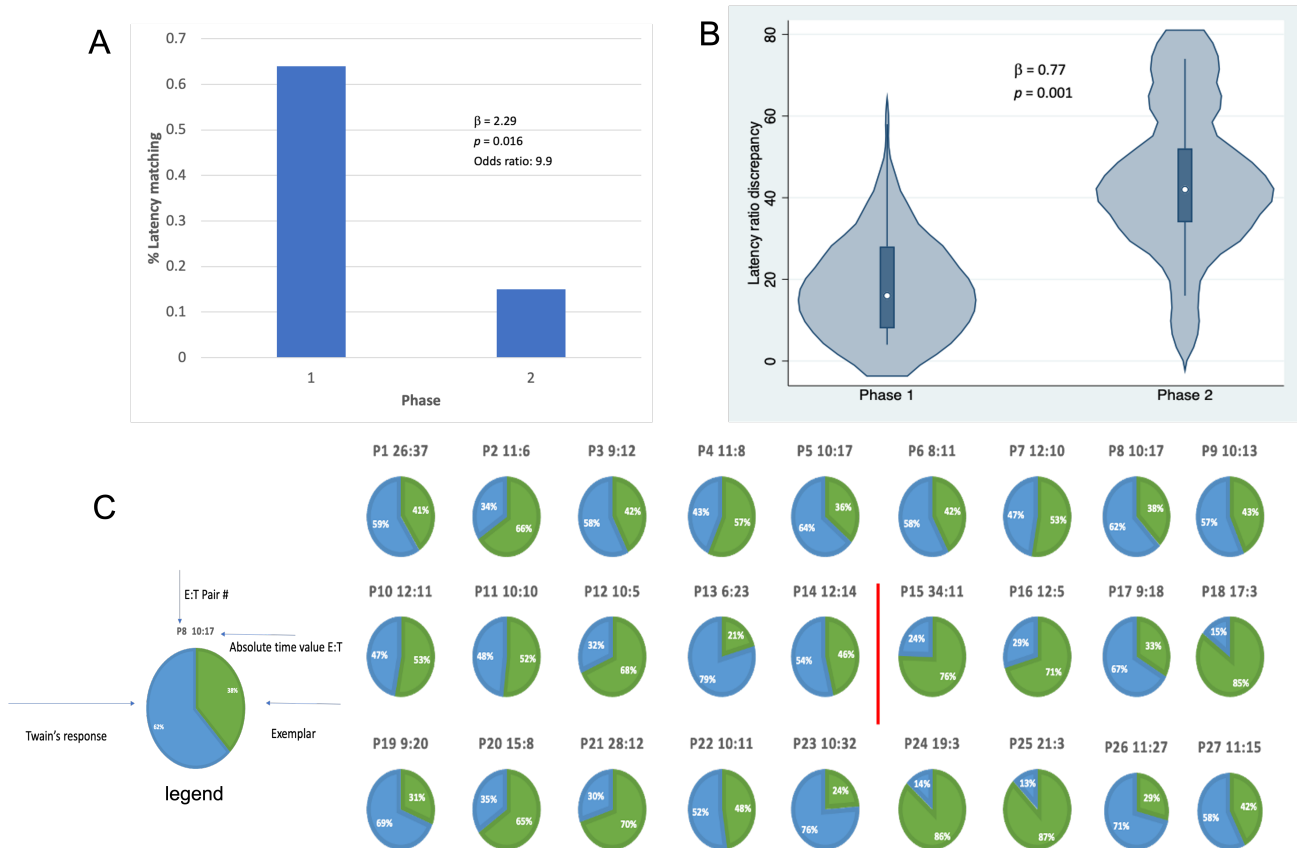


Figure 4. Latency matching between Twain’s calls and exemplars. **A.** Logistic regression indicating a greater proportion of matching between Twain’s calls and the exemplar was found (9.9x more likely) in Phase 1 during “Engagement” than in Phase 2 during “Agitation”. **B.** Negative binomial regression on the difference or discrepancies between the latency ratios of Twain’s calls and the exemplars where discrepancy was significantly lower for Phase 1 than Phase 2. **C.** Visualization (pie charts) of the latency ratio among successive pairs of Twain’s calls and exemplars as a time series. Red line separates Phase 1 from Phase 2.

for species-specific signals where recent variants may be more salient for signals that are being modified at a faster rate. For example, a playback experiment conducted with Campbell’s monkeys found that individuals responded more often to current variants of a signal compared with older versions of the same signal (Lemasson et al., 2005).

Second, appropriate equipment to record and broadcast stimuli during playback experiments with high fidelity is certainly desirable. We acknowledge that the quality of our playback suffered from low-end frequency drop out and a mismatch between the recorded and ambient background noise (see Supplementary Materials for details); however, it still managed to solicit an extended exchange/interaction with Twain indicating that the components retained and broadcast in the playback were salient enough to elicit a response. Diagnostic tests prior to playback experiments should be conducted to ensure that recorded playback stimuli retain a high fidelity of acoustic components (with particular attention to features known to be salient to the study species) and cover most of the frequency or amplitude ranges of the study animal’s natural range.

Third, we recommend that multiple hydrophones (or microphones) be used and deployed at different depths (locations) to capture the configuration of acoustic interactions. The data from each specific hydrophone’s recording did not accurately capture the full interaction event. Indeed, the data from the starboard-side hydrophone revealed additional responses by Twain not detected by the port-side hydrophone, which was deployed adjacent to the speaker. These differences in the detection of sounds from these two hydrophones could be due to differences in signal sensitivity, location of deployment on the vessel, and other behavioral factors of the subject including distance from or orientation to each hydrophone. Deploying multiple hydrophones (an array would be ideal) provides additional information with which to corroborate acoustic data and/or behavioral observations.

Finally, we want to stress the value of collecting data on multiple modalities of behavior in the pursuit of understanding

the function and flexibility of animal signals. Many studies on animal communication seek to understand signal function by characterizing the context in which the signal is solicited. Independent or behavioral observations that are blind to experimental procedures are particularly useful in providing contextual clues to acoustic behavior. These types of behavioral clues allowed our study team to break up our encounter into different phases based on the behavior of the targeted recipient of the experiment. Identifying these behavioral changes can make adaptive playback design more engaging and salient for experimental targets as well as provide contextual clues for scientists focused on analyzing the function of animal signals. Although the use of a handheld video recorder was sufficient for our group to record behavioral observations on the top-deck of the vessel, we would suggest using cameras with a wider lens (such as a 360° camera or drone) to facilitate observations over a more expansive area. Drones are particularly promising for documenting these types of interactions because they can follow the animals and provide a closer view or more detailed account of behavioral sequences and interactions. Wide angle underwater cameras such as Go Pros would also be useful in the case of aquatic animals to capture the underwater activities of the targeted individual(s).

More importantly, we suggest that an “interactive” approach to playback where the broadcasting of the stimuli should be modified in response to the animals’ acoustic or behavioral response during the playback period, including but not limited to the timing of playbacks. While our results certainly indicate the importance of dynamical control over the temporal element in a playback, a major limitation in our playback study was the inability to modify anything other than the timing of the playback. This limitation may have either frustrated or disinterested the whale over the course of the three phases as indicated by her surface behavior. This type of more complex manipulation can likely be accomplished with readily available software such as Audacity (<http://audacityteam.org/>) and sufficient computer resources to process recorded signals in almost real-time. Such an adaptive versus static approach would permit real-time recording, manipulation, and subsequent playback of calls to be conducted, allowing modification of both temporal and spectral features that can elicit differential responses in real time.

We therefore conclude that although considerably challenging, adaptive interactive acoustic playback is possible to conduct given recent innovations in technology. Such technology allows immediate and adaptable manipulation of playback signals as a function of the study subject’s current and changing behavior (King, 2015; King and McGregor, 2016; Butkowski et al., 2010; McGregor, 2000; McGregor et al., 1992; Douglas and Mennill, 2010; Miller et al., 2009; Nielsen and Vehrencamp, 1995; Burt et al., 2001). Such an adaptive playback design incorporates dynamics into this well-established experimental framework and allows previously intractable questions to be addressed about signal function and meaning by probing the system and its subjects in real-time. As such, this approach offers a more powerful tool for examining animal communication than traditional passive approaches by enabling the exploration of the interactive and dynamical features of intelligent communicative behavior.

ACKNOWLEDGEMENTS

We thank Marc Choquette (Captain, *R/V Glacier Seal*, Juneau), Don and Denise Bermant (M/Y Blue Pearl, Vancouver), Tony Gilbert (Program Director, Seakeepers), Collette Costa, and Debbie Kolyer for logistical support. We also thank Clark Snodgrass, Katie Zacarian, Aza Raskin and Britt Selvitelle for support in observational data collection and whale identification, and James Crutchfield for his participation in the playback trial as well as initial discussions during the 2021 field season. All field investigations were conducted under National Marine Fisheries Service Research Permit No. 19703 to Fred Sharpe. Author contributions: B.M. recorded, prepared and selected the whup exemplar for playback and determined the timing and extent of the playbacks. B.M. and L.W. acoustically recorded the playback trial. J.A.H video recorded the playback trial. L.W., J.A.H., J.F., F.S. collected observational data. B.M. and L.W. analyzed the data. All authors contributed to the overall study design and the writing of the manuscript. Authors denoted with an asterisk (*) contributed equally. Funding: This research was supported by Templeton World Charity Foundation Diverse Intelligences grant TWCF0440 (P.I. L. Doyle). The opinions expressed in this publication do not necessarily reflect the views of Templeton World Charity Foundation, Inc. Competing Interests: None declared. Data and materials availability: Data will be made available upon reasonable request.

SUPPLEMENTARY MATERIALS

Under construction

REFERENCES

Bowling, D. L., Herbst, C. T., and Fitch, W. T. (2013). Social origins of rhythm? Synchrony and temporal regularity in human vocalization. *PLoS ONE*, 8(11):1–7.

- Bradbury, J. W. and Vehrencamp, S. L. (2011). *Principles of animal communication (2nd ed.)*. Sinauer Associates.
- Briefer, E. (2012). Vocal Expression of Emotions in Mammals: Mechanisms of Production and Evidence. *Communication Skills Collection*, pages 1–29.
- Burt, J. M., Campbell, S. E., and Beecher, M. D. (2001). Song type matching as threat: A test using interactive playback. *Anim Behav*, 62(6):1163–1170.
- Butkowski, T., Yan, W., Gray, A., Cui, R., Verzijden, M., and Rosenthal, G. G. (2010). Automated interactive video playback for studies of animal communication. *J Vis Exp*, 2010.
- Cerchio, S. and Dahlheim, M. (2001). Variation in feeding vocalizations of humpback whales *Megaptera novaeangliae* from southeast Alaska. *Bioacoustics*, 11(4):277–295.
- Clink, D. J., Lau, A. R., and Bales, K. L. (2019). Age-related changes and vocal convergence in titi monkey duet pulses. *Behaviour*, 156(15):1471–1494.
- Cusano, D. A., Indeck, K. L., Noad, M. J., and Dunlop, R. A. (2020). Humpback whale (*Megaptera novaeangliae*) social call production reflects both motivational state and arousal. <https://doi.org/10.1080/09524622.2020.1858450>, 31(1):17–40.
- D'Amelio, P. B., Trost, L., and ter Maat, A. (2017). Vocal exchanges during pair formation and maintenance in the zebra finch (*Taeniopygia guttata*). *Frontiers in Zoology*, 14(1):1–12.
- De Rosa, A., Castro, I., and Marsland, S. (2022). The acoustic playback technique in avian fieldwork contexts: a systematic review and recommendations for best practice. *Ibis (London)*, 164:371–387.
- Deecke, V. B. (2007). Studying Marine Mammal Cognition in the Wild: A Review of Four Decades of Playback Experiments. *Aquatic Mammals*, 32(4):461–482.
- Demartsev, V., Gersick, A. S., Jensen, F. H., Thomas, M., Roch, M. A., Manser, M. B., and Strandburg-Peshkin, A. (2022). Signalling in groups: New tools for the integration of animal communication and collective movement. *Methods in Ecology and Evolution*, pages 1–12.
- Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., and Manser, M. (2018). Vocal Turn-Taking in Meerkat Group Calling Sessions. *Current Biology*, 28(22):3661–3666.
- Douglas, S. B. and Mennill, D. J. (2010). A review of acoustic playback techniques for studying avian vocal duets. *Journal of Field Ornithology*, 81(2):115–129.
- Ducouret, P., Dreiss, A. N., Gémard, C., Falourd, X., and Roulin, A. (2018). Barn owl nestlings vocally escalate when interrupted by a sibling: evidence from an interactive playback experiment. *Animal Behaviour*, 145:51–57.
- Dunlop, R. A. (2017). Potential motivational information encoded within humpback whale non-song vocal sounds. *The Journal of the Acoustical Society of America*, 141(3):2204–2213.
- Dunlop, R. A., Noad, M. J., Cato, D. H., and Stokes, D. (2007). The social vocalization repertoire of east Australian migrating humpback whales (*Megaptera novaeangliae*). *The Journal of the Acoustical Society of America*, 122(5):2893–2905.
- Ferrer-i Cancho, R. and McCowan, B. (2009). A law of word meaning in dolphin whistle types. *Entropy*, 11(4):688–701.
- Fischer, J., Noser, R., and Hammerschmidt, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *Am J Primatol.*, 2013(75):643–663.
- Fournet, M. and Szabo, A. (2013). Vocal repertoire of Southeast Alaskan humpback whales (*Megaptera novaeangliae*). *The Journal of the Acoustical Society of America*, 134(5):3988–3988.
- Fournet, M. E., Gabriele, C. M., Culp, D. C., Sharpe, F., Mellinger, D. K., and Klinck, H. (2018). Some things never change: multi-decadal stability in humpback whale calling repertoire on Southeast Alaskan foraging grounds. *Scientific Reports*, 8(1).
- Fournet, M. E., Szabo, A., and Mellinger, D. K. (2015). Repertoire and classification of non-song calls in Southeast Alaskan humpback whales (*Megaptera novaeangliae*). *The Journal of the Acoustical Society of America*, 137(1):1–10.
- Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., Poole, M. M., Robbins, J., and Noad, M. J. (2011). Dynamic Horizontal Cultural Transmission of Humpback Whale Song at the Ocean Basin Scale. *Current Biology*, 21(8):687–691.
- Greenfield, M. D., Honing, H., Kotz, S. A., and Ravignani, A. (2021). Synchrony and rhythm interaction: from the brain to behavioural ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1835).
- Handel, S., Todd, S. K., and Zoidis, A. M. (2009). Rhythmic structure in humpback whale (*Megaptera novaeangliae*) songs: Preliminary implications for song production and perception. *The Journal of the Acoustical Society of America*, 125(6):EL225.
- Hausberger, M., Giacalone, A., Harmand, M., Craig, A. J., and Henry, L. (2020). Calling rhythm as a predictor of the outcome of vocal interactions: flight departure in pale-winged starling pairs. *The Science of Nature*, 107(1).
- Healy, K., McNally, L., Ruxton, G. D., Cooper, N., and Jackson, A. L. (2013). Metabolic rate and body size are linked with perception of temporal information. *Animal Behaviour*, 86(4):685–696.
- Henry, L., Craig, A. J., Lemasson, A., and Hausberger, M. (2015). Corrigendum to: Social coordination in animal vocal

- interactions. Is there any evidence of turn-taking? The starling as an animal model [(2015). *Front. Psychol.* 6:1416. doi: 10.3389/fpsyg.2015.01416]. *Frontiers in Psychology*, 6(DEC):1924.
- Herzing, D. L. (2015). Synchronous and rhythmic vocalizations and correlated underwater behavior of free-ranging Atlantic spotted dolphins (*Stenella frontalis*) and bottlenose. *Anim Behav Cogn*, 2(1):14–29.
- Hof, P. R. and Van der Gucht, E. (2007). Structure of the cerebral cortex of the humpback whale, *Megaptera novaeangliae* (Cetacea, Mysticeti, Balaenopteridae). *Anat Rec (Hoboken)*, 290(1):1–31.
- Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., and Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of Counseling Psychology*, 61(1):146–153.
- Katsu, N., Yamada, K., Okanoya, K., and Nakamichi, M. (2019). Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Current zoology*, 65(1):99–105.
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akcay, C., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Căsar, C., Coen, M., Deruiter, S. L., Doyle, L., Edelman, S., Ferrer-i Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., Huetz, C., Hughes, M., Hyland Bruno, J., Ilany, A., Jin, D. Z., Johnson, M., Ju, C., Karnowski, J., Lohr, B., Manser, M. B., Mccowan, B., Mercado, E., Narins, P. M., Piel, A., Rice, M., Salmi, R., Sasahara, K., Sayigh, L., Shiu, Y., Taylor, C., Vallejo, E. E., Waller, S., and Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, 91:13–52.
- King, S. L. (2015). You talkin’ to me? Interactive playback is a powerful yet underused tool in animal communication research. *Biol Letters*, 11.
- King, S. L. and McGregor, P. K. (2016). Vocal matching: the what, the why and the how. *Biology Letters*, 12(10).
- Kipper, S., Kiefer, S., Bartsch, C., and Weiss, M. (2015). Female calling? Song responses to conspecific call playbacks in nightingales, *Luscinia megarhynchos*. *Animal Behaviour*, 100:60–66.
- Lemasson, A., Hausberger, M., and Zuberbühler, K. (2005). Socially meaningful vocal plasticity in adult Campbell’s monkeys (*Cercopithecus campbelli*). *Journal of comparative psychology*, 119(2):220–229.
- Levinson, S. C. and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6(JUN).
- Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., and Atkins, D. C. (2015). More Than Reflections: Empathy in Motivational Interviewing Includes Language Style Synchrony Between Therapist and Client. *Behavior Therapy*, 46(3):296–303.
- Maynard, D. F., Ward, K. A. A., Doucet, S. M., and Mennill, D. J. (2012). Calling in an acoustically competitive environment: Duetting male long-tailed manakins avoid overlapping neighbours but not playback-simulated rivals. *Animal Behaviour*, 84(3):563–573.
- McCowan, B., Hanser, S. F., and Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, 57(2):409–419.
- McCowan, B., Hanser, S. F., and Doyle, L. R. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology*, 116(2):166–172.
- McCowan, B. and Newman, J. D. (2000). The role of learning in chuck call recognition by squirrel monkeys (*Saimiri sciureus*). *Behaviour*, 137:279–300.
- McGregor, P. K. (2000). Playback experiments: Design and analysis. *Acta Ethol.*, 2000(3):1.
- McGregor, P. K., Dabelsteen, T., Shepherd, M., and Pedersen, S. B. (1992). The signal value of matched singing in great tits: evidence from interactive playback experiments. *Anim Behav.*, 1992(43):6.
- Miksis-Olds, J. L., Buck, J. R., Noad, M. J., Cato, D. H., and Dale Stokes, M. (2008). Information theory analysis of Australian humpback whale song. *The Journal of the Acoustical Society of America*, 124(4):2385–2393.
- Miller, C. T., Beck, K., Meade, B., and Wang, X. (2009). Antiphonal call timing in marmosets is behaviorally significant: Interactive playback experiments. *J Comp Physiol A Neuroethol Sensory, Neural, Behav Physiol.*, 195.
- Mondémé, C. (2022). Why study turn-taking sequences in interspecies interactions? *Journal for the Theory of Social Behaviour*, 52(1):67–85.
- Nakahara, F. and Miyazaki, N. (2010). Vocal exchanges of signature whistles in bottlenose dolphins (*Tursiops truncatus*). *Journal of Ethology*, 29(2):309–320.
- Nielsen, B. M. and Vehrencamp, S. L. (1995). Responses of song sparrows to song-type matching via interactive playback. *Behav Ecol Sociobiol.*, 37:109–117.
- Oesch, N. (2019). Music and language in social interaction: Synchrony, antiphony, and functional origins. *Frontiers in Psychology*, 10.
- Pace, F., Benard, F., Glotin, H., Adam, O., and White, P. (2010). Subunit definition and analysis for humpback whale call classification. *Applied Acoustics*, 71(11):1107–1112.
- Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009). Studying Synchronization to a Musical Beat in Nonhuman Animals. *Annals of the New York Academy of Sciences*, 1169(1):459–469.

- Payne, R. and Webb, D. (1971). ORIENTATION BY MEANS OF LONG RANGE ACOUSTIC SIGNALING IN BALEEN WHALES*. Technical report, Woods Hole Oceanographic Institute, Contribution No. 2317, Woods Hole.
- Payne, R. S. and McVay, S. (1971). Songs of humpback whales. *Science*, 173(3997):585–597.
- Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences*, 285(1880).
- Pitman, R. L., Deecke, V. B., Gabriele, C. M., Srinivasan, M., Black, N., Denking, J., Durban, J. W., Mathews, E. A., Matkin, D. R., Neilson, J. L., Schulman-Janiger, A., Shearwater, D., Stap, P., and Ternullo, R. (2017). Humpback whales interfering when mammal-eating killer whales attack other species: Mobbing behavior and interspecific altruism? *Mar Mam Sci*, 33:7–58.
- Poesel, A. and Dabelsteen, T. (2005). Territorial responses of male blue tits to simulated dynamic intrusions: effects of song overlap and intruder location. *Animal Behaviour*, 70(6):1419–1427.
- Pougnault, L., Levréro, F., Leroux, M., Paulet, J., Bombani, P., Dentressangle, F., Deruti, L., Mulot, B., and Lemasson, A. (2022). Social pressure drives “conversational rules” in great apes. *Biological Reviews*, 97(2):749–765.
- Powell, D. L. and Rosenthal, G. G. (2017). What artifice can and cannot tell us about animal behavior. *Current Zoology*, 63(1):21–26.
- Ravignani, A. (2015). Evolving perceptual biases for antisynchrony: A form of temporal coordination beyond synchrony. *Frontiers in Neuroscience*, 9(SEP).
- Ravignani, A. (2018). Rhythm and synchrony in animal movement and communication. *Current Zoology*, 65(1):77–81.
- Reiss, D., McCowan, B., and Marino, L. (1997). Communicative and other cognitive characteristics of bottlenose dolphins. *Trends in Cognitive Sciences*, 1(4):140–145.
- Rekdahl, M. L., Dunlop, R. A., Goldizen, A. W., Garland, E. C., Biassoni, N., Miller, P., and Noad, M. J. (2015). Non-song social call bouts of migrating humpback whales. *The Journal of the Acoustical Society of America*, 137(6):3042–3053.
- Rekdahl, M. L., Dunlop, R. A., Noad, M. J., and Goldizen, A. W. (2013). Temporal stability and change in the social call repertoire of migrating humpback whales. *The Journal of the Acoustical Society of America*, 133(3):1785–1795.
- RStudio Team (2020). RStudio: Integrated Development for R.
- Silber, G. K. (1986). The relationship of social vocalizations to surface behavior and aggression in the Hawaiian humpback whale (*Megaptera novaeangliae*). *Canadian Journal of Zoology*, 64(10):2075–2080.
- StataCorp. (2017). Stata Statistical Software: Release 15. .
- Thompson PO, Cummings WC, and Ha SJ (1986). Sounds, source levels, and associated behavior of humpback whales, Southeast Alaska. *Journal of the Acoustical Society of America*, 80(3):735–740.
- Tyack, P. (1983). Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds. *Behavioral Ecology and Sociobiology*, 13(1):49–55.
- Vonk, J. (2021). Nonhuman Intelligence. *Encyclopedia of Evolutionary Psychological Science*, pages 5432–5440.
- Watkins, W. (1977). Air-borne sounds of the humpback whale *Megaptera novaeangliae*. *J of Mammology*, 48:573–578.
- Wild, L. A. and Gabriele, C. M. (2014). Putative contact calls made by humpback whales (*Megaptera novaeangliae*) in Southeastern Alaska. *Canadian Acoustics*, 42:23–31.
- Wiley, R. (1983). The evolution of communication: information and manipulation. *Animal Behaviour*, pages 157–189.
- Wood, A., Lipson, J., Zhao, O., and Niedenthal, P. (2021). Forms and Functions of Affective Synchrony. *Handbook of Embodied Psychology*, pages 381–402.
- Xiao, B., Georgiou, P. G., Imel, Z. E., Atkins, D. C., and Narayanan, S. S. (2013). Modeling therapist empathy and vocal entrainment in drug addiction counseling. *Interspeech*.