View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by St Andrews Research Repository

Imitation of novel conspecific and human speech sounds in the killer whale *(Orcinus orca)*

José Z. Abramson[1,2,3], Mª Victoria Hernández-Lloreda[3,4], Lino García[5], Fernando Colmenares[3,6], Francisco Aboitiz[1] & Josep Call[7,8]

[1]Departamento de Psiquiatría, Facultad de Medicina and Centro Interdisciplinario de Neurociencias, Pontificia Universidad Católica de Chile, Marcoleta 391, Santiago, Chile [2]Facultad de Ciencias, Escuela de Medicina Veterinaria, Universidad Mayor, Santiago de Chile [3]Grupo UCM de Psicobiología Social, Evolutiva y Comparada, Universidad Complutense de Madrid, 28223, Madrid, Spain [4]Departamento de Metodología de las Ciencias del Comportamiento, Facultad de Psicología, Campus de Somosaguas, Universidad Complutense de Madrid, 28223, Madrid, Spain [5]Dpto. Teoría de la Señal y Comunicaciones ETSIST, UPM, Spain [6]Departamento de Psicobiología, Facultad de Psicología, Campus de Somosaguas, Universidad Complutense de Madrid, 28223, Madrid, Spain [7]School of Psychology and Neuroscience, University of St Andrews, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP, United Kingdom [8]Department of Developmental and Comparative Psychology, Max-Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

Abstract

Vocal imitation is a hallmark of human spoken language, which, along with other advanced cognitive skills, has fuelled the evolution of human culture. Comparative evidence has revealed that although the ability to copy sounds from conspecifics is mostly uniquely human among primates, a few distantly related taxa of birds and mammals have also independently evolved this capacity. Remarkably, field observations of killer whales have documented the existence of group-differentiated vocal dialects that are often referred to as traditions or cultures and are hypothesized to be acquired non-genetically. Here we use a -*Do as I do*- paradigm to study the abilities of a killer whale to imitate novel sounds uttered by conspecific (vocal imitative learning) and human models (vocal mimicry). We found that the subject made recognizable copies of all familiar and novel conspecific and human sounds tested and did so relatively quickly (most during the first 10 trials and three in the first attempt). Our results lend support to the hypothesis that the vocal variants

observed in natural populations of this species can be socially learned by imitation. The capacity for vocal imitation shown in this study may scaffold the natural vocal traditions of killer whales in the wild.

Keywords: Vocal learning; imitation; mimicry; cetacean culture; do as I do; killer whale

1. Introduction

Learning a previously unknown behaviour by observation from another individual [1] enables the non-genetic transfer of information between individuals and constitutes a potential driver for the diffusion and consolidation of group-specific behavioural phenotypes (i.e., traditions and cultures) [2,3]. Imitation of novel sounds, also referred to as vocal production learning [4] and defined as the ability to learn to produce a novel sound just from hearing it, is a core property of human speech which, along with other cognitive skills, has fuelled the evolution of another adaptation uniquely evolved in our species, human culture [5]. Comparative evidence has revealed that although the ability to copy sounds from conspecifics is widespread in birds, it is strikingly rare in mammals [4,6], and among primates it is unique to humans [7,8, but see 9]. Cetaceans are one of the few mammalian taxa shown to have evolved the ability for vocal production learning. Studies of several cetacean species in the wild have revealed that they exhibit substantial behavioural diversity between sympatric groups in terms of the acoustic features of their vocal repertoires (songs, calls) [10]. Bottlenose dolphins (*Tursiops truncatus*) display individual recognition signature whistles, humpback whales (*Megaptera novaengliae*) produce songs where some elements persist while others evolve over generations, and sperm whales (*Physeter macrocephalus*) possess group-specific coda repertoires [4,10,11]. It has been suggested that imitative learning can underpin these behaviours with experimental evidence for the ability for sound imitation demonstrated mainly in the bottlenose dolphin [11,12,13] and recently in the beluga (*Delphinapterus leucas*) [14,15].

Among cetaceans, the killer whale (*Orcinus orca)* stands out regarding the study of vocal

dialects in the wild [16]. Each family or matrilineal unit or pod within a population has been documented to have a unique vocal dialect, including a combination of unique and shared call types [17-19]. These dialects are believed to be transmitted via social learning [16-18], not only from mother to offspring (vertical transmission), but also between matrilines (horizontal transmission) [18-21]. Moreover, the similar acoustic features found between different populations in the same area do not correlate with geographic distance [22]. Since many of these group-differentiated signatures are not accounted for by ecological factors or genetic inheritance, the hypothesis that they may have been acquired through social learning, particularly imitation, appears plausible [16-24].

Elucidating the precise mechanism of social learning involved is difficult, however, particularly for acoustic communication in wild populations. Although killer whales have been shown to be capable of learning novel motor actions from conspecifics through imitation [25], the experimental evidence for production imitation skills in the vocal domain is still scarce in this species. There are reports on killer whales in the field and in captive settings indicating that they can copy novel calls from conspecifics [26,27] and even from heterospecifics such as bottlenose dolphins [28] or sea lions [24]. One Icelandic female was found to match novel calls from a Northern Resident female with whom she had been housed together for several years [26]. Two juvenile killer whales, separated from their natal pods, were observed to mimic the barks of sea lions in a field study [24]. Crance et al. [27] and Musser et al. [28] took advantage of two unplanned cross-socializing experimental situations to show that two juvenile males learned novel calls from an unrelated but socially close adult male, and three individuals learned novel whistles from a dolphin, respectively.

However, as suggestive as these reports of killer whales copying sounds from other individuals are, the lack of experimental controls curtails the interpretation about the underlying acquisition mechanisms. Experimental data are needed to ascertain whether vocal learning is a plausible mechanism underlying the complexity of vocal traditions in killer whales in the wild. However, to

77 the best of our knowledge, not even anecdotal reports exist about killer whales spontaneously
78 mimicking human speech similar to those reported in some birds (e.g., parrots [29], mynahs [30])
79 and mammals (elephants [31] seals [32], belugas [14]).

80 In most mammals, sound production occurs in the vocal folds within the larynx (the sound
81 source) and the supralaryngeal vocal tract, consisting of pharyngeal, oral, and nasal cavities (the
82 filter) [33]. In humans, this apparatus increases in complexity due to the unusual neurological and
83 motor control that we can exert on these structures [33,34]. In contrast, toothed cetaceans (e.g.,
84 killer whales, belugas and dolphins) have evolved a pneumatic sound production in the nasal
85 complex passages (instead of the larynx) involving bilateral structures such as a pair of phonic lips,
86 that can operate as two independent sound sources and filters [35,36]. This difference in the sound
87 production system between toothed cetaceans and humans make the investigation of cetacean vocal
88 production particularly valuable for comparative analyses of flexible vocal production.

89 Here we report an experimental study of sound learning and mimicry abilities of a killer whale
90 listening to familiar or novel sounds uttered by a conspecific or a human model and requested to
91 reproduce them on command ('Do this!'). The *Do-as-I-do* paradigm [37] involves the copying of
92 another's untrained (familiar or novel) motor or vocal actions under a specific previously trained
93 signal in the absence of results-based cues. The Do-as-I-do training method has been successfully
94 used in studies of primates, birds, dogs, and two species of cetaceans [12, 25, 38]. In fact, this
95 method was previously used in a study of production imitation of novel motor actions in the group
96 of killer whales studied here [25]. Ultimately, we wanted to test the long-standing hypothesis that
97 the group-specific vocal patterns documented in wild populations of killer whales can be learned
98 socially and, more specifically, through production imitation learning.

99 **2. Methods**

100 **(a) Subjects**

We tested a 14 year-old female killer whale (*Orcinus orca*), named Wikie, housed at Marineland Aquarium in Antibes, France. The conspecific model was her own 3 years old calf, named Moana, born in Marineland. Wikie had been trained for a variety of examination and exercise behaviours with standard operant conditioning procedures and fish/tactile positive reinforcement. Also, she had participated in a previous experimental study of action imitation [25], so she was already trained with the 'copy' command.

(b) Procedure

The study comprised three phases. *Phase 1* involved retraining and reinforcing the subject to respond to the gesture-based command 'copy' ('Do that!') given by the trainer, that had been used 4 years earlier in the previous study of action imitation aforementioned [25]. *Phase 2* involved testing the subject's response to the trainer's copy command when the model uttered familiar vocalizations (n = 3 different sounds), that is, vocalizations that the subject had already performed herself, either because she had been trained with them or because they were part of her natural repertoire (see Table 1). Finally, *Phase 3* involved testing the subject with novel sounds (n = 11 different sounds), that is, sounds that were unknown to the subject in terms of neither having heard them, nor having been uttered by her previously. To ensure that the unfamiliar sounds (conspecific and humans) were as different as possible from what they had produced before we compared them with 278 sound samples extracted from 'Hodgins' sound recording baseline of the vocal repertoire in this same group of killer whales [39], where she had identified up to eleven distinct discrete call types and we found no matching with our sample of novel conspecific or human sounds. In addition, before running the experiment we recorded 28 hours of in-air spontaneous sounds produced by the killer whales during their free time to see if the subject (or any other killer whale in the group) uttered sounds similar to the novel sounds in our sample. (Further details are given in the ESM.) *Phase 3* comprised two testing conditions: a conspecific model (condition 1) and a human model (condition 2). In condition 1, the subject first listened to a conspecific model's performance that included three

126 familiar sounds and five novel sounds (test trials), and then was signalled to copy them. The sounds
127 were presented in two formats: 1) performed by a killer whale model live and 2) played through a
128 speaker (e.g. conspecific sounds like airy atonal sounds as 'Breathy' and 'Strong' raspberries, or
129 tonal whiny sirens sounds like 'Wolf'). In condition 2, the subject also listened to three familiar and
130 six other novel sounds (test trials), but now they were produced by a human model (e.g. human
131 sounds like a human laugh 'Ah Ah' or human words like 'One Two' (Table S1 gives the complete
132 description of each sound). In the two conditions, the sounds were presented with the constraint
133 that no more than three consecutive test trials of the novel sound could occur in a row. In each
134 session, a single novel sound was presented to the subject at a time. We also interspersed the three
135 familiar sounds that had been used in the previous phases and control trials consisting of 'non-copy'
136 trials during which the subject's trainer did not make the copy sign and asked for any other trained
137 action that the subject regularly was requested to perform during the aquarium shows. Therefore,
138 sessions consisted of several familiar sounds and control trials and from six to ten test trials of the
139 novel sound. The subject was positively reinforced with fish and/or tactile and voice reinforcement
140 signals whenever she yielded a correct response as judged in real time by two observers (Wikie's
141 trainer and one experimenter), but only when she was asked to copy familiar sounds or perform
142 familiar actions (control trials). During the test trials (novel sounds from conspecific and human
143 models), the subject received no rewards irrespective of whether she responded correctly or not, so
144 that the experimenter did not provide any cues, thus making real time judgments unnecessary.
145 Altogether, phase 1 lasted one session, phase 2 lasted seven sessions and phase 3 fifty- two
146 sessions. All the sounds were asked and performed when the subject's head was above the water
147 surface with her blowhole exposed.

148 Three different set-ups were used. (a) *Conspecific live condition*: The two trainers (T_M and T_S;
149 M for model and S for subject) were positioned on different sides of a wooden panel 2m long x 1.90
150 cm high placed in a position in which S and M could see each other and their own trainer, but could

not see the other trainer's commands. T_M was positioned on the right side of the panel, and T_S was on the left side; thus, the trainers were in a position from which they were not able to see each other's signals either (see figure 1). (b) *Conspecific speaker condition*: two trainers were also required, one trainer held the speaker and another (T_S) gave the copy command to the subject; and (c) *Human live condition*; just one trainer was needed, as he both uttered the sound and gave the 'copy' signal (see figure 1). Table 1 gives the complete list of sounds by phase examined in this study and Table S1 gives the description of sounds. Audio samples of each demonstrated sound and of the subject's copy are available in the electronic supplementary material.

All sessions were videotaped and were recorded with Fostex Fr2 and Zoom H-4N digital recorders and a Rode NTG-2 condenser shotgun microphone. To play the sounds in the speaker condition a sound launcher app for iOS 'SoundPad Live' was developed. The sounds were played through an Ipad to an Ik Multimedia 'I Loud' portable Bluetooth speaker.

(c) Coding and data analysis

The analysis comprised two steps. In the first step we used a traditional method of categorization that consisted of using acoustic inputs and making a selection of the sounds that looked more similar [23,26,39-41]. That is, one experimenter listened to each test trial, and scored whether the subject's vocal response correctly matched the sound uttered by the model. Then, for reliability analysis, six naïve judges, blind to the model's true sound were presented with pair of sounds (model and candidate copies) and were asked to judge if the copy matched the model sample (scoring Yes for correct matching or No for non-matching) across 6 samples (3 correct and 3 incorrect, the latter chosen randomly from the pool of sounds emitted by the subject) for each demonstrated sounds.

Next, using a visual inspection of the wave form we analysed two time domain-related parameters, namely, the number and duration of bursts, of a random sample of 5 copies of each novel vocalization using Adobe Audition and then we calculated the intraclass correlation coefficient

(ICC) as a measure of concordance between model and copy sounds. The ICC for absolute agreement was estimated using a two-way random effects model.

We also run an objective detailed analysis in which the demonstrated and imitated sounds selected in the first step were subjected to an analysis of matching accuracy using algorithms implemented in Matlab version 2014a, using the signal processing toolbox Version 6.21 (R2014a) and the additional code and scripts designed by Lersch 2011 [42] available at http://www.audiocontentanalysis.org/code/. These analyses went through several steps.

First, we selected and extracted a subset of acoustic features (e.g. statistics, timbre or quality of sound, intensity-related, tonal or temporal) of both model-copy time-variant sounds that allowed us to compare sounds produced with the remarkably different acoustic modes of production aforementioned. These features form a compact informative set with respect to the desired properties of the original data [42]. All of these features were implemented using a 20 ms time window, hamming windowing, with an overlap of 50 % (hop 10 ms).

The challenge was to select in an exploratory approach a subset of these features in time and frequency domains that a priori seemed suitable for comparing sounds made by two species that use totally different production mechanisms. The main features selected were as follows: 1) *Spectral Pitch Contour ACF* (Autocorrelation Function of the Magnitude Spectrum), that shows the evolution of the fundamental frequency over time; 2) *Time Energy Evolution*, that allows to compare the evolution of the energy pattern over time between the model's and the subject's acoustic signals (temporal regularity and rhythm)**;** 3) *Pitch Class Profile,* a histogram-like 12-dimensional vector (corresponding to the 12 notes of the diatonic musical scale) with each dimension representing both the number of occurrences of the specific pitch class in a time frame and its energy or velocity throughout the analysis block [42]. Figure 2 presents an example of a *Wave form*, *Spectrogram* and *Pitch Class Profile* of the demonstrated and the copy of the human (tonal) novel sound 'Hello', and of the conspecific (atonal) novel sound 'Breathy Raspberry'

acoustic analyses. (See figures S2-S4 in the electronic supplementary material for one example for each spectral analysis for each of the main features selected and for a complete list of all features selected.)

Second, once these features were selected all the characteristics of each frame were compacted into a single vector. Finally, for the comparison it was necessary to then take into account that these signals were of different duration. We utilized a *Dynamic Time Warping (DTW)* method to deal with the alignment task, that is, with the operations of stretching and compressing audio parts allowing similar shapes to match even if they are out of phase in the time domain. DTW represents a family of algorithms developed for the automated recognition of human speech that allows for limited compression and expansion of the time axis of a signal to maximize frequency overlap with a reference signal [42]. DTW is a more robust distance measure for time series capable of quantifying similarity (or dissimilarity) in an optimal way [42] as, typically, dissimilarity function is a Euclidean distance measure that calculates and cumulates a cost according to a correspondence function (where a zero cost indicates a perfect match). That is, the higher the matching cost, the more dissimilar (less similar) the two sequences.

DTW has been widely documented and used in digital signal processing, artificial intelligence tasks such as pattern recognition (e.g., sign and gestural language), music information retrieval and signal processing, audio forensic or machine learning [42] and has recently proven to be an excellent technique for assessing matching accuracy between sounds produced by marine mammals and in particular for automatic classification of killer whale call types [43-,45]. In the present study, DTW was used to measure dissimilarity of the aforementioned acoustic subset of features that were previously selected between the audio signal of the demonstrated sound and that of the subject, revealing the extent of alignment or synchronization between both signals.

Finally, in order to establish relative comparisons between any model-copy sound pair a 'dissimilarity index' scale was constructed, which allowed us to calibrate the distance measures

obtained in the DTW analyses and thus establish how similar or dissimilar were the two sounds (demonstrated sound and that of the subject) in all the subsets of features selected. Since the dissimilarity index does not have a fixed upper limit, we rescaled the index into an interval from 0 to 1 to quantitatively assess the degree of dissimilarity. As in the non-rescaled version, 0 in this scale represents a perfect copy (i.e., a sound compared with a copy of itself) and 1 represents maximum dissimilarity. To establish this ceiling value (the top of the scale) we chose a main benchmark value, technically referred to as 'anchor'. Since the value depends on the particular vocalizations analysed, indices of dissimilarity were calculated between four randomly chosen demonstration sounds and copies uttered by the subject that corresponded to other different demonstrated sound. The benchmark value chosen was the round score closest to the maximum found (940378 score for 'Amy' paired with 'One Two Three'), which accordingly in this case was rounded to 1000000 (See ESM for a complete list of DTW dissimilarity index scores.) The rescaled dissimilarity index represents the division of the accumulated distance in relation to the distance value of the anchor of dissimilarity. Among these same four pairs of different sounds we also took the lowest score (the more similar) as another benchmark for what could be considered bad and good copies. Finally, another benchmark was included to serve as a reference point for what could be considered a 'high quality match' (i.e. a human copying another human known word). For this we calculated the dissimilarity index between the sound 'Hello' produced by the trainer and the experimenter copy of the same sound (see figure 4).

3. Results

Inter-observer reliability of whether model and subject sounds matched was high (Fleiss'weighted kappa: 0.8; $p < 0.001$; Observed agreement = 0.90).

(a) Familiar sounds

The subject correctly copied all of the trained sounds, either demonstrated by a conspecific or by a human. In *Phase 1* the subject recalled the copy command given by the trainer 4 years before as

indicated by her response in the first trial. *Phase 2* involved testing the subject's response to the trainer's copy command when the model uttered familiar sounds. With the copy signal alone the sound 'Song' was copied in the 1st trial, 'Blow' was copied in the 2nd trial (first session) and 'Birdy' was accurately matched in the 34th trial (sixth session). The criterion required for moving to the final experimental phase, i.e., 90 % of correct trials, on these three intermixed familiar sounds was achieved by Wikie in the seventh session. In *Phase 3* the subject also copied correctly all of the trained conspecific sounds performed by a human model in the transfer sessions (n = 2) and in the first trial. In sum, the subject made recognizable copies of the demonstrated sound judged in real time by two observers, Wikie's trainer and one experimenter, and then confirmed by both of them listening to the recordings.

(b) Novel sounds

The subject produced recognizable copies of all of the untrained sounds, either demonstrated by a conspecific or by a human (as judged by two experimenters that listened to the sound recordings after the test and then confirmed by 6 independent observers). In the *live conspecific condition* the novel sounds (n = 3) were copied before the 10th trial ('Strong Raspberry'), with one sound copied in the 2nd trial ('Creaking Door'), and the other in the 3rd trial ('Breathing Raspberry'). In the *conspecific through speaker condition,* the novel sounds (n = 2) were copied before the 17th trial ('Wolf'), with the other sound copied in the 6th trial ('Elephant'). In the *conspecific through human model condition* the novel sound tested (n = 1) was copied in the first trial ('Strong Raspberry'). Finally, in the *human sound condition* the novel sounds (n = 6), although they weren't perfect copies, Wikie produced recognizable copies of the human model sounds before the 17th trial ('Ah Ah'), with two sounds copied in the first trial ('Hello ' and 'One, Two, Three').

Visual examination of spectral patterns revealed a good matching of the demonstrated sound and the subject's copy in several of the acoustic features analysed. For all sound parameters tested, no differences were observed between the model's sound and the subject's match in the *total*

number of bursts (Cohen's kappa = 1, p < .0005). When tested with novel conspecific sounds, a high concordance was found between *burst duration* of the model's sound and the subject's copy (ICC: 0.79; $p < 0.001$, N=31 bursts). When tested with human sounds, a very high concordance between *burst duration* of model's sound and subject's copy was found (ICC: 0.89; $p < 0.001$, N=65 bursts) showing better performance compared to killer whale sounds.

In the automated quantitative analysis, the DTW showed an optimal overlap represented by a diagonal line alignment between both sounds (demonstrated and copy) in all the examples for each sound judged by the experimenters as correct imitations in phase 1. This diagonal line alignment of the 'shortest line' between both signals indicated similarity in all features selected [42]. Figure 3 presents an example of a DTW analysis in the matching of the subject's and the human model's for the sound 'Hello' (tonal); the conspecific's novel sound 'Breathy Raspberry' (atonal) and the familiar sounds 'Birdy' (tonal) and 'Blow' (atonal). (See figure S1 in the electronic supplementary material for one DTW example of all the others novel sounds tested). Although the fundamental frequency of copies made by human and killer whale models was remarkably dissimilar, the outline F0 contours turned out to be very similar. Figure 4 shows a representation of a DTW distance dissimilarity index between the demonstrated sound and the best match (the lowest DTW value) among the random sample of 5 copies of each vocalization type of the subject for each and every sound tested plus four 'incorrect' reference control points (corresponding to randomly chosen demonstrated sounds paired with copies that corresponded to other different subject's sound and another 'high quality copy' reference control point (human copying another human known word), (see ESM for a complete list of DTW dissimilarity index scores). Overall, expected matches (when demonstration and copy were of the same sound type) did match, while expected non-matches (when demonstration and copy were of different sound types) did not. Specifically, we found that copies of *familiar* conspecific sounds fell below a dissimilarity index threshold (horizontal red dotted line below the lowest incorrect random pair copy) that divided our results in good or bad copies and most of them were close to the 'high quality match' score (human imitating human

anchor), with one score being below this value ('Blow'). Copies of *novel* conspecific sounds were located very close to this 'high quality match' score and *novel* speech sounds demonstrated by humans were distributed across the whole range of good copies with one even below this 'high quality match' benchmark. If we take as a criterion of matching accuracy the values obtained with familiar sounds from conspecifics, we observe that except for the sound 'Blow', which is the simplest untrained sound consisting only of a single burst of atonal voiceless breath (see electronic supplementary material second example on Sound File Nº1), the copies of *novel* conspecific sounds and three of *novel* speech sounds ('Amy', 'Hello' and 'Ah Ah') were even more closely matched than were tonal *familiar* conspecific sounds.

Finally, analysing the features selected for the DTW analysis separately, the spectrogram analysis revealed that the subject produced harmonics when exposed to tonal sounds, but not when exposed to atonal or noisy sounds (see figure 2 and electronic supplementary material figures S2-S4). This pattern held even for the human tonal sounds.

4. Discussion

Although the subject did not make perfect copies of all novel conspecific and human sounds, nonetheless, they were recognizable copies as assessed by both external independent blind observers and the acoustic analysis. There was great variability in the number of good copies produced after a sound was copied for the first time (Table 1). Possible factors that could explain this variability are the difficulty in producing novel sounds and some uncontrolled factors such as variation on motivational levels and social dynamics across sessions. Additionally, our non-differential reinforcement regime (good copies of novel sounds were not reinforced to avoid shaping) may have also contributed to this variability. Consequently, it is conceivable that our data represent a conservative estimate of the killer whale's capacity for vocal imitation.

According to the DTW dissimilarity scale (figure 4), all the copies of novel conspecific utterances fell below the dissimilarity index threshold for good and bad copies (pairs of different

demonstrated and copied sounds randomly chosen) and most of them were close or even fell below the 'high quality match' score, as represented by the human-copying-human anchor. Similarly, although three of the copies of human sounds were only close to the dissimilarity index threshold for good and bad copies (incorrect randomly paired copies), the other three fell close to the 'high quality match' score (human imitating human anchor); that is, they were very accurate copies, with one falling even below this benchmark. This level of accuracy is particularly remarkable given that the subject possessed a very different sound production system compared to humans. Some parameters such as the fundamental frequency were sometimes drastically different between the human model and the killer whale copies, but the outline F0 contours were nonetheless quite similar (figure 4).

Overall, the DTW analyses revealed that the accuracy of copies was much higher when these were of the same sound than when they involved a different sound, which strongly suggests that the copies were specific to the demonstrated sound. We believe that the subject's responses represent a case of *vocal imitation* rather than response facilitation, as the latter form of social learning does not apply to individuals reproducing a model's novel sound [46]. Moreover, the subject's perfect performance in the control 'non-copy' trials in which she was requested to perform a trained action or sound different from that of the model, ruled out automatic response facilitation (i.e., copying the model's sound spontaneously) [46] because she only copied was she was requested to do so.

DTW analyses also revealed that the subject's copies of *novel* conspecific and human sounds were in most cases even more accurate than were the copies of *familiar* sounds. Thus, in three of the novel speech sounds ('Hello', 'Amy' and 'Ah Ah'), the accuracy of the copies was even greater than the matching accuracy of some of the familiar sounds uttered by the conspecific model. Moreover, four copies of novel sounds were found to be high quality matches, as they were close to the benchmark score of a human copy of the human sound, and one was even a better match (see 'Breathy Raspberry' in figure 4). A greater copying accuracy for novel compared to familiar sounds

might suggest that the cognitive mechanisms responsible for producing familiar and novel sounds do not fully overlap. It is possible that the matching of familiar sounds relies more heavily on response facilitation than imitation where the subject's copy is mainly shaped by the *general* characteristics of the stored representation than by the sound's *specific* individual components. In contrast, learning to match a novel action or sound might require the subject to carefully process the individual components of the auditory experience, which might generate a better match. The subject's matching accuracy is all the more remarkable as she was able to accomplish it (a) in the absence of extensive trial-and-error across all the experimental conditions, (b) in response to sounds presented in-air and not in-water (the species' usual medium for acoustic communication), and (c) in the case of her matching of speech sounds, through the use of a sound production system that greatly differs from that of the model's [35, 36]. Note that the subject readily matched the harmonic quality of human tonal sounds (see figure 2 and electronic supplementary material figures S2-S4). The anatomical structures involved in sound production of cetaceans differ from those used by terrestrial mammals and birds in that cetaceans are adapted to an aquatic lifestyle where the sound producing organs compress while diving because of water pressure related changes [35]. This has been hypothesized to have favoured the development of vocal learning in marine mammals as they need to have a substantial voluntary control over sound production in order to successfully meet the demands of reliably generating the same sounds at different depths [47].

Our experimental findings lend support to the hypothesis that the group-differentiated acoustic dialects that have been documented in many field studies of killer whales [16-23] and other cetaceans [10] can be acquired and maintained through social learning and, more specifically, through imitation. These results add to the growing database of socially learned sounds reported in previous non-experimental and experimental studies of killer whales and other cetaceans (dolphins [11-13]; belugas [14,15]). As a mammalian order, cetaceans stand out for their complex sociality, elevated encephalization, and advanced cognitive skills [48]. Compared to the fission-fusion societies of bottlenose dolphins, however, the social systems of killer whales are reported to be

more strongly structured and closed [10,16]. Thus, the well-developed propensity of killer whales to copy what others are doing, that is, to translate visual or auditory input into motor responses that conform to the group's norm would be consistent with the body of observations on group-specific acoustic dialects, synchronized behaviour, and sophisticated cooperative strategies documented in this species [10].

The results reported here show that killer whales have evolved the ability to control sound production and qualify as open-ended vocal learners. It can be argued that since our experimental design included in-air (rather than in-water) sounds, the positive results obtained cannot directly reflect the killer whales' capacity for learning to copy underwater sounds in their natural environment. However, our main objective was to test whether the killer whales were capable of learning novel sounds through imitative learning, regardless of the type of sound (in-air vs. in-water) and the model (conspecifics vs. heterospecifics). The atypical nature of the sounds that we used represents a strength rather than a weakness in relation to our main question because it demonstrates flexibility not just on *what* is copied but on *how* is copied. With regard to what is copied, our data demonstrate that killer whales can copy sounds outside their usual repertoire – which is an important piece of information if one wants to know not only know what a species does, but also what it can do, under a variable set of circumstances. With regard to the issue of how it is copied, our data might indicate that the sensory-perceptual and cognitive skills recruited in imitating in-air sounds are ancestral traits, dating back to the terrestrial ancestors of cetaceans. Moreover, given the highly derived state of the sound producing apparatus uniquely evolved by cetaceans, the imitative capacities found in this study also underscore the fine-tuned ability of this species to flexibly produce accurate matches of heterospecific in-air sounds.

Future experimental studies of imitation of in-water sounds demonstrated by conspecifics are needed to firmly establish the role of social learning in the killer whale's vocal dialects documented in the wild. Another challenge for future research is to ascertain whether the neural and cognitive

403 scaffolding for vocal learning in cetaceans and humans (and other taxa) are homologous or
404 analogous, and whether they are adaptations or have been co-opted for new fitness-enhancing
405 functions in the unique suite of environmental challenges they encounter in the seascape they
406 inhabit [6,10,11,47,48]. Finally, we extended DTW analysis used in previous studies [39,44,45] by
407 incorporating several additional features of killer whales' demonstrated and imitated sounds into the
408 algorithm. However, these results must be taken with caution because the choice of features was
409 exploratory. Further studies are thus needed to standardize the assessment of the matching accuracy
410 of different sound features as well as the validation of the dissimilarity index. Although we see
411 great potential in this analytical approach for comparative studies of vocal learning, its applicability
412 may vary depending on the study's objectives, the sounds investigated, and the species' vocal
413 production system.

414

References

1. Zentall TR. 2012 Perspectives on observational learning in animals. *J. Comp. Psychol*. **126**, 114. (doi: 10.1037/a0025381)
2. Boyd R, Richerson PJ, Henrich J. 2011 The cultural niche: why social learning is essential for human adaptation. *Proc. Natl. Acad. Sci. USA* 108 (Suppl 2), 10918–10925. (doi: 10.1073/pnas.1100290108)
3. Whiten A, Van Schaik CP. 2007 The evolution of animal 'cultures' and social intelligence. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **362**, 603-620. (doi: 10.1098/rstb.2006.1998)
4. Janik VM, Slater PJB. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1-11. (doi:10.1006/anbe.2000.1410)
5. Tomasello M. 2011 Human culture in evolutionary perspective. *Adv. Cult. Psychol.* **1**, 5-51.
6. Colbert-White EN, Corballis MC, Fragaszy DM. 2014. Where apes and songbirds are left behind: A comparative assessment of the requisites for speech. *Comp. Cog. Behav. Rev*. **9**, 1-28. (doi: 10.3819/ccbr.2014.90004)
7. Hauser M, Chomsky N, Fitch WT. 2002 The Language Faculty: What is it, who has it, and how did it evolve? *Science* **298**, 1569-1579. (doi: 10.1126/science.298.5598.1569)
8. Seyfarth RM, Cheney DL. 2014 The evolution of language from social cognition. *Curr. Opin. Neurobiol.* **28**, 5-9. (doi: 10.1016/j.conb.2014.04.003)
9. Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW. 2016. Vocal fold control beyond the species-specific repertoire in an orang-utan. *Sci. Rep*, **6**.
10. Whitehead H, Rendell L. 2014 *The cultural lives of whales and dolphins*. University of Chicago Press.
11. Janik VM. 2014 Cetacean vocal learning and communication. *Curr. Opin. Neurobiol.***28**, 60-65. (doi: 10.1016/j.conb.2014.06.010)
12. Herman LM. 2002 Vocal, social, and self-imitation by bottlenosed dolphins. In *Imitation in animals and artifacts* (eds K. Dautenhahn C. Nehaniv), pp. 63-108.Cambridge, UK: MIT Press.
13. 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)*. *J. Comp. Psychol.* **98**, 10
14. Ridgway S, Carder D, Jeffries M, Todd M. 2012 Spontaneous human speech mimicry by a cetacean. *Curr. Biol.* **22**, R860-R861. (doi: 10.1016/j.cub.2012.08.044)
15. Murayama T, Iijima S, Katsumata H, Arai K. 2014 Vocal imitation of human speech, synthetic sounds and beluga sounds, by a beluga (*Delphinapterus leucas*). *Int. J. Comp. Psychol.* **27**.
16. Filatova OA, Samarra FI, Deecke VB, Ford JK, Miller PJ. Yurk, H. 2015 Cultural evolution of killer whale calls: background, mechanisms and consequences. *Behaviour* **152**, 2001-2038. (doi:10.1163/1568539X-00003317)
17. Ford JK. 1991 Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Can. J. Zool*. **69**, 1454-1483. (doi:10.1139/z91-206)
18. Deecke VB, Ford JK, Spong P. 2000 Dialect change in resident killer whales: implications for vocal learning and cultural transmission. *Anim. Behav*. **60**, 629-638. (doi: 10.1006/anbe.2000.1454)
19. Miller PJ, Bain DE. 2000 Within-pod variation in the sound production of a pod of killer whales, Orcinus orca. *Anim. Behav.* **60**, 617-628.
20. Filatova OA, Miller PJO. 2015 An agent-based model of dialect evolution in killer whales. *J. Theor. Biol.* **373**, 82-91. (doi: 10.1016/j.jtbi.2015.03.020)

21. Weiß BM, Symonds H, Spong P, Ladich F. 2011 Call sharing across vocal clans of killer whales: Evidence for vocal imitation? *Mar. Mam. Sci.* **27**, E1-E13. (doi:10.1111/j.1748-7692.2010. 00397)

22. Filatova OA, Deecke VB, Ford JK, Matkin CO, Barrett-Lennard LG, Guzeev MA., ... Hoyt, E. 2012 Call diversity in the North Pacific killer whale populations: implications for dialect evolution and population history. *Anim. Behav.***83**, 595-603.

23. Yurk H, Barrett-Lennard L, Ford JKB, Matkin CO. 2002 Cultural transmission within maternal lineages: vocal clans in resident killer whales in southern Alaska. *Anim. Behav.* **63**, 1103–1119. (doi:10.1006/ anbe.2002.3012)

24. Foote AD, Griffin RM, Howitt D, Larsson L, Miller PJ, Hoelzel AR. 2006 Killer whales are capable of vocal learning. *Biol. Lett.* **2**, 509-512.

25. Abramson JZ, Hernández-Lloreda MV, Call J, Colmenares F. 2013 Experimental evidence for action imitation in killer whales (*Orcinus orca*). *Anim. Cogn.* **16**, 11-22.

26. Bain DE. 1986 Acoustic behavior of Orcinus: sequences, periodicity, behavioral correlates and an automated technique for call classification. *Behavioral biology of killer whales*, 335-371.

27. Crance JL, Bowles AE, Garver A. 2014 Evidence for vocal learning in juvenile male killer whales, (*Orcinus orca*), from an adventitious cross-socializing experiment. *J. Exp. Biol*. **217**, 1229-1237.

28. Musser WB, Bowles AE, Grebner DM, Crance JL. 2014 Differences in acoustic features of vocalizations produced by killer whales cross-socialized with bottlenose dolphins. *The Journal of the Acoustical Society of America*, **136**, 1990-2002.

29. Pepperberg IM. 2010 Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain. Lang.* **115**, 81-91.

30. Klatt DH, Stefanski RA. 1974. How does a mynah bird imitate human speech? *J. Acoust. Soc. Am.* **55**, 822–832.

31. Stoeger AS, Mietchen D, Oh S de Silva S, Herbst CT, Kwon S, Fitch WT. 2012 An Asian elephant imitates human speech. *Curr. Biol.***22**, 2144-2148.

32. Ralls K, Fiorelli P, Gish S. 1985 Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina. Can. J. Zool. **63**, 1050–1056.

33. Fitch WT. 2006 Production of vocalizations in mammals. *Vis. Commun.* **3,**145.

34. Fitch WT, de Boer B, Mathur N, Ghazanfar AA. 2016. Monkey vocal tracts are speech-ready. *Sci. Adv*. **2**, e1600723.

35. Cranford TW, Amundin M, Norris KS. 1996 Functional morphology and homology in the odontocete nasal complex: implications for sound generation. *J. Morph.* **228**, 223-285.

36. Tyack PL, Miller EH. 2002. Vocal anatomy, acoustic communication and echolocation. Marine mammal biology: An evolutionary approach, 142-184.

37. Hayes KJ, Hayes C. 1952. Imitation in a home-raised chimpanzee. -*J. Comp. Physiol. Psychol.* **45**, 450-459.

38. Huber L, Range F, Voelkl B, Szucsich A, Viranyi Z, Miklosi A. 2009 The evolution of imitation: what do the capacities of non-human animals tell us about the mechanisms of imitation? *Phil. Trans. R. Soc. B.* **364**, 2299‑2309. (doi:10.1098/rstb.2009.0060)

39. Hodgins-Davis A. 2004 'An Analysis of the Vocal Repertoire of the Captive Killer Whale Population at Marineland of Antibes, France,' Thesis, Wellesley College,

40. Ford JK. 1989 Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia. *Can. J. Zool.,* **67**, 727-745.

41. Deecke VB, Ford JK, Spong P. 1999 Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (*Orcinus orca*) dialects. *J. Acoust. Soc. Am*. **105**, 2499-2507.

42. Lerch, A. 2012 Temporal Analysis. An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics, 119-137.

43. Deecke VB, Janik VM. 2006 Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. *J. Acoust. Soc. Am*. **119**, 645-653.

44. Brown JC, Hodgins-Davis A, Miller PJ. 2006 Classification of vocalizations of killer whales using dynamic time warping. *J. Acoust. Soc. Am*. **119**, EL34-EL40. (doi:10.1121/1.2166949)

45. Brown JC, Miller PJ. 2007 Automatic classification of killer whale vocalizations using dynamic time warping. *J. Acoust. Soc. Am.* **122**, 1201-1207. (doi:10.1121/1.2747198)Byrne RW. 2002 Imitation of novel complex actions: What does the evidence from animals mean? *Adv. Study Behav*. **31**, 77-105. (doi: 10.1016/S0065-3454(02)80006-7)

46. Hoppitt W, Laland KN. 2008. Social processes influencing learning in animals: a review of the evidence. *Adv. Study Behav*. **38**, 105-165.

47. Tyack PL, Sayigh LS. 1997. Vocal learning in cetaceans. Social influences on vocal development. **11,** 208.

48. Marino L, Connor RC, Fordyce RE, Herman LM, Hof PR et al (2007) Cetaceans have complex brains for complex cognition. PLoS Biol **5** 966–972. (doi:10.1371/journal.pbio.0050139)

Electronic Supplementary Material is available in the online content of the paper and at https://figshare.com/s/2991d28752ca0690e843. This includes methods details, raw data, figures S1-S5, and 12 audio files examples (ESM Audio File S1: 3 conspecific familiar sounds; ESM Audio File S2.1-S2.5: 5 conspecific novel sounds; ESM Audio File S3.1 – S3.6: 6 human novel sounds).

Ethics. The Ethics and Animal Welfare Committee (CEBA-MEDUC) of the School of Medicine, Pontifical Catholic University of Chile, have approved this research. This research adhered to the legal requirements of the country (France) in which the work was carried out and Marineland institutional guidelines.

Data accessibility. The datasets supporting this article have been uploaded as part of the supplementary material and at https://figshare.com/s/2991d28752ca0690e843.

Author Contributions J. Z.A. conceived the study. J.Z.A., M.V.H.LL and J.C. designed the experiment, which was conducted by J.Z.A. and M.V.H.LL. M.V.H.LL designed and carried out the data analyses and interpretation. J.Z.A. and L.G. performed the sound analyses and interpretation. J.Z.A. and M.V.H.LL drafted the paper. J.Z.A. and F.C. co-wrote the paper. J C. and F.A. helped to write and provided critical revisions and some of the ideas in the paper. J.ZA., F.C., F.A. and J.C. provided financial support. All authors have read and approved the final manuscript.

Acknowledgements Special thanks go to Andrea Hodgins for sharing her data and making her thesis available to us and to Manuel Castellote for his advice on sound analyses. We are grateful to the directors of Parques Reunidos and the Marineland Aquarium of Antibes, Jesús Fernández and Jon Kershaw for allowing us to conduct this research. Furthermore, we appreciate the work of head coaches Lindsay Rubinacam and Duncan Versteegh, and all orca trainers team for their help and support. Special thanks go to Amy Walton for her dedication and training of the killer whale model Moana. Finally we want to thank Francisco Serradilla from Universidad Politécnica de Madrid for developing the sound launcher app for iOS.

526 **Funding** This project was conducted at the Marineland Aquarium Antibes, France and supported by a Postdoctoral Scholarship
527 FONDECYT Nº 3140580 to J.Z. Abramson. This study was partly funded by project grants PSI2011-29016-C02-01, PSI2014-
528 51890-C2-1-P (MINECO, Spain) and UCM-BSCH GR3/14-940813 (Universidad Complutense de Madrid y Banco Santander
529 Central Hispano) to F. Colmenares.

530 **Competing interests:** The authors declare that they have no competing interests.

531 **Author Information** Correspondence and requests for materials should be addressed to J.F.Z.A (zabramson@psi.ucm.es)

532

533

Figure 1. Experimental set up. (a) Conspecific live condition: The two trainers (T_M and T_S; M for model and S for subject) were positioned on different sides of a wooden panel 2m long x 1.90cm high placed in a position in which S and M could see each other and their own trainer, but could not see the other trainer's commands; (b) conspecific speaker condition: One trainer holds the speaker and another (T_S) gave the copy command to the subject; and (c) human live condition; Just one trainer was needed, as he both uttered the sound and gave the 'copy' signal.

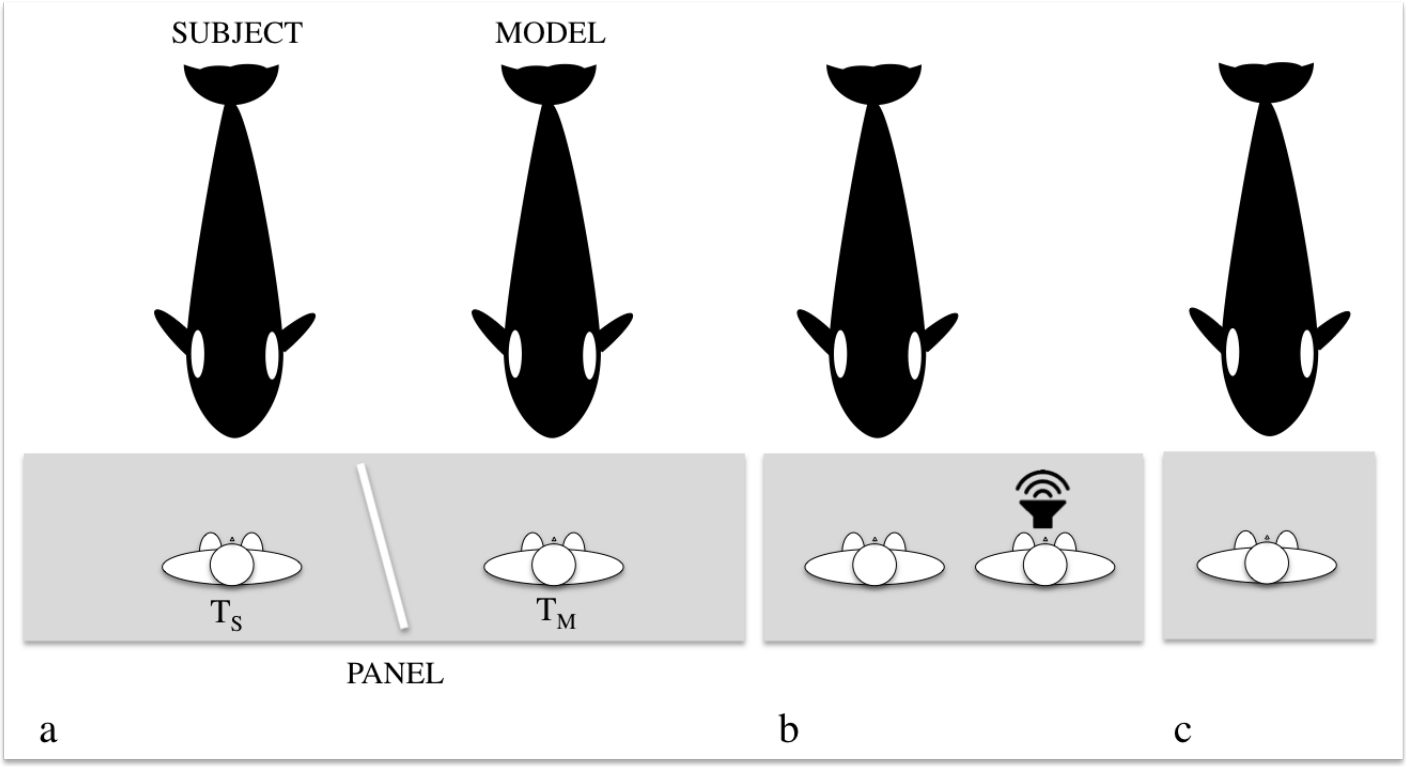
Figure 2. *Wave form and spectrogram* of the model (a1) and the copy (a2) of the human (tonal) novel sound 'HE'. Note the harmonic pattern in both signals. 'HE' *Pitch Class profile* of the model (b1) and the copy (b2) *Wave form and spectrogram* of the model (c1) and the copy (c2) of the conspecific (atonal) novel sound 'BR'. Note the in harmonic pattern in both signals. 'BR' *Pitch Class profile* of the model (d1) and the copy (d2)

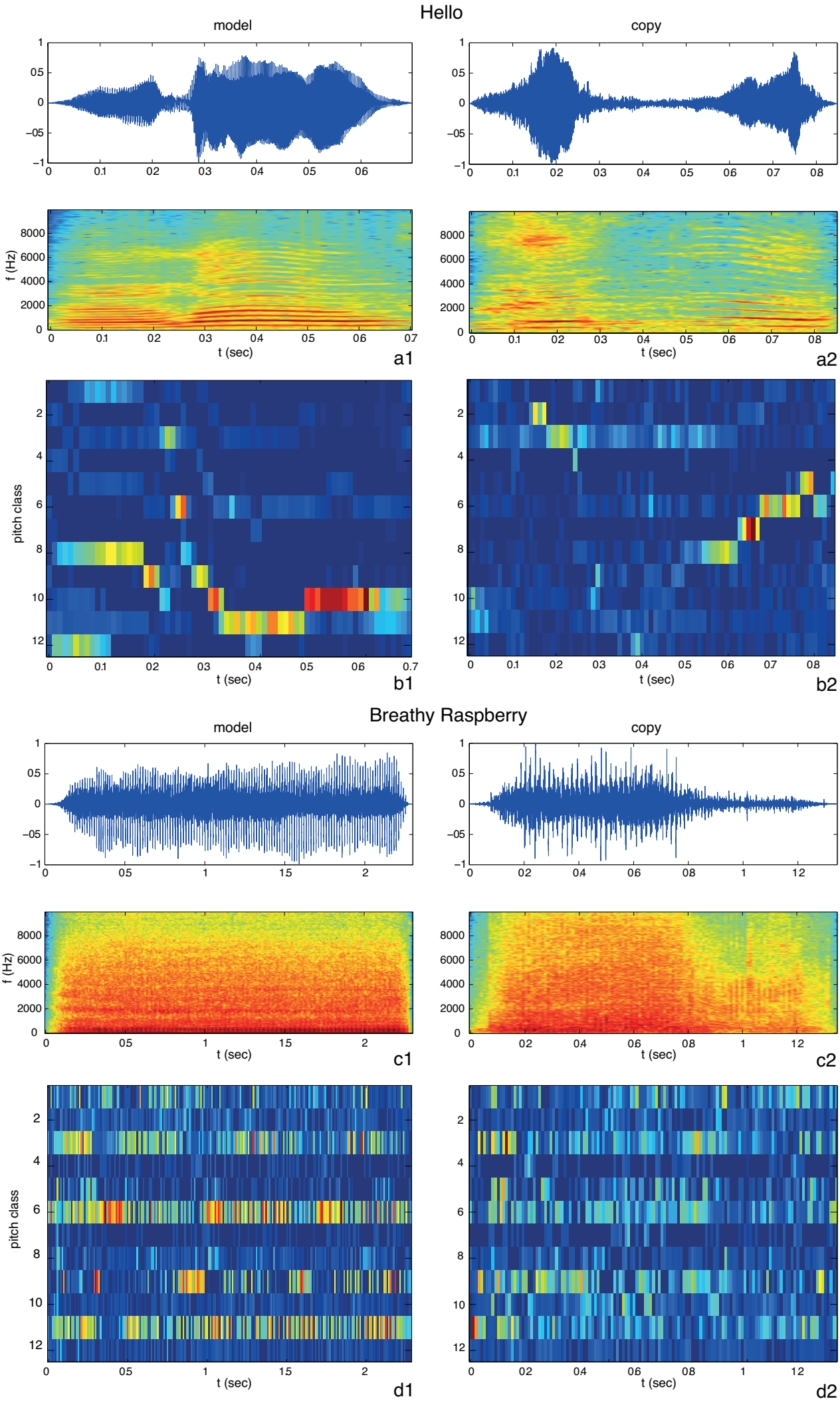
Figure 3. Dynamic Time Warping Familiar and Novel Conspecific and Human Sounds (Tonal and Atonal). In both axes all the characteristic features of the signals are aligned and the black line shows the. shortest path (minimum distance) between the model and the observer sounds streams.. (a); DTW familiar sound 'BL' (atonal) of the model and the copy (b); DTW familiar sound 'BI' (tonal) of the model and the copy (c); DTW novel sound 'HE' (tonal) of the model and the copy (d); DTW novel sound 'BR' (atonal) of the model and the copy.

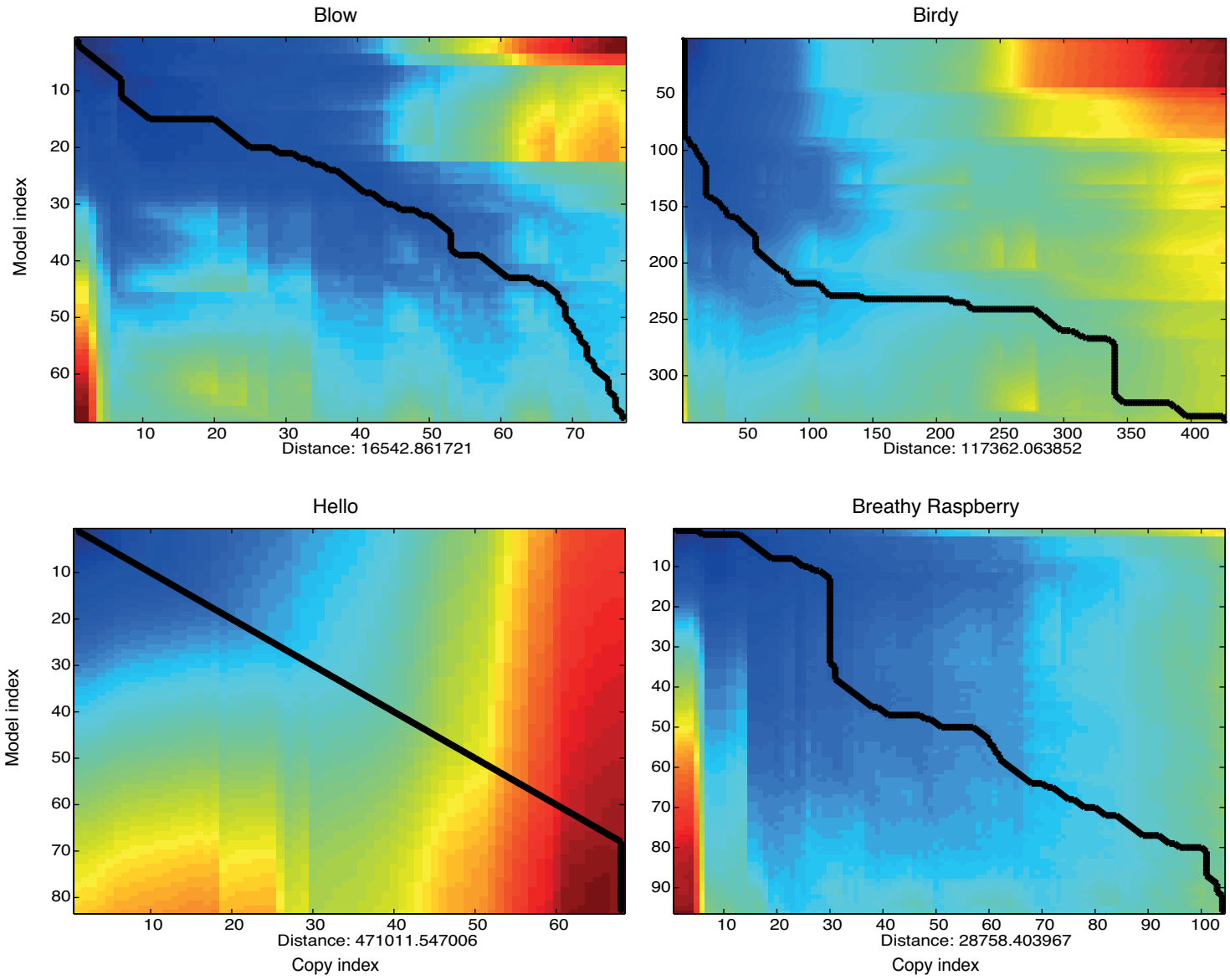
Figure 4 Dynamic Time Warping dissimilarity index distribution. Distribution of the DTW dissimilarity index between the model and the copy for each vocalization; familiar (blue dots), killer whale novel (green dots) and human novel (turquoise dots). Five control benchmarks (red dots) separated by a red vertical dotted line are also represented, where the first one correspond to the 'high quality match' score (human imitating human benchmark) and the others correspond to the four randomly chosen incorrect copies (model sounds paired with copies that corresponded to other different models). The horizontal red dotted line below the lowest incorrect random pair copy serves as a benchmark for dividing the results between good and bad copies.

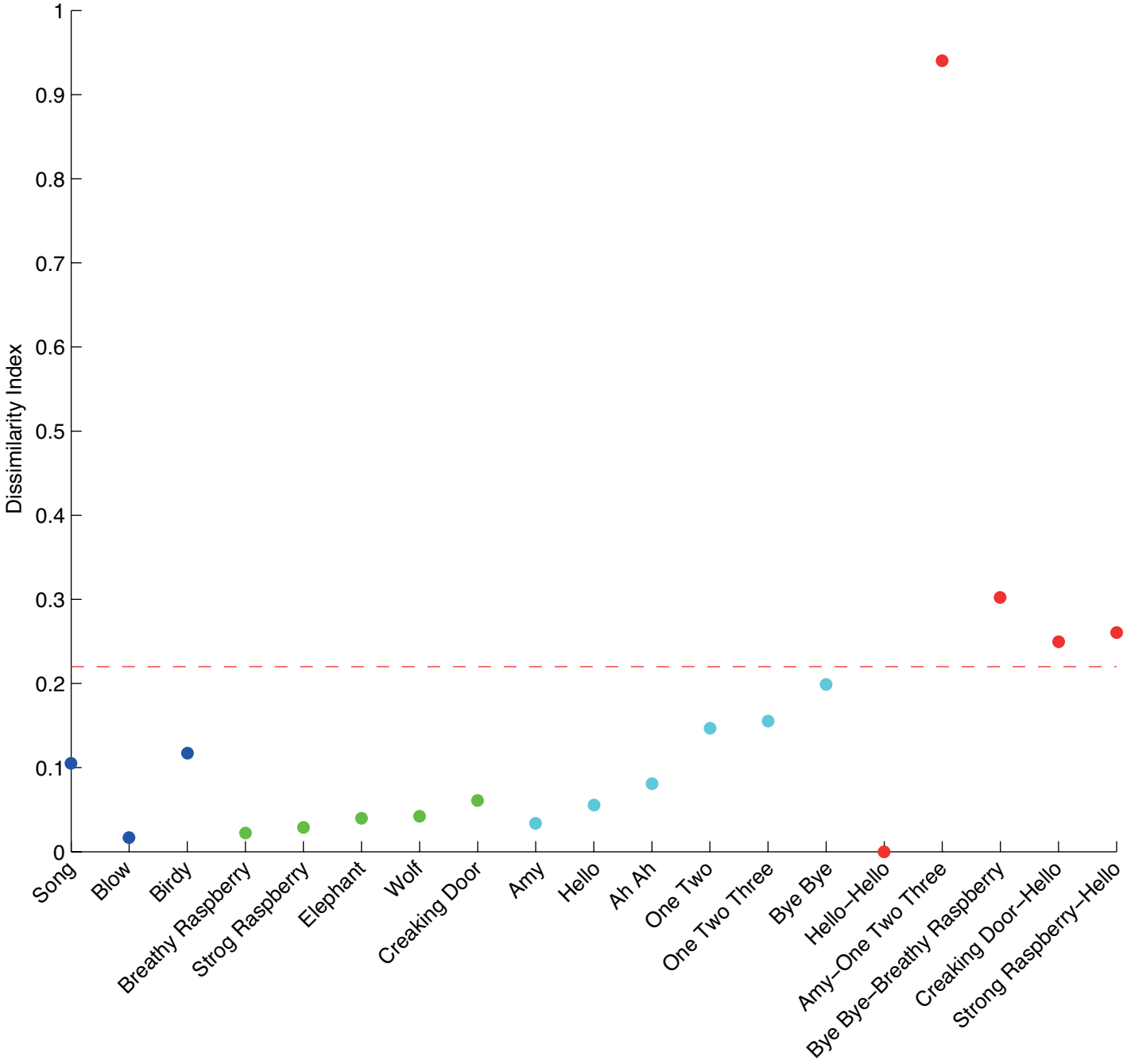
Table 1. Total number of trials for each sound tested, number of trials until the model's sound was judged to be copied by the subject (according to two experimenters that listened to the sound recordings after the test and then confirmed by 6 independent observers), and percentage of correct trials since the first full copy

	No. of trials	First trial copied	% correct since the 1st copy
FAMILIAR SOUNDS			
Song (SO)	394	1	100
Birdy (BI)	316	34	98
Blow (BL)	371	2	99
Through human model (transfer sessions)			
SO	30	1	100
BL	30	1	100
NOVEL SOUNDS			
Conspecific Alive Model			
Strong Raspberry (SR)	30	10	19
Creaking Door (CD)	30	2	100
Breathy Raspberry (BR)	30	3	30
Conspecific through speaker			
SR	30	1	100
CD	30	4	44
BR	30	1	57
Wolf (WO)	30	17	36
Elephant (EL)	30	6	28
Conspecific through human model (transfer sessions)			
SR	30	1	100
Human			
Ah Ah (AA)	30	17	14
Hello (HE)	30	1	55
Bye Bye (BB)	30	12	21
Amy (AM)	30	8	26
One Two (OT)	30	3	36
One Two Three (OTT)	30	1	23


SUBJECT
MODEL
T_S
T_M
PANEL
a
b
c


Hello
model
copy
f (Hz)
t (sec)
a1
a2
pitch class
b1
b2
Breathy Raspberry
c1
c2
d1
d2


Blow
Birdy
Hello
Breathy Raspberry
Model index
Model index
Distance: 16542.861721
Distance: 117362.063852
Distance: 471011.547006
Distance: 28758.403967
Copy index
Copy index


Dissimilarity Index
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Song
Blow
Birdy
Breathy Raspberry
Strog Raspberry
Elephant
Wolf
Creaking Door
Amy
Hello
Ah Ah
One Two
One Two Three
Bye Bye
Hello–Hello
Amy–One Two Three
Bye Bye–Breathy Raspberry
Creaking Door–Hello
Strong Raspberry–Hello