Exploration of Aural & Visual Media About Birds Informs Lessons for Citizen Science Design

Jessica L Oliver[1], Margot Brereton[1], Selen Turkay[1], David M Watson[2], Paul Roe[1]
[1] Queensland University of Technology, Brisbane, Australia
[2] Charles Sturt University, Albury-Wodonga, Australia
jl.oliver@hdr.qut.edu.au; [m.brereton, selen.turkay, p.roe]@qut.edu.au; DWatson@csu.edu.au

ABSTRACT
Acoustic sensing has been hailed as a game-changer for detecting furtive wildlife, but uptake has been constrained by the laborious process of reviewing resultant torrents of audio data. To inform the design of interactive interfaces for reviewing audio recordings, we explored how people interact with aural and visual media about birds. We observed how twelve participants with different levels of interest in birds engaged with vocalization recordings, visualizations of bird calls, photographs, and range maps of three species. By conducting thematic analysis, we identified a variety of *Challenges of Exploration* and *Benefits of a Media Assortment*. We contribute lessons for designing to *Bridge Knowledge & Context* and to *Facilitate Long-term Engagement* with audio in ways that are fun, accessible, and informative. We provide explicit guidance for designers to diversify how citizen scientists interact with nature through audio as they move from engagement to conservation action.

Author Keywords
Audio interactions; aural; birds; citizen science; multisensory; nature connection; visual; wildlife.

CCS Concepts
• Human-centered computing~Human Computer Interaction

INTRODUCTION
As the field of Interaction Design expands to address *more than human* concerns, we investigate how to design more engaging technologies that elicit curiosity and desire to understand nature. Recent advances in technology allow members of the public to engage with collection, analysis, and interpretation of scientific data in new ways as citizen scientists [30, 45]. However, relatively little design research investigates interaction in citizen science. Human Computer Interaction (HCI) explorations of citizen science have been restricted to hobbyist practices [11, 14, 46], gamification [5, 6, 17, 22], interface relationships [21, 23, 55], and online data analysis [62]. Still, the role of design in engaging and retaining participants is unclear. We investigated how people engage with multisensory information, to inform the design of compelling interactions to elicit and grow passions about the sights and sounds of nature.

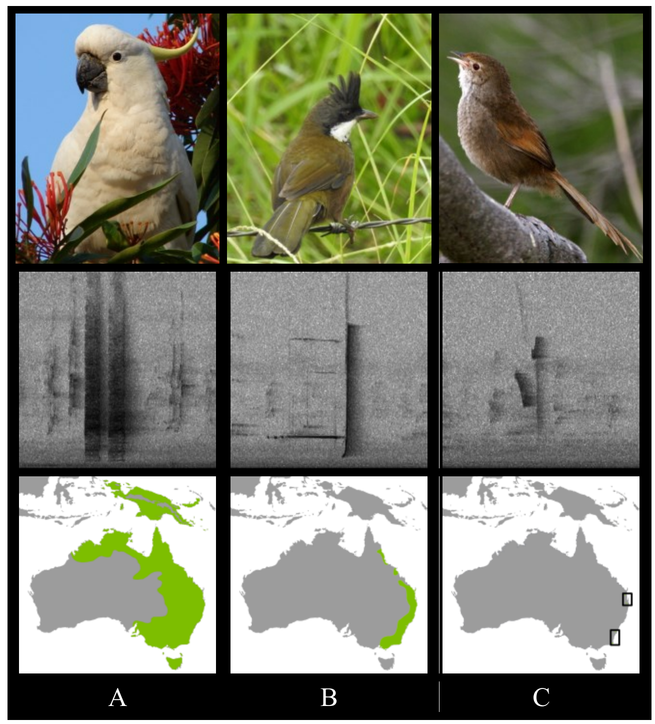

Figure 1. Participants matched printed visual & aural media for three bird species: A) Sulphur-crested cockatoo; B) Eastern whipbird; & C) Eastern bristlebird. Media included photographs, call visualizations (i.e. spectrograms), call audio, & distribution maps. Boxes in Map C outline regions where Eastern bristlebirds live in small, isolated populations. Bird photos: A) © Tatiana Gerus; B) © Philip Venables; & C) © David Cook. Spectrograms: © Jessica L Oliver. Map data [4].

Birds are ubiquitous and can often act as indicators of ecosystem health [40]. Some species are difficult to find via traditional field surveys due to being furtive and/or rare (e.g. Australia's night parrot [43]). One of the most promising ways to find birds that are difficult to see is to collect audio recordings from their presumed habitats (i.e. acoustic sensing), then search the audio for calls made by the bird (e.g. [32]). However, manual searching through audio is often required, and the process is slow requires expertise, which limits use acoustic sensing for wildlife conservation.

DIS '20, July 6–10, 2020, Eindhoven, Netherlands

ACM ISBN 978-1-4503-6974-9/20/07…$15.00
https://doi.org/10.1145/3357236.3395478

Several recent DIS and HCI studies highlight design opportunities and challenges with scaling up citizen science participation to support wildlife conservation [19, 38, 42]. Both conservationists and citizen scientists need ways to explore audio, exchange knowledge, and cooperatively identify animal calls [42]. However, the few citizen science projects that have included reviewing audio have suffered relatively low participation [16, 62]. We explore how people interact with and match diverse media about birds, to inform the design of more engaging citizen science (Figure 1).

Although reviewing calls in an online environment is far less engaging for people than transcribing notebooks or sorting camera-trap photographs [16], citizen scientists have enjoyed recording calls of wild animals *in situ* [49]. While recording calls in the wild, people are surrounded by nature with the potential of seeing animals. Online citizen science projects, by contrast, often involve simplistic tasks [48], not designed to embody nature broadly. According to [31, pg 265], "*… integration of information from different sensory modalities has many advantages for human observers, including increase of salience, resolution of perceptual ambiguities, and unified perception of objects and surroundings*". We posit that audio will be more relatable if augmented with other types of media.

We aimed to explore how people interact with different types of bird information, such as what birds look and sound like, as well as where they live, to inform the design of audio interactions for future citizen science projects. Our exploratory study was guided by the multi-faceted research question: *How are interactions with bird information influenced by 1) information type; 2) media modality; 3) prior experience; 4) social dynamics; 5) characteristics of different bird species?* To investigate this, we created a multi-sensory activity, including calls, spectrograms (visuals of audio), photos, and distribution maps of three bird species (Figure 1). Each media type was provided to 12 participants in stepwise stages, either independently or in small groups. We then identified patterns of media interaction.

We contribute two themes and lessons for those interested in designing for audio, citizen science, and/or nature engagement. The main themes identified include 1) *Challenges of Exploration* and 2) *Benefits of a Media Assortment*, with three and four subthemes per respective theme highlighting interaction processes. We then educed two key lessons by exploring how themes relate to existing technologies that engage people with nature sounds and broader literature. *Lesson 1* is to design to *Bridge Knowledge, Context, & Goals*. Subthemes describe how this can be done via 1) *Media Salience*; 2) *Accessible Associations*; 3) *Species Information & Perceptions*. *Lesson 2* is to design to *Facilitate Long-term Engagement*. Subthemes demonstrate how this can be done by designing for 1) *Growth of Knowledge with Purpose*, 2) *Diversity in Task Difficulty*, 3) *Customizable Complexity*, 4) *Collaborative Puzzle Solving*. Our work defines new pathways of interaction with audio in meaningful ways for those interested in designing for citizen science, nature engagement, and other audio interactions.

BACKGROUND

Audio Has Potential to Support Wildlife Studies

Wildlife can be difficult to find if they are furtive, rare, nocturnal, and/or inhabit inaccessible terrain, but acoustic sensing holds promise as part of an alternative way of locating where such species live. Acoustic sensors (i.e. audio recorders) are often deployed to record sounds made by animals [65] and use signatures in audio to confirm and understand their occurrence patterns. Sounds made by animals that are captured on recordings may come from vocalizations, stridulating (i.e. rubbing body parts together), movements, and other mechanisms. With audio recorders becoming less expensive, more efficient, durable, smaller, and more useable (e.g. AudioMoth [24, 25]), more environmental audio is being recorded. Analysis of these large audio files, however, remains largely intractable. Technical expertise is required to manage large datasets (e.g. [36, 63]) and to develop automated recognition algorithms, which may still have limited effectiveness (e.g. [47]). Extensive call knowledge of the desired species is also required to accurately annotate a dataset (e.g. even to train algorithms). Innovative ways to review environmental audio recordings are needed to facilitate wildlife studies.

Citizen Science with Audio Analysis is Limited

Many online citizen science projects are hosted on the rapidly expanding Zooniverse platform [16, 71]. Originally the platform was designed to classify galaxies in photographs [34]), and today, most projects are visually oriented and tasks are designed so that no prior knowledge is required for participation [16, 62]. Tasks include annotating, classifying, counting, drawing, outlining, and transcribing different types of media [48]. Image classification projects often have high rates of participation, whereas the two projects that involved the classification of audio both visually and aurally had comparatively lower levels of engagement [16]. The identification accuracy of calls was variable for both bats [35] and whales [54]. The low engagement with environmental audio, and variable accuracy identifying sounds, indicates a need to design audio interactions that are both fun and informative for people.

Design May Uncover Audio Interaction Barriers

The DIS and HCI community has begun to explore how to facilitate engagement and sensemaking with audio interactions. Exploratory work highlights that playful interactions with devices that include aural and visual representations of bird calls (i.e. via audio recordings and spectrograms respectively) and photographs are engaging to people [8, 51, 57]. Stories and knowledge were also elicited from birders reviewing soundscape visualizations [41]. These birders, however, craved more environmental context to interpret spectrograms, such as a list of species that reside where recordings were obtained. Cooperatively exploring audio aurally, and visually by providing spectrograms at multiple resolutions, is likely to aide interpretation [18, 41].

Recent work has begun to analyze sound interpretation and interactions quantitatively. At the soundscape level, people can distinguish between recordings from different but very similar areas (i.e. different urban streets [29]). Specific sounds, however, are less consistently annotated in more complex soundscapes that include a variety of sounds [10]. Analysis of audio is also more efficient, in terms of time and accuracy, with the use of visual representations called spectrograms [10]. Considerations when designing for annotating audio include weighing the need to recall a sound versus precisely identifying it [9]. Such quantitative studies would be complemented by design studies that explore how people relate to and make sense of audio aurally and visually.

Design Needed to Support Conservation with Audio

The DIS/HCI community has begun exploring interactions with audio to support conservation efforts. Researchers have explored practices of people interested in learning about birds, such as amateur birders [14, 41], an endangered species recovery team [42], and a community group [19]. Social groups of birders often cooperatively listen to calls, search for birds, and identify species. The ability to listen to audio online has the potential to extend a birder's hobby, for example, by allowing them to confirm and track what was heard *in situ*, listen to new birds or habitats, and communicate with others about aural observations [14]. Fun and engaging designs are needed that allow conservationists and citizen scientists to easily review audio, exchange knowledge, and cooperatively learn [42]. Interactions with audio are also likely to be more meaningful if designs foster community building [41] and integrate culturally significant aspects [19]. Despite progress in our understanding of communities, interactions with audio are understudied.

METHODS

This section details our activity design rationale, participant recruitment, collection of data, and analysis for this study.

Activity Design Rationale

We created and trialed an exploratory paper- and audio-based activity to investigate the experiences of participants interacting with media about three species of birds. We used bird photos, bird call audio, spectrograms, and species distribution maps (Figure 1), which allowed for the exploration of using multiple senses and crossmodal perception.

A researcher's role and positionality are key considerations, as they influence the research method and participant interactions [28]. The primary author has a background in ecology and extensive experience teaching people about birds in the wild, which she drew upon to create the activities. To explore people's familiarity with birds and interpretation of diverse types of media, three Australian species were selected based on differences in behaviors, appearance, and ubiquity. The Eastern whipbird was selected as a bird more recognizable by its distinctive whip-like call than by sight, given its furtive nature, for people familiar with forests outside the city. The Sulphur-crested cockatoo was chosen as a species that can be easily seen and heard both within and outside the city. Lastly, to explore how people would interact with unfamiliar species, the locally critically endangered Eastern bristlebird was selected. We anticipated people would relate to media from each species in a unique way.

Participants were provided media in a stepwise fashion, receiving one type of media at a time, until all media had been acquired. Typically, all three spectrograms were provided first, followed by three photographs, then three audio recordings, and lastly the three distribution maps. Two exceptions made to the order of media were provided to participants to account for different participant backgrounds, with the intent for the activity to have an appropriate difficulty level. In one case, for example, E1 had researched bristlebird ecology extensively, so photographs were not revealed to her until the last step, given the likelihood she would too easily identify the species by sight. In the other instance, by contrast, photographs were shown to F1 and F2 first, rather than spectrograms, given their newness to Australia, and people's propensity to become familiar with birds by sight before by calls. Adapting media order proved effective, given all participants completed the activity while finding it sufficiently challenging.

Participant Recruitment

We recruited members of the public with varying levels of experience taking notice of birds or engaging with audio. Purposive sampling was conducted to recruit a diverse group of participants, representative of people who have a broad interest in the subject matter and propensity to become citizen scientists (Table 1). Given that many citizen science projects require high levels of participation to deliver rigorous scientific outcomes, this study aimed to investigate how to design technologies that would be sufficiently enticing to people, with a focus on offering alluring aural and visual interactions with media about nature. Two participants with knowledge of the Eastern bristlebird's plight were recruited from the recovery team for the species (i.e. D1 and E1), to consider whether prior knowledge of bristlebird conservation needs would influence interactions with the activity.

Most participants learned of the project through the lead author speaking at public events (e.g. A1, A2, and A3 joined after hearing a talk to birdwatchers) and advertising through university courses (e.g. F1 and F2 learned of the project through their classes; Table 1). Originally, we planned for participants to individually complete the activity; however, several requested to participate with peers or brought peers to the interview. We welcomed this as an opportunity to explore not only how people interacted with media, but also with each other. Particularly given that people interested in birds are known to 1) identify animals cooperatively [14], 2) share information with peers using citizen science technology (e.g. eBird [58]), and 3) help each other make sense of nature media [41]. Three participants completed the activity individually (i.e. C1, D1, and E1); whereas, nine completed the activity in groups of two or three (i.e. A, B, F, and G indicate group membership; Table 1).

ID	Age	Professions	Experiences with Birds
A1 A2 A3	31 38 28	Biologist; Veterinarian; Artist;	A1 has expertise identifying cockatoos calls in audio; A2 is more practice birding then A3, though both are novices
B1 B2	73 70	Retired statistician; Retired tradesman	B1 is an experienced birder & develops call recognizers; B2 is an experienced birder who searches audio for calls
C1	27	Honors student (ecology)	C1 is an experienced birder who searched audio for calls
D1	49	Wildlife sanctuary manager	D1 enjoys leisurely birding & oversees captive breeding of wildlife (e.g. bristlebirds)
E1	26	Ph.D. student (ecology)	E1 researches ecology & conservation of bristlebirds
F1 F2	35 26	Master's students (info technology)	F1 & F2 are international students with interest in birds but no experience
G1 G2	37 39	Master's student (tech & innovat. management) Sound designer & audio engineer	Neither G1 or G2 actively go birding, but both casually observe birds in the wild (without using tools such as binoculars or bird guides)

Table 1. Twelve participants, their respective professional backgrounds, and experiences with birds.

Data Collection & Analysis

For this study, we employed "think-aloud protocols" [27], asking participants to verbalize their thoughts about the experience of completing each stage of the activity and decision rationale for actions taken. When progress ceased, the researcher asked and answered questions to elicit a discussion of issues encountered. Dialogs elicited during the completion of the activity were audio-recorded and transcribed. Transcripts were thematically analyzed using inductive methods [7]. We coded each session line-by-line, then examined codes across sessions to identify indicative patterns relevant to informing design for audio analysis. As an exploratory study of a specific context, we do not claim our findings to be generalizable but suggest readers will be able to judge applicability to their study contexts.

FINDINGS

We analyzed transcripts and observational notes, which included participant inquiries, rationale, and actions, to assess how people related to media about birds. The two primary themes identified include 1) *Challenges of Exploration* and 2) *Benefits of a Media Assortment*, with respective subthemes highlighting interaction processes.

Theme 1: Challenges of Exploration

As participants explored media, they often had challenges, which are organized into subthemes: *Interpreting Spectrograms*, *Deciphering Call Diversity*, and *Connecting Media with Bird Lives*.

Interpreting Spectrograms

Prior experience interacting with spectrograms helped the sensemaking process, though most participants had difficulty interpreting spectrogram variables (Figure 1 &Figure 2A). G1, who is unfamiliar with visualizing sound, thought that the y-axis represented amplitude (i.e. loudness). G2 responded by correctly speculating that this axis was in frequency (i.e. kHz), and clarified that loudness was represented by the "*intensity of the darkness*". G2 works with music audio, which likely explains his more nuanced understanding of the spectrogram (Table 1). In another session, D1 was unfamiliar with spectrograms and had difficulty interpreting visuals literally. Instead, he related the shape of the cockatoo call in the spectrogram to "*... a highway, or car tracks*" (Figure 1A & Figure 2B). His idea demonstrates that spectrograms can elicit imaginative, metaphorical thinking. Interpreting spectrogram attributes is challenging regardless of experience, and creative interactions may be one way to help people relate bird calls to spectrogram variables.

Figure 2. Participants interacting with bird media printouts & audio recordings. A) A1, A2, & A3 inspecting spectrograms with the researcher; B) D1 describing a spectrogram looking like "car tracks"; C) B1 & B2 taking turns matching media; D) E1 matching maps to spectrograms.

Often participants did not notice and struggled to decipher specific call features. Parsing one sound from another visually was difficult for some. E1, for example, asked whether shadows were part of an individual bird's call or background noise. At times, certain characteristic features of focal calls also went unnoticed, such as when F1 did not realize that the horizontal line in the spectrogram was part of the whipbird call (Figure 1B). A variety of factors influence how sounds are captured in audio recordings and shown in spectrograms. Participants rarely considered these factors when exploring spectrograms, such as how a bird’s ability to control the volume of its call, how close the bird is to the audio recorder, and the conditions the environment that impact the travel of calls (e.g. call attenuation in dense vegetation). Only B2 questioned how the closeness of a bird to the recorder would influence its shape. Future designs need to illuminate both diagnostic features of target calls and variation in call appearance for specific species via training tutorials to support citizen scientists learning to identify calls accurately.

Deciphering Call Diversity

Some participants had extensive prior knowledge of the call diversity of specific species (i.e. vocal repertoires), which sometimes lead to incorrect assumptions of what calls were included in the activity. B1, for example, knew that female whipbirds respond to male calls immediately with two loud, sharp notes when present (Figure 3A & B). He was unsure how a female whipbird call would look in a spectrogram, however, which prompted him to ask whether there was an antiphonal female response call in the spectrogram. Similarly, D1 originally assumed a female whipbird would have been replying to a male in the spectrogram (Figure 3A & B). Conversely, he initially thought of a single, loud “RARRK” call for the Sulphur-crested cockatoo, rather the two-note call we used (Figure 1A &Figure 3C). His knowledge of broader call diversity for both species caused him incorrectly pair whipbird and cockatoo spectrograms with incorrect photos. Despite prior knowledge of call diversity leading to incorrect assumptions, participants enjoyed reasoning their decisions and recognizing their misidentifications. Such mistakes led to participants with prior knowledge gaining a greater appreciation and familiarity with commonly confused patterns and species. Understanding the knowledge and assumptions that prospective participants have allows designers to create interactions that provide prompts and information to avoid potential confusion.

This exploratory study exemplified that not all bird calls are equally memorable or recognizable. When selecting species, we suspected would be most recognizable by sight and sneaky whipbirds would be most memorable by their distinctive call. Likewise, as we anticipated, only participants involved in efforts to ensure the survival of bristlebirds as a species (i.e. D1 and E1) had any familiarity with their media (Table 1). Spectrograms were new to both D1 and E1, and neither person attributed a spectrogram showing a bristlebird call. Recordings for this activity were obtained from captive bristlebird aviaries within earshot of D1’s office, so he was familiar with their calls. In contrast, E1 studied wild habitats of bristlebirds, only having serendipitously hearing calls of sparsely populated wild bristlebirds. As such, E1 said the call sounded familiar. Upon learning the call was made by a captive bristlebird, she questioned whether that was a “*normal*” call, and remembers the wild bristlebird calls she had heard were usually a three-note call. E1’s observation aligned with a prior study, as the person most knowledgeable about wild bristlebirds also noted that calls from captive birds seemed “corrupt” or different to calls of wild birds [42]. This reinforces that when possible, understanding the vocal repertoire of a species (within and across populations) is important, as is considering what influences a people’s propensity to see, hear, recognize, and remember a bird.

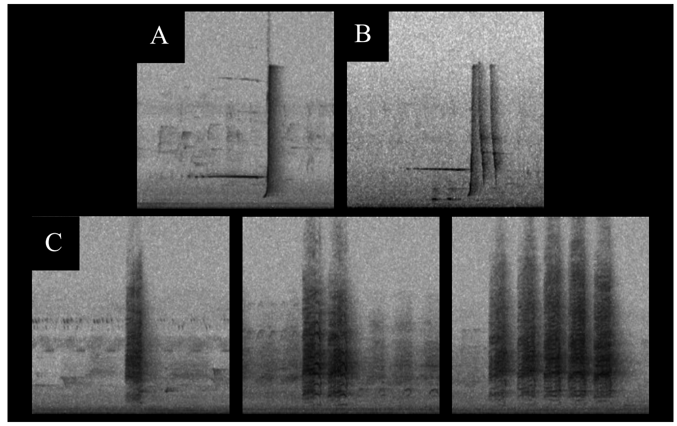

Figure 3. Spectrograms showing call variation for two species. A) a male Eastern whipbird call; B) a male whipbird vocalization, followed by a two-note, female response call; & C) a series of three Sulphur-crested cockatoo calls [x = Time (5 to 7 secs); y = frequency (0 to 11 kHz)] © Jessica L Oliver

Connecting Media with Bird Lives

While participants struggled to relate to the lives of birds through inspecting media, the activity elicited interests in learning about birds. Looking at the same photo, F1 assumed the bristlebird was “*really young*”, because of its brown body coloration. She was fascinated to learn, however, bristlebirds and a myriad of other birds stay brown throughout their lives, and that it’s the bristlebird’s eye irises that change color with age, from bright red to brown. F1 also inquisitively asked whether the frequency of a bird’s song is related to its body size. She thought if visible physical aspects of bird anatomy influenced a “*different kind of singing…*”, and were indicated, then she “*would be able to match them*”. Providing broad information, such as similarities in calls, plumage changes, and distinguishing physical attributes, for groups of species, may evoke more nuanced inspection and understanding of media.

Map matching proved difficult for most participants, though the process revealed map utility, as well as elicited critical thinking and curiosity from participants. When attempting to match maps of the cockatoo and whipbird, E1 deliberated in detail what types of habitat she had seen each species, distribution of those habitat types, and other information she knew about broader species of birds in the (Figure 1A & B). While she eventually mismatched the maps, maps engaged her with considering the environmental context that influences birds' lives more than other media. While E1 initially felt that maps were tangential and “*too hard*” for most people, she later envisioned maps could be a useful mechanism to sort birds likely to be in audio by habitat types.

The meaning of the green coloration of maps was also ambiguous to those who had not seen distribution maps, but an exploration of this also evoked broader ideation for future map use. While inspecting the bristlebird map, for example, F1 questioned whether birds were abundant or rare in respective colored areas (Figure 1). She imagined being able to use maps in quizzes that would challenge her to learn associations between bird species and their habitats. While matching was difficult, maps interactions revealed how people grapple these unfamiliar media, elicited ideas for potential use, and unveiled that maps are useful when interested in gaining deeper context for the lives of birds.

Theme 1 Recap

Challenges of Explorations became apparent with all types of media the activity included. *Interpreting Spectrograms* was difficult largely due to a lack of familiarity with figure variables and with deciphering calls. Challenges with *Deciphering Call Diversity r*esulted from either people not having enough knowledge of calls or incorrectly applying prior bird behavior and call knowledge to provide from the vocal repertoire of a species. *Connecting Media with Bird Lives* was difficult was particularly for participants with the least prior knowledge of bird physiology, behaviors, ecology, and distribution. newest to taking notice of birds. We gained insights into participant knowledge differences and found that regardless of prior knowledge levels all media was interesting to participants. Interactions with media often elicited promising creative, critical, and curious thinking from participants. Further work is needed to identify fun and effective ways to illuminate nuanced differences between calls, appearance, and habitat uses of birds, whether related or different species.

Theme 2: Benefits of a Media Assortment

We identified there were several benefits of using an assortment of media, which are organized into subthemes: *Eliciting Enthusiasm*, *Conjuring Diverse Knowledge*, *Resolving Ambiguity*, and *Solving Through Cooperation*.

Eliciting Enthusiasm

Many positive expressions and interactions indicated participants were enjoying the activity. Statements of interest and enjoyment, such as "*Nice*", "*Sweet*", "*Oh, goody gumdrops*" (A1), "*Cool*" (C1), and "*Yay*" (E1), were enthusiastically and regularly made by participants when excited by a task or receiving affirmation via teammates or the interviewer. Others were interested in knowing answers, like D1 when asking, "*Hang on, Hang on, am I right? …*". Lighthearted sarcasm also indicated engagement, particularly with A1, A2, and A3, who had different levels of experience birding (Table 1 & Figure 2A). Given her experiences with audio, A1 accurately identified one spectrogram shape as "*cockatoo*" and the other as a "*passerine*" (i.e. songbird; Figure 1A & C, respectively). A2 jokingly said to the interviewer, "*you asked which bird…. She's just giving you genres, man*". Jovial, rapid exchanges (e.g. G1 and G2) and methodical puzzling through each stage (e.g. B1 and B1) indicated participants being engrossed. E1 enjoyed learning she could see three dots in the spectrogram for the male whipbird, and hearing the sound for the first time, saying *That's quite cool*" (Figure 1B & Figure 2D). Such positive feedback signaled that, while the task is challenging, people enjoyed interacting with diverse media.

Novices were provided more context about birds to facilitate advancement, and this provided an enjoyable experience. Upon receiving maps, F1 shared that matching maps was interesting since she learned "*a little bit*" about the three bird species from previous steps and that she would not have been interested if asked to match maps at the beginning of the activity. Both F2 and F1 were compelled to share broad familiarity with birds. Activity interactions evoked F2's memories of pet cockatoos being noisier than other birds and producing similar sounds to humans. This prompted F1 to describe how some birds are "*very talented*" at mimicking sounds heard, including even camera shutters. Superb lyrebirds are likely the species F1 described given that this species has a propensity to mimic even mechanical sounds commonly heard, which is regularly featured on popular nature documentaries (e.g. [2]). Once finished, F1 also reflected that the activity is enjoyable and challenging and that completing it evoked competitive feelings even though she has not engaged with identifying birds before. Interacting with bird media was engaging for novices when provided enough context to relate to birds.

Conjuring Diverse Knowledge

Several participants accurately identified species by the spectrogram alone, by drawing from their diverse prior nature and/or audio experiences (Table 1). Both A1 and C1 were immediately able to identify the cockatoo call by the spectrogram having a "broadband" call showing (Table 1). And likewise, C1 also immediately identified the whipbird call from the spectrogram given his previous experiences. By contrast, E1 had no prior experience with spectrograms, but she was able to interpret the shape based on her experiences hearing the distinctive call of the whipbird when in the forests of South East Queensland. Others Most others who identified the whipbird call by spectrogram did so through collaborative discussion (e.g. G1 and G2; A1, A2, and A3). Participants recognizing calls from prior audio and outdoor explorations provide a promise that with the right experiences, calls can be visually and aurally memorable.

Participants similarly successfully applied prior experiences in-the-wild to other media types as well (Table 1). While not identifying the cockatoo from the spectrogram alone, E1 immediately identified the call audio, saying, "*Oh, so that's a parrot… is that just a cockatoo?*", and then correctly guessed it was a Sulphur-crested cockatoo. With species identifications being clarified when needed to allow progression, participants with prior experience successfully matched distribution maps, as individuals (i.e. D1 and C1) and collaboratively (i.e. A1 and A2; G2 and G1; B1 and B1).

Despite B1 asking about the bristlebird photo, "*That's a grassbird isn't it?…*", and both B1 and B2 mismatching spectrograms with photos of a whipbird and a bristlebird, B2 confidently said, "*I do know the distribution*", and the pair quickly matched maps with photos correctly then demonstrated their expertise (Figure 2C). G2 described his ability to match maps from having travel the country widely and added, "*Knowing the distribution of the cockatoo I can say with relative comfort that that's what that is*", tapping on the correct map, and G1 concurred. Diverse media interactions elicited varied forms of knowledge, whether the person is a casual bird observer or a birder who actively searches for birds in leisure time.

Resolving Ambiguity

Ambiguity arising in the early stages could be resolved when audio recordings were heard. Initially, D1 mismatched spectrograms and photos of the cockatoo and whipbird, for example, but with the addition of call audio, he realized the in-the-wild calls from his memories were different than calls included in the activity (Figure 3). D1 concluded, "*hearing the call and seeing that, I can clearly say 'well that's clearly the whipbird'… there's that constant sound*". Similarly, for cockatoo media, D1 said, "*hearing how loud that particular call was, that double call, it makes sense. And they're close together, 'Raak, raak', very clearly to me spells out that call*". He also said he would have "*instantly*" identified calls if hearing audio first, which reflects our media order was appropriate to maximize interactions and engagement. Being able to interpret and resolve ambiguity with the addition of audio demonstrates the benefits of crossmodal interactions.

Similarly, photographs and maps helped participants correct mismatches and resolve uncertainty. Despite researching and surveying habitats of bristlebirds, for example, E1 did not identify the bird by spectrogram or audio of its call, but maps provided the clue she needed. When seeing maps, she said, "*… I still don't know what this one is… it's not bristlebird is it?*". She laughed and continued, "*That would be pretty bad if it was bristlebird… But the distribution looks very much like bristlebird*". Being provided photographs last, given her familiarity with captive bristlebirds, she exclaimed, "*Oh, so it is a bristlebird?!*". The photos validated the map hunch in this instance. Providing diverse forms of media allowed people to relate to birds in ways that resonated most for them.

Solving Through Cooperation

Cooperative discussions often resulted in correct answers being reached based on diverse bird experiences (Table 1). G2 described, for instance, one of the spectrograms as "*a low, like a mid-sustained note, followed by a burst*", which prompted G1 to whistle the sound, and then to recognize what he had whistled as being a whipbird call. G2 then agreed and pointed out the bird was an Eastern whipbird.

People with mixed levels of skill cooperatively solved tasks as well, such as A1, A2, and A3 also demonstrated through interpreting the whipbird spectrogram. A2 initially wasn't able to interpret the spectrogram, but A1 described the features of the spectrogram shape, which prompted both A3 and A2 to ask A1 more about loudness and frequency. Immediately after A1 shared more of her knowledge, A2 excitedly asked, "*Is that a whipbird?*". Exchanging knowledge and impressions allowed participants to cooperatively correctly solve matches.

Not all instances of cooperative problem solving resulted in correct identifications, but interactions were still informative. B1 and B2 were one group who exhibited such cooperation (Figure 2C). B1 correctly matched cockatoo media but inadvertently mismatched the whipbird and bristlebird spectrograms and photos. Questioningly, B2 replied, "*You reckon that's a cockatoo?*" B1 replied, "*Yeah, that's my matching…*" (Figure 2C). Referring to the B2's correctly matched cockatoo, B2 deliberated, "*I would have thought that one might have been whipbird…*". B2 inquired if the single-note call of a male or the two-note call indicative of a female call (Figure 3B). With the interviewer confirming a male call, B2 reasoned that it might be the cockatoo. Regardless of the accuracy of the identifications, cooperation elicited deep thinking on how media and experiences relate.

Theme 2 Recap

Benefits of a Media Assortment became evident as participants interacted with diverse types of media with aural and visual modalities. Doing so was effective in *Eliciting Enthusiasm* as evidenced by expressions of excitement, explicit statements of enjoyment, and relating aspects of the activity to popular media. Interactions were effective in *Conjuring Diverse Knowledge* of participants that have experiences interacting with audio of bird calls and/or engaging with observing birds in the wild. Providing diverse forms of media allowed for *Resolving Ambiguity*. Such interactions allowed people opportunities to relate interactions to their own experiences in diverse ways. We also learned the potential for *Solving Through Cooperation*.

DISCUSSION

Citizen science projects involving audio classification have had low levels of often brief participation [16, 62], and those that involved classifying spectrograms had variable accuracy levels [35, 54]. An audio-focused project that is slow to acquire annotations of wildlife calls can still be useful for biological and ecological studies (e.g. [52]). When furtive vocal wildlife is under threat of extinction, however, time is of the essence, and species can only be saved if where they reside can be rapidly identified and protected. (e.g. [42]). This exploratory study has begun to unpack the complex challenges and opportunities for diversifying how citizen scientists interact with audio and other media to learn about wildlife and support conservation. Next, we explored how our findings relate to broader literature and technologies focused on interactions with audio, birds, and nature more broadly. From this exploration, we offer lessons highlighting the importance of designing to 1) *Bridge Knowledge & Context*, and 2) *Long-term Engagement*. We offer the lessons to designers interested in exploring interactions for designing citizen science, nature engagement, and audio exploration.

Lesson 1: Bridge Knowledge & Context

Interactions with nature media to review audio, should *Bridge Knowledge, Context, & Goals*. This can be done by designing for 1) *Media Salience*; 2) *Accessible Associations*; 3) *Species Information & Perceptions*.

Media Salience

Interactions with each type of media revealed challenges. Making sense of spectrograms represented the most complex and difficult component. Achieving a consensus on call identification is less likely when soundscapes are complex [10]. We found that even people who were familiar with spectrograms benefitted from prompts, such as hints and questions, to assist aural and visual interpretation of audio. In several instances, specific nuanced features of calls would have gone unnoticed without cues being provided. Applications like Cornell Lab of Ornithology’s Bird Song Hero [12] (Figure 4) engage people with matching birdcall audio recordings with associated spectrograms. The spectrograms included in Bird Song Hero, however, lack salience that participants in our study regularly sought. The application’s introduction to the spectrogram variables of frequency, time, and amplitude in an induction video, is somewhat cursory, and the calls in spectrograms are modified, cropped, and centered, which makes comparing call frequencies impossible (Figure 4). In Bird Song Hero, people are prompted to focus on recognizing patterns to associate calls aurally and visually and have no need to notice the photograph of the bird. Presenting interrelated media is key to growing curiosity and allowing people to link their knowledge, deeding to deeper interest in birds.

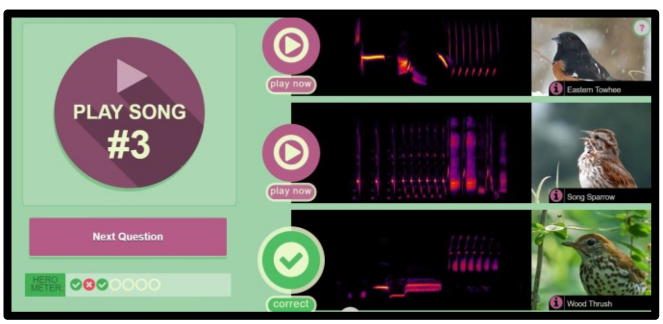

Figure 4. Screenshot of Bird Song Hero’s second level.

In reviewing raw spectrograms, our participants had difficulty interpreting how call variables interacted. They also found it hard to decipher focal calls from other calls and background noises, which is necessary when working with sound from the wild. Making sense of spectrograms would be easier if people were provided with interactive tools that allowed them to experiment with how variables, such as frequency, time, and amplitude, influence call shape. Illuminating specific features of target calls would also help to ensure key features did not go unnoticed.

Beyond interpretation of spectrograms, interactions with bird photos and species maps also revealed ways to increase salience that allows for more meaningful engagement. Designs need to facilitate making linkages between different bird media, whether comparing songbirds to parrots, different cockatoo species, or the song repertoire of a specific species. Prototypes can also explore how best to introduce additional types of media about birds, such as photographs of plumage variation, habitats, nests, or behaviors.

Accessible Associations

Playful designs for engaging with bird call audio can evoke enjoyable experiences [51, 57]. Our findings highlighted how the experience shaped enjoyable interactions. Participants with *in situ* experience observing birds related more easily to at least one media type, and they were most interested in being provided nuanced call information. By contrast, those with little to no experience birding or reviewing audio needed a great deal more context regarding bird appearance, calls, anatomy, and distribution, as well as how these attributes interrelate. When exploring media in groups with varied experience levels, those with the most knowledge largely dominated, with inexperienced people deferring to them. Relevant designs are needed for people with differing levels of experience with birds and their calls. If designing to facilitate cooperation between people with differing levels of experience, providing experienced people with moderation roles may foster long-term engagement. The influence of bird familiarity, impact of audio knowledge, and design of cooperation needs further research.

Species Information & Perceptions

While we purposefully selected three different species that we thought would pose different kinds of challenges, a paradox was revealed. As anticipated, participants were most able to relate to a spectrogram showing the call of a male Eastern whipbird given the common, frequent, and distinctive “*whip-crack*” call [3, includes a URL to audio of a male and female calling]. People with a broad familiarity with birds in the wild often interpreted calls more easily than those with in-depth knowledge of bird calls. Experienced birders in our study made assumptions about what calls would be included based on their more nuanced knowledge. However, calls included in the activity for Sulphur-crested cockatoos and Eastern whipbirds differed from presumed calls, though we did not attempt to select unusual calls. Even a person familiar with Eastern bristlebirds calls from the wild (i.e. E1) expected the call included to be different. Experiences with individual bird species shaped how people interacted with bird media provided. How people relate to calls of a specific species, as well as similar-sounding species that reside in the same region need careful consideration.

Knowledge of threatened species vocalization can be limited to a few individuals, and cooperation is key to acquiring essential knowledge [42] Collaborative research between species experts and designers is essential to ensure the call repertoire is captured and the most informative calls are included in designs (e.g. the most common and/or unique calls for the species; [42]). Our study also exemplified the importance of designing with novices, who we see as prospective citizen scientists, to understand common knowledge, assumptions, and confusion when aurally and visually exploring audio about birds. Understanding what calls are most memorable and resonate with people is crucial. Awareness of difficulties with deciphering calls can inform design of playful and gameful [20] training approaches that allow citizen scientists to become skillful at identifying calls from audio accurately. Including conservationists and citizen scientists in the design process is likely to remove ambiguities and ease the process of searching for calls [15].

Lesson 2: Facilitate Long-Term Engagement

Designs to interact with environmental audio and other media should *Facilitate Long-term Engagement*. Subthemes indicate that this can be done by designing for 1) *Growth of Knowledge with Purpose*, 2) *Diversity in Task Difficulty*, 3) *Customizable Complexity*, 4) *Collaborative Puzzle Solving*.

Growth of Knowledge with Purpose

Prospective citizen scientists were interested in learning complicated information, and we see similarities with amateur musicians. Birders and amateur musicians, for example, are both intrinsically motivated and invest time in growing complex skills [26, 37]. Playing the video game Guitar Hero, which allows people to simulate being musicians [1], may spark novices to have an interest in music. Engagement with the application, however, is likely to be transient and frivolous for amateur musicians interested in gaining virtuosity [37]. Bird Song Hero is analogous in its goals being extraneous, with the aural and visual matching of only a single call per species needed to earn "hero metric" points (Figure 4) [12]. Long-term engagement with identifying calls in audio is most likely to occur when offered interesting ways to interpret spectrograms, along with an explicit purpose for how calls will be used when identified.

Similarly, to musicians, our participants were interested in acquiring broader knowledge, which isn't possible with existing environmental audio interactions. People are interested in contextualizing calls with broader animal lives and having information that facilities this. While citizen science projects including audio review, Bat Detectives [35, 70] and Whale FM [53, 69] had explicit scientific goals, such projects lacked means for citizen scientists to expand their knowledge. People could only interact with audio-visual and aural data without broader wildlife context. Identifying project goals, citizen scientist interests, needs for long-term skill acquisition, and associated data are essential when designing for sustained audio interactions.

Diversity in Task Difficulty

Diversity in online tasks can increase engagement in online citizen science. Citizen science tasks do not need to be simple, and out of necessity, some must be complex. Birders invest substantial time into learning what birds look like and sound like in their leisure time. The application eBird is designed to augment birders' social practices, knowledge, and other cultural practices, to encourage birders to submit bird observations for ecological research [58, 67, 68]. The application is designed for those with prior knowledge of birds, as indicated by features such as species being organized taxonomically. As the world's most popular citizen science application that has led to substantial scientific outcomes, eBird exemplifies the power of a particular group. Data contributed to eBird has also been used to train machine learning algorithms for a new application that assists novice birders learn by narrowing down potential species observed (e.g. Merlin for birds [13]). Additional popular applications allow people to share observations, photos (e.g. iNaturalist for flora and fauna [66]), and audio recordings (e.g. FrogID for frogs [49]). Online citizen science does not have nearly the interaction diversity to accommodate people with diverse ranges of experience to complete tasks with different levels of difficulty and to discover interesting facts and grow knowledge. Simplistic tasks on Zooniverse have been very effective for classifying images of galaxies or animals [16, 34, 48, 59]. Analogous simplistic tasks for classifying wildlife calls from audio, however, have proven inadequate. Our findings highlight the promise in creating rich training tools that engage people with learning about bird calls aurally and visually by providing more complex and engaging information about birds, which may lead to effective analysis of audio online.

Customizable Complexity

Each participant had a unique way of interacting with media that made differences in their experiences with birds and audio evident. Some people may need to start with broad audio comparisons, such as contrasting mechanical to animal sounds and gradually integrate more nuanced comparisons. Others with more familiarity with audio and bird calls may be inclined to dive into more challenging comparisons, such as similar calls from different species (e.g. compare all cockatoo species), the call repertoires of a specific species (Figure 3), or variability in a specific call (e.g. how a bird's distance from the audio recorder impacts amplitude and call shape). Customizable interactions or level design is needed to support people with varied experiences.

When interacting with diverse multisensory media, several people asked questions about bird anatomy, behaviors, habitat, and ecological aspects. Interactions with the bird media likely elicited a holistic curiosity about nature. Lalanne and Lorenceau [31] suggested that exploring information via multiple sensory modalities can improve salience, reveal ambiguities, and allow perceptions to be contextualized. When hearing and seeing audio together several participants started noticing sound in more nuanced ways. Interaction with bird media has been found to be engaging as well, whether exploring bird photos and calls with playful devices (e.g. [51, 57]) and a structured game focused on bird ecology [50].

A variety of applications are also used by birders to improve their ability to identify birds. Thayer's Birding Software is one of the more elaborate examples, which provides birders with a wealth of information to explore in flexible, self-determined ways. The application includes, for instance, bird photos, call audio (i.e. aurally only), behavior videos, bird taxonomy, maps of species range and abundance, quizzes created automatically or customized, and extensive field guide materials [60]. However, the use of gameful interaction [20] to build knowledge is minimal; it is more a multimedia reference resource. Most birding applications provide audio recordings of a limited number of calls for each species that can only be passively played (e.g. [39, 56]). Some web platforms and applications now provide spectrograms with calls, but there are no other ways to interact with calls visually or aurally [44]. Identifying birds by calls, whether in the wild or online is an invaluable skill that birders often want to learn (e.g. [42]). Investigations are needed to explore interactions that are more gameful and elicit crossmodal perceptions that provide more holistic understandings of birds, their calls, and where they live. This will allow people to meaningfully relate to nature in meaningful ways that are applicable in the wild.

Collaborative Puzzle Solving

Previous studies involving interactions with spectrograms revealed that cooperative exploration advances the sensemaking process [18, 41], which our study further confirms. Our participants with prior experience interpreting spectrograms for whipbird and cockatoo found interpreting their spectrograms relatively easy. On several occasions, people with in-the-wild species experience, but no prior knowledge of spectrograms identified calls in spectrograms cooperatively by reasoning through how shapes may sound. Whether experience levels were equivalent or different, there were instances where cooperative sensemaking helped people to disentangle ambiguity. Cooperative problem-solving builds collective intelligence, "*... a form of universal intelligence, constantly enhanced, coordinated in real-time, and resulting in affective mobilization of skills*" [33, pg 13]. Exploring such interactions in a cooperative online environment would be a worthwhile endeavor.

Most people find it difficult to learn calls aurally, so there is value in considering how calls can be made more relatable and memorable. We noted a particularly unique and innovative way that one participant, who lacked any experience with audio, related the call of the cockatoo to the familiar shape of "car tracks". Future studies may explore if naming calls based on ideas elicited from the shape (akin to a Rorschach, or inkblot test) may prove useful for remembering calls. People have explored creating a *folksonomy* – "*the emergent labeling of lots of things by people in a social context*" [61, pg 3], for verbal descriptions of calls. However, these proved largely intractable [64]. Visual forms of folksonomy, whether through metaphors or otherwise, may help people relate to spectrograms and support learning about calls into the future.

CONCLUSIONS

We found that when people are exploring media they draw upon their prior experiences, whether interacting with audio, actively searching for wild birds, or opportunistically observing birds when outdoors. Interacting with bird media elicited curiosity and people became keen to learn about birds beyond their calls. Diverse salient interactions with visual and aural audio media are needed to decipher calls, in conjunction with media introducing people to lives of birds and their ecosystems, to contextualize calls and sustain interest. Merely providing people with a set of isolated call spectrograms and audio recordings will not facilitate nuanced observation of variation of wildlife calls. The audio recordings of wildlife calls, whether provided by professional recordists in apps or citizen scientists, are typically lack means for people with a quizzical mindset to interpret audio media in relation to their behaviors and broader interactions with their environment. In general, call spectrograms are usually presented in a way that does not allow for the interactive learning that deepens sensemaking about the nature of calls themselves. We offer a variety of ideas and paths forwards to expand on what we know about interaction with audio and other nature media, in the hopes that future work will unveil a suite of interaction strategies and processes that foster people becoming interested in birds and their calls, building skill, and identifying calls of the wild as citizen scientists to support survival of species.

Our exploratory study examined how people interact with media that reflect how birds look, sound and live. Our findings revealed important challenges and opportunities for citizen science projects that include the review of audio aurally and visually to find and identify wildlife sounds, or projects that include interaction with broader media about wildlife. Challenges included participants having difficulty with interpreting spectrograms, deciphering diversity of bird calls, and connecting media to the lives of birds. Interacting with an assortment of media visually and aurally had a variety of benefits associated with using multiple senses when interacting with bird media and crossmodal perception. Doing so, helped to elicit enthusiasm and knowledge that allowed for complex, and sometimes cooperative, problem-solving. We contribute design lessons and future directions for those interested in citizen science, sounds of nature, or eliciting passion for wildlife and its preservation.

ACKNOWLEDGMENTS

We are grateful to our participants for eagerly taking part and giving us permission to publish photos of them in action (Figure 2)! Sincere thanks to photographers for permission to use their gorgeous images (Figure 1). Comments by Sarah Webber, Anne Bowser, Tamara Homburg, and DIS reviewers substantially improved our work and we are grateful. Thank you to Ruben Venegas Li for map assistance, as well as Mark Cottman-Fields & Anthony Truskinger for technical support. The Australian Research Council (DP140102325) and the Wildlife Preservation Society of Queensland supported this research.

REFERENCES

[1] Activision. n.d. Guitar Hero. Retrieved 28 Dec 2019 from https://www.guitarhero.com/

[2] BBC Earth. 2009. Attenborough: The Amazing Lyre Bird Sings Like a Chainsaw! Now in High Quality | Bbc Earth. Retrieved 23 Dec 2019 from https://www.youtube.com/watch?v=mSB71jNq-yQ

[3] Birdlife Australia. n.d. Eastern Whipbird. Retrieved 16 Jan 2020 from http://www.birdlife.org.au/bird-profile/eastern-whipbird

[4] BirdLife International and Handbook of the Birds of the World. 2019. Bird Species Distribution Maps of the World. Version 2019.1. Retrieved 24 April 2019 from http://datazone.birdlife.org/species/requestdis

[5] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell and Jennifer Preece. 2013. Using Gamification to Inspire New Citizen Science Volunteers. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Gamification '13), 18–25. http://dx.doi.org/10.1145/2583008.2583011

[6] Anne Bowser, Derek Hansen, Jennifer Preece, Yurong He, Carol Boston and Jen Hammock. 2014. Gamifying Citizen Science: A Study of Two User Groups. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 137-140. http://dx.doi.org/10.1145/2556420.2556502

[7] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative research in psychology*. 3, 2, 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

[8] Margot Brereton, Malavika Vasudevan, Tshering Dema, Jessica L. Cappadonna, Cara Wilson and Paul Roe. 2017. The Ambient Birdhouse: Bringing Birds inside to Learn About Birds Outside. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems* (DIS '17), 321-324. http://dx.doi.org/10.1145/3064857.3079184

[9] Mark Cartwright, Graham Dove, Ana Elisa Méndez Méndez, Juan P Bello and Oded Nov. 2019. Crowdsourcing Multi-Label Audio Annotation Tasks with Citizen Scientists. In *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '19), 1-11. http://dx.doi.org/10.1145/3290605.3300522

[10] Mark Cartwright, Ayanna Seals, Justin Salamon, Alex Williams, Stefanie Mikloska, Duncan MacConnell, Edith Law, Juan P. Bello and Oded Nov. 2017. Seeing Sound: Investigating the Effects of Visualizations and Complexity on Crowdsourced Audio Annotations. *Proceedings of ACM Human-Computer Interactaction*. Vol 1, No CSCW, 1-21. http://dx.doi.org/10.1145/3134664

[11] Alan Chamberlain and Chloe Griffiths. 2013. Moths at Midnight: Design Implications for Supporting Ecology-Focused Citizen Science. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (MUM '13), 25. http://dx.doi.org/10.1145/2541831.2541858

[12] Cornell Lab of Ornithology. n.d. Bird Song Hero. Retrieved 17 Dec 2019 from https://academy.allaboutbirds.org/bird-song-hero/

[13] Cornell Lab of Ornithology. n.d. Merlin Id. Retrieved 30 May 2019 from http://merlin.allaboutbirds.org/

[14] Mark Cottman-Fields, Margot Brereton and Paul Roe. 2013. Virtual Birding: Extending an Environmental Pastime into the Virtual World for Citizen Science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 2029-2032. http://dx.doi.org/10.1145/2470654.2466268

[15] Mark Cottman-Fields, Margot Brereton, Jason Wimmer and Paul Roe. 2014. Collaborative Extension of Biodiversity Monitoring Protocols in the Bird Watching Community. In *Proceedings of the 13th Participatory Design Conference* (PDC '14), 111-114. http://dx.doi.org/10.1145/2662155.2662193

[16] Joe Cox, Eun Young Oh, Brooke Simmons, Chris Lintott, Karen Masters, Anita Greenhill, Gary Graham and Kate Holmes. 2015. Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects. *Computing in Science & Engineering*. 17, 4, 28-41. http://dx.doi.org/10.1109/MCSE.2015.65

[17] K. Crowston and N. R. Prestopnik. 2013. Motivation and Data Quality in a Citizen Science Game: A Design Science Evaluation. In *Proceedings of the 2013 46th Hawaii International Conference on System Sciences* (HICSS '13), 450-459. http://dx.doi.org/10.1109/HICSS.2013.413

[18] Tshering Dema, Margot Brereton, Jessica L. Cappadonna, Paul Roe, Anthony Truskinger and Jinglan Zhang. 2017. Collaborative Exploration and Sensemaking of Big Environmental Sound Data. *Computer Supported Cooperative Work*. 26, 4-6, 693-731. http://dx.doi.org/10.1007/s10606-017-9286-9

[19] Tshering Dema, Margot Brereton and Paul Roe. 2019. Designing Participatory Sensing with Remote Communities to Conserve Endangered Species. In *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '19), 1-16. http://dx.doi.org/10.1145/3290605.3300894

[20] Sebastian Deterding, Dan Dixon, Rilla Khaled and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (MindTrek '11), 9–15. http://dx.doi.org/10.1145/2181037.2181040

[21] Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan and Anna L Cox. 2014. Designing for Dabblers and Deterring Drop-Outs in Citizen Science. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems (CHI '14), 2985-2994. http://dx.doi.org/10.1145/2556288.2557262

[22] Alexandra Eveleigh, Charlene Jennett, Stuart Lynn and Anna L. Cox. 2013. "I Want to Be a Captain! I Want to Be a Captain!": Gamification in the *Old Weather* Citizen Science Project. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Gamification '13), 79-82. http://dx.doi.org/10.1145/2583008.2583019

[23] Yurong He, Jennifer Preece, Carol Boston, Anne Bowser, Derek Hansen and Jen Hammock. 2014. The Effects of Individualized Feedback on College Students' Contributions to Citizen Science. In *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 165-168. http://dx.doi.org/10.1145/2556420.2556484

[24] Andrew P. Hill, Peter Prince, Evelyn Piña Covarrubias, C. Patrick Doncaster, Jake L. Snaddon and Alex Rogers. 2017. Audiomoth: Evaluation of a Smart Open Acoustic Device for Monitoring Biodiversity and the Environment. *Methods in Ecology and Evolution*. 9, 5, 1199-1211. http://dx.doi.org/10.1111/2041-210X.12955

[25] Andrew P. Hill, Peter Prince, Jake L. Snaddon, C. Patrick Doncaster and Alex Rogers. 2019. Audiomoth: A Low-Cost Acoustic Device for Monitoring Biodiversity and the Environment. *HardwareX*. 6, e00073. http://dx.doi.org/10.1016/j.ohx.2019.e00073

[26] Glen T. Hvenegaard. 2002. Birder Specialization Differences in Conservation Involvement, Demographics, and Motivations. *Human Dimensions of Wildlife*. 7, 1, 21-36. http://dx.doi.org/10.1080/108712002753574765

[27] Riitta Jääskeläinen. 2010. Think-Aloud Protocol, In *Handbook of Translation Studies*, Yves Gambier and Luc van Doorslaer (eds.). John Benjamins Publishing Company, Amsterdam / Philadelphia, 371-374.

[28] Rose Johnson, Yvonne Rogers, Janet van der Linden and Nadia Bianchi-Berthouze. 2012. Being in the Thick of in-the-Wild Studies: The Challenges and Insights of Researcher Participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 1135-1144. http://dx.doi.org/10.1145/2207676.2208561

[29] Dani Korpi, Toni Heittola, Timo Partala, Antti Eronen, Annamaria Mesaros and Tuomas Virtanen. 2013. On the Human Ability to Discriminate Audio Ambiances from Similar Locations of an Urban Environment. *Personal and Ubiquitous Computing*. 17, 761–769. http://dx.doi.org/10.1007/s00779-012-0625-z

[30] Christopher Kullenberg and Dick Kasperowski. 2016. What Is Citizen Science? – a Scientometric Meta-Analysis. *PLoS One*. 11, 1, e0147152. http://dx.doi.org/10.1371/journal.pone.0147152

[31] Christophe Lalanne and Jean Lorenceau. 2004. Crossmodal Integration for Perception and Action. *Journal of Physiology-Paris*. 98, 1, 265-279. http://dx.doi.org/10.1016/j.jphysparis.2004.06.001

[32] Nicholas P Leseberg, Stephen A Murphy, Nigel A Jackett, Bruce R Greatwich, Jamie Brown, Neil Hamilton, Leo Joseph and James EM; Watson. 2019. Descriptions of Known Vocalisations of the Night Parrot 'Pezoporus Occidentalis'. *Australian Field Ornithology*. 36, 79-88. http://dx.doi.org/10.20938/afo36079088

[33] Pierre Levy. 1997. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Perseus Books.

[34] Chris J Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M Jordan Raddick, Robert C Nichol, Alex Szalay and Dan Andreescu. 2008. Galaxy Zoo: Morphologies Derived from Visual Inspection of Galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*. 389, 3, 1179-1189.

[35] Oisin Mac Aodha, Rory Gibb, Kate E. Barlow, Ella Browning, Michael Firman, Robin Freeman, Briana Harder, Libby Kinsey, Gary R. Mead, Stuart E. Newson, Ivan Pandourski, Stuart Parsons, Jon Russ, Abigel Szodoray-Paradi, Farkas Szodoray-Paradi, Elena Tilova, Mark Girolami, Gabriel Brostow and Kate E. Jones. 2018. Bat Detective—Deep Learning Tools for Bat Acoustic Signal Detection. *PLOS Computational Biology*. 14, 3, e1005995. http://dx.doi.org/10.1371/journal.pcbi.1005995

[36] Richard Mason, Paul Roe, Michael Towsey, Jinglan Zhang, Jennifer Gibson and Stuart Gage. 2008. Towards an Acoustic Environmental Observatory. In *Proceedings of the 2008 Institute of Electrical and Electronics Engineers Ninth International Conference on eScience* (eScience '08), 135-142. http://dx.doi.org/10.1109/eScience.2008.16

[37] James McDermott, Toby Gifford, Anders Bouwer and Mark Wagy. 2013. Should Music Interaction Be Easy? (1st ed. 2013.), In *Music and Human-Computer Interaction*, Simon Holland et al. (eds.). Springer London, London, 29-47. http://dx.doi.org/10.1007/978-1-4471-2990-5

[38] Stuart Moran, Nadia Pantidi, Tom Rodden, Alan Chamberlain, Chloe Griffiths, Davide Zilli, Geoff Merrett and Alex Rogers. 2014. Listening to the Forest and Its Curators: Lessons Learnt from a Bioacoustic Smartphone Application Deployment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 2387-2396. http://dx.doi.org/10.1145/2556288.2557022

[39] Michael Morcombe and David Stewart. n.d. Michael Morcombe & David Stewart Eguide to the Birds of Australia. Retrieved 20 Jan 2020 from http://www.mydigitalearth.com/dproducts/morcombeinfo.html

[40] Federico Morelli. 2015. Indicator Species for Avian Biodiversity Hotspots: Combination of Specialists and Generalists Is Necessary in Less Natural Environments.

Journal for Nature Conservation. 27, 54-62. http://dx.doi.org/10.1016/j.jnc.2015.06.006

[41] Jessica L. Oliver, Margot Brereton, David M. Watson and Paul Roe. 2018. Visualisations Elicit Knowledge to Refine Citizen Science Technology Design: Spectrograms Resonate with Birders. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (OZCHI '18), 133-144. http://dx.doi.org/10.1145/3292147.3292171

[42] Jessica L. Oliver, Margot Brereton, David M. Watson and Paul Roe. 2019. Listening to Save Wildlife: Lessons Learnt from Use of Acoustic Technology by a Species Recovery Team. In *Proceedings of the 2019 Designing Interactive Systems Conference* (DIS ‘19), 1335-1348. http://dx.doi.org/10.1145/3322276.3322360

[43] Penny Olsen. 2018. *Night Parrot: Australia's Most Elusive Bird*. CSIRO PUBLISHING.

[44] Bob Planqué, Willem-Pier Vellinga, Sander Pieterse and Jonathon Jongsma. n.d. Xeno-Canto: Sharing Bird Sounds from around the World. Retrieved 20 Jan 2020 from www.xeno-canto.org

[45] Michael J. O. Pocock, John C. Tweddle, Joanna Savage, Lucy D. Robinson and Helen E. Roy. 2017. The Diversity and Evolution of Ecological and Environmental Citizen Science. *PLoS One*. 12, 4, e0172579. http://dx.doi.org/10.1371/journal.pone.0172579

[46] Jennifer Preece, Carol Boston, Tom Yeh, Jacqueline Cameron, Mary Maher and Kazjon Grace. 2016. Enticing Casual Nature Preserve Visitors into Citizen Science Via Photos. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (CSCW '16), 373-376. http://dx.doi.org/10.1145/2818052.2869104

[47] Nirosha Priyadarshani, Stephen Marsland and Isabel Castro. 2018. Automated Birdsong Recognition in Complex Acoustic Environments: A Review. *Journal of Avian Biology*. 49, 5, 1-27. http://dx.doi.org/10.1111/jav.01447

[48] Holly Rosser and Andrea Wiggins. 2018. *Tutorial Designs and Task Types in Zooniverse*. Association for Computing Machinery, Jersey City, NJ, USA. http://dx.doi.org/10.1145/3272973.3274049

[49] Jodi JL Rowley, Corey T Callaghan and Timothy Cutajar. 2019. Frogid: Citizen Scientists Provide Validated Biodiversity Data on Frogs of Australia. *Herpetological Conservation and Biology*. 14, 1, 155-170.

[50] Prasad Sandbhor and Priti Bangal. 2018. Game System Based on Mixed-Species Bird Flocks: A Science Outreach Approach. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (CHI PLAY '18), 595-601. http://dx.doi.org/10.1145/3270316.3271509

[51] Mangalam Sankupellay, Anna Kalma, Sean Magin, Jessica L Cappadonna, Paul Roe and Margot Brereton. 2017. Birdsound: Enticing Urban Dwellers to Engage with Local Birds around Their Home. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (OZCHI '17), 172-181. http://dx.doi.org/10.1145/3152771.3152790

[52] Laela Sayigh, Nicola Quick, Gordon Hastie and Peter Tyack. 2013. Repeated Call Types in Short-Finned Pilot Whales, *Globicephala Macrorhynchus*. *Marine Mammal Science*. 29, 2, 312-324. http://dx.doi.org/10.1111/j.1748-7692.2012.00577.x

[53] Scientific American. 2011. The Whale Song Project (Whale Fm). Retrieved 20 Jan 2020 from https://www.scientificamerican.com/citizen-science/the-whale-song-project-whale-fm/

[54] Lior Shamir, Carol Yerby, Robert Simpson, Alexander M von Benda-Beckmann, Peter Tyack, Filipa Samarra, Patrick Miller and John Wallin. 2014. Classification of Large Acoustic Datasets Using Machine Learning and Crowdsourcing: Application to Whale Calls. *Acoustical Society of America*. 135, 2, 953-962. http://dx.doi.org/10.1121/1.4861348

[55] Nirwan Sharma. 2016. Species Identification in Citizen Science: Effects of Interface Design and Image Difficulty on User Performance and Workload. In *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '16), 128-133. http://dx.doi.org/10.1145/2851581.2890382

[56] David Allen Sibley. n.d. . Sibley Eguide to Birds. Retrieved 30 Jun 2020 from https://www.sibleyguides.com/about/the-sibley-eguide-to-birds-app/

[57] Alessandro Soro, Margot Brereton, Tshering Dema, Jessica L Oliver, Min Zhen Chai and Aloha May Hufana Ambe. 2018. The Ambient Birdhouse: An Iot Device to Discover Birds and Engage with Nature. In *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '18), 1-13. http://dx.doi.org/10.1145/3173574.3173971

[58] Brian L Sullivan, Jocelyn L Aycrigg, Jessie H Barry, Rick E Bonney, Nicholas Bruns, Caren B Cooper, Theo Damoulas, André A Dhondt, Tom Dietterich, Andrew Farnsworth, Daniel Fink, John W Fitzpatrick, Thomas Fredericks, Jeff Gerbracht, Carla Gomes, Wesley M Hochachka, Marshall J Iliff, Carl Lagoze, Frank A La Sorte, Matthew Merrifield, Will Morris, Tina B Phillips, Mark Reynolds, Amanda D Rodewald, Kenneth V Rosenberg, Nancy M Trautmann, Andrea Wiggins, David W Winkler, Weng-Keen Wong, Christopher L Wood, Jun Yu and Steve Kelling. 2014. The Ebird Enterprise: An Integrated Approach to Development and Application of Citizen Science. *Biological Conservation*. 169, 31-40. http://dx.doi.org/10.1016/j.biocon.2013.11.003

[59] Alexandra Swanson, Margaret Kosmala, Chris Lintott, Robert Simpson, Arfon Smith and Craig Packer. 2015. Snapshot Serengeti, High-Frequency Annotated Camera Trap Images of 40 Mammalian Species in an

African Savanna. *Scientific Data*. 2, 1-14. http://dx.doi.org/10.1038/sdata.2015.26

[60] Thayer's Birding Software. n.d. . Thayer's Birding Software Retrieved 17 Dec 2019 from https://www.thayerbirding.com/

[61] Gruber Thomas. 2007. Ontology of Folksonomy: A Mash-up of Apples and Oranges. *International Journal on Semantic Web and Information Systems*. 3, 1, 1-11. http://dx.doi.org/10.4018/jswis.2007010101

[62] Ramine Tinati, Max Van Kleek, Elena Simperl, Markus Luczak-Rösch, Robert Simpson and Nigel Shadbolt. 2015. Designing for Citizen Data Analysis: A Cross-Sectional Case Study of a Multi-Domain Citizen Science Platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 4069-4078. http://dx.doi.org/10.1145/2702123.2702420

[63] Anthony Truskinger, Mark Cottman-Fields, Philip Eichinski, Michael Towsey and Paul Roe. 2014. Practical Analysis of Big Acoustic Sensor Data for Environmental Monitoring. In *Proceedings of the Institute of Electrical and Electronics Engineers Fourth International Conference on Big Data and Cloud Computing* (BDCLOUD '14), 91-98. http://dx.doi.org/10.1109/BDCloud.2014.29

[64] Anthony Truskinger, Ian Newmarch, Mark Cottman-Fields, Jason Wimmer, Michael Towsey, Jinglan Zhang and Paul Roe. 2013. Reconciling Folksonomic Tagging with Taxa for Bioacoustic Annotations, In *Web Information Systems Engineering – Wise 2013: 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part I*, Xuemin Lin et al. (eds.). Springer, Berlin, Heidelberg, 292-305. http://dx.doi.org/10.1007/978-3-642-41230-1_25

[65] Woody Turner. 2014. Sensing Biodiversity. *Science*. 346, 6207, 301. http://dx.doi.org/10.1126/science.1256014

[66] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona and Serge Belongie. 2018. The Inaturalist Species Classification and Detection Dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR '18), 8769-8778. http://dx.doi.org/10.1109/CVPR.2018.00914

[67] Andrea Wiggins. 2011. Ebirding: Technology Adoption and the Transformation of Leisure into Science. In *Proceedings of the 2011 iConference* (iConference '11), 798-799. http://dx.doi.org/10.1145/1940761.1940910

[68] Chris Wood, Brian Sullivan, Marshall Iliff, Daniel Fink and Steve Kelling. 2011. Ebird: Engaging Birders in Science and Conservation. *PLoS biology*. 9, 12, e1001220. http://dx.doi.org/10.1371/journal.pbio.1001220

[69] Zooniverse. 2011. Whale Fm – the Whale Song Project. Retrieved 20 Jan 2020 from https://blog.zooniverse.org/2011/11/29/whale-fm-the-whale-song-project/

[70] Zooniverse. 2015. Bat Detective. Retrieved 20 Jan 2020 from https://www.batdetective.org/

[71] Zooniverse. n.d. Zooniverse. Retrieved 13 Jan 2020 from https://www.zooniverse.org/