Evaluating citizen science for dialect research on the nightingale song (*Luscinia megarhynchos*)

Dissertation

zur Erlangung des akademischen Grades:

Doctor rerum naturalium

(*Dr. rer. nat.*)

eingereicht an der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von:

M. Sc., Denise Jäckel, geb. Bock

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Peter A. Frensch (komm.)

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin:

Prof. Dr. Dr. Christian Ulrichs

Gutachter/innen:
1. Prof. Dr. Gerlind Lehmann

2. PD Dr. Mirjam Knörnschild

3. Prof. Dr. Aletta Bonn

Tag der mündlichen Prüfung:
06.09.2022

Table of contents

Abstract

Citizen science (CS) is a method that has been increased in science worldwide in recent years. Although many studies have compared these data with those of academic researchers, there are still concerns about their quality. Data quality depends on various factors such as the studied species, research question and scientific field. Songbird studies frequently benefited from valid CS data generated simultaneously around the world for species with a small repertoire.

In my doctoral thesis I aimed to evaluate the method of CS for a bird species with a large repertoire, the nightingale (*Luscinia megarhynchos*), as a use case based on dialect research. I investigated whether the three main assumed reasons for poor quality (anonymity, inexperience and lack of standardisation) led to incomplete, temporal or spatial biassed and inaccurate bioacoustic data. Therefore, I analysed non-standardised CS recordings, which were generated with a smartphone via the '*Naturblick*' app, which contained an in-built pattern recognition algorithm.

My co-authors and I compared in the first study (**Chapter II**) this non-standardised CS data with standardised ones from academic researchers. The data scope was by way larger in the CS recordings due to the contribution of all participants, independent of their experience. Differences in quality could be attributed to the devices used (smartphone actuality, brand) and not to the users *per se*. Citizen scientists generated data with a better temporal distribution, which covered all calendar weeks and times of day during the breeding season for the first time. The accuracy was lower in CS data than in ones from academics.

In the second study (**Chapter III**), my co-authors and I conducted a dataset comparison between non-standardised CS, open and standardised shared data from academic researchers. Citizen scientists provided the largest data scope. The wide geographical coverage of the CS recordings far exceeded that of open and shared ones. All datasets showed a temporal and geographical bias. The combination of different datasets resulted in the largest temporal and spatial dialect investigation to date and revealed the most extensive stability in the song of the nightingale and birds in general.

My co-authors and I compared in the third study (**Chapter IV**) the development of the project before (2018, 2019) and during the COVID-19 pandemic (2020). The data scope increased over time. The spatial distribution extended continuously and closely followed the natural occurrence of the nightingale over the whole of Germany. Misidentifications were very low, decreased steadily and were systematically affected by other bird species with vocal similarities. Anonymity and inexperience had no impact on the data scope or accuracy, the former even being a potential success factor for the project. This study revealed that data quality was positively affected by the project duration.

In summary (**Chapter V**), my doctoral thesis showed that the method CS could be used to generate valid data for bioacoustic research. My findings showed that anonymity, lack of experience and standardisation did not lead to low quality but in fact to a large dataset, which was as valuable as ones from academic researchers. The results are of great relevance for future CS projects to improve the quality and the trust in these data.

Zusammenfassung

Citizen Science (CS) ist eine Methode, die in den letzten Jahren in der Wissenschaft weltweit an Bedeutung gewonnen hat. Obwohl viele Studien diese Daten mit denen von akademischen Forschenden verglichen, gibt es immer noch Bedenken hinsichtlich ihrer Qualität. Datenqualität hängt von verschiedenen Faktoren wie der untersuchten Spezies, der Forschungsfrage und dem wissenschaftlichen Bereich ab. Singvögeln Studien profitierten zunehmend von validen CS-Daten, die gleichzeitig auf der ganzen Welt für Arten mit einem kleinen Repertoire erhoben wurden.

In meiner Doktorarbeit zielte ich darauf ab die Methode CS für eine Vogelart mit einem großen Repertoire, der Nachtigall (*Luscinia megarhynchos),* als Anwendungsfall auf der Grundlage der Dialektforschung zu evaluieren. Ich untersuchte, ob die drei vermeintlichen Hauptgründe für schlechte Qualität (Anonymität, Unerfahrenheit und fehlende Standardisierung) zu unvollständigen, zeitlich oder räumlich verzerrten und ungenauen bioakustischen Daten führten. Dazu analysierte ich nicht-standardisierte CS-Aufnahmen, die mit einem Smartphone über die '*Naturblick*' App erstellt wurden, welche einen eingebauten Mustererkennungsalgorithmus enthielt.

Meine Co-Autoren und ich haben in der ersten Studie (**Kapitel II**) diese nicht standardisierten CS-Daten mit standardisierten von akademischen Forschenden verglichen. Der Datenumfang war bei den CS-Aufnahmen aufgrund des Beitrags aller Teilnehmer, unabhängig von ihrer Erfahrung, deutlich größer. Qualitätsunterschiede konnten auf die verwendeten Geräte (Aktualität, Marke des Smartphones) zurückgeführt werden und nicht auf die Nutzer an sich. Citizen Scientists generierten Daten mit einer besseren zeitlichen Verteilung, die erstmalig alle Kalenderwochen und Tageszeiten während der Brutsaison erfassten. Die Genauigkeit der CS-Daten war geringer als jene von Akademikern.

In der zweiten Studie (**Kapitel III**) haben meine Co-Autoren und ich einen Datensatzvergleich zwischen nicht standardisierten CS, offenen und standardisierten geteilten Daten von akademischen Forschenden durchgeführt. Citizen Scientists lieferten den größten Datenumfang. Die breite geografische Abdeckung der CS-Aufzeichnungen übertraf bei weitem die der offenen und geteilten Daten. Alle Datensätze wiesen eine zeitliche und geografische Verzerrung auf. Die Kombination verschiedener Datensätze führte zur bisher größten zeitlichen und räumlichen Dialektstudie und zeigte die größte Stabilität im Gesang der Nachtigall und Vögeln im Allgemeinen.

In der dritten Studie (**Kapitel IV**) verglichen meine Co-Autoren und ich die Entwicklung des Projekts vor (2018, 2019) und während der COVID-19-Pandemie (2020). Der Datenumfang nahm im Laufe der Zeit zu. Die räumliche Verbreitung erweiterte sich kontinuierlich und folgte eng dem natürlichen Vorkommen der Nachtigall in ganz Deutschland. Die Zahl der Fehlidentifikationen war sehr gering, nahm stetig ab und wurde systematisch durch andere Vogelarten mit ähnlichen Gesängen beeinflusst. Anonymität und Unerfahrenheit hatten keinen Einfluss auf den Umfang und die Genauigkeit der Daten, ersteres war sogar ein möglicher Erfolgsfaktor für das Projekt. Die Studie ergab, dass die Datenqualität durch die Projektdauer positiv beeinflusst wurde.

Zusammenfassend (**Kapitel V**) konnte ich in meiner Doktorarbeit zeigen, dass mit der Methode CS valide Daten für die bioakustische Forschung gewonnen werden können. Meine Ergebnisse zeigten, dass Anonymität, mangelnde Erfahrung und Standardisierung nicht zu geringer Qualität führten, sondern zu einem großen Datensatz, der genauso wertvoll war wie jene von akademischen Forschenden. Die Ergebnisse sind von großer Bedeutung für künftige CS-Projekte zur Verbesserung der Qualität und des Vertrauens in diese Daten.

Chapter I

General Introduction

1. Scientific background

Citizen science (CS) is a method in which non-professional volunteers (citizen scientists) work participative with scientists to gain scientific knowledge through the combination of research, education and action (Bonney et al., 2009). Participants can take part in a growing number of areas either with their expertise, tools or resources (Vohland et al., 2021). This allows them to be involved in various stages of scientific processes, mostly for data collection, categorization, transcription or analysis (Dickinson et al., 2012; Bonney et al., 2016; Hecker et al., 2018).

The practice of involving citizen scientists began as early as the 18th and 19th centuries with Benjamin Franklin and Charles Darwin (Miller-Rushing, Primack, & Bonney, 2012; Silvertown, 2009). In bioacoustic, particularly in the field of the conservation of birds, this method has been a long tradition (Bonney et al., 2009). One of the first ornithology CS projects started in 1749 in Finland with the collection of migration information (Greenwood, 2007). In 1880, the US Geological Survey began to involve the public in monitoring, which became an important part of the Breeding Bird Survey, which was established in 1966 (Dickenson, Zuckerberg & Barter, 2010). The Christmas Bird Count, started in 1900 and is the currently longest running CS project. It was initiated by ornithologists at the American Museum of Natural History (New York) as an alternative to regular holiday bird-shooting contests (Forrester et al., 2015). Over time CS projects established as a complement to traditional data collection methods (e.g. Diblíková et al., 2019) for monitoring biodiversity, population trends and migration patterns (e.g Roy et al., 2012; Palacin-Silva et al., 2016; Planillo et al., 2021). CS also evolved to a worldwide reaching method, ranging from local to global projects (e.g. ebird, Xeno-Canto) due to access to mobile internet and the widespread use of smartphones with data upload capabilities (Theobald et al., 2015; Kosmala et al., 2016). The development of apps within smartphones allowed for participation in real time from anyone, at any time and anywhere around the world which would otherwise be unachievable (Silvertown, 2009). This also enabled scientists and citizen to communicate, collect and analyse data simultaneously, leading to many scientific discoveries and enlarged understandings in the fields of ecological and environment.

Many studies have examined whether these data have the equivalent quality to those generated by academic researchers. Some studies have perceived CS data as less good and unreliable (Burgess et al., 2017). Anonymity, citizen scientist's lack of experience and the absence of standardised protocols (Scott & Hallam, 2003; Dickinson, Zuckerberg & Bonter, 2010; Ahrends et al., 2011) were considered as the main reasons for incomplete data, temporally or spatially bias and inaccuracy (Dickinson, Zuckerberg & Bonter, 2010). Other studies demonstrated equivalent (Flesch & Belt, 2017), better (e.g. Moyer-Horner, Smith & Belt, 2012) or improved data of citizen scientists over time (van der Wal et al., 2016). Thus, quality differences are not yet fully understood – because they are often based on the species studied, the research question or the scientific field (Lewandowski & Specht, 2015).

Citizen science in birdsong research

Songbird studies increasingly benefited from the cooperation with citizen scientists due to the simultaneous generation of data around the world with valid data quality. This doctoral thesis evaluated the method of CS for dialect research as a use case. Dialects are song variations over time and space (Catchpole & Slater, 2008). Findings on the cultural evolution of a sexually selected behavioural trait could potentially help to improve our understanding of the socially learned, complex vocalisations of other animal groups, such as whales, bats, elephants and

humans (Beecher & Brenowitz, 2005; Goller & Shizuka, 2018). Previous CS dialect projects successfully studied bird species with a small repertoire (Yellowhammer: *Emberiza citrinella*; Diblíková et al., 2019; Chipping Sparrow's: *Spizella passerine;* Searfoss et al., 2020), but not yet with a large one, such as the common nightingale (*Luscinia megarhynchos*).

The Nightingale as use case

The nightingale is a particularly suitable model species for a dialect CS project due to its charismatic song, large breeding range (Asia to Europe;) and territorial behaviour (Glutz von Blotzheim, 1988). Males return annually to the same breeding area where they were born (Roth & Amrhein, 2010). At the beginning of the breeding season and before the arrival of females, male nightingales sing during the day to establish and defend territories (Ratcliffe & Otter, 1996). Nocturnal song begins with the arrival of females and stops as soon as a male is mated (Amrhein et al., 2004). Beginning in 1980 (doctoral thesis Hultsch), the song of the nightingale was studied intensively in the laboratory and later in the wild (e.g. Kiefer et al., 2014, Landgraf, Weiss & Kipper, 2015). In this species, males have a very versatile repertoire (up to 200 different song types; Kipper et al., 2004) which expands during the first two years ("open ended learners"; Kiefer et al., 2006). Song types are unique sequences of elements and frequencies with a given duration (Gil &Gahr, 2002), which are learned very stereotypic with many individual variations in nightingales (Kiefer et al., 2006). Their song is very complex and suggests the presence of spatial vocal variations. Previous studies (Hultsch & Todt, 1981; Sprau & Mundry, 2010) demonstrated decreasing song similarity with increasing distance. Comparisons between Germany and France indicated the potential for dialects in the frequency with which certain song types and their sequences were sung (doctoral thesis Weiss, 2012) and also showed evidence for temporal stability (diploma thesis Schehka, 2004).

Based on this research and a pilot study, Nightingale City Berlin in 2016, the citizen science nightingale project (*Forschungsfall Nachtigall*) was developed and established at the Museum für Naturkunde Berlin (MfN), Germany. Between 2018 and 2020 citizen scientists in German-speaking countries were invited to generate nightingale recordings with their smartphones through the '*Naturblick*' app (for app details see Stehle et al., 2020). Participants were supported by an in-built pattern recognition algorithm (Lasseck, 2018, 2019), but did not receive any protocols or specific instructions as to how, where, and when they should record nightingales. The low-threshold approach was used to reach not only bird specialists but also the broad public. Citizen scientists could participate anonymously or with an individual username. The recordings were not standardised in terms of the technique used, time, place, or duration (maximum of two minutes).

2. Conceptual framework

Thesis structure

The doctoral thesis contains findings from a cumulative dissertation. Based on previous scientific knowledge (**Chapter I**), three studies were developed which have been successfully published in peer-reviewed journals (**Chapters II, III and IV**). The last chapter (**V**) serves as a summary of the findings, resulting conclusions and further scientific questions.

Research questions and goals

My thesis aimed to evaluate the method CS for bioacoustic research using dialect studies based on the nightingale song as a use case. The identification of possible quality differences is an important process to further establish trust amongst the science community in this method. Quality is not an absolute value, but multidimensional (reviewed in Lewandowski & Specht, 2015) and depends on the respective study requirements based on structure, design, resources and goals (Munzinger, 2015). I investigated in a systematic analysis the aspects of completeness (data scope), bias (temporal or spatial) and accuracy (misidentifications). In order to examine whether valid data can be produced despite the three assumed main causes of poor quality (anonymity, inexperienced, non-standardisation), I conducted three studies. The first two studies (**Chapter II and III**) were based on comparisons of CS data with other datasets (e.g. from academic researchers). The third study (**Chapter IV**) analysed the development over the course of the project. Subsequently, I will explain the investigated scientific questions and their underlying studies in more detail.

Are CS data as valid as those of academic researchers?

(see **Chapter II**: Jäckel, D., Mortega, K. G., Sturm, U., Brockmeyer, U., Khorramshahi, O., & Voigt-Heucke, S. L. (2021). Opportunities and limitations: A comparative analysis of citizen science and expert recordings for bioacoustic research. Plos one, 16(6), e0253763.)

The quality assessment of CS data is generally based on comparisons with ones of academic researchers. Most of the differences found in previous reviews (e.g. Lewandowski & Specht 2015), were based on unequal collected datasets (e.g. Kosmala et al., 2016), but rarely on the same conditions (e.g. Freitag, Meyer & Whiteman, 2016). This makes it impossible to determine whether data quality from citizen scientists and scientists varies or not (Specht & Lewandowski, 2018). Standardisation has long been seen as a reason for high quality (Kosmala, et al., 2016), but has recently been rejected by Danielsen and colleagues (2021). Some studies indicated that increased experience improved data quality (e.g. seabirds; Haywood, Parrish & Dolliver, 2016), while others showed that participants provided valid data right from the beginning (reviewed in Kosmala et al., 2016).

In the first study (**Chapter II**) my co-authors and I examined data that were temporally (2018 and 2019) and spatially (Berlin) equivalent generated. Recordings made in a non-standardised CS project using a smartphone without any experience were compared to those generated standardised by academic researchers using professionally calibrated equipment. To determine if CS data were valid for dialect research as a use case, we investigated the following aspects: 1) completeness (data scope) in terms of the recording number and duration (mean, cumulative) between academic researcher and citizen scientist in general and with different experience 2) potential temporal biases in the calendar weeks and times of day 3) accuracy with respect to misidentifications (nightingales vs non-nightingales) 4) influence of the devices used on measurements of acoustic parameters (durations, frequencies). This was a crucial first step to understand opportunities or limitations between standardised data from academic researchers and non-standardised ones from citizen scientists with different experiences.

Can dialect studies be conducted with CS data?

(see **Chapter III**: Jäckel, D., Mortega, K. G., Brockmeyer, U., Lehmann, G. U., & Voigt-Heucke, S. L. (2022) Unravelling the Stability of Nightingale Song Over Time and Space Using Open, Citizen Science and Shared Data. Frontiers in Ecology and Evolution, 89.)

For a long time, dialect research was based on data collected by scientific researchers (Poulsen, 1958) as it has been assumed that CS projects generates incomplete data with temporal and spatial bias (Dickinson, Zuckerberg & Bonter, 2010; Fischer, Gerber & Wentz, 2021). Recently, scientists have been increasingly supported by CS data from different regions around the world (Silvertown, 2009). These dialect studies have been successfully conducted on species with a small repertoire on the individual level (Diblíková et al., 2019; Searfoss et al., 2020). So far, it is unknown whether the method CS can also be used for species with many song types, such as the nightingale, on the population level. There was first evidence in the literature for temporal stability (Schehka, 2004) as well as spatial differences on a small scale in the nightingale (Weiss, 2012).

In the second study (**Chapter III**), my co-authors and I conducted the largest temporal (nearly a century) and geographical (almost the whole of Europe) dialect study to date. We examined how valid non-standardised CS data were for dialect research on the population level for the following aspects: 1) data scope completeness in terms of the number of songs, song types and their diversity 2) temporal or spatial bias with regard to the time periods and regions 3) percentages of certain parts of songs, the occurrence and relative frequency of song types. The study was the basis for a further understanding about whether birdsong analysis over time and space is possible with both non-standardised and standardised data.

How does a CS project evolve over time?
(see **Chapter IV**: Jäckel, D., Mortega, K. G., Brockmeyer, U., Darwin, S., Sturm, U., Lasseck, M., Lehmann, G. U.C., Voigt-Heucke, S. L. (2022) Community engagement and data quality: Best practices and lessons learned from a citizen science project on birdsong. Journal of Ornithology.)
Many quality evaluations of CS data were based on analyses over a short time period such as one year (Wiggins et al., 2011) and few studies examined potential changes over time (e.g. Edgar & Stuart-Smith 2009). Long-term projects can be influenced by external factors such as the COVID-19 outbreak. The pandemic changed data scopes and distributions worldwide (Phillips et al., 2021), but also increased accuracy, despite reduced participants (Smith & Hamed, 2020). Dickinson and colleagues (2010) suggest that anonymity and inexperience were the main reasons for differences in the data scope and accuracy and not external factors.

In the third study (**Chapter IV**), my co-authors and I compared data from the nightingale CS project before (2018, 2019) and during (2020) the COVID-19 pandemic. We evaluated the development for 1) completeness of the data scope with regard to the recording number and cumulative duration 2) spatial distribution to test for possible biases und 3) accuracy in terms of underlying patterns of misidentifications (temporal, similarities) 4) data contributed by anonymous citizen scientists and those who participated with a username in one, two or three years. This study served to identify best practices and lessons learned about the influence of anonymity and experience on data quality over the course of the project.

Chapter II

Opportunities and limitations: A comparative analysis of citizen science and expert recordings for bioacoustic research
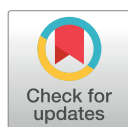
PLOS ONE

Opportunities and limitations: A comparative analysis of citizen science and expert recordings for bioacoustic research

Denise Jäckel[1,2], Kim G. Mortega[1], Ulrike Sturm[1], Ulrich Brockmeyer[1], Omid Khorramshahi[1], Silke L. Voigt-Heucke[1,3]*

1 Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany, 2 Life Sciences Faculty, Humboldt-Universität zu Berlin, Berlin, Germany, 3 Animal Behaviour, Institute of Biology, Freie Universität Berlin, Berlin, Germany

* voigt.heucke@gmail.com

Abstract

Citizen science is an approach that has become increasingly popular in recent years. Despite this growing popularity, there still is widespread scepticism in the academic world about the validity and quality of data from citizen science projects. And although there might be great potential, citizen science is a rarely used approach in the field of bioacoustics. To better understand the possibilities, but also the limitations, we here evaluated data generated in a citizen science project on nightingale song as a case study. We analysed the quantity and quality of song recordings made in a non-standardized way with a smartphone app by citizen scientists and the standardized recordings made with professional equipment by academic researchers. We made comparisons between the recordings of the two approaches and among the user types of the app to gain insights into the temporal recording patterns, the quantity and quality of the data. To compare the deviation of the acoustic parameters in the recordings with smartphones and professional devices from the original song recordings, we conducted a playback test. Our results showed that depending on the user group, citizen scientists produced many to a lot of recordings of valid quality for further bioacoustic research. Differences between the recordings provided by the citizen and the expert group were mainly caused by the technical quality of the devices used—and to a lesser extent by the citizen scientists themselves. Especially when differences in spectral parameters are to be investigated, our results demonstrate that the use of the same high-quality recording devices and calibrated external microphones would most likely improve data quality. We conclude that many bioacoustic research questions may be carried out with the recordings of citizen scientists. We want to encourage academic researchers to get more involved in participatory projects to harness the potential of citizen science—and to share scientific curiosity and discoveries more directly with society.

Introduction

Citizen science (hereinafter abbreviated as CS) flourishes globally and has received significant recognition from diverse stakeholders in recent years. It is acknowledged for its potential to contribute to the transformation of the scientific system [1], promote global biodiversity monitoring [2], inform policies [3] as well as educate and promote scientific research in society [4]. In contrast to the traditional scientific research process, volunteers are involved in various activities of knowledge production for science and society [5]. CS is not a new hype: it has, especially in ornithology, a long tradition. For example, as early as 1749, one of the first CS projects in Finland collected data on migratory birds [6]. Today volunteers contribute large amounts of data in ornithological monitoring [7], which provide invaluable data for identifying trends in population numbers over time [8]. With the expansion of the internet and the increasing availability of user-friendly, cost-effective technology, citizen scientists got access to sophisticated data collection and transmission technology [9, 10]. Smartphone-based applications (mobile apps) allow citizen scientists to easily send photos, video, audio recordings, observation data and GPS positions [11]. This opened up new opportunities for CS in the field of bioacoustics, which otherwise depended on expensive equipment. In the field of ornithology, songbird dialects have been studied for decades [e.g. 12–14], with as well as without citizen scientists. So far, there are only a few CS projects with a focus on geographic variation in birdsong. One prominent European example is the Yellowhammer, *Emberiza citrinella*, with a detailed large-scale mapping of geographic variation of song dialects based on acoustic data collected by citizen scientists [14, 15]. Recently, a CS project based in North America successfully investigated the variation in chipping sparrow's song [16].

Although CS data are increasingly recognized as both a complement to and a replacement for conventional data sources [17, 18], there is still an ongoing intense debate about challenges such as data quality [e.g. 19]. Both the lack of knowledge, skills and motivation of the participants [20, 21] and insufficient study design of CS projects [22, 23] are discussed as potential reasons for poor data quality. Interestingly, some studies have however shown that citizen scientists were more careful in their measurements and annotations because they were quite aware of their novice status [e.g. 24, 25]. Other studies have found a learning effect among citizen scientists and an increase in data quality over time [e.g. 26]. Nevertheless, this has led some scientists to generalise and thus, to *per se* consider CS data to be inferior to expert data (hereinafter abbreviated as EX; [25, 27]). At times, studies using CS data faced problems in being published in peer-reviewed journals [28]. Data quality, however, is a multidimensional measurement of accuracy, completeness, consistency and timeliness [29] that consists of a variety of attributes [30]. How data quality can be assessed strongly depends on the research question and thus on the parameters under consideration [for an overview see 31 or 32]. Comparisons between CS and EX data often focus on ecological aspects, i.e. the quality of species distribution maps [e.g. 33] or the occurrence of species for monitoring data [34]. In most cases, experts provided better monitoring data, because citizen scientists underrepresented [35] or overrepresented the species to be studied [36, 37]. Bernard and colleagues [38] found that monitoring data do not differ between citizen scientists and experts when frequent species with high detection probabilities were investigated. However, other studies have shown that—regardless of the frequency of occurrence of the species under investigation—citizen scientists produce equivalent data to experts, which were considered reliable and comparable [e.g. 26, 39, 40]. In this comparison, however, it is important to note that there are projects in which the knowledge, skills and accuracy of the citizen scientist are crucial for the validity and quality of data (e.g. eBird https://ebird.org/home). Additionally, there are projects in which the knowledge and skills of the citizen scientists are less important for the quality of the collected data, as

13

these are generated for example with an app and subsequently checked by scientists (e.g. Bird-NET [41], our study).

At present, studies on data quality of CS recordings in the field of bioacoustics, for example for song dialect research, are missing. To conduct dialect research, in particular, a large dataset with many recordings from many different males and regions is important as well as a high number of included songs and song types within the recordings. Especially in the field of song dialects, where the regional song variations between populations are studied over geographical distances, there is a great potential to use the power of CS. A high recording quality is required to be able to examine spectrograms, a visual way of representing the signal over time at various frequencies. For the investigation of regional variations, mainly the occurrence of song types is considered [40]. This song type classification can be performed semi-automatically by using cross-correlation or visual inspection of the recordings, which requires a high signal-to-noise ratio. Both approaches have already been successfully conducted in the song analysis of the common nightingale, *Luscinia megarhynchos* [42]. In nightingale song research, mainly nocturnal recordings have been used as these are easier to generate due to the continuous singing of the males at night. Further, the nocturnal song is more diverse due to its function of attracting females than the diurnal song, which males use for territorial defence [43]. There is yet no indication that certain song types are sung merely at night or during the day (personal observation). Nocturnal singing is also easier for humans to hear because of the largely low or absent background noise. The resulting higher recording quality also makes the nocturnal song more suitable for semi-automatic cross-correlation measurements.

Nightingale song consists of several song categories which have, due to different volumes and spectral characteristics, different range characteristics and thus different signal-to-noise ratios. Whistle songs (Fig 1A) for example, have a long-range transmission [44, 45] whereas rapid trills (Fig 1B) degrade quickly over distance [45, 46], which means that their usability for semi-automatic cross-correlation measurements might be different. Thus, to better understand the impact of CS and EX recording devices on the recording quality, all song categories need to be tested. In addition, measurements of frequencies [e.g. 47] and durations [e.g. 48] have already been used in dialect studies with other bird species (MacGillivray's Warbler, *Geothlypis tolmiei*, and grey-breasted wood-wren, *Henicorhina leucophrys*), although these have not yet been examined in CS recordings to assess the quality. Moreover, it has not yet been systematically investigated whether the assumption is valid that the use of different recording devices in the analysis of nightingale songs can be neglected due to their stereotypical song learning [49].

To contribute to the further development of CS in bioacoustics, we here compared the quality of nightingale song recordings collected either via a smartphone app by citizen scientists or with professional recording devices by EX in a case study. In the nightingale CS project, all citizen scientists were called upon to participate without restriction through various public channels (radio, newspaper, etc.). They did not receive any detailed briefing or protocols before or during the breeding season, nor did they receive any information or feedback for the exact generation of the recordings (time, place, duration, orientation of the smartphone, etc.). It can be assumed that due to the German species name "Nachtigall" (nightingale), which contains the word "Nacht" (night), many participants thought that the nightingale sings mainly or only at night. Furthermore, the nightingale is better heard at night due to the low (a)biotic and anthropogenic background noises, signifying to citizen scientist that nightingale males only or at least mainly sing at night. Midnight excursions offered as part of the project between 23:00 and 1:00 hours might have further confirmed these assumptions. In addition, we did not specify to the CS when in the breeding season they should generate recordings since, in the case of the nightingale changes in the breeding season such as declining song performance [50], lower

14
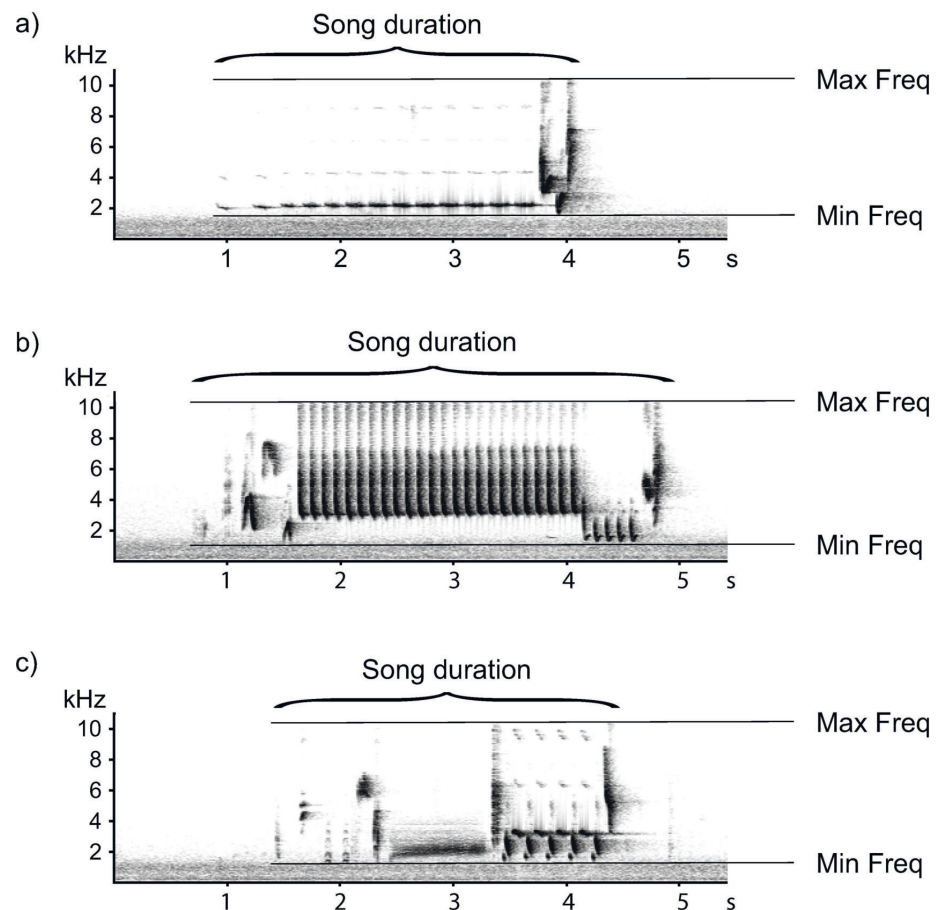
Fig 1. Exemplary spectrograms demonstrating the measured temporal and spectral parameters of the song duration, maximum frequency (Max Freq) and minimum frequency (Min Freq) in nightingale song for a) whistle songs, b) trill songs and c) buzz songs.

number of different song types and higher repetition rate of the same song type (personal observation) have been found. This is why continuous recordings over the breeding season are also important for dialect studies in the nightingale.

We were particularly interested in the question of whether app recordings collected by citizen scientists are valid for identifying dialects in nightingales. To inquire this, we first examined the timing, quantity and quality of CS and EX recordings and then compared standard parameters in recordings generated simultaneously under identical conditions with either professional or mobile recording devices. We hypothesised that CS recordings made with smartphones are by large as valid for dialect studies in the nightingale as EX recordings. We predicted that 1) the CS recordings differ in the temporal coverage to EX recordings (H1: time of day/calendar week), 2) the CS recordings are of comparable quantity and quality to the EX recordings (H2: data quality), 3) the CS recordings are more likely to be valid with an increasing number of individual recordings (H3: improvement) and 4) the CS and EX recordings

15

differ in recording quality because of the technical differences of the devices (H4: microphone comparison).

Material and methods

The nightingale as model species

The common nightingale (*Luscinia megarhynchos*) is a well-suited model species for a CS project in Berlin, as it is omnipresent in spring from around mid-April to late June during day and night. Its charismatic song is easy to identify even for laypeople and recordings of nocturnal songs easily reach a good recording quality, as there is hardly any background noise or other bird species to be heard. In addition, the song of the nightingale is so loud (74 dB (A) at 1 m distance [51]) that it is easily perceived by humans even from a distance and can thus be well recorded. Males possess an extraordinary large song-type repertoire (approx. 180 different song types per male) and differ considerably in repertoire size [49]. The song is mostly examined and classified on the song type level. Noteworthy categories of songs include whistles, trills and buzzes (Fig 1). Due to its highly complex song, the nightingale is also an interesting prospect model species for dialect studies. However, this research question has not yet been investigated for this species.

Citizen science recordings—the 'nightingale citizen science project'

We conducted the nightingale citizen science project at the Museum für Naturkunde Berlin (MfN), Germany. In spring, the project invited participants to generate nightingale audio recordings with their smartphone via the mobile app 'Naturblick' (see more details [52]). The app's pattern recognition supported citizen scientists in identifying audio recordings by presenting the top three candidates of each classification run [53, 54]. We designed the project with a very low threshold of participation in order to engage as many people as possible. The CS project therefore only set the target to record singing nightingales with the app. We did not provide any explicit specifications as to where, when, how long and how often to record a nightingale song. The 'Naturblick' app also encouraged users to share a nightingale recording with the CS project if it was taken by chance. The maximum duration of the recordings was limited to two minutes due to technical reasons. The citizen scientists could decide for themselves whether they wanted to share their recordings anonymously or with an individual username. The project was conducted over a two year period (2018–2019). Since twice as many nightingale song recordings were generated in the second project year, we used only the CS data from 2019 from Berlin for all further analyses. All CS data was based on recordings with the smartphone app 'Naturblick' (sampling rate = 44,1 kHz, bitrate: 256 Kbit/s, audio encoder: AAC Low Complexity (AAC-LC) audio codec). GPS coordinates were automatically included in the metadata for all recordings.

Expert recordings

For the EX recordings, we used different datasets, which were created in the same time period as the CS data with expert equipment by academic researchers. First, students of the Freie Universität Berlin (FU) recorded nine nightingale males as part of a master course in four locations of Berlin 'Volkspark Friedrichshain' (52°31'39.9648", 13°25'58.656"), 'Dreipfuhl' (52° 26'49.272", 13°16'19.6752"), 'Rehberge' (52°35'7.7244", 13°11'6.1512") and 'Tiergarten' (52° 30'51.0804", 13°21'38.3076") between 22 May and 04 June 2018. The FU data were recorded using a Sennheiser microphone (Sennheiser ME66/K6 directional microphones; 44,100 Hz, 16-bit resolution) connected to a Tascam Dr-40 4-Track Portable Digital Recorder. Second,

16

twelve additional one-hour long nightingale recordings were generated by academic experts of the MfN between 28 April and 07 May 2019 in Berlin. We recorded spontaneous nocturnal songs for individual, non-banded males in the field during the established recording time between 23:00 and 2:00 hours. We recorded three males at about the same time at night in the same area. This resulted in six recordings in the communal park 'Volkspark Friedrichshain' (52˚31'39.9648", 13˚25'58.656") and six recordings in a green space in the area of 'Altglienicke' (52˚24'27.524", 13˚31'3.1476"). We used three professional recording devices (two Zoom H2n recorder and a Marantz solid-state recorder PMD660 (sampling frequency: 44,1 kHz; resolution: 16 bit) with a Sennheiser ME66/K6 Microphone (Georgsmarienhütte, Germany). The microphones were equipped with windbreakers. For later analyses, we randomly selected six recordings from 2018 and six recordings from 2019. Here we did not aim for a comparison between the years, but we rather aimed to use a wide selection of different EX recordings.

Verification of recordings

All audio recordings were visualised for further analyses using Avisoft SASLab Pro 5.2 (R. Specht, Berlin, Germany). As the recordings via the app were generated in MP3 and m4a formats the recordings were transferred into the WAV format to be opened by Avisoft. For this purpose, we used the program WaveLab 7. Audio analyses were conducted using the same settings (sampling rate = 22,050 Hz, FFT = 1024 points, Hamming-Window, overlap 93,75%). The CS recordings were analysed visually and acoustically for nightingale songs, nightingale calls, the song of another bird species but which was not a nightingale, and no birds. A very small number of well-trained citizen scientists (n = 4) supported this step of recording classification. We only used nightingale songs for further analysis.

Comparison between the recording times of the CS and EX group

We determined the time of day and calendar weeks for all recordings. As recommended in the literature [49], the EX recordings were made at standardised times (between 23:00–3:00 hours) when nightingales are particularly reliably singing—especially in the beginning of the season—and the SNRs are most likely high due to the low background noise. The citizen scientists, on the other hand, had no instructions as to when they should record. Since the probability of making many and valid recordings on the one hand at night and on the other hand at the beginning of the season is high, we assumed that the times for recording CS and EX would therefore overlap. The time and date of the CS recordings were recorded directly via the 'Naturblick' app, actively shared with the CS project and delivered as metadata.

Comparison between the relative percentage of valid recordings in CS and EX data

For the comparison between approaches, we used the CS recordings from 2019 (n = 5679) and the EX recordings from 2018 (n = 6) and 2019 (n = 6). We evaluated the relative percentage of recordings of nightingale song, other bird species and no birds (all recordings = 100%; number of real nightingales / 100% = relative percentage). Furthermore, we categorized the nightingale song recordings as to whether at least one song type in its entirety was recognizable by both, syllables and elements in the spectrograms (in the following abbreviated as 'ist' = identifiable song types) or to a lesser degree, i.e. some syllables or elements were not clearly shown in the spectrograms ('nist' = non-identifiable song types). The former were seen as indicators of a valid recording quality, the latter of a reasonable recording quality that could however not be used for dialect research based on the identification of song types. We examined the cumulative duration of recordings in order to determine the scope of the dataset. The duration of the

17

CS recordings was supplied directly by means of the metadata. This includes the entire duration of the recording, but not the start and end of singing within the recording.

Comparison between the relative percentage of valid recordings in CS data among different user types

Based on their username and the number of recordings that they shared with the project, citizen scientists were divided into three user groups: 1) one-time users who had generated only one recording (one recording), 2) frequent users who made several recordings (two to nineteen recordings) and 3) power users who made many recordings (more than 19 recordings). This classification was based on the graphical distribution of the number of recorders and the number of recordings. This curve flattened out at 20 recordings per participant. For the quantitative analysis, we used the parameters described above. Furthermore, we examined the number of songs within a recording, since a recording's duration does not indicate how many songs are included.

Comparison between the signal-to-noise ratio (SNR) in CS and EX recordings

We examined all song categories for potential differences, as the nightingale's song categories have different transmission characteristics and thus different signal strengths. We selected three different song types from each song category (three whistles, three buzzes, three trill songs = in total nine different song types) for 'ist' CS nocturnal recordings from 2019 and the EX recordings from 2018 and 2019. For each of these song types, we randomly selected a sample of 10 recordings out of each data source from the Berlin 'Volkspark Friedrichshain'. We used the R-package warbleR [55] to automatically determine the signal-to-noise ratios (SNR) of recordings. For this purpose, the start and the end of a song were selected via an interactive spectrogram display in R using the mouse cursor. The SNR values were then automatically determined for the marked area. Referring to Araya-Salas and colleagues [56], we defined recordings to be of valid quality if they had a SNR over 10 dB. However, other authors recommend lower thresholds for the SNRs, such as Barmatz and colleagues [57, 58].

For both CS and EX recordings, we lack information about exact distances to the singing bird. We assumed that the citizen scientists approached singing males as closely as possible. The EX recordings were conducted by placing a professional recorder underneath a song post of a prospective male (see Fig 2). At night, nightingales hardly move but remain sitting on their song posts. During the day, males move around more often (personal observation); they are marking their territory by singing and therefore make use of several song posts located on the border of their territory. In previous studies, SNR values were also obtained without direct distance measurement to the bird [57, 58]. These studies evaluated the usability of monitoring recordings in terms of their SNR values. Likewise, we here aimed to evaluate via a SNR analysis whether the CS recordings were valid to determine song types by semi-automatic cross-correlation.

Comparison between the playback test recordings of CS and EX recording devices (smartphones vs. professional equipment)

To test whether measurements of spectral and temporal parameters would be influenced depending on the very different devices used by citizens or experts, we performed a standardized playback test. In September 2020, we simulated a singing nightingale and recorded it with several devices that differed considerably in terms of both, recording quality and price. In this

18
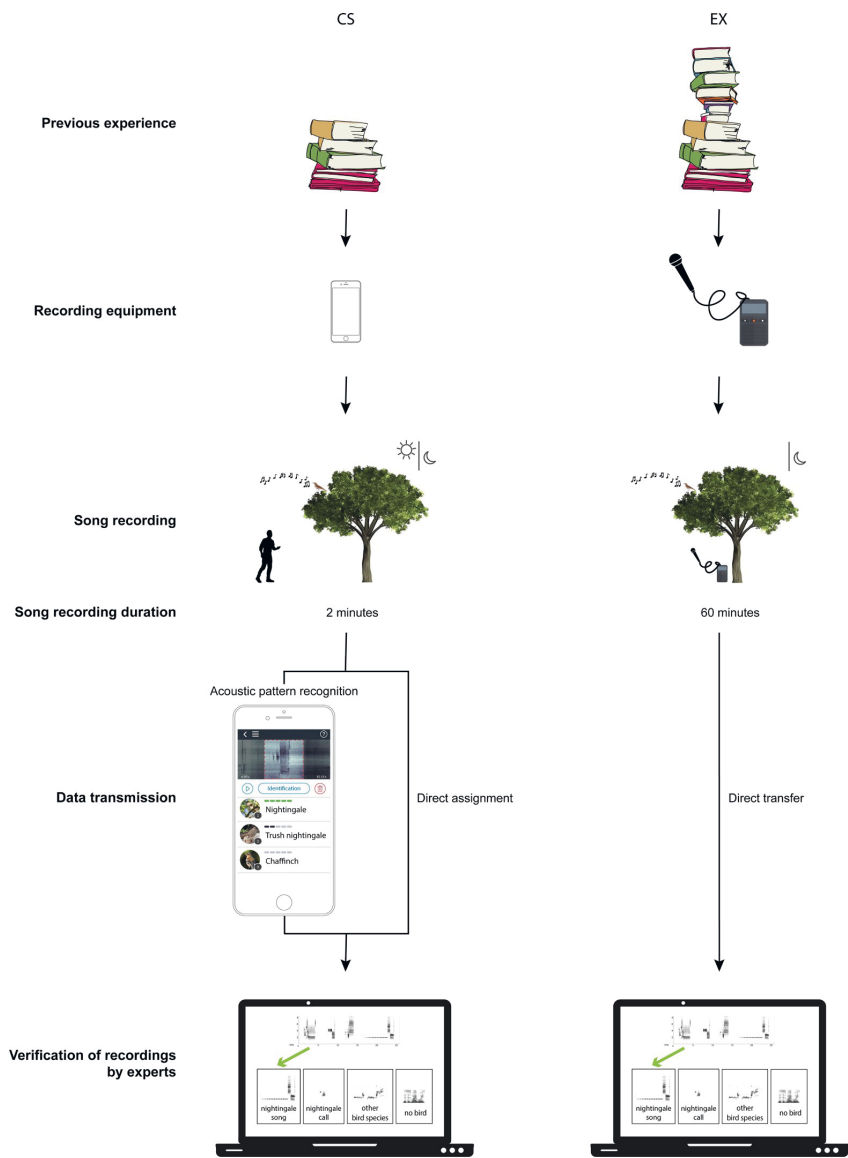
Fig 2. Process of generating recordings of CS (left) and EX (right). The similarities and differences in previous experience, recording equipment, song recording, song recording duration, data transfer and verification of recordings by experts are pointed out (from top to bottom).

simulation, a loudspeaker (JBL Charge 3) was placed on a chair 8 m away from a bench on a windless and sunny day. The playback was done at 8 m spacing, as in our personal experience this is a good average for the natural distance in the field when recording a nightingale. On the bench, in total 12 different recording devices were placed side by side and the microphones of the smartphones were aligned facing the loudspeaker. The devices were positioned in a

19

horizontal position, as this by our experience seemed to be the most common position by citizen scientists when recording with a smartphone.

We based the choice of smartphone brands used for the playback on those that were most frequently used for the CS app recordings. The sensitivity of the smartphones was not standardised via the gain settings. As expert equipment, we tested recording devices that have been used for generations for the EX recordings (Zoom H2n recorder and Marantz solid state recorder PMD660 with a Sennheiser ME66/K6 Microphone (Georgsmarienhütte, Germany)). As smartphones, we tested 10 different devices widely used: the smartphone brands Apple (three devices), Google (two devices), HTC (two devices) and Samsung (three devices). The loudspeaker was set to 74 dB (A) at a distance of 1 m using a calibration device (TEAC Df-1). This corresponds to the natural source level of a nightingale [51]. The loudspeaker was used to broadcast a recording that contained three whistles, buzz and trill song types each (duration of the audio file = 96 seconds). We chose different song categories and within these three different song types to cover different frequencies and to check the frequency response of the various devices. We did not perform a standardised test where different frequencies are independently assessed, as we intended to test the devices under natural conditions. To compare the recording quality of all devices and their built-in/external microphones, the audio was then simultaneously recorded: for the smartphones with the 'Naturblick' app and in the case of the two recording devices on the built-in data carrier. As a standard, all recorders had two built-in microphones. By default, we used only the first channel of the recordings for our following measurements. Subsequently, we performed standard measurements of spectral properties of the song type in the spectrogram, i.e. the minimum and maximum frequency, as well as the duration of the song types, was measured (Fig 1). We defined the song type duration as the duration from the beginning of the first to the end of the last element (in seconds). The measurements of acoustic parameters (frequencies and song duration) were done for the original recordings, which were played back in the test, as well as for the generated recordings of the different devices. One person carried out measurements twice manually in Avisoft. The spectrogram settings were adjusted in advance. For the three parameters (minimum frequency, maximum frequency and duration) we then averaged the values obtained for all song types and categories, determined the deviation from the measured values of the original recording and compared the results among the devices.

Statistical analysis

To test for differences in the quality (SNR) of recordings between the CS and EX approach, we used Welch tests for normally distributed data and Wilcoxon signed rank tests for not normally distributed data: We used Friedman tests to compare differences among parameters (minimum frequencies, maximum frequencies and durations) and subsequent Nemenyi-Wilcoxon-Wilcox tests as a post hoc test. We set statistical significance at $p \leq 0.05$. All statistical analyses were performed using R (version 4.0.0).

Ethics statement

This study compares the quantity and quality of CS and EX recordings of nightingale song. The data of the citizen scientists were shared with our project with their approval via the 'Naturblick' app. The EX recordings were made during a university course at the 'Freie Universität Berlin'. For both types of recordings, we obtained the consent of participants to analyse their data. In Germany, the approval of an ethics committee is not required for such research questions and was therefore not obtained. We have therefore received all the necessary permissions required in Germany.

20

Results

In total, more than 3,000 citizen scientists recorded a cumulative 82 hours of song and 35,462 songs without exact specifications as to when and how often to collect data. The EX recordings contributed a cumulative 12 hours of song and 4,921 songs to the study's dataset. The CS and EX recordings cannot be compared in terms of these overall figures, as they were recorded with different specifications: CS—no time specifications when, how and how long they recorded; recording limit of two minutes, and EX—time specifications when, how and how long they recorded; recording limit of one hour. For this reason, in all further comparisons, we used the relative percentage of valid quality recordings rather than the total number and focused on the quality of the data that could be used for further bioacoustic analysis.

Recording times of citizen scientists and experts

Most CS and EX recordings were generated between 23:00 and 00:00 hours (Fig 3A). Overall, CS recordings were made during all times of day without any temporal gaps. The fewest CS recordings were made between 3:00 and 4:00 hours. In total, citizen scientists generated recordings between the 16th and the 26th calendar week. Most recordings were generated between the 17th and 21st calendar week by both, the CS and EX group (Fig 3B). The fewest CS recordings were made in the 13th, 25th and 26th calendar week.

The relative percentage of recordings with valid quality in CS and EX data

The comparison between the CS and EX recordings showed that the EX group produced 100% 'ist' recordings and the CS group 53% of 'ist' recordings (S1 Table). In addition, nightingale recordings were sent from twelve countries within the CS project. The EX recordings come from one country. The CS group also generated 37% 'nist' recordings, 2% call recordings, 4% other bird species and 4% 'no bird' recordings (for 2019 see Fig 4). The mean duration of the recordings was higher for the EX recordings (60 minutes) than for the CS recordings (54 seconds). The cumulative recording time was higher for CS recordings (89 hours) than for EX recordings (6 hours).

The relative percentage of CS recordings with valid quality for single, frequent and power users

The comparison among user types showed that the frequent users had the highest number of all categories of recordings, the second-longest average recording time and the longest cumulative recording time (S2 Table). Power users had the highest percentage with 85% of 'ist' recordings, whereas the percentage of single and frequent users was similar with 50% and 47% (Fig 5A). Conversely, power users generated the lowest percentage of 'nist' recordings (13%), followed by one-time users (40%) and frequent users (46%). Within all the user groups, 'no bird' song recordings made up the lowest percentage of recordings (power users = 2%, frequent users = 1%, one-time users = 3%). Power users had the longest mean duration of 'ist' recordings with 99 seconds, followed by frequent users with 72 seconds and single users with 59 seconds (Fig 5B). Among power users, the 'nist' recordings had nearly the same total duration (67 seconds) as the category 'other bird song recordings' (70 seconds). The 'nist' recordings were longer in their total duration than the total duration of 'no bird' song recordings (6 seconds) for the power users, and this was the other way round for the single users (46 seconds to 5 seconds). For all user groups, the 'no bird' song recordings were the shortest (single users: 5 seconds, frequent users: 24 seconds, power users: 6 seconds). The cumulative number of songs differed between the user groups with frequent users having the largest cumulative number of songs (n = 11,845), followed by power users (n = 3,602) and single users (n = 1,288; Fig 5C).
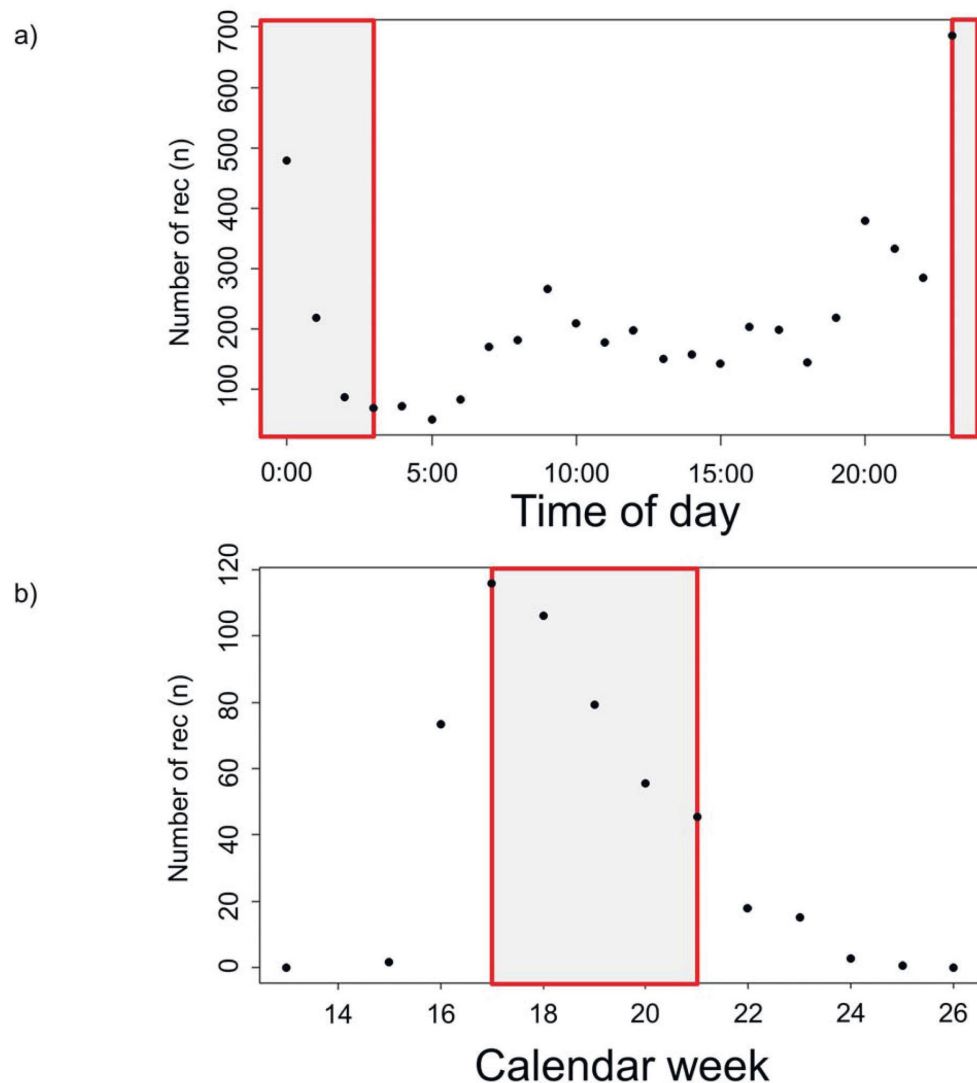
Fig 3. Temporal distributions of 5,679 citizen science recordings (abbreviated as rec) with the 'Naturblick' app for 2019. In red, the times of previous six nightingale studies are shown (2004–2017; [61–64, 79, 80]). a) Representation of the number of recordings in relation to the time of day. b) Presentation of the weekly number of recordings in the course of the breeding season. https://doi.org/10.5281/zenodo.4817236.

https://doi.org/10.1371/journal.pone.0253763.g003

The signal-to-noise ratio (SNR) in CS and EX recordings

The EX recordings had in all song type categories a mean SNR that was higher than 10 dB. The CS recordings only had a median higher than 10 dB for the song type category whistle and trill. The SNRs of CS data differed significantly among all song type categories (whistle, trill and buzz) from the EX data (whistle: Wilcoxon signed rank test: W = 287, p-value = 0.016, trill: Wilcoxon signed rank test: W = 316, p-value = 0.048, buzz: Welch test: t = -5.5705,

22

a)   b)

CS   EX
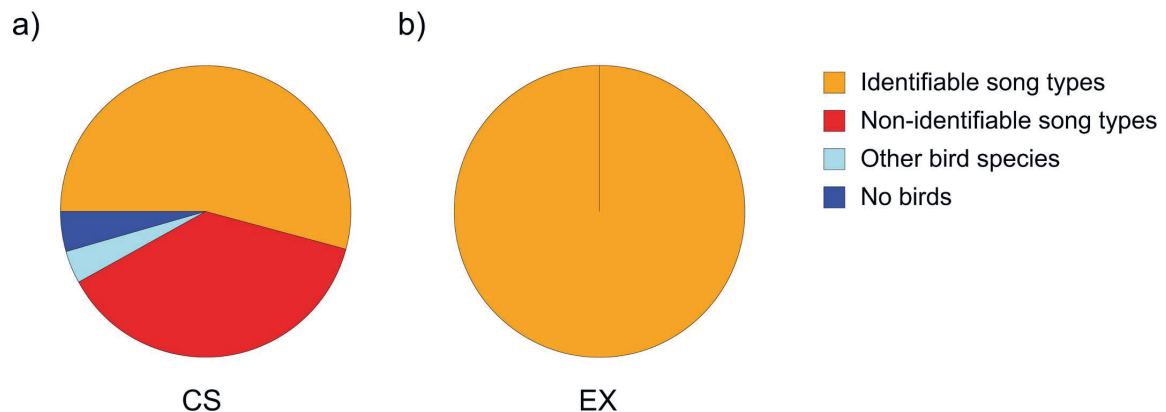
Fig 4. Comparison of defined recording categories for 2019, based on a) 5,679 citizen science recordings (CS) with smartphones via the 'Naturblick' app and b) six expert recordings (EX) with high-quality microphones. The categories are displayed in different colours (red: one song type was in its entirety recognizable by syllables and elements in the spectrograms, orange: some syllables or elements were not clearly shown in the spectrograms, light blue: other bird species and dark blue: no birds). https://doi.org/10.5281/zenodo.4817236.

df = 56.704, p-value <0.001). The number of valid recordings determined by SNR values above 10 dB (according to Araya-Salas and colleagues [55]) was higher for EX recordings (in total—whistle: 23, trill: 27, buzz: 24) than for CS recordings (in total—whistle: 17, trill: 11, buzz: 18, Fig 6B).

Playback test for recordings of CS and EX recording devices (smartphones vs. professional equipment)

We found significant differences between the recording quality of different recording devices, expressed by the deviation from the minimum and maximum frequency as well as duration from the original playback file (Figs 7A–7C and 8). Overall, the professional PMD recording device showed the lowest deviation from the original recording in the minimum and maximum frequency measurements, while the also professional Zoom H2n and the HTC smartphones showed the highest deviation. We detected significant differences among recording devices in the deviation from the minimum frequencies of the test playback file (Fig 7A; Friedman test: F = 26.079, n = 6, df = 5, p-value <0.001). We detected the greatest deviation from the original recording in the minimum frequency for the professional Zoom Hn2 and the smallest deviation was measured for the professional PMD recording device. Post-hoc test revealed that measurements for the professional PMD recording device differed significantly from the smartphone brands HTC (Nemenyi-Wilcoxon-Wilcox, n = 2, p = <0.001) and Apple (Nemenyi-Wilcoxon-Wilcox, n = 2, p = 0.004) as well as from the professional Zoom H2n (Nemenyi-Wilcoxon-Wilcox, n = 2 p = <0.001). We found significant differences in the deviation from the maximum frequencies of the test playback file among the recording devices (Fig 7B; Friedman test: F = 37.444, n = 6, df = 5, p-value <0.001). The smartphone brand HTC had the largest deviation at the maximum frequency from the original recording and the professional PMD had the smallest deviation. Post-hoc tests showed that the deviation of the maximum frequencies of the HTC smartphones differed significantly from the smartphone brands Google (Nemenyi-Wilcoxon-Wilcox, n = 2, p = 0.004) and Apple (Nemenyi-Wilcoxon-Wilcox, n = 2, p = 0.002) as well as the professional PMD recording device (Nemenyi-Wilcoxon-

23
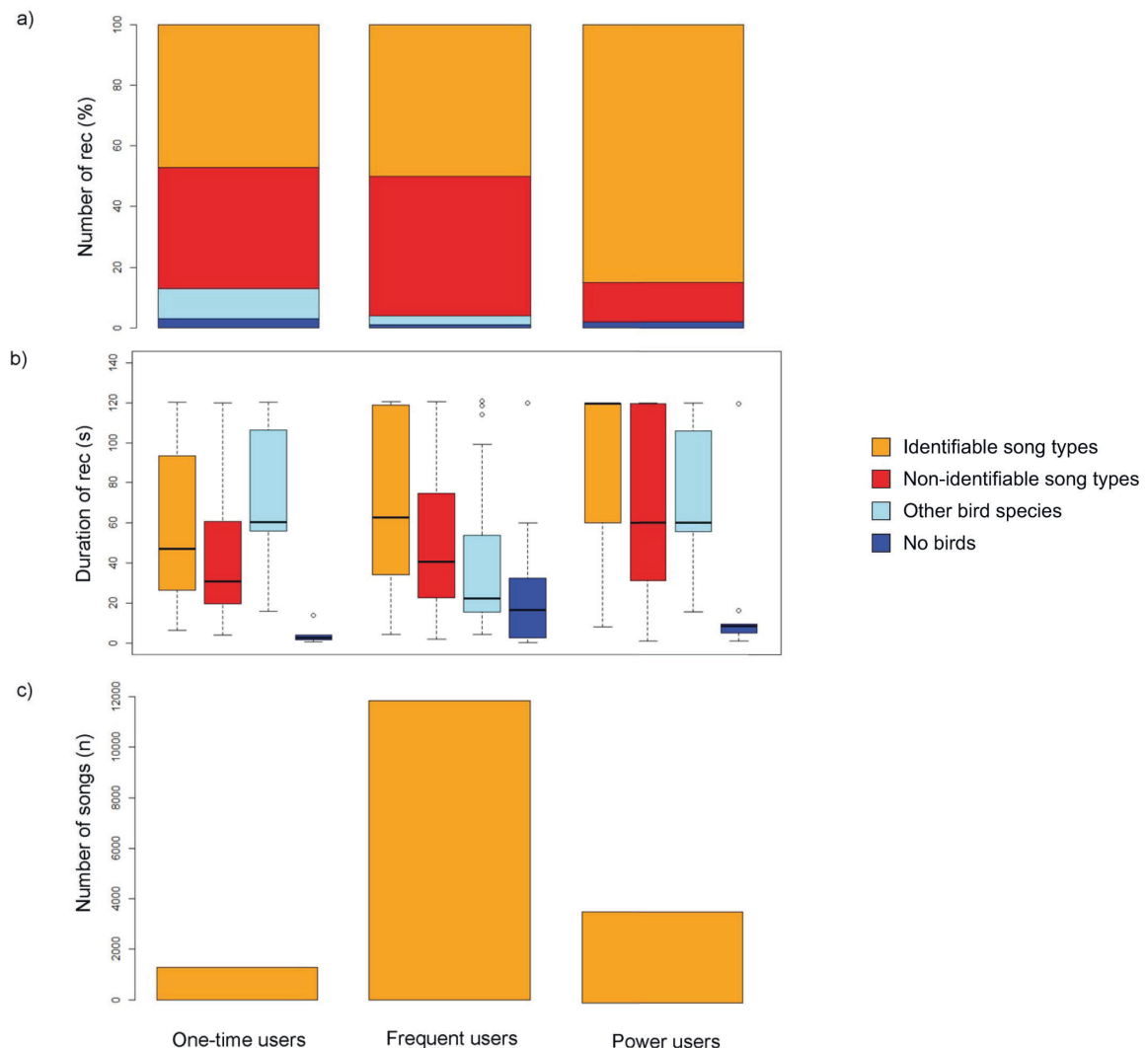
Fig 5. Comparison of the number of citizen science recordings (rec) in the year 2019, based on recordings with the 'Naturblick' app of 245 one-time users, 361 frequent users and 18 power users (one-time users = 1 recording; multiple users = 2–19 recordings; frequent users ≥ 20 recordings). In red: one song type was in its entirety to be identified by syllables and elements in the spectrograms, in orange: some syllables or elements were not clearly shown in the spectrograms, in light blue: other bird species and in dark blue: no birds. a) The number of recordings. b) The duration of recordings shown in the boxplots. The median is represented by a solid black line and the mean by a dashed black line within a box. The borders of boxes are 25 and 75 percentiles. The bars above box plots indicate significant differences between two stimulus categories. c) Shows the cumulative number of songs in all user groups for the first category. https://doi.org/10.5281/zenodo.4817236.

Wilcox, n = 2, p < 0.001). Moreover, post-hoc tests revealed that the deviations of the professional Zoom H2n differed significantly from the smartphone brands Google (Nemenyi-Wilcoxon-Wilcox, n = 2, p = 0.044) and Apple (Nemenyi-Wilcoxon-Wilcox, n = 2, p = 0.03) and the professional PMD recording device (Nemenyi-Wilcoxon-Wilcox, n = 2, p < 0.001). The

24
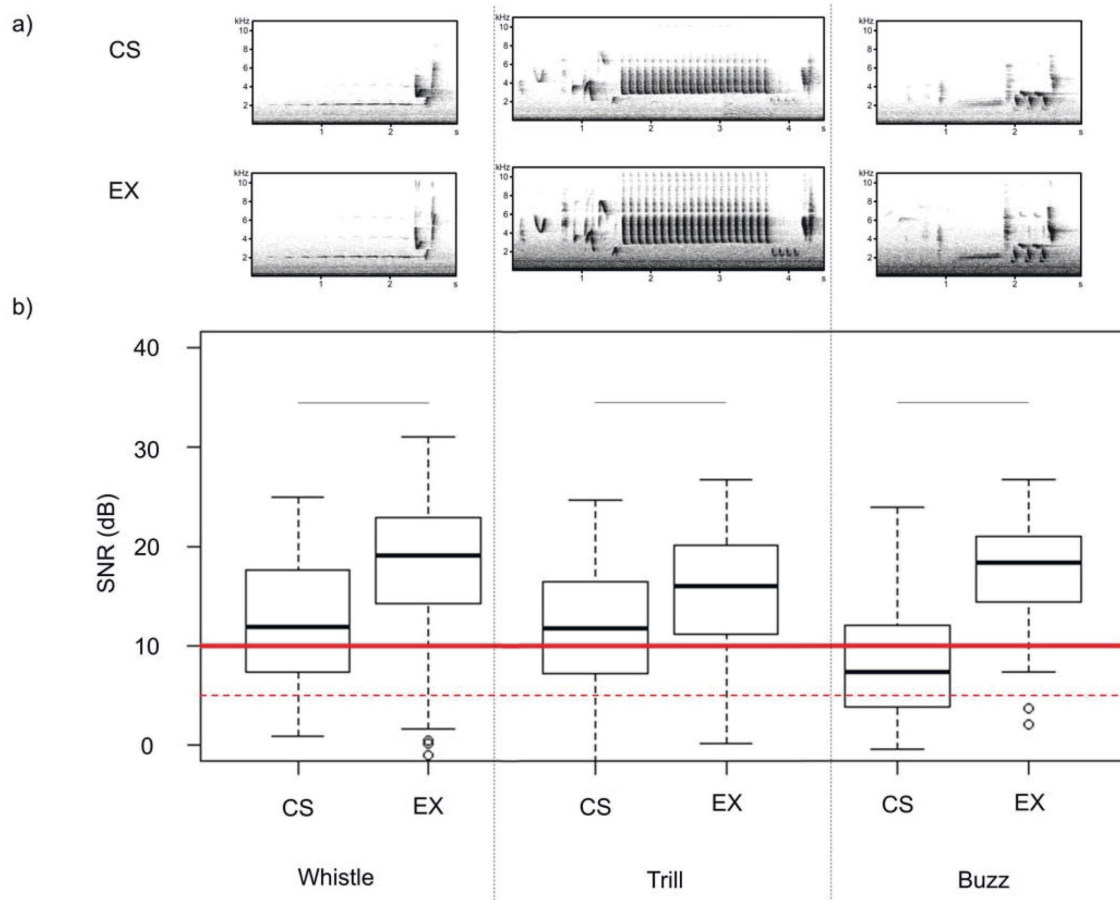
Fig 6. Comparison of citizen science (CS) recordings with the 'Naturblick' app and expert recordings with equipment using professional microphones (EX) in the year 2018 and 2019. a) Spectrograms of CS recordings (top) compared to EX recordings (below) for the one example per song categories (whistle, trill and buzz). b) Comparison the signal to noise ratio (SNR) as boxplots for three of these selected song types per each category: whistle (left), trill (middle) and buzz (right). The median of the boxplot is represented by a solid black line and the mean by a dashed black line within a box. The borders of boxes are 25 and 75 percentiles. The bars above box plots indicate significant differences between two stimulus categories. SNRs over 10 dB (red line) were defined in this study as valid quality according to Fitzpatrick and colleagues [58]. The red dotted line indicates a value of 5 dB, a threshold used by Barmatz and colleagues [57, 58] for valid quality. The black lines indicate significant differences. All recordings were examined and displayed under the same settings (sample rate = 22,050 Hz, FFT = 1024 points, Hamming-Window, overlap 93,75%). https://doi.org/10.5281/zenodo.4817236.

https://doi.org/10.1371/journal.pone.0253763.g006

recording devices did not differ among their deviations from the original duration of the playback file (Fig 7C; Friedman df = 5, n = 2, p-value = 0.063; for details see S3 Table).

Discussion

This study highlights the potential of citizen science for bioacoustic research. We found that bioacoustic research—for instance here, dialect research on the nightingale—could be carried out both with the recordings of citizen scientists and experts. The frequently discussed lack in the overall data quality of CS data could not be confirmed in this case study. Instead, we were able to show that the quality of CS recordings was in large parts equivalent and not *per se* inferior to EX recordings. Furthermore, our study confirms the notion that CS has the advantage
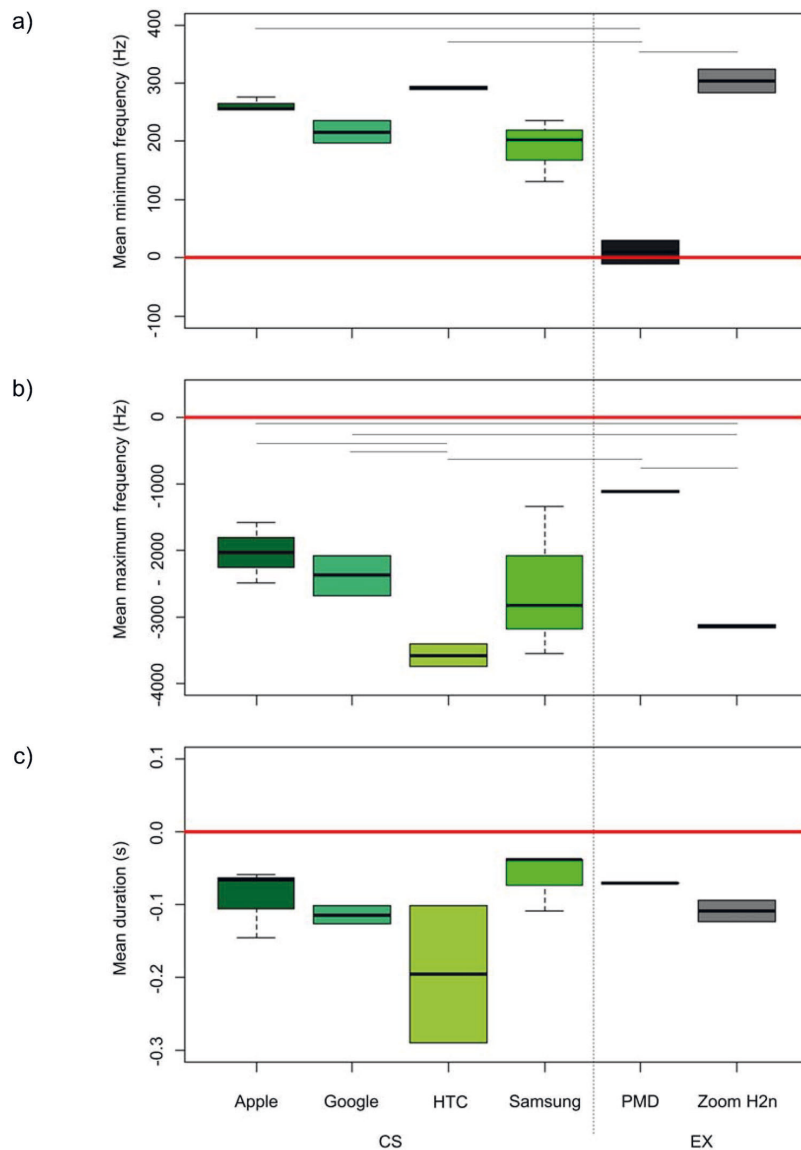
Fig 7. Comparison of deviations from original playback files of the different recording devices, i.e. smartphone brands (CS) and expert recordings equipment with professional microphones (EX) in a test. Depicted in boxplots is the deviation in a) the minimum frequency, b) the maximum frequency and c) song duration from playing back a test file of nightingale song. The median is represented by a solid black line within a box. The borders of boxes are 25 and 75 percentiles. The red line shows a zero line. The closer a deviation to the zero line is, the smaller it was. The black lines above boxplots indicate significant differences between the recording devices tested.

https://doi.org/10.1371/journal.pone.0253763.g007

to generate large datasets. In the following, we discuss the two aspects that we believe may have influenced our results: the human and the technical factor.

26
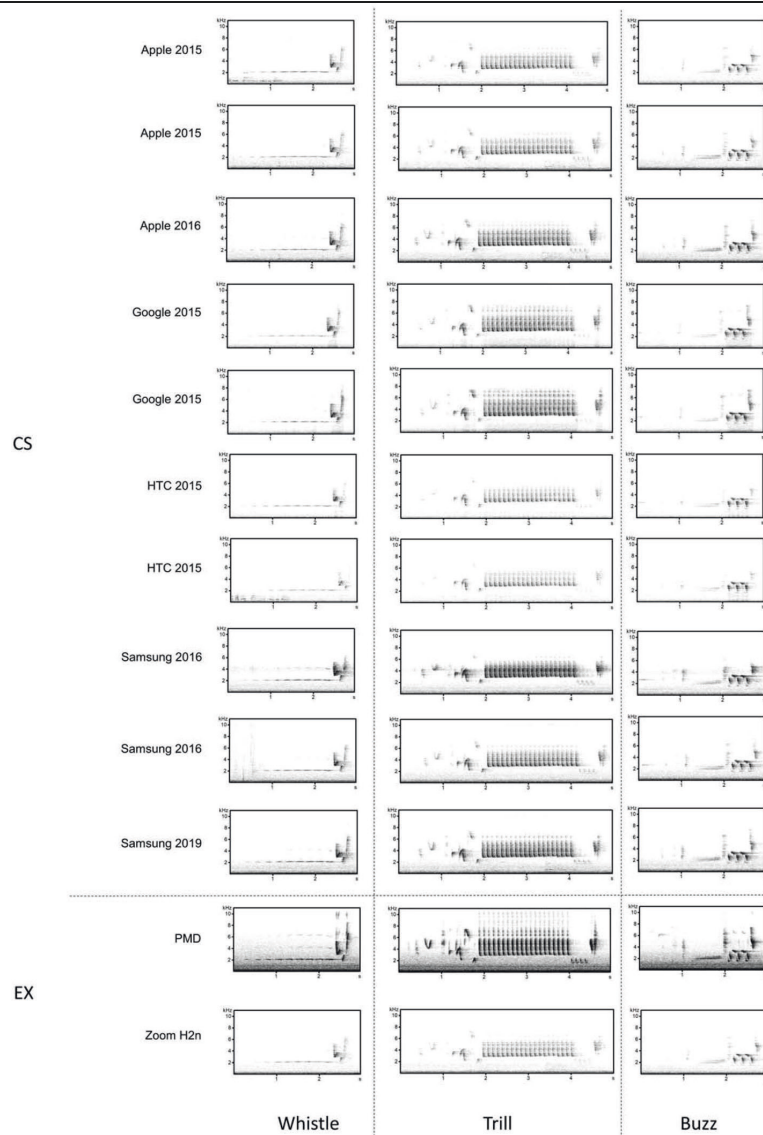
Fig 8. Comparison of citizen science (CS) recordings with the 'Naturblick' app and expert recordings with equipment using professional microphones (EX) during a playback test. Spectrograms of CS recordings (top) compared to EX recordings (below) for the one example per song categories (whistle, trill and buzz).

The human influence on CS data

Many studies comparing citizen scientists with experts assume that the latter *by proxy* generate better data due to their extensive experience and scientific background [59, 60]. However, in our study, the knowledge and skills of the citizen scientists were less decisive for the recording quality, as participants were able to validate their recordings with a pattern algorithm and data was additionally verified by experts. We infer from our results that previous experience is not

27

always required when collecting data that is mainly relying on good technical equipment. Our comparison of the recording times between the two approaches (non-standardised CS times vs. standardised expert times) showed that citizens' studies without a specification achieved a much higher temporal and spatial resolution than previous nightingale expert studies. Academic researchers are usually looking for valid quality as well as long recordings; thus, they usually specify research designs with standard recording times in which they can expect valid singing performances of their study objects. In consequence, this leads to a sometimes narrow temporal recording window and consequently, the bandwidth of recording times and possible variations in singing behaviour of study species may be lost. Earlier nightingale studies mainly used the nocturnal song for their analyses, as this signals an unpaired status of the males and allows a comparison between the males under study and their songs [e.g. 61]. Nevertheless, recordings outside the night (thus dawn and day) and over the whole breeding season are of crucial value for dialect research. Firstly, the number of different song types decreases over the breeding season and the repetition rate of the same song type increases (personal observation). And secondly, it has not yet been investigated whether the song types differ between night and day. The CS recordings covered not only the whole breeding season but the whole day. Most CS recordings were without specifications made during the same time (day and week) as previous nightingale studies [e.g. 49, 61, 62]. The majority of recordings for both approaches were made at night. We detected two recording peaks at 9 am and 8 pm. The first peak was probably created when most people were on their way to work and the second one when people actively went out to listen to nightingales in the beginning of the night. It is worth mentioning that there is no time of the day during which a song recording of a nightingale was not made in the course of the CS project. This indicates that the CS recordings have much greater temporal coverage because they were made throughout the day. Also, the weekly coverage by citizen scientists was greater than by the experts in earlier nightingale studies [e.g. 49, 61, 62].

Without having precise specification, citizen scientists generated the most recordings between the 17th and the 25th calendar week. Standardized recording times in previous nightingale studies were usually between the 17th and 18th calendar week [63], up to 21st calendar week [64] or up to 22nd calendar week [62]. Thus, without having any concrete guidelines, following the singing pattern of the nightingale resulted in similar and partly overlapping recording periods. The decreasing number of CS recordings as the breeding season progressed is certainly also due to the fact that with more male nightingales being paired, the nocturnal singing decreases, which is generally associated with female attraction [43]. Our results confirm that both, the CS and the EX approach are suited to documenting the song behaviour of the nightingale over the course of its breeding season. We hence propose that CS recordings are a valuable addition to conventional nightingale studies and this approach should also be applied to studies on other vocal bird and animal species.

EX recordings had a higher percentage of data that was valid for further analysis, yet CS data had a higher total number and total duration of valid nightingale song recordings. Despite the given limitation in a maximum recording duration of two minutes, the cumulated recording time of the CS data was significantly higher than the EX data. This highlights that also with many short recordings a large dataset may be generated. The required recording length certainly depends on the specific research questions. For example, long recordings may be needed for song analyses on an individual level and also depend on the species' song repertoire, i.e. the number of song types. Based on our definition, non-identifiable song type recordings have a reasonable recording quality that could not, however, be used for dialect research since complete song types without any missing elements should be considered for analysis. But they might however be useful for further structural analyses, e.g. based on elements and syllables. Citizens have also recorded other bird sounds that were no nightingales. On the one hand, a song very similar to the

28

nightingale could have caused these false identifications, on the other hand, many citizen scientists may have suspected that only nightingales sing at night. However, due to light pollution in urban areas, species such as the blackbird, *Turdus merula*, and the robin, *Erithacus rubecula*, also sing at night [65, 66]. During the quality check of all recordings, we noticed that citizen scientists often only tested the 'Naturblick' app without the intention to record a nightingale at this particular moment. On other occasions, nightingales may have ceased singing, when the recording started. These circumstances, among others, might have led to 'no bird' recordings.

The citizen scientists in our case study created a large dataset without any form of an external incentive to take extensive recordings of nightingales (e.g. badges or other award systems). The widespread notion that there is a lack of intrinsic motivation in participants without an external reward system [21, 22], was in our case not true. Citizen scientists generated many 'ist' recordings even without video instruction or assistance from a scientist and lack of knowledge and skills. Yet, with an instructional video and a detailed explanation of which kind of recordings should be generated and why, the number of recordings possibly would have been even higher in number, mean duration and/or quality. For dialect research, many recordings with many song types from different males as well as from different regions is more helpful than a few recordings of selected individuals at one or few location(s). Previous research has demonstrated that the repertoire of a nightingale male is well represented with one-hour recordings [e.g. 64]. Long-term stationary recordings of an individual would be, for example, advantageous when studying changes in song behaviour in the course of the breeding season. However, large numbers of geographically widespread citizen scientists offer novel potential for dialect research. For instance, different populations with a large distribution range can be investigated in a short period of time, which a single research group would not be able to achieve. Our study underlines that citizen scientists facilitate large-scale data collection. Furthermore, our results are in contrast to the notion that special knowledge is crucial to generate a large valid dataset [21, 22]. Additionally, our results contradict the assumption that it is always necessary to instruct citizen scientists by scientists to obtain better data [58]. At least not for the investigation of nightingale song dialects that would be based on visual inspection of spectrograms; the bioacoustic research question that we chose as background for a criterion for data quality.

Our comparison among the three user groups within the CS project (single, frequent and power users) showed that as predicted, user groups generated a dataset of different recording quality. We demonstrated that with the higher number of recordings in the group of power users, the percentage of recordings with valid quality for further analysis was also higher—without any special instructions or training by scientists. Furthermore, 'no bird' song recordings made up the lowest percentage in all user groups followed by other bird species recordings. The latter indicates that with the support of the pattern-recognition of the 'Naturblick' app, all user types were equally good at distinguishing nightingale from vocalizations of other bird species. Although the power users generated the most high-quality nightingale recordings, the larger group of frequent users in terms of the number of participants contributed recordings with the longest cumulative recording time and the largest cumulative number of song types. We conclude that even without extensive training, citizen scientists are able to generate recordings of valid quality. The assumption that citizen scientists must be trained and instructed over a longer period of time [11] was hence not true for this case study. We believe that in the science of citizen science [67] a general assessment of the data quality, training needs as well as required knowledge and skills of participants is not possible. We feel that it is particularly important to consider the skill and knowledge requirements specific to the research questions when planning CS activities.

29

The technical influence on CS data

Several differences that we found between CS and EX recordings are most likely not due to training deficits in citizens, but down to technical differences in the recording devices used. Experts had a higher relative percentage of valid recordings determined via SNRs than citizen scientists. We assume, however, that this was not exclusively caused by the fact that the citizen scientists produced less and poorer nightingale recordings with an unqualified recording behaviour, but was also due to the technical limitations of some smartphone brands. For example, studies also showed that microphones with a low SNR lead to noisy recordings in which weak and/or distant signals are no longer to be identified clearly [68]. Darras and colleagues [69] showed that even using different professional microphones resulted in quite different SNRs. Thus, the quality of audio recordings is not only influenced by training but mostly by the choice of device i.e. microphone quality. Some of the EX recordings in our data generated with professional equipment were yet marked via the SNR values as poor in quality. This underlines that low recording quality is not a phenomenon that can and should be attributed exclusively to CS data.

We found lower SNRs in CS recordings for all song categories than in EX recordings because smartphones have limitations in frequency. Smartphone microphones are optimized for the frequency range of human speech (300–5,000 kHz), and their directional characteristic suppresses surrounding noise, especially in the bass range [70]. An earlier study showed similar results to ours: Smartphones only generated reliable recordings in a range of 300–3,400 Hz and uncontrollable compression levels occurred at higher frequencies [71]; in fact, exactly the range of maximum frequencies (up to 10 kHz) found in the nightingale's song. Furthermore, Yousefian and Loizou [72] demonstrated that some smartphones use several microphones to separate the ambient noise from the speech during phone calls, and Martın-Donas and colleagues [73] found that the more microphones a smartphone has, the more background noise it filters out. This kind of audio pre-processing is further perturbing the frequency sensitivity of the recording devices used by CS. Moreover, most of the EX recordings were generated with the PMD and an external Sennheiser microphone.

We expected that due to their long-range transmission characteristics [44], SNRs of whistles would be better than the SNRs of trills, since their signal strength decreases faster over distance [45, 46]. We detected this in the EX recordings, but not in the CS recordings. In the former, the SNR values of the whistles and trills were equal (12 dB). In the latter, the SNR values of the whistlers were higher (19 dB) than those of the trills (16 dB). Furthermore, in the EX recordings, the SNRs of the buzz were 18 dB higher than the trills and almost as good as the whistles. However, the song category of the buzz had the lowest SNRs (below 10 dB) in CS recordings and was therefore by definition of [56] not of valid quality. However, referring to other sources such as [51, 59], which define valid quality recordings above at an SNR of 5 dB, all CS recordings would be of valid quality whereby e.g. durations can be measured. Thus, depending on how strict the threshold is, either only the whistles and trills (at 10 dB) or also the buzz songs (at 5 dB) are of valid quality. In the nightingale, buzz song types are an indicator of the quality of a male [50] and may therefore be presented at high pitched volume. Trill song types are used in aggressive interactions [74] and as an indicator of male quality [75], which may have led to a greater range and thus to better SNRs in general. Nevertheless, all CS recordings showed a significant lower SNR value than the EX recordings. Thus, the CS recordings were not equal to EX recordings, but still of valid quality. From personal experience, we can say that the significantly worse SNRs of CS recordings did not, however, lead to the fact that song types could be assigned to categories or types more poorly. Therefore, we believe that our CS recordings of nightingales can be used for further research questions, such as dialect research since

the study of regional variations in bird song is usually based on the relative occurrence of song types [75], rather than spectral parameters.

In our test recordings, quite large deviations from the original values were found in the parameters minimum and maximum frequency as well as for song durations. The device producing the least deviation from the original's minimum and maximum frequencies in its recordings was the professional PMD. The Zoom H2n, which is also used as professional equipment however, showed the greatest deviation from the original recording at both the minimum and maximum frequency. This shows that even a professional recorder without an external microphone may provide even worse measurement values than smartphones. One reason for this could be that the Zoom recorder is in particular designed to be used for long-term monitoring recordings and not for fine structure analyses. This is because long-term monitoring surveys use pattern recognition algorithms to determine the potential occurrence of bird species by analyzing vocalizations. Here mainly frequency and duration ranges are used instead of precise measurements, which also allows the use of recordings with lower SNRs [76]. This is in line with our data, which showed that the SNR values of the Zoom recorder were valid for our further analyses. Out of the smartphones, the recordings of the HTC (a low-cost brand) had very large deviations from the acoustic parameters of the original recordings. The duration most likely showed large deviations, because not all frequencies were recorded and thus the song type was not represented in its entirety with all elements and syllables. The smartphone brands Apple, Nexus and Samsung showed a significantly larger deviation in the frequencies than the PMD, but were comparable in the durations. Interestingly, in terms of song duration, the Samsung smartphone devices performed better than the PMD. Our test in comparing the recording quality of the devices showed that the quality of the brand, and thus ultimately the price, actually played an important role here in the frequency measurements, but not in the measurement of durations. Hence, the statement that the use of different recording devices can be neglected in the analysis of nightingale songs because of their very stereotypical song learning [49], does not apply when comparing measurements of frequencies of recordings that were made with different recording devices. Clare and colleagues [77] already described that measurements alone cannot accurately determine the effectiveness and usability of a dataset. The authors recommended that data quality should be presented as a kind of threshold value, which is derived from both data accuracy and the intended analyses. They suggested that how data quality is assessed, indeed depends on the research question. The question investigated here was whether CS recordings of nightingale may be used for song dialect research. Our prediction that the quality would be first, valid for this research question and second, comparable to EX data, could by large be confirmed.

Conclusion and recommendations

Our study shows that nightingale recordings generated via a smartphone app are valid to investigate dialects at the song type level. Based on our results of the poor recording quality of low-cost smartphone brands, we would recommend the use of external and regularly calibrated microphones for projects relying on the analysis of fine structures. Kardous and colleagues [78] already recommended the utilization of external, calibrated microphones to improve the overall accuracy and precision of sound recordings. They showed that this eliminated much of the variability and limitations of built-in smartphone microphones. Furthermore, when measuring frequencies and durations, we suggest to ideally always use the same brand of recording devices, so that any differences found are due to song variations and not to discrepancies in the microphone used. Standardizations with regard to citizen scientists'

31

devices, e.g. by equipping them through the project or recommending the use of certain brands, could provide a solution for small projects and without continental-scale.

Despite the limited academic experience of citizen scientists, we strongly advocate that CS can make valuable contributions to science itself. In view of our results, we believe that CS recordings offer the potential to support bioacoustic and in particular dialect research with extensive datasets. Our case study demonstrated that dialect research on the song type level of the nightingale can be carried out with both CS and EX recordings. We support the notion of Butcher and Niven [17] as well as Lisjak and colleagues [18], stating that CS may complement and potentially replace conventional data sources. Based on our findings, we thus want to encourage bioacoustic researchers to first, use data made available by volunteers and non-academics, such as the recordings in open databases like XenoCanto (https://www.xeno-canto.org/) for instance in dialect research and second, to further establish CS as a research approach that has the dual benefit of providing large, and with newer technology also valid data, as well as opening science to society.

Supporting information

S1 Table. Comparison of functional types of recordings by CS recordings with the 'Naturblick' app and six EX recordings with traditional microphones in the years 2018 and 2019. The representation of the raw data are given as subsumed numbers. Comparison between the relative percentage of recordings with valid quality for further analysis between CS and EX data. https://doi.org/10.5281/zenodo.4817236.
(PDF)

S2 Table. Comparison of functional types of recordings in the years 2019, based on 245 one-time users, 361 frequent users and 18 power users recordings with the 'Naturblick' app. One-time users have generated one recording, multiple users shared on average 2–19 recordings and frequent users shared on average over 20 recordings. Comparison between the relative percentage of recordings with valid quality for further analysis between CS data among different user types. https://doi.org/10.5281/zenodo.4817236.
(PDF)

S3 Table. Comparison of the deviation of acoustic parameters for the minimum frequency, the maximum frequency and song duration from a sample of nine nightingale song types between citizen science recordings (CS) using different smartphone brands and expert recordings (EX) using equipment with professional microphones. Comparison between the playback test recordings of CS and EX recording devices (smartphones vs. professional equipment).
(PDF)

Acknowledgments

32

Author Contributions

Conceptualization: Denise Jäckel, Kim G. Mortega, Silke L. Voigt-Heucke.

Data curation: Denise Jäckel.

Formal analysis: Denise Jäckel, Ulrich Brockmeyer.

Investigation: Denise Jäckel, Kim G. Mortega, Silke L. Voigt-Heucke.

Methodology: Denise Jäckel, Kim G. Mortega, Ulrich Brockmeyer, Silke L. Voigt-Heucke.

Project administration: Silke L. Voigt-Heucke.

Software: Ulrike Sturm, Omid Khorramshahi.

Supervision: Kim G. Mortega, Silke L. Voigt-Heucke.

Validation: Denise Jäckel, Kim G. Mortega, Ulrike Sturm.

Visualization: Denise Jäckel, Ulrich Brockmeyer.

Writing – original draft: Denise Jäckel.

Writing – review & editing: Kim G. Mortega, Ulrike Sturm, Omid Khorramshahi, Silke L. Voigt-Heucke.

References

1. Hecker S, Haklay M, Bowser A, Makuch Z, Vogel J. Citizen science: innovation in open science, society and policy: UCL Press; 2018.

2. Chandler M, See L, Copas K, Bonde AMZ, Lopez BC, Danielsen F, et al. Contribution of citizen science towards international biodiversity monitoring. Biological Conservation. 2017; 213:280–94.

3. Fritz S, See L, Carlson T, Haklay MM, Oliver JL, Fraisl D, et al. Citizen science and the United Nations sustainable development goals. Nature Sustainability. 2019; 2(10):922–30.

4. Bonney R, Ballard H, Jordan R, McCallie E, Phillips T, Shirk J, et al. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Online Submission. 2009.

5. Haklay M, Motion A, Balázs B, Kieslinger B, Greshake TB, Nold C, et al. ECSA's Characteristics of Citizen Science. 2020. Zenodo. Available from: http://doi.org/10.5281/zenodo.3758668 PMID: 33158395

6. Greenwood JJD. Citizens, science and bird conservation. Journal of Ornithology. 2007; 148(1):S77–S124.

7. Johnston A, Hochachka W, Strimas-Mackey M, Gutierrez VR, Robinson O, Miller E, et al. Best practices for making reliable inferences from citizen science data: case study using eBird to estimate species distributions. BioRxiv. 2019:574392.

8. Magurran AE, Baillie SR, Buckland ST, Dick JM, Elston DA, Scott EM, et al. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology & Evolution. 2010; 25(10):574–82. https://doi.org/10.1016/j.tree.2010.06.016 PMID: 20656371

9. Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, et al. Citizen science. Next steps for citizen science. Science. 2014; 343(6178):1436–7. https://doi.org/10.1126/science.1251554 PMID: 24675940

10. Reed J, Raddick MJ, Lardner A, Carney K, editors. An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. 2013 46th Hawaii International Conference on System Sciences; 2013: IEEE.

11. Starr J, Schweik CM, Bush N, Fletcher L, Finn J, Fish J, et al. Lights, Camera . . . Citizen Science: Assessing the Effectiveness of Smartphone-Based Video Training in Invasive Plant Identification. Plos One. 2014; 9(11):e111433. https://doi.org/10.1371/journal.pone.0111433 PMID: 25372597

12. Lemon RE. How birds develop song dialects. The Condor. 1975; 77(4):385–406.

13. Bjerke TK, Bjerke TH. Song dialects in the Redwing Turdus iliacus. Ornis Scandinavica. 1981:40–50.

14. Diblíková L, Pipek P, Petrusek A, Svoboda J, Bílková J, Vermouzek Z, et al. Detailed large-scale mapping of geographical variation of Yellowhammer *Emberiza citrinella* song dialects in a citizen science project. Ibis. 2019; 161(2):401–14.

33

15. Hidayat T, Kurniawan I, Tapilow F, editors. Bird on Your Smartphone: How to make identification faster? IOP Conference Series: Materials Science and Engineering; 2018: IOP Publishing.

16. Searfoss AM, Liu WC, Creanza N. Geographically well-distributed citizen science data reveals range-wide variation in the chipping sparrow's simple song. Animal Behaviour. 2020; 161:63–76.

17. Butcher GS, Niven DK. Combining Data From the Christmas Bird Count and the Breeding Bird Survey to Determine the Continental Status and Trends of North America Birds; National Audubon Society: New York, NY,USA. 2007. Available from: http://www.audubon.org/sites/default/files/documents/report_1.pdf

18. Lisjak J, Schade S, Kotsev A. Closing data gaps with citizen science? Findings from the Danube region. ISPRS International Journal of Geo-Information. 2017; 6(9):277.

19. Lukyanenko R, Wiggins A, Rosser HK. Citizen science: An information quality research frontier. Information Systems Frontiers. 2019:1–23.

20. Whitelaw G, Vaughan H, Craig B, Atkinson D. Establishing the Canadian Community Monitoring Network. Environ Monit Assess. 2003; 88(1–3):409–18. https://doi.org/10.1023/a:1025545813057 PMID: 14570426

21. Conrad CC, Hilchey KG. A review of citizen science and community-based environmental monitoring: issues and opportunities. Environ Monit Assess. 2011; 176(1–4):273–91. https://doi.org/10.1007/s10661-010-1582-5 PMID: 20640506

22. Paulos E. Designing for Doubt Citizen Science and the Challenge of Change. 2009: Available from: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=9C47580863FBFC0B4391A26615E75F5A?doi=10.1.1.187.5824&rep=rep1&type=pdf

23. Bonter DN, Cooper CB. Data validation in citizen science: a case study from Project FeederWatch. Frontiers in Ecology and the Environment. 2012; 10(6):305–9.

24. Delaney DG, Sperling CD, Adams CS, Leung B. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions. 2008; 10(1):117–28.

25. Newman C, Buesching CD, Macdonald DW. Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation—"Sed quis custodiet ipsos custodies?". Biological Conservation. 2003; 113(2):189–97.

26. Falk S, Foster G, Comont R, Conroy J, Bostock H, Salisbury A, et al. Evaluating the ability of citizen scientists to identify bumblebee (Bombus) species. PLoS One. 2019; 14(6):e0218614. https://doi.org/10.1371/journal.pone.0218614 PMID: 31233521

27. Riesch H, Potter C. Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. Public understanding of science. 2014; 23(1):107–20. https://doi.org/10.1177/0963662513497324 PMID: 23982281

28. Theobald EJ, Ettinger AK, Burgess HK, DeBey LB, Schmidt NR, Froehlich HE, et al. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation. 2015; 181:236–44.

29. Wand Y, Wang RY. Anchoring data quality dimensions in ontological foundations. Communications of the Acm. 1996; 39(11):86–95.

30. Pipino LL, Lee YW, Wang RY, Data quality assessment. Commun. ACM. 2002; 45: 211–218.

31. Wang RY, Strong DM. Beyond accuracy: What data quality means to data consumers. Journal of management information systems. 1996; 12(4):5–33.

32. Lewandowski E, Specht H. Influence of volunteer and project characteristics on data quality of biological surveys. Conserv Biol. 2015; 29(3):713–23. https://doi.org/10.1111/cobi.12481 PMID: 25800171

33. Aceves-Bueno E, Adeleye AS, Bradley D, Brandt WT, Callery P, Feraud M, et al. Citizen Science as an Approach for Overcoming Insufficient Monitoring and Inadequate Stakeholder Buy-in in Adaptive Management: Criteria and Evidence. Ecosystems. 2015; 18(3):493–506.

34. Brown ED, Williams BK. The potential for citizen science to produce reliable and useful information in ecology. Conserv Biol. 2019; 33(3):561–9. https://doi.org/10.1111/cobi.13223 PMID: 30242907

35. Kremen C, Ullman KS, Thorp RW. Evaluating the quality of citizen-scientist data on pollinator communities. Conserv Biol. 2011; 25(3):607–17. https://doi.org/10.1111/j.1523-1739.2011.01657.x PMID: 21507061

36. Steger C, Butt B, Hooten MB. Safari Science: assessing the reliability of citizen science data for wildlife surveys. Journal of Applied Ecology. 2017; 54(6):2053–62.

37. Gardiner MM, Allee LL, Brown PMJ, Losey JE, Roy HE, Smyth RR. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment. 2012; 10(9):471–6.

34

38. Bernard ATF, Gotz A, Kerwath SE, Wilke CG. Observer bias and detection probability in underwater visual census of fish assemblages measured with independent double-observers. Journal of Experimental Marine Biology and Ecology. 2013; 443:75–84.

39. Paul K, Quinn MS, Huijser MP, Graham J, Broberg L. An evaluation of a citizen science data collection program for recording wildlife observations along a highway. J Environ Manage. 2014; 139:180–7. https://doi.org/10.1016/j.jenvman.2014.02.018 PMID: 24705097

40. Lovell S, Hamer M, Slotow R, Herbert D. An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. Biodiversity and Conservation. 2009; 18(12):3295–307.

41. Kahl S, Wood CM, Eibl M, Klinck H. BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics. 2021; 61: 101236.

42. Vokurkova J, Petruskova T, Reifova R, Kozman A, Morkovsky L, Kipper S, et al. The causes and evolutionary consequences of mixed singing in two hybridizing songbird species (Luscinia spp.). PLoS One. 2013; 8(4):e60172. https://doi.org/10.1371/journal.pone.0060172 PMID: 23577089

43. Amrhein V, Kunc HP, Naguib M. Seasonal patterns of singing activity vary with time of day in the nightingale (Luscinia megarhynchos). Auk. 2004; 121(1):110–7.

44. Richards DG. Alerting and message components in songs of rufous-sided towhees. Behaviour. 1981; 76(3–4):223–49.

45. Naguib M. Reverberation of rapid and slow trills: implications for signal adaptations to long-range communication. The Journal of the Acoustical Society of America. 2003; 113(3): 1749–1756. https://doi.org/10.1121/1.1539050 PMID: 12656407

46. Naguib M, Schmidt R, Sprau P, Roth T, Flörcke C, Amrhein V. The ecology of vocal signaling: male spacing and communication distance of different song traits in nightingales. Behavioral Ecology. 2008; 19(5): 1034–1040.

47. Hamao S. Asymmetric response to song dialects among bird populations: the effect of sympatric related species. Animal Behaviour. 2016; 119: 143–150.

48. Pitocchelli J, Guerra D, Kender J. Macrogeographic variation in song of the MacGillivray's Warbler (Geothlypis tolmiei). The Wilson Journal of Ornithology. 2018; 130(3): 716–729.

49. Kiefer S, Sommer C, Scharff C, Kipper S. Singing the popular songs? Nightingales share more song types with their breeding population in their second season than in their first. Ethology. 2010; 116: 619–626.

50. Weiss M, Kiefer S, Kipper S. Buzzwords in females' ears? The use of buzz songs in the communication of nightingales (Luscinia megarhynchos). PLoS One. 2012; 7(9):e45057. https://doi.org/10.1371/journal.pone.0045057 PMID: 23028759

51. Brumm H. The impact of environmental noise on song amplitude in a territorial bird. Journal of Animal Ecology. 2004; 73(3):434–40.

52. Sturm U, Tscholl M. The role of digital user feedback in a user-centred development process in citizen science. Journal of Science Communication. 2019; 18(1): A03.

53. Stehle M, Lasseck M, Khorramshahi O, Sturm U. Evaluation of acoustic pattern recognition of nightingale (Luscinia megarhynchos) recordings by citizens. Research Ideas and Outcomes. 2020; 6: e50233.

54. Lasseck M. Improving Bird Identification using Multiresolution Template Matching and Feature Selection during Training. In CLEF. 2016: 490–501.

55. Araya-Salas M, Smith-Vidaurre G. warbleR: an R package to streamline analysis of animal acoustic signals. Methods in Ecology and Evolution. 2017; 8(2): 184–191.

56. Araya-Salas M, Smith-Vidaurre G, Webster M. Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. Bioacoustics. 2019; 28(1): 57–73.

57. Barmatz H, Klein D, Vortman Y, Toledo S, Lavner Y. Segmentation and Analysis of Bird Trill Vocalizations. In 2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE). IEEE. 2018: 1–5.

58. Barmatz H, Klein D, Vortman Y, Toledo S, Lavner Y. A Method for Automatic Segmentation and Parameter Estimation of Bird Vocalizations. In 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE. 2019: 211–216

59. Fitzpatrick M, Preisser E, Ellison A, Elkinton J. Observer bias and the detection of low density populations. Ecol. Appl. 2009; 19: 1673–79. https://doi.org/10.1890/09-0265.1 PMID: 19831062

60. Galloway A, Tudor M, Vander Haegen W. The reliability of citizen science: a case study of Oregon White Oak stand surveys. Wildl. Soc. Bull. 2006; 34: 1425–29.

61. Landgraf C, Hultsch H, Scharff C, Kipper S. What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. Journal of Ornithology. 2016; 157(1): 49–60.61

35

62. Kiefer S, Scharff C, Kipper S. Does age matter in song bird vocal interactions? Results from interactive playback experiments. Frontiers in Zoology. 2011; 8(1): 1–8. https://doi.org/10.1186/1742-9994-8-1 PMID: 21223557

63. Landgraf C, Wenchel R, Kaiser A, Kipper S. Singing onstage: female and male common nightingales eavesdrop on song type matching. Behavioral Ecology and Sociobiology. 2014; 68(7): 1163–1171.

64. Kiefer S, Spiess A, Kipper S, Mundry R, Sommer C, Hultsch H, et al. First-year common nightingales (Luscinia megarhynchos) have smaller song-type repertoire sizes than older males. Ethology. 2006; 112: 1217–1224.

65. Russ A, Rüger A, Klenke R. Seize the night: European Blackbirds (Turdus merula) extend their foraging activity under artificial illumination. Journal of Ornithology. 2015; 156(1): 123–131.

66. Kempenaers B, Borgström P, Loës P, Schlicht E, Valcu M. Artificial night lighting affects dawn song, extra-pair siring success, and lay date in songbirds. Current Biology. 2010; 20(19): 1735–1739. https://doi.org/10.1016/j.cub.2010.08.028 PMID: 20850324

67. Vohland K, Land-Zandstra A, Ceccaroni L, Lemmens R, Perelló J, Ponti M, et al. The Science of Citizen Science. Springer. 2021

68. Darras K, Batary P, Furnas B, Celis-Murillo A, Van Wilgenburg SL, Mulyani YA, et al. Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. Journal of Applied Ecology. 2018; 55(6):2575–86.

69. Darras KF, Deppe F, Fabian Y, Kartono AP, Angulo A, Kolbrek B, et al. High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. PeerJ. 2020: 8: e9955. https://doi.org/10.7717/peerj.9955 PMID: 33150056

70. Rahman T, Adams AT, Zhang M, Cherry E, Zhou B, Peng H, et al. BodyBeat: a mobile system for sensing non-speech body sounds. In MobiSys. 2014; 14: 2594368–2594386.

71. Krump G. Akustische Messmöglichkeiten mit Smartphones. 2012; 1. Available from: http://pub.dega-akustik.de/DAGA_2012/data/articles/000160.pdf

72. Yousefian N, Loizou PC. A dual-microphone speech enhancement algorithm based on the coherence function. IEEE Transactions on Audio, Speech, and Language Processing. 2011; 20(2): 599–609. https://doi.org/10.1109/TASL.2011.2162406 PMID: 22207823

73. Martın-Donas JM, López-Espejo I, Gomez AM, Peinado AM. A postfiltering approach for dual-microphone smartphones. 2018: Available from: https://www.researchgate.net/profile/Ivan_Lopez-Espejo/publication/329102476_A_postfiltering_approach_for_dual-microphone_smartphones/links/5d95ed51a6fdccfd0e72a64f/A-postfiltering-approach-for-dual-microphone-smartphones.pdf

74. Podos J. Motor constraints on vocal development in a songbird. Animal Behaviour. 1996; 51(5): 1061–1070.

75. Sprau P, Schmidt R, Roth T, Amrhein V, Naguib M. Effects of rapid broadband trills on responses to song overlapping in nightingales. Ethology. 2010; 116(4): 300–308.

76. Frommolt KH, Tauchert KH. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. Ecological Informatics.2014; 21:4–12.

77. Clare JD, Townsend PA, Anhalt-Depies C, Locke C, Stenglein JL, Frett S, et al. Making inference with messy (citizen science) data: when are data accurate enough and how can they be improved?. Ecological Applications. 2019; 29(2): e01849. https://doi.org/10.1002/eap.1849 PMID: 30656779

78. Kardous CA, Shaw PB. Evaluation of smartphone sound measurement applications (apps) using external microphones—A follow-up study. The Journal of the acoustical society of America. 2016; 140(4): 327–333. https://doi.org/10.1121/1.4964639 PMID: 27794313

79. Kipper S, Mundry R, Sommer C, Hultsch H, Todt D. Song repertoire size is correlated with body measures and arrival date in common nightingales, Luscinia megarhynchos. Animal Behaviour. 2006; 71 (1): 211–217.

80. Landgraf C, Wilhelm K, Wirth J, Weiss M, Kipper S. Affairs happen-to whom? A study on extrapair paternity in common nightingales. Current zoology. 2017; 63(4): 421–431. https://doi.org/10.1093/cz/zox024 PMID: 29492002

36

Chapter III

Unravelling the Stability of Nightingale Song Over Time and Space Using Open, Citizen Science and Shared Data

Unravelling the Stability of Nightingale Song Over Time and Space Using Open, Citizen Science and Shared Data

Denise Jäckel[1,2]*, Kim G. Mortega[1], Ulrich Brockmeyer[1], Gerlind U. C. Lehmann[3] and Silke L. Voigt-Heucke[1,4]

[1] Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany, [2] Faculty of Life Sciences, Humboldt-Universität zu Berlin, Berlin, Germany, [3] Evolutionary Ecology, Department of Biology, Humboldt-Universität zu Berlin, Berlin, Germany, [4] Animal Behavior, Institute of Biology, Freie Universität Berlin, Berlin, Germany

Open science approaches enable and facilitate the investigation of many scientific questions in bioacoustics, such as studies on the temporal and spatial evolution of song, as in vocal dialects. In contrast to previous dialect studies, which mostly focused on songbird species with a small repertoire, here we studied the common nightingale (*Luscinia megarhynchos*), a bird species with a complex and large repertoire. To study dialects on the population level in this species, we used recordings from four datasets: an open museum archive, a citizen science platform, a citizen science project, and shared recordings from academic researchers. We conducted the to date largest temporal and geographic dialect study of birdsong including recordings from 1930 to 2019 and from 13 European countries, with a geographical coverage of 2,652 km of linear distance. To examine temporal stability and spatial dialects, a catalog of 1,868 song types of common nightingales was created. Instead of dialects, we found a high degree of stability over time and space in both, the sub-categories of song and in the occurrence of song types. For example, the second most common song type in our datasets occurred over nine decades and across Europe. In our case study, open and citizen science data proved to be equivalent, and in some cases even better, than data shared by an academic research group. Based on our results, we conclude that the combination of diverse and open datasets was particularly useful to study the evolution of song in a bird species with a large repertoire.

Keywords: birdsong, cultural evolution, geographic variation, *Luscinia megarhynchos*, repository, song dialects, vocal learning

INTRODUCTION

Open science practices such as open data (OD), citizen science (CS), and data sharing may open new avenues for answering novel kinds of research questions. Open data are databases and platforms that are freely available to the public and can be reused without copyright restrictions (Kassen, 2013). Acoustic datasets, for example, are increasingly shared publicly both on museum

archives, e.g., the animal sound archive at the Museum für Naturkunde Berlin (MfN)[1], and in open platforms, e.g., Xeno-Canto[2]. Biologists are increasingly using these large open datasets of observations to answer a broad range of questions on evolution and ecology across time and space (Aplin et al., 2021; Dominguez et al., 2021). CS as a concept describes research conducted by non-professional scientists who are not institutionally employed in this field of science (as defined in the Greenbook Citizen Science; Bonn et al., 2017). The use of CS data is rapidly growing, especially in ornithology, and has been successfully used, e.g., for bird monitoring (de Camargo Barbosa et al., 2021), geographical distribution (de Jesus et al., 2021), and population trends (Neate-Clegg et al., 2020). Nevertheless, usage of CS data is not yet well established in ornithology (Weisshaupt et al., 2021) due to a lack of information and thus quality concerns about data availability, data bias, and data generation (Hochachka et al., 2012; Lukyanenko et al., 2016). This quality concern is why CS data are often not made publicly available in the sense of open science. The concerns can be overcome when different datasets are connected, in particular when CS data are used as a complement to fill in gaps (Kirchhoff et al., 2021) or when they are directly compared with data gathered by academic researchers. In an earlier study, we were able to show that citizen scientists made a surprising number of recordings of valid quality when compared to expert recordings and that differences between the recordings were mainly due to the technical quality of the equipment used. Recordings were valid to distinguish songs visually and conduct duration measurements. Frequency measurements showed significant differences depending on the cost, recentness, and brand of the smartphone. Thus, quality is multidimensional and always depends on the research question. In the current study, we conducted a quality dataset comparison based on the same parameters to ensure that open and CS data are considered as sufficiently valuable and are used more often in ornithology. To aid in establishing such a representative case study, we investigated temporal and geographical song variations, so-called dialects, in a bird species.
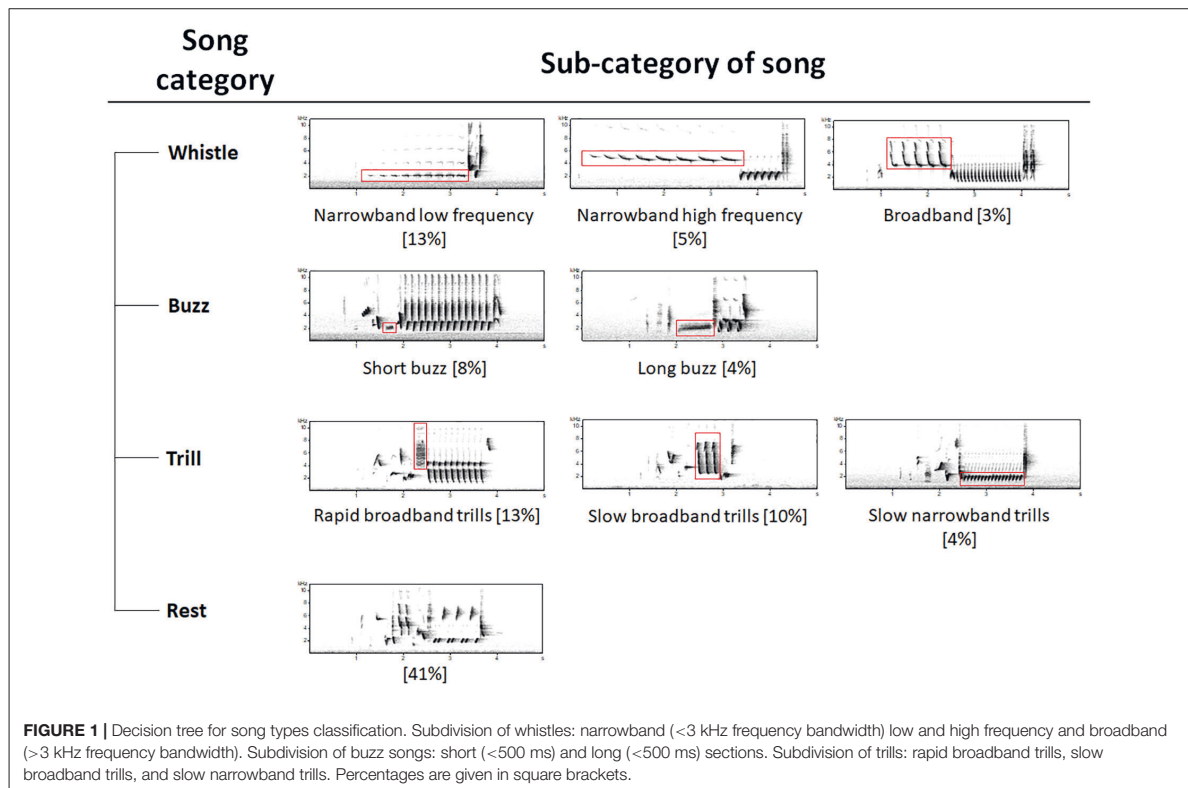
Studies of dialects have historically been based mostly on academic-collected data (in our case shared data) and began as early as (Poulsen, 1958), when Poulsen studied evolutionary processes leading to the formation of dialects, the shape and extent of spatial variation within a species, and the coding of dialects. Dialects have been found with academic-collected data in cetaceans (Sharpe et al., 2019), hummingbirds (Gaunt et al., 1994), parrots (Wright and Dahlin, 2018), suboscine passerines (Kroodsma et al., 2013), bats (Prat et al., 2017) and with CS data in songbirds (oscine, Searfoss et al., 2020). Song in birds is a sexually selected and culturally transmitted trait that plays a critical role, particularly in breeding ecology as dialects can restrict gene flow among conspecific populations (Baker and Mewaldt, 1978) or contribute to reproductive isolation and speciation resulting in ring species (Irwin et al., 2001). In the well-studied, mostly temperate model species, males sing to defend their territory and attract females (Catchpole and Slater, 2008).

Songbirds acquire their song mainly through social learning via cultural imprinting. In this process, young males learn their song from their parents or neighbors and pass those songs onto the next generation (Mennill et al., 2018). This process can lead to copying errors and modifications over time and space, resulting in a cultural evolution of song (Lynch, 1996). Sexual selection, cultural drift, environment, and population adaptation (immigration, emigration) influence whether song types are particularly common or rare in individual songs and populations (Derryberry et al., 2011). A song type is a unique sequence of elements with defined frequencies and durations, as described by Gil and Gahr (2002). If song types are beneficial to individuals in terms of environmental adaptation, female choice – or their own fitness – some songs types persist over time (Lynch, 1996). Unbeneficial song types can be loosed or modified, which lead to geographic variation that may evolve into distinct, local dialects over time (Beecher and Brenowitz, 2005). Local dialects reduce song variation at the microgeographic level compared to the macrogeographic level, resulting in greater differences between populations than within populations (Catchpole and Slater, 2008). Investigating dialects over time and space allows insight into the processes of song evolution (Rendell and Whitehead, 2005) and improves the understanding of songs as well as repertoires and thus vocal learning (Pozzi et al., 2010). Previous dialect studies have increasingly benefited from the use of open databases (Wright and Dahlin, 2018) and CS recordings (Diblíková et al., 2019). Due to the internet and GPS-based app development (Dickinson et al., 2012), these studies generated considerable data on a spatial and temporal scale that would be unachievable for traditional scientific studies due to financial or logistical constraints (Diblíková et al., 2019). These studies showed that songs and parts of them remain stable for up to three decades (Nelson et al., 2004) or change geographically especially in bird species with small repertoires, and typically with one to three song types (Diblíková et al., 2019; Searfoss et al., 2020). Nevertheless, it is not yet known how the song or parts of it evolve over a century or Europe-wide across the entire breeding range and in a bird species with a large repertoire.

In our study, we examined the common nightingale (*Luscinia megarhynchos*), a species with approximately 190 song types (Hultsch and Todt, 1982) and a Europe-wide breeding range (from the Mediterranean to Central Asia in the East and Central Europe in the North; Glutz von Blotzheim, 1988). The nightingale is particularly interesting for studies on song dialects due to its complex and well-studied song behavior in the field (e.g., Kiefer et al., 2006) and in the laboratory (e.g., Hultsch and Todt, 1996). Males are territorial, return to their natal breeding area, and have been found to also occupy the same territory in consecutive years (Roth and Amrhein, 2010). Songs can be classified into four song categories (whistle, buzz, trill, and so-called "other songs," **Figure 1**) as well as song types which form the song repertoire (doctoral thesis Hultsch, 1980; unpublished data). Nightingale males, as so-called open-ended learners, are among the bird species in which 1-year-old males can significantly expand and adjust their repertoire for the following breeding season, but not thereafter (Kipper et al., 2004). In this process, song types are learned in a very stereotypical way (Hultsch et al., 1999), which

---

[1] https://www.tierstimmenarchiv.de/

[2] https://xeno-canto.org/

FIGURE 1 | Decision tree for song types classification. Subdivision of whistles: narrowband (<3 kHz frequency bandwidth) low and high frequency and broadband (>3 kHz frequency bandwidth). Subdivision of buzz songs: short (<500 ms) and long (<500 ms) sections. Subdivision of trills: rapid broadband trills, slow broadband trills, and slow narrowband trills. Percentages are given in square brackets.

has the capacity to make the song repertoire very diverse with many individual song type variations (Landgraf et al., 2015). These different song types are not repeated directly, but rather after a variety of other song types, with some shared commonly and others rarely within and between individuals (Hultsch and Todt, 1989). Earlier studies have already shown that neighboring nightingales in intermediate distances share a higher proportion of songs than close or distant ones (Hultsch and Todt, 1981) due to song type adjustment to the breeding population (Kiefer et al., 2010, 2014). Based on knowledge of song ontogeny and migration behavior, temporally stable and geographically variable nightingale song can be expected and would be consistent with previous studies. Weiss (2012; unpublished data) noted in his doctoral thesis that dialects are most likely coded in the number of times a song type appears (relative frequency). Schehka (2004) found in her master thesis that common song types have been proven to be stable in their occurrence over two decades and in two countries, while she did not study rare song types. Thus, several lines of evidence suggest that stability and dialects may be occurring on a larger scale, although their extent has not been fully investigated yet. We addressed this question in our study by examining the stability (occurrence of common and rare song types), and dialects (relative frequency) on an expanded temporal (almost a century) and geographical (whole Europe) scale. Because certain parts of songs (thereafter defined as sub-categories of song) showed stability in some and coded for

dialects in other species with a small repertoire, we investigated their function in the large repertoire of the nightingale. OD and CS data approaches have been used successfully for dialect research in other species with few song types (Wright and Dahlin, 2018; Diblíková et al., 2019), but never before in a songbird with so many song types, such as the nightingale. Therefore, we examined whether the datasets showed different results for the sub-categories of song, the occurrence or relative frequency of common song types to assess their comparability and derive their respective usefulness.

To investigate the potential for new avenues for research, we specifically addressed two biological and one technical hypothesis: (1) Common song types occur in a stable manner over nearly a century in the nightingale song (H1: stability). Since stable common song types have previously been found over two decades in the nightingale, we predicted stability over several decades, as has been detected in other species. (2) The relative frequency of common song types is encoded for dialects in nightingales across Europe (H2: dialects). As the relative song type frequency has been shown to differ in the nightingale between two neighboring countries, we predicted that such variance occurs also across the entire breeding range of the nightingale. (3) The sub-categories of song, occurrence, and relative frequency do not differ between the datasets (H3: datasets). Since OD, CS, and shared data have been used successfully for dialect research in other species and CS

40

recordings were determined to be as valid as scientific recordings in an earlier comparison, we expected that all datasets used will be appropriate for nightingale dialect research.

MATERIALS AND METHODS

The Nightingale Song as a Model in Song Research

The nightingale is a recognized model species in song research due to its clearly structured song with durations of about 2–4 s alternating with silent intervals (pauses) of similar length (Hultsch, 1980; unpublished data). The basal unit is an element with a characteristic frequency (bandwidth) and duration, containing pauses less than 10 ms. Up to four elements are grouped together to form syllables which are separated by pauses of 10 ms and produced once or several times repeated in a stereotypic order, form phrases. These are grouped into four sections (α, β, γ, and Ω) which can be characterized by analyzing element features and repetition rate (Hultsch, 1980; unpublished data; **Figure 1**). Male nightingales can produce a variable number of distinct song types whereas different individual birds can sing the same song type with minor variations (between them, or from rendition to rendition), resulting in extraordinarily large song repertoires (on average 190 with a maximum of 250 different song types; e.g., Hultsch and Todt, 1982; Kipper et al., 2004). As it is therefore challenging to study the song and with regard to the work economy, we focused on 20 song types (approx. 10% of the song types of a nightingale repertoire, **Supplementary Figure 1**). For this species, it is a standard procedure (e.g., Kipper et al., 2015) and has also been used in a previous dialect study of the nightingale due to the complex song (Schehka, 2004; unpublished data). We examined the ten most common and the ten rarest song types (occurred only once) from Berlin-Brandenburg within all datasets and across years. As there were 276 rare song types we selected ten of them randomly. Using these 20 song types (ten common, ten rare ones with Berlin-Brandenburg as a reference point) we analyzed their geographical occurrence across Europe. The ten most common song types were used for the temporal and dataset occurrence analysis and the relative frequency estimation for time, space and datasets.

Databases

To unravel the potential of OD, CS, and data sharing, we used and compared four datasets (**Figure 3**): one open archive, two CS datasets, and one shared dataset. In total, we analyzed 14,140 nightingale songs. Since the length as well as the recording quality varies greatly, songs are a better measure to indicate the amount of data material than recording numbers. No recording originated from more than one dataset. Recording standards varied, depending on the species, research question, and context. For studies on individual song differences in the nightingale, 1-h nocturnal recordings are the standard (Kiefer et al., 2014), as they contain the most songs as well as no disturbing background noise in terms of other singing birds (bachelor thesis Jäckel, 2013; unpublished data). Only the shared recordings were standardized. The other datasets were unstandardized in terms
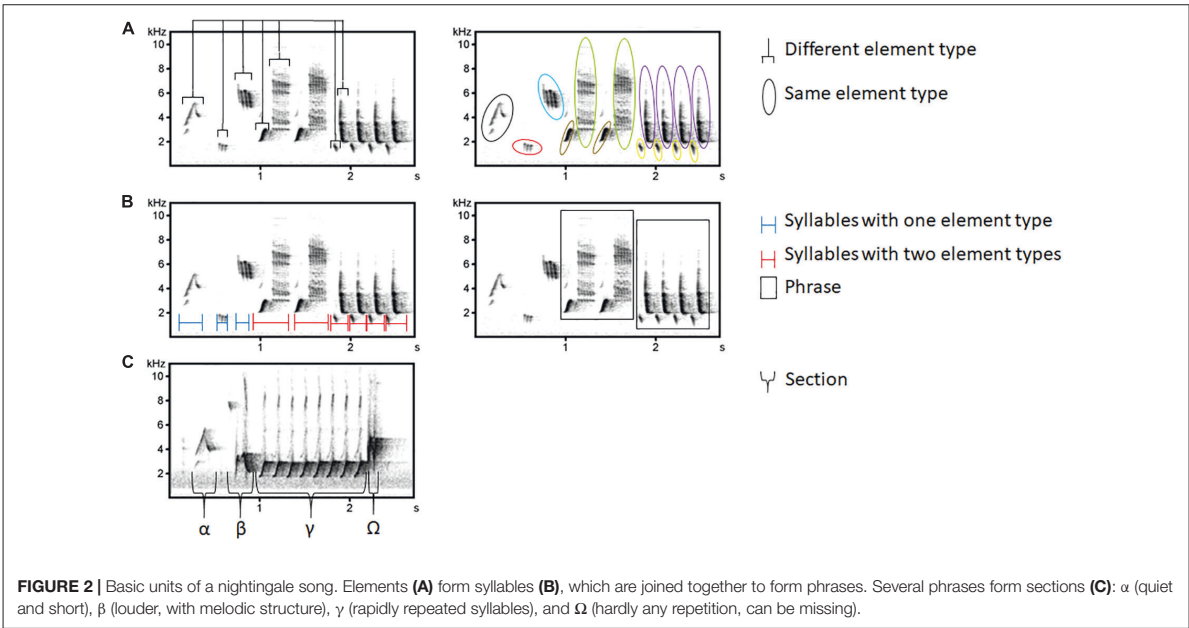
of technique (smartphones to professional equipment), time (day and night), location (Europe-wide), and durations (few seconds to 90 min). However, standardized recordings were not required for our study, as we were working on the population rather than the individual level. Individual differences are used to define signature and not geographical variations. We cannot completely rule out the possibility that certain individuals were recorded repeatedly and therefore their songs occur multiple times in our dataset, in particular this is the case for the citizen science data. However, based on the GPS data we assume that the vast majority of recordings were not from a few individuals, but that the recordings were well distributed in space and time. To emphasize that we are not making statements about individuals, we worked at the population level.

We used the open animal sound archive of the MfN (hereinafter abbreviated as ASA) which contained recordings from the 1930s until 2015 of which we examined 3,262 songs from Germany. These recordings were generated by different experts with different recording devices (sampling frequency: 96 kHz; resolution: 16 bit). The data were distributed differently over the decades, with a significant decrease from 2006 onward. We analyzed 3,833 songs from a nightingale citizen science project (project name: "Forschungsfall Nachtigall," hereinafter abbreviated as FFN; eng. "nightingale citizen science project"), led by a team of academic researchers at the MfN. These recordings were generated with diverse smartphones (sampling frequency: 44.1 kHz; resolution: 16 bit) within 2018 and 2019 mainly in Berlin. Additionally, we examined 4,751 songs from the CS website Xeno-Canto of 13 countries and spanning over 26 years. These recordings were generated by both laypeople and experts, using a wide range of technology (sampling frequency: 8–48 kHz; resolution: 16 bit). The shared dataset of 2,294 songs from Berlin was generated by academic researchers (hereinafter abbreviated as AR) in 2018 and 2019 by students of Freie Universität Berlin (Berlin, Germany) and staff of the FFN. Recordings were generated with Sennheiser ME66/K6 directional microphones (Georgsmarienhütte, Germany) connected to a Tascam Dr-40 4-Track Portable Digital Recorder or a Marantz solid-state recorder PMD660 as well as Zoom H2n recorders (sampling frequency: 44.1 kHz; resolution: 16 bit).

A European Nightingale Song Type Catalog

The high precision of song learning, copying, and delivering in male nightingales within and between individuals allows for a reliable assignment of a song to the same song type (Hultsch et al., 1999). We used Avisoft SASLab Pro 5.2 (R. Specht, Berlin, Germany) for visual inspection of songs (sampling rate = 22.05 kHz, FFT = 1024 points, Hamming-Window, overlap 93.75%). As Avisoft can only open WAV files, the MP3 and m4a recordings were converted into WAV files by using the program WaveLab 7. We started our cataloging process with recordings from Berlin-Brandenburg and extended across Germany and Europe. During the last 2 years, a single person (DJ) visually compared songs using spectrograms with catalog song types and categorized a song as either known or new. We did a self-check

FIGURE 2 | Basic units of a nightingale song. Elements **(A)** form syllables **(B)**, which are joined together to form phrases. Several phrases form sections **(C)**: α (quiet and short), β (louder, with melodic structure), γ (rapidly repeated syllables), and Ω (hardly any repetition, can be missing).

FIGURE 3 | Overview of the data material for open science [animal sound archive (ASA)], citizen science [Xeno-Canto and "Forschungsfall Nachtigall" (FFN)] and shared data [academic researcher (AR)]. Information was separated into specification of the recordings and raw data.

at the beginning to evaluate the accuracy of the cataloged song types. In this cataloging process, comparable to the nightingale song type catalog of Hultsch and Todt (1981) and Kipper et al. (2004), we based the definition of a song type on an identical γ-section. But our analysis differed from their previous catalog: (a) As we worked with different types of recordings (unknown individuals), we defined variations of a song type as all differences in the elements in the α, β, and in the Ω-section (**Figure 2**). Variations in the previous catalog occurred when more than three of ten elements (80% of the song type) differed in the α or β

section. Variations are thus minor modifications of a song type that do not occur in the repetitive γ-section. We therefore based the structure within our song type catalog on the γ-section since any differences in that section were used to define a new song type. (b) We added all newly found song types to the catalog. For the previous catalog, which was based on 1-h song recordings of individual known nightingales, it was necessary that song types had to be sung at least twice by the same male in the same recording to be included. (c) We created our catalog on a basis of recordings across European datasets without any knowledge of

the individual. Doing so, we created a European song type catalog on the population level.

Sub-Categories of Song Classification

We established previously that recordings made with qualitatively and technically different recording devices presented significantly different acoustic parameters in subsequent measurements. Therefore, we based the classification of sub-categories of song on visually distinguishable parameters, i.e., frequency ranges, durations, pauses, and element rates – and not on measuring frequencies. Sub-categories of song were extended divisions of the known four song categories of the nightingale: whistles, buzzes, trills, and "other songs" (**Figure 1**; Hultsch, 1980; unpublished data). "Other songs" were not further subdivided. Whistle, buzz and trill elements occur in the γ-section. Whistle elements never occur with buzz elements in the same song type simultaneously. It is possible that whistle or buzz elements are followed by trill elements in different sections, whereas these are no trill songs.

Whistles have been shown to have frequency-dependent functions (Hultsch, 1980; unpublished data) and are used for female attraction (Naguib et al., 2002). On this basis, we classified whistles according to the frequency bandwidth of the last whistle element into narrowband (<3 kHz) and broadband (>3 kHz) ones. We divided the narrowband ones corresponding to the frequency of the last whistle element into low (<4 kHz) and high (>4 kHz) whistles. Buzz songs have been found to indicate the quality of a male by the length of the buzz section (Weiss et al., 2012). We subdivided the buzz songs into ones with short (<500 ms) and long (>500 ms) sections. The relationship between frequency bandwidth and trill rate has been demonstrated to serve as an aggressive signal in nightingales (Kunc et al., 2006). In some bird species, this relationship has been found to indicate the quality of a male to females (Podos, 1996). Based on this finding, we divided the trills into rapid broadband (bandwidth of >9 kHz and an element repetition rate without visible pauses between the trills), slow broadband (bandwidth of >9 kHz and element repetition rate with visible pauses between the trills), and slow narrowband trills (bandwidth <6 kHz and element repetition rate with visible pauses between the trills).

Comparison of Time Periods

We specified four time periods (TP) for the temporal analysis. TP1 was defined by us as the 1930s, followed by TP2 (1960–1980), TP3 (1981–2000), and TP4 (2001–2019). As the recordings were not evenly distributed over time and some years were missing, we had to combine two decades as one TP (**Figure 4A**). See for an overview of the songs per TP and datasets in **Supplementary Table 1**.

Comparison of Geographical Regions

For the geographical analysis, we selected the region Berlin-Brandenburg as a reference for a location since the majority of our recordings (58%) were generated there and both, the FFN and the AR recordings, largely originated from this region (99–100%). We did not have GPS tags of a large amount of the recordings for the countries Italy, France, Spain, and Portugal, which did not allow for a statistical method to classify the regions. We merely had the information about the respective countries but did refrain from a comparison based on national borders. Thus, due to the missing metadata, we based the geographic distribution on the approximate estimated distance from Berlin-Brandenburg and defined five geographical regions (**Figure 5A**). An overview of the songs per region separated according to the datasets is available in **Supplementary Table 2**.

Comparison of Different Databases

In order to compare the different datasets, we chose the same time period (TP4: 2001–2019) and region (Berlin-Brandenburg) to exclude temporal and regional differences in the analysis. We chose this period because the last recordings from the ASA were from 2015. In addition, the AR and FFN recordings were generated in 2018 and 2019. We thus had sufficient recordings to study and compare data from all datasets (AR: 2,294 songs; ASA: 1,669 songs; FFN: 3,366 songs; Xeno-Canto: 99 songs; **Figure 6A**).
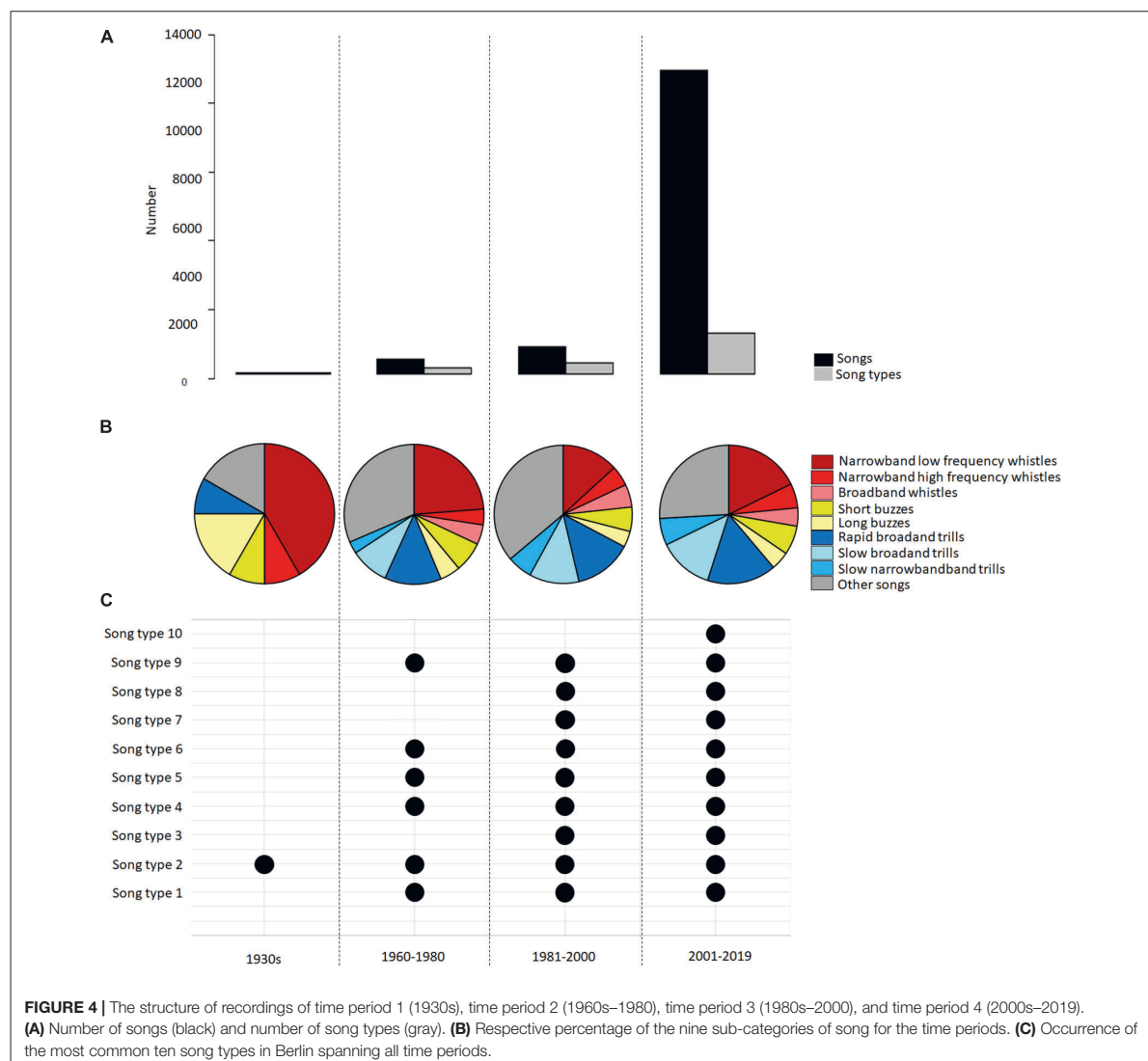
Statistical Analyses

For the occurrence analysis of a given song type, we recorded whether it was present or absent in the recordings that we analyzed over time, space or dataset. We defined the relative frequency from the number of times a song type appears divided by the total number of songs. Two one-sided tests (TOST) were used to test for the effects of time, space, and datasets for sub-categories of song and the relative frequency of song types. We set the upper and lower equivalence boundary at 0.5 and the statistical significance at $p \leq 0.05$. All statistical analyses were performed using R version 4.0.0 (R-Team, 2020; unpublished data).

RESULTS

European Song Type Catalog

Based on 14,140 songs from four datasets, we created a European nightingale song type catalog at the population level, which currently contains 1,868 different song types. As of September 2021, the catalog comprises 20% whistles, 12% buzz, 27% trills, and 41% so-called "other songs." The nine sub-categories of song were significantly equivalent between the datasets in their percentages (TOST test: equivalent p-values of <0.01). The occurrences of song types differed among the four datasets. 78% of songs types appeared in only one of the datasets, 9% of songs types were found in two of the datasets, 5% of songs types occurred in three datasets and 8% of songs types were in all four datasets. Song types occurred with different frequencies in all datasets (1×: 56%; ≤10×: 30%; >100×: 1%). The rate of new song types that have not yet been sung (song type diversity) was different between the datasets (AR: each sixth song; ASA: each fourth song; FFN: each thirteenth song; Xeno-Canto: each fourth song). The agreement between the song types varied (highest between AR and FFN: 61; lowest between Xeno-Canto and ASA: 26; **Supplementary Table 3**).

FIGURE 4 | The structure of recordings of time period 1 (1930s), time period 2 (1960s–1980), time period 3 (1980s–2000), and time period 4 (2000s–2019).
(A) Number of songs (black) and number of song types (gray). **(B)** Respective percentage of the nine sub-categories of song for the time periods. **(C)** Occurrence of the most common ten song types in Berlin spanning all time periods.

Temporal Comparison of Nightingale Song

When comparing the sample years between 1930 and 2019 among all datasets, we detected that the nine sub-categories of song were distributed in a stable way with similar percentages over time (TOST test: equivalent p-values of <0.01). In most time periods, the narrowband low-frequency whistles, the rapid broadband trills, and the "other songs" accounted for the largest percentage of the nine sub-categories of song (**Figure 4B** and **Supplementary Table 4**). An analysis with regard to song type stability showed that the second most common song type in our dataset – a narrowband low-frequency whistle – occurred in all time periods. Six common song types were found in the recordings from the 60s until 2019 (**Figure 4C**). An analysis with

regard to the relative frequency of the ten most common song types revealed that they accounted for 2% of the total number of songs and were significantly equivalent between the time periods (TOST test: equivalent p-values of <0.01).

Geographical Comparison of Nightingale Song

A comparison between the five geographic regions in all datasets showed a stable and similar distribution in the percentage of the nine sub-categories of song over space (TOST test: equivalent p-values of <0.01). The highest percentage of the nine sub-categories of song was accounted for by the low-frequency narrowband whistles, the rapid broadband trills and the "other songs" in all regions (**Figure 5B** and

FIGURE 5 | The distribution of the recordings across the five regions we defined. The regions were based on the approximate estimated distance from the Berlin-Brandenburg region. **(A)** Numbers of songs and song types (in brackets). **(B)** Respective percentage of the nine sub-categories of song for the other regions. **(C)** Occurrence of the most common ten song types from the Berlin-Brandenburg region in the other regions. The different shades of red each represent a song type. The most common song type is described by the top red left square. The tenth most common song type is the bottom red right square. A white square means that none of the song types in question were found. **(D)** Occurrence of the ten rare song types found only once in the Berlin-Brandenburg region in the other regions. The different shades of green and blue each represent a song type. A white square means that none of the song types in question were found.

Supplementary Table 5). An analysis with regard to the stability of the ten most common song types revealed that nine of them were found in four of our geographic regions (**Figure 5C**). Six of the ten rare song types occurred in four regions (**Figure 5D**). An analysis with regard to the relative frequency of the ten most common song types showed that they accounted for 2% of the total number of songs and were significantly equivalent between the regions (TOST test: equivalent p-values of <0.01).

Comparison of Nightingale Song Among the Four Different Datasets

We found an equal distribution of the nine sub-categories of song when comparing the four datasets in the time between 2001 and 2019 from Berlin-Brandenburg (TOST test: equivalent p-values of <0.01). The low-frequency whistles had the highest percentage, followed by the rapid broadband trills and "other songs" (**Figure 6B** and **Supplementary Table 6**). Concerning the ten most common song types, four of them occurred in all datasets (**Figure 6C**), six of them occurred in the AR, ASA, and FFN recordings. An analysis with regard to the relative frequency of the ten most common song types revealed that they accounted for 3% of the total number of songs and were significantly equivalent between the datasets (TOST test: equivalent p-values of <0.01).

DISCUSSION

The aim of this study was to investigate whether nightingale song is stable in terms of song type occurrence over nearly a century and whether song dialects are found in the relative frequency of song types across Europe. We also examined if sub-categories of song codes for stability or dialects. Additionally, we investigated if open, citizen science and shared data are an advantageous complement for dialect studies in a bird species with a large repertoire. Our study encompassed a period twice as large as previous studies (e.g., Grant and Grant, 1996) with recordings from the 1930s until 2019 and covered a linear geographic range of 2,652 km across 13 European countries, twice as large as previous studies (e.g., Weiss, 2012; unpublished data). In the following, we will evaluate the results from the biological (temporal stability and spatial dialects) and the technical (comparison of different datasets) perspectives.

Vocal Dialect Research Across Time and Space

Our European population-based nightingale song type catalog was three times larger than a previous one (623 song types; Weiss et al., 2014), which included songs of 271 individual nightingale males from a Berlin and 272 from a Brandenburg population. In addition, the catalogs differed in two more respects. First, we included all song types, instead of only

45

FIGURE 6 | The structure of the recordings within the datasets (AR, academic researchers; ASA, animal sound archive; FFN, "Forschungsfall Nachtigall; Xeno-Canto) from Berlin spanning the period between 2001 and 2019. **(A)** Number of songs (black) and number of song types (gray). **(B)** Respective percentage of the nine sub-categories of song for the datasets. **(C)** Occurrence of the ten most common song types from Berlin in all time periods.

those that appeared at least twice because otherwise that would have resulted in the loss of 1,048 song types. Second, in the definition of new song types: any differences in γ-section (our study) vs. less than 80% similarity or differences in more than three of ten elements in the α- and β-section (Weiss et al., 2014). However, we would like to emphasize that even if the total number of song types varies between the different catalogs, the results should remain unaffected. This notion is supported by the number of song types for the Berlin-Brandenburg region in two populations (623, Weiss et al., 2014) compared to multiple populations (893, our method).

Despite our large song type catalog, we found, for the first time, stability both in time (nearly a century) and space (geographical area — Europe) within the percentage of sub-categories of song and the occurrence of song types (common and rare) in the song of a bird species with a

large repertoire. The stability of two frequent sub-categories of song could be explained by the fact that they may indicate the quality of a male, as low-frequency whistles and rapid broadband trills are difficult to produce as has been shown in other species (Podos, 1996; Kirschel et al., 2009), and degrade less in the nightingale habitat (Wiley and Richards, 1978). Thus, these sub-categories of song indicate stability over time and space in bird species with large (the nightingale) and small repertoires (Nelson et al., 2004) and cetaceans (Deecke et al., 2000). Temporal stability in the occurrence of song types have also been found in songbird species with a small repertoire (Grant and Grant, 1996), hummingbirds (González and Ornelas, 2009), parrots (Wright et al., 2008) and in cetaceans (Rekdahl et al., 2013), where the most common vocalizations were likewise stable. Although geographic persistence of certain song types has been noted in songbirds (Searfoss et al., 2020) and parrots (Wright et al., 2008), suggesting an important function in animal

communication, we expected temporal stability and spatial drift: (a) Nightingales have been shown to adapt their song to the population from the first to the second breeding season (Kiefer et al., 2010, 2014). (b) Nightingales have been demonstrated to share song types among males of the same population to facilitate communication during interactions between males (Todt and Naguib, 2000). (c) Different song types are performed with different frequencies, while those shared with neighboring males are common (Hultsch, 1980; unpublished data). The stable occurrence of common and rare song types detected in our large dataset is consistent with previous playback experiments (Weiss, 2012; unpublished data). Moreover, this stability seems to be consistent with the nightingale behavior in the wintering grounds (Glutz von Blotzheim, 1988). Spatially separated nightingales probably interact on wintering grounds, which could lead to adaptation and stability of song types. These stable song types could have positive effects (adaptation to the environment, choice of female) and an important communicative function in social interactions, such as in cetaceans (Rekdahl et al., 2013). All this could have led to the fact that the song characteristics we studied were stable across time and space. Future studies should investigate exactly which and why these stable song types exist in nightingale song, and how they mediate communication.

In contrast to our findings on the stability of nightingale song, we observed no evidence of song dialects in the relative frequency of the ten most common song types across Europe. As Weiss (2012; unpublished data) found differences in the relative frequency of song types between populations from Berlin, Brandenburg, and France, we expected the same for Europe in general. However, the relative frequency of common song types was so low that no geographical differences could be inferred. This low relative frequency could be related to the size of our song type catalog but not to the datasets used, as they showed the same results. However, as the exact relative frequency of song types of a male can only be determined after approximately 1 h (533 songs; Kipper et al., 2004), we suggest that it is not the relative frequency of a single song type that codes for dialects, but the composition of song type delivery of a bird, i.e., repertoire or syntax (Weiss, 2012; unpublished data). This has already been suggested for orcas (Riesch et al., 2006; unpublished data) and is in line with previous bird studies showing that even species with a large repertoire recognize foreign males from short songs without the need to listen to the complete repertoire (e.g., Jaška et al., 2015; Moser-Purdy and Mennill, 2016). Since nightingales sing with immediate variety, i.e., the same song type is not sung repeatedly, but only after a certain number of songs (Hultsch and Todt, 1989), this pattern allows them to hear several song types in a short time – and thus may enable nightingales to determine whether a singer is a resident or a stranger. This result is indeed consistent with studies suggesting that song dialects are unlikely to emerge in species with a large vocal repertoire (Goodfellow and Slater, 1986; Lachlan et al., 2016; Tchernichovski et al., 2017), such as the nightingale. However, we need to stress that we in fact only looked at a rather small part of the nightingale song type catalog determined in our case study. Thus, future studies could extend the analysis to examine if potentially other common or rare song types code for dialects.

Combining Datasets for Research Over Time and Space

We evaluated OD, CS, and shared data for dialect research over time and space in direct comparison of sub-categories of song, as well as the occurrence and relative frequency of the ten most common song types in nightingale song. The similarity in the three most frequent sub-categories of song and variations in the other six sub-categories of song probably resulted from some song types occurring in only one dataset, leading to different song type diversities and thus percentage distribution. Because all ten common song types were found in the datasets, except Xeno-Canto, we suspect that the absence of the common song types was due to the number of analyzed songs. Since these 99 Xeno-Canto songs were only sufficient for finding four common song types, about three times the amount of songs is needed for a valid analysis. We showed that the datasets are not different *per se*, but our evaluation was influenced by other aspects.

The differences between the datasets were evident in their temporal and spatial validity for our analysis, as all exhibited either a temporal or spatial bias. Our temporal analysis was mainly conducted with the OD and the CS (especially Xeno-Canto) recordings. The spatial analysis was mainly possible with the CS recordings as these had the largest spatial distribution and the most metadata, since these were generated with an app (FFN) and Xeno-Canto requires a minimum of metadata for the recordings (Vellinga and Planqué, 2015). CS projects can provide more data with a greater geographic coverage than academic researchers alone, which are limited in number of simultaneously generated recordings due to the (mostly) stationary location and the restricted availability of (expensive) equipment (e.g., Vellinga and Planqué, 2015; Steen et al., 2021). The shared data were the least suited for our temporal and geographic analysis as they were only from 2 years from Berlin. The importance of the geographic distribution of the recordings was evident in two ways: First, song type diversity was mainly affected by the spatial distribution of the recordings and not by the number of songs and thus the size of the dataset nor the duration or the time of the recording as no song types only occur at night or day (Jäckel, 2013; unpublished data). Second, the degree of concordance of song types in each dataset was also influenced by spatial distribution, with datasets from the same region (e.g., AR and FFN) accounting for the greatest share. We conclude that to study song stability and dialects (1) the recording location is more important than recording length and that (2) many short, non-standardized OD and CS recordings may achieve the same diversity of song types within a dataset as long, standardized professional recordings. Thus, for our study using the nightingale as an example, standardized recordings were not required in terms of technology (professional equipment), time (nocturnal song), duration (1 h), or great experience of the recordist.

Our study showed that even if longer recordings would have been helpful, considering the large repertoire of the nightingale, it was more critical for our temporal and spatial analysis to have access to diverse datasets. To make access also possible for future nightingale researchers, we are in the process of making the recordings from the FFN online available in an open archive

47

at the MfN. Currently only half of the academic researchers publish their data in open archives (Tenopir et al., 2011), as such access is not encouraged widely, which reduces the opportunities for reuse (Mills et al., 2015; Federer, 2020). Missing open science practices are particularly disadvantageous for evolution and ecology research, where datasets from longitudinal studies contribute greatly to scientific understanding but are expensive and time-consuming (Mills et al., 2015; Hughes et al., 2017). We would therefore like to encourage more song recordings to be made publicly available with sufficient metadata so that they can be effectively used in future research. Open science, and thus the availability of diverse datasets, allowed us to conduct the largest temporal and geographic study of song evolution to date. By directly comparing the datasets, we have shown that OD and CS recordings are well suited – and in some cases better suited – for studying song variation in bird species with a large repertoire than data collected by an academic research group, because they were collected over a shorter period of time and in a much smaller geographic range. Therefore, we recommend that OD and CS datasets should be used more frequently in ornithology and that song recordings should be openly accessible. Given that each dataset has different advantages and disadvantages, we further recommend combining diverse datasets for dialect research in bird species with a large repertoire.

CONCLUSION

In our study, we found the song of common nightingales to be stable over time and space. We did not detect evidence of dialects. However, the nightingale has a highly sophisticated syntax and a complex song, only a small part of which we studied, so further analysis is needed to rule out the absence or presence of dialects. Our results showed that non-standardized OD and CS recordings offer the potential to support ornithology and in particular dialect research. The general quality concerns about OD and CS could not be confirmed in our study. Based on our findings, when studying a bird species with a large repertoire such as the nightingale, we recommend combining different and diverse datasets as quality is multidimensional and always depends on the research question. As more and more (CS) recordings are currently being made available on open platforms (Sanderson et al., 2021; unpublished data), our study is of great use to increase trust in them. We hope to encourage fellow bird enthusiasts to share their song recordings in open repositories and promote the use of these data to the scientific community.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval were not required for the animal study because, in Germany, permission from an ethics committee is not required for such research questions.

AUTHOR CONTRIBUTIONS

DJ, KM, and SV-H wrote the concept and design of the study and analyzed and interpreted the data. DJ and UB performed the statistical analysis and designed the figures. DJ wrote the first draft of the manuscript. KM, GL, and SV-H revised the manuscript. All authors contributed to the manuscript and revised, read, and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.778610/full#supplementary-material

REFERENCES

Aplin, L. M., Major, R. E., Davis, A., and Martin, J. M. (2021). A citizen science approach reveals long-term social network structure in an urban parrot, *Cacatua galerita*. *J. Anim. Ecol.* 90, 222–232. doi: 10.1111/1365-2656.13295

Baker, M. C., and Mewaldt, L. R. (1978). Song dialects as barriers to dispersal in white-crowned sparrows, *Zonotrichia leucophrys nuttalli*. *Evolution* 32, 712–722. doi: 10.1111/j.1558-5646.1978.tb04624.x

Beecher, M. D., and Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends Ecol. Evol.* 20, 143–149. doi: 10.1016/j.tree.2005.01.004

Bonn, A., Richter, A., Vohland, K., Pettibone, L., Brandt, M., Feldmann, R., et al. (2017). *Grünbuch Citizen Science Strategie 2020 für Deutschland*. Cenate Sotto: Bürger schaffen Wissen (GEWISS).

Catchpole, C. K., and Slater, P. J. B. (2008). *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge University Press.

de Camargo Barbosa, K. V., Develey, P. F., Ribeiro, M. C., and Jahn, A. E. (2021). The contribution of citizen science to research on migratory and urban birds in Brazil. *Ornithol. Res.* 29, 1–11. doi: 10.1007/s43388-020-00031-0

de Jesus, M. D., Zapelini, C., and Schiavetti, A. (2021). Can citizen science help delimit the geographical distribution of a species? The case of the Callistoctopus sp. ("eastern octopus") on the Brazilian coast. *Ethnobiol. Conserv.* 10, 1–15. doi: 10.15451/ec2020-09-10.03-1-15

Deecke, V. B., Ford, J. K. B., and Spong, P. (2000). Dialect change in resident killer whales: implications for vocal learning and cultural transmission. *Anim. Behav.* 40, 629–638. doi: 10.1006/anbe.2000.1454

Derryberry, E. P., Claramunt, S., Derryberry, G., Chesser, R. T., Cracraft, J., Aleixo, A., et al. (2011). Lineage diversification and morphological evolution in a large-scale continental radiation: the Neotropical ovenbirds and woodcreepers (Aves: Furnariidae). *Evolution* 65, 2973–2986. doi: 10.1111/j.1558-5646.2011.01374.x

Diblíková, L., Pipek, P., Petrusek, A., Svoboda, J., Bílková, J., Vermouzek, Z., et al. (2019). Detailed large-scale mapping of geographical variation of yellowhammer *Emberiza citrinella* song dialects in a citizen science project. *Ibis* 161, 401–414. doi: 10.1111/ibi.12621

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., et al. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Front. Ecol. Environ.* 10, 291–297. doi: 10.1890/110236

Dominguez, M., Lapido, R., Gorrindo, A., Archuby, D., Correa, E., Llanos, F., et al. (2021). A citizen science survey discloses the current distribution of the endangered Yellow Cardinal *Gubernatrix cristata* in Argentina. *Bird Conserv. Int.* 31, 139–150. doi: 10.1017/S0959270920000155

Federer, L. (2020). Measuring and mapping data reuse: findings from an interactive workshop on data citation and metrics for data reuse. *Harv. Data Sci. Rev.* 2, 1–8. doi: 10.1162/99608f92.ccd17b00

Gaunt, S. L., Baptista, L. F., Sịnchez, J. E., and Hernandez, D. (1994). Song learning as evidenced from song sharing in two hummingbird species (*Colibri coruscans* and *C. thalassinus*). *Auk* 111, 87–103. doi: 10.2307/4088508

Gil, D., and Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. *Trends Ecol. Evol.* 17, 133–141. doi: 10.1016/S0169-5347(02)02410-2

Glutz von Blotzheim, U. N. (1988). *Handbuch der Vögel Mitteleuropas*. Wiesbaden: Aula.

González, C., and Ornelas, J. F. (2009). Song variation and persistence of song neighborhoods in a lekking hummingbird. *Condor* 111, 633–640. doi: 10.1525/cond.2009.090029

Goodfellow, D., and Slater, P. (1986). A model of bird song dialects. *Anim. Behav.* 34, 1579–1580. doi: 10.1016/S0003-3472(86)80233-0

Grant, B. R., and Grant, P. R. (1996). High survival of Darwin's finch hybrids: effects of beak morphology and diets. *Ecology* 77, 500–509. doi: 10.2307/2265625

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., and Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends Ecol. Evol.* 27, 130–137. doi: 10.1016/j.tree.2011.11.006

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., et al. (2017). Long-term studies contribute disproportionately to ecology and policy. *Bioscience* 67, 271–281. doi: 10.1093/biosci/biw185

Hultsch, H., Mundry, R., and Todt, D. (1999). "Learning, representation and retrieval of rule-related knowledge in the song system of birds," in *Learning*, eds A. Friederici and R. Menzel (Berlin: Walter de Gruyter), 89–115. doi: 10.1515/9783110803488.89

Hultsch, H., and Todt, D. (1981). Repertoire sharing and song-post distance in nightingales (*Luscinia Megarhynchos B.*). *Behav. Ecol. Sociobiol.* 8, 183–188. doi: 10.1007/BF00299828

Hultsch, H., and Todt, D. (1982). Temporal performance roles during vocal interactions in nightingales (*Luscinia megarhynchos B.*). *Behav. Ecol. Sociobiol.* 11, 253–260. doi: 10.1007/BF00299302

Hultsch, H., and Todt, D. (1989). Memorization and reproduction of songs in nightingales (*Luscinia megarhynchos*): evidence for package formation. *J. Comp. Physiol. A* 165, 197–203. doi: 10.1007/BF00619194

Hultsch, H., and Todt, D. (1996). Discontinuous and incremental processes in the song learning of birds: evidence for a primer effect. *J. Comp. Physiol. A.* 179, 291–299. doi: 10.1007/BF00194983

Irwin, D. E., Bensch, S., and Price, T. D. (2001). Speciation in a ring. *Nature* 409, 333–337. doi: 10.1038/35053059

Jaška, P., Linhart, P., and Fuchs, R. (2015). Neighbour recognition in two sister songbird species with a simple and complex repertoire – a playback study. *J. Avian Biol.* 45, 151–158. doi: 10.1111/jav.00482

Kassen, M. (2013). A promising phenomenon of open data: a case study of the Chicago open data project. *Govern. Inf. Q.* 30, 508–513. doi: 10.1016/j.giq.2013.05.012

Kiefer, S., Scharff, C., Hultsch, H., and Kipper, S. (2014). Learn it now, sing it later? field and laboratory studies on song repertoire acquisition and song use in nightingales. *Naturwissenschaften* 101, 955–963. doi: 10.1007/s00114-014-1236-5

Kiefer, S., Sommer, C., Scharff, C., and Kipper, S. (2010). Singing the popular songs? nightingales share more song types with their breeding population in their second season than in their first. *Ethology* 116, 619–626. doi: 10.1111/j.1439-0310.2010.01774.x

Kiefer, S., Spiess, A., Kipper, S., Mundry, R., Sommer, C., Hultsch, H., et al. (2006). First-Year common nightingales (*Luscinia Megarhynchos*) have smaller song-type repertoire sizes than older males. *Ethology* 112, 1217–1224. doi: 10.1111/j.1439-0310.2006.01283.x

Kipper, S., Kiefer, S., Bartsch, C., and Weiss, M. (2015). Female calling? Song responses to conspecific call playbacks in nightingales, *Luscinia megarhynchos*. *Anim. Behav.* 100, 60–66. doi: 10.1016/j.anbehav.2014.11.011

Kipper, S., Mundry, R., Hultsch, H., and Todt, D. (2004). Long-term persistence of song performance rules in nightingales (*Luscinia megarhynchos*): a longitudinal field study on repertoire size and composition. *Behaviour* 141, 371–390. doi: 10.1163/156853904322981914

Kirchhoff, C., Callaghan, C. T., Keith, D. A., Indiarto, D., Taseski, G., Ooi, M. K., et al. (2021). Rapidly mapping fire effects on biodiversity at a large-scale using citizen science. *Sci. Total Environ.* 755:142348.

Kirschel, A. N., Blumstein, D. T., Cohen, R. E., Buermann, W., Smith, T. B., and Slabbekoorn, H. (2009). Birdsong tuned to the environment: green hylia song varies with elevation, tree cover, and noise. *Behav. Ecol.* 20, 1089–1095. doi: 10.1093/beheco/arp101

Kroodsma, D., Hamilton, D., Sánchez, J. E., Byers, B. E., Fandiño-Mariño, H., Stemple, D. W., et al. (2013). Behavioral evidence for song learning in the suboscine bellbirds (*Procnias spp.*; *Cotingidae*). *Wilson J. Ornithol.* 125, 1–14. doi: 10.1676/12-033.1

Kunc, H. P., Amrhein, V., and Naguib, M. (2006). Vocal interactions in nightingales, *Luscinia megarhynchos*: more aggressive males have higher pairing success. *Anim. Behav.* 72, 25–30. doi: 10.1016/j.anbehav.2005.08.014

Lachlan, R. F., van Heijningen, C. A. A., ter Haar, S. M., and ten Cate, C. (2016). Zebra finch song phonology and syntactical structure across populations and continents-a computational comparison. *Front. Psychol.* 7:980. doi: 10.3389/fpsyg.2016.00980

Landgraf, C., Weiss, M., and Kipper, S. (2015). Multiple song features are related to paternal effort in common nightingales. *BMC Evol. Biol.* 15:115. doi: 10.1186/s12862-015-0390-5

Lukyanenko, R., Parsons, J., and Wiersma, Y. F. (2016). Emerging problems of data quality in citizen science. *Conserv. Biol.* 30, 447–449. doi: 10.1111/cobi.12706

Lynch, A. (1996). "The population memetics of birdsong," in *Ecology and Evolution of Acoustic Communication in Birds*, eds D. E. Kroodsma and E. H. Miller (New York: Cornell University Press), 181–197. doi: 10.7591/9781501736957-017

49

Mennill, D. J., Doucet, S. M., Newman, A. E. M., Williams, H., Moran, I. G., Thomas, I. P., et al. (2018). Wild birds learn songs from experimental vocal tutors. *Curr. Biol.* 28, 3273–3278. doi: 10.1016/j.cub.2018.08.011

Mills, J. A., Teplitsky, C., Arroyo, B., Charmantier, A., Becker, P. H., Birkhead, T. R., et al. (2015). Archiving primary data: solutions for long-term studies. *Trends Ecol. Evol.* 30, 581–589. doi: 10.1016/j.tree.2015.07.006

Moser-Purdy, C., and Mennill, D. J. (2016). Large vocal repertoires do not constrain the dear enemy effect: a playback experiment and comparative study of songbirds. *Anim. Behav.* 118, 55–64. doi: 10.1016/j.anbehav.2016.05.011

Naguib, M., Mundry, R., Hultsch, H., and Todt, D. (2002). Responses to playback of whistle songs and normal songs in male nightingales: effects of song category, whistle pitch, and distance. *Behav. Ecol. Sociobiol.* 52, 216–223.

Neate-Clegg, M. H., Horns, J. J., Adler, F. R., Aytekin, M. ÇK., and Şekercioğlu, ÇH. (2020). Monitoring the world's bird populations with community science data. *Biol. Conserv.* 248:108653. doi: 10.1016/j.biocon.2020.108653

Nelson, D. A., Hallberg, K. I., and Soha, J. A. (2004). Cultural evolution of puget sound white-crowned sparrow song dialects. *Ethology* 110, 879–908. doi: 10. 1111/j.1439-0310.2004.01025.x

Podos, J. (1996). Motor constraints on vocal development in a songbird. *Anim. Behav.* 51, 1061–1070.

Poulsen, H. (1958). The Calls of the Chaffinch (*Fringilla coelebs L.*) in Denmark. *Dan. Ornithol. Foren. Tidsskr.* 52, 89–105.

Pozzi, L., Gamba, M., and Giacoma, C. (2010). The use of Artificial Neural Networks to classify primate vocalizations: a pilot study on black lemurs. *Am. J. Primatol.* 72, 337–348. doi: 10.1002/ajp.20786

Prat, Y., Azoulay, L., Dor, R., and Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biol.* 15:e2002556. doi: 10.1371/journal.pbio.200 2556

Rekdahl, M. L., Dunlop, R. A., Noad, M. J., and Goldizen, A. W. (2013). Temporal stability and change in the social call repertoire of migrating humpback whales. *J. Acoust. Soc. Am.* 133, 1785–1795. doi: 10.1121/1.4789941

Rendell, L., and Whitehead, H. (2005). Spatial and temporal variation in sperm whale coda vocalizations: stable usage and local dialects. *Anim. Behav.* 70, 191–198. doi: 10.1016/j.anbehav.2005.03.001

Roth, T., and Amrhein, V. (2010). Estimating individual survival using territory occupancy data on unmarked animals. *J. Appl. Ecol.* 47, 386–392. doi: 10.1111/ j.1365-2664.2010.01786.x

Searfoss, A. M., Liu, W.-C., and Creanza, N. (2020). Geographically well-distributed citizen science data reveals range-wide variation in the chipping sparrow's simple song. *Anim. Behav.* 161, 63–76. doi: 10.1016/j.anbehav.2019. 12.012

Sharpe, D. L., Castellote, M., Wade, P. R., and Cornick, L. A. (2019). Call types of Bigg's killer whales (*Orcinus orca*) in western Alaska: using vocal dialects to assess population structure. *Bioacoustics* 28, 74–99. doi: 10.1080/09524622. 2017.1396562

Steen, V. A., Tingley, M. W., Paton, P. W., and Elphick, C. S. (2021). Spatial thinning and class balancing: key choices lead to variation in the performance of species distribution models with citizen science data. *Methods Ecol. Evol.* 12, 216–226. doi: 10.1111/2041-210X.13525

Tchernichovski, O., Feher, O., Fimiarz, D., and Conley, D. (2017). How social learning adds up to a culture: from birdsong to human public opinion. *J. Exp. Biol.* 220, 124–132. doi: 10.1242/jeb.142786

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS One* 6:e21101. doi: 10.1371/journal.pone.0021101

Todt, D., and Naguib, M. (2000). Vocal interactions in birds: the use of song as a model in communication. *Adv. Stud. Behav.* 29, 247–296. doi: 10.1016/S0065-3454(08)60107-2

Vellinga, W. P., and Planqué, R. (2015). "The Xeno-canto collection and its relation to sound recognition and classification," in *Proceedings of the 2015 CLEF*, (Toulouse: CLEF).

Weiss, M., Hultsch, H., Adam, I., Scharff, C., and Kipper, S. (2014). The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proc. R. Soc. B Biol. Sci.* 281:20140460. doi: 10.1098/rspb.2014. 0460

Weiss, M., Kiefer, S., and Kipper, S. (2012). Buzzwords in females' ears? the use of buzz songs in the communication of nightingales (*Luscinia megarhynchos*). *PLoS One* 7:e45057. doi: 10.1371/journal.pone.0045057

Weisshaupt, N., Lehikoinen, A., Mäkinen, T., and Koistinen, J. (2021). Challenges and benefits of using unstructured citizen science data to estimate seasonal timing of bird migration across large scales. *PLoS One* 16:e0246572. doi: 10. 1371/journal.pone.0246572

Wiley, R. H., and Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: implications for the evolution of animal vocalizations. *Behav. Ecol. Sociobiol.* 3, 69–94.

Wright, T. F., and Dahlin, C. R. (2018). Vocal dialects in parrots: patterns and processes of cultural evolution. *Emu* 118, 50–66. doi: 10.1080/01584197.2017. 1379356

Wright, T. F., Dahlin, C. R., and Salinas-Melgoza, A. (2008). Stability and change in vocal dialects of the yellow-naped amazon. *Anim. Behav.* 76, 1017–1027. doi: 10.1016/j.anbehav.2008.03.025

50

Chapter IV

Community engagement and data quality: Best practices and lessons learned from a citizen science project on birdsong

ORIGINAL ARTICLE

Community engagement and data quality: best practices and lessons learned from a citizen science project on birdsong

Denise Jäckel[1,2] · Kim G. Mortega[1] · Sarah Darwin[1] · Ulrich Brockmeyer[1] · Ulrike Sturm[1] · Mario Lasseck[1] · Nicola Moczek[3] · Gerlind U. C. Lehmann[4] · Silke L. Voigt-Heucke[1,5]

Abstract

Citizen Science (CS) is a research approach that has become popular in recent years and offers innovative potential for dialect research in ornithology. As the scepticism about CS data is still widespread, we analysed the development of a 3-year CS project based on the song of the Common Nightingale (*Luscinia megarhynchos*) to share best practices and lessons learned. We focused on the data scope, individual engagement, spatial distribution and species misidentifications from recordings generated before (2018, 2019) and during the COVID-19 outbreak (2020) with a smartphone using the 'Naturblick' app. The number of nightingale song recordings and individual engagement increased steadily and peaked in the season during the pandemic. 13,991 nightingale song recordings were generated by anonymous (64%) and non-anonymous participants (36%). As the project developed, the spatial distribution of recordings expanded (from Berlin based to nationwide). The rates of species misidentifications were low, decreased in the course of the project (10–1%) and were mainly affected by vocal similarities with other bird species. This study further showed that community engagement and data quality were not directly affected by dissemination activities, but that the former was influenced by external factors and the latter benefited from the app. We conclude that CS projects using smartphone apps with an integrated pattern recognition algorithm are well suited to support bioacoustic research in ornithology. Based on our findings, we recommend setting up CS projects over the long term to build an engaged community which generates high data quality for robust scientific conclusions.

Keywords Community science · Public engagement · COVID-19 · Pandemic · Birdsong · Dialect

Zusammenfassung

Gesellschaftliches Engagement und Datenqualität: Bewährte Praktiken und Erfahrungen aus einem bürgerwissenschaftlichen Projekt zum Vogelgesang
Citizen Science (CS) ist eine Forschungsmethode, die in den letzten Jahren an Bedeutung gewonnen hat und innovatives Potenzial für die Dialektforschung in der Ornithologie bietet. Da die Vorbehalte gegenüber CS-Daten immer noch

✉ Denise Jäckel
denise.jaeckel@mfn.de

1 Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

2 Life Sciences Faculty, Humboldt-Universität zu Berlin, Berlin, Germany

3 PLAN Institute for Architectural and Environmental Psychology, Berlin, Germany

4 Evolutionary Ecology, Department of Biology, Humboldt-Universität zu Berlin, Berlin, Germany

5 Animal Behaviour, Institute of Biology, Freie Universität Berlin, Berlin, Germany

⚫ Springer

weit verbreitet sind, haben wir die Entwicklung eines dreijährigen CS-Projekts zum Gesang der Nachtigall (*Luscinia megarhynchos*) analysiert, um bewährte Praktiken und gewonnene Erfahrungen darzustellen. Wir fokussierten uns auf den Datenumfang, das individuelle Engagement von Teilnehmenden, die räumliche Verteilung und die Fehlbestimmungen von Arten aus Aufnahmen, die vor (2018, 2019) und während des COVID-19-Ausbruchs (2020) mit einem Smartphone unter Verwendung der "Naturblick" App erstellt wurden. Die Anzahl der Aufnahmen von Nachtigallgesängen und das individuelle Engagement stiegen stetig an und erreichten ihren Höhepunkt in der Saison während der Pandemie. 13.991 Aufnahmen von Nachtigallgesängen wurden von anonymen (64%) und nicht-anonymen Teilnehmenden (36%) erstellt. Im Laufe des Projekts weitete sich die räumliche Verteilung der Aufnahmen aus (von Berlin auf bundesweit). Die Rate der Fehlbestimmungen war gering, ging im Laufe des Projekts zurück (von 10% auf 1%) und wurde hauptsächlich mit gesanglichen Ähnlichkeiten mit anderen Vogelarten beeinflusst. Unsere Studie zeigte außerdem, dass das gesellschaftliche Engagement und die Datenqualität nicht direkt von den durchgeführten Disseminationsaktivitäten beeinflusst wurden, sondern dass erstere von externen Faktoren abhingen und letztere von der App profitierte. Wir schließen daraus, dass CS-Projekte, die Smartphone-Apps mit einem integrierten Mustererkennungsalgorithmus verwenden, gut geeignet sind, um die bioakustische Forschung in der Ornithologie zu unterstützen. Auf der Grundlage unserer Ergebnisse empfehlen wir, CS-Projekte langfristig zu etablieren, um eine aktive Teilnehmergemeinschaft (Community) aufzubauen, die qualitativ hochwertige Daten für fundierte wissenschaftliche Schlussfolgerungen generiert.

Introduction

Citizen science (CS) is a research approach in which volunteers (non-professional scientists) and academic researchers (professional scientists) work together in one or several research processes to gain scientific knowledge (Bonney et al. 2014). In recent years, CS has been increasingly used in ecological and environmental research (e.g., Planillo et al. 2021). With the development of technical tools for citizen scientists such as interactive platforms or smartphone apps, the data collection increased recently (Falk et al. 2019). This technological improvement expanded local (e.g., Urban Birds Conservation Program of Vitoria-Gasteiz) and national projects (e.g., Dialects of Czech Yellowhammers: Diblíková et al. 2019) to global ones with real time data transmission (e.g., Xeno-Canto, Cornell Lab of Ornithology). Citizen scientists generated data on a large temporal and spatial scale that could otherwise not be obtained (Diblíková et al. 2019; Searfoss et al. 2020). The aim of CS studies can be quite diverse, e.g., they can serve conservation, raise awareness or monitor environmental efforts (e.g., Roy et al. 2012; Planillo et al. 2021). In the field of ornithology, CS has a long tradition and contains some long-running projects; for example, the National Audubon Society's Christmas Bird Count, the North American Breeding Bird Survey and the Pan-European Common Bird Monitoring Scheme. The number of long-term projects has increased ever since, while professional and lay bird enthusiasts are continuing to upload data on online platforms and archives (e.g., eBird: Sullivan et al. 2009, ornitho.de: Frick and Jaehne 2013, Project Feeder-Watch). Indeed, Randler (2021) showed that people using ornitho.de have better birding skills compared to other birders. Citizen scientists with a longer engagement provide a large amount of valid data e.g., to document the worldwide presence or absence of birds (Lepczyk al. 2005).

Ecological estimates of diversity or abundance based on CS data can be affected by project structure and thus community engagement as well as data quality (Dickinson et al. 2010). Community engagement, in the sense of generated data, have been found to decrease with project duration (Bruckermann et al. 2021). In contrast, the COVID-19 outbreak starting in 2020 increased individual engagement and data scope in CS projects (Hochachka et al. 2021). The pandemic also changed human activities around the world in terms of data distribution, probably due to a greater desire to spend time in nature (Venter et al. 2021), and quality (Phillips et al. 2021). Data quality is multidimensional and may be expressed by more than a dozen factors (reviewed in Lewandowski and Specht 2015) such as anonymity or inexperience of citizen scientists and project duration (Dickinson et al. 2010). In bird studies, this can lead to misidentifications. Identification is a complex task that relies on several factors such as vocal and visual similarities between co-occurring species or species richness (Kelling et al. 2015).

We here aimed to gain a better understanding of the community engagement and data quality in a CS project that was based on the song of the Common Nightingale (*Luscinia megarhynchos*) before (2018, 2019) and during the COVID-19 pandemic (2020). The nightingale is an ideal study object to address scientific questions using a CS approach. Nightingales possess a very memorable melodic and complex song which can frequently be heard from mid-April to late June in parks and gardens (Glutz von Blotzheim 1988) in Berlin, Germany. The literal translation in German ('*Nacht*'igall) as well as in English ('*night*'ingale) is a "night singer" and suggests a purely nocturnal song. Yet, nightingales are known to sing during both the night and the day-time (Amrhein et al. 2004). In the course of three nightingale breeding seasons, we invited all citizens to record nightingales with

their smartphones. Based on previous CS projects which showed that difficult tasks reduce data quality (Kosmala et al. 2016), we decided against strict protocols or specific instructions (duration, number, place or time of the recordings). Instead, we chose a low-threshold approach to target citizen scientists with little or no prior ornithological knowledge. The intent was to reach a wide audience, resulting in a diverse, high level of community engagement. The project, which started in Berlin in 2018, has been expanded to cover all of Germany from 2019 to 2020. We allowed the participants to engage anonymously or non-anonymously. As it has been shown that many dissemination activities increase data quality (Bryant and Oliver 2009), we aimed to provide information about nightingale song on the project website, in press coverage, and at scientific or cultural face-to-face events, mostly in Berlin and mainly before the COVID-19 pandemic.

The nightingale citizen science project (*Forschungsfall Nachtigall*) was based on a 2016 pilot project and previous dialect findings in the nightingale song (master thesis: Schehka 2004; doctoral thesis: Weiss 2012). Dialects are song variations between different populations and time periods (Catchpole and Slater 2008). For dialect studies, many recordings with a wide spatial distribution are needed. A growing body of literature demonstrates that this can easily be obtained through the power of worldwide participating citizen scientists (e.g., Diblíková et al. 2019; Searfoss et al. 2020). Our opportunistic approach indeed led to a large collection of geo-referenced nightingale song recordings. We previously showed that the majority of our CS data were valid enough (Jäckel et al. 2021) and of high value for dialect studies (Jäckel et al. 2022). The development of the project over three nightingale breeding seasons has not yet been studied.

Here, we thus focused on the community engagement (data scope, spatial distribution) and data quality (species misidentifications) before and during the COVID-19 pandemic. We investigated (1) the data scope in terms of the number of participants, cumulative duration and number of recordings from participants who took part either anonymously or non-anonymously, (2) the spatial and temporal distribution of recordings, and (3) species misidentifications in total and from anonymous users and non-anonymous ones and underlying patterns. In 2020 during the COVID-19 pandemic, we predicted to find a decrease in community engagement due to our reduced dissemination activities, yet an increase in individual engagement. The project has been promoted with a wider geographic outreach after the first year, whereby we expected that recordings would be more widely distributed over the last 2 years. We predicted that particularly common and other melodious bird species were mistaken for nightingales—especially during the night—as

lay people often assume that only nightingales have a nocturnal song.

Methods

The nightingale citizen science project and its cooperation with the 'Naturblick' app

The nightingale citizen science project was launched in 2018 as a collaboration with the 'Naturblick' app at the Museum für Naturkunde (MfN) in Berlin, Germany. The app has been available since 2016 and has already been widely used in 2017 with almost 50,000 downloads (Sturm and Tscholl 2019). As a special feature, the app includes a pattern recognition algorithm (PRA) which automatically identifies bird species based on cross-correlation via template matching of spectrogram segments (Lasseck 2016; Stehle et al. 2020). During the last years, the PRA improved due to the use of neural networks as well as deep learning (Lasseck 2018).

Using diverse dissemination activities such as events both inside and outside the MfN, midnight excursions and press coverage, the public was invited to download the 'Naturblick' app on their smartphones and record nightingale songs (for details see Jäckel et al. 2021). Public events were free of charge, included two to 180 participants and by large took place in Berlin. Press coverage occurred mainly before and during the breeding season in the form of radio interviews, newspaper articles and social media posts. The app featured a bespoke button showing a nightingale, to highlight the citizen science project. By clicking on this button, participants could transmit their recordings directly or make use of the PRA to aid in species identification (Fig. 1). As part of this process, the three bird species whose vocalisations most closely match the recording were presented (Stehle et al. 2020). Participants were also allowed to choose whether they wanted to submit the recordings anonymously or non-anonymously with an individual username. Because of technical limitations, it was not possible to submit recordings more than once or with a duration of more than two minutes per recording. Temporal (day, time) and spatial (GPS coordinates) information were automatically captured in the metadata, if permitted by the participant.

Development of the data scope over the study period

To get a better understanding of the community development over the course of the project, we descriptively analysed the data scope before (2018, 2019) and during
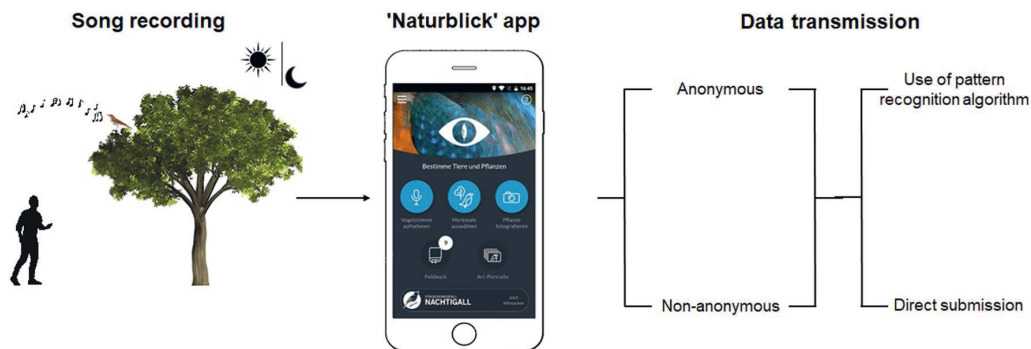
Fig. 1 Left: Process of song recording transmission. Middle: Display of the 'Naturblick' app. Right: For data transmission, users could participate anonymously or non-anonymously and either make use of the pattern recognition algorithm or directly submit the recording

the COVID-19 pandemic (2020). We used the number of participants, cumulative duration and number of all recordings by participants who took part anonymously or non-anonymously in 1, 2, or 3 years as parameters. As dissemination activities are supposed to have an impact (Bryan and Oliver 2009), we examined whether their reduction in 2020 had any effect. Less dissemination activities were provided during the 2020 nightingale breeding season. That year was dominated by the COVID-19 pandemic lockdown and face-to-face events were not permitted in Berlin. Instead, online field trips and events were undertaken, a method which has been found to be equally reliable (Rögele et al. 2022).

Development of spatial distribution of CS recordings in Germany

To enable comparisons between populations, it is important to obtain recordings from different places. Citizen scientists, therefore, have to cover a wide spatial distribution. We aimed to reach people from Berlin, where the MfN is based, and all over Germany, since participation in this project was possible from any location with a smartphone and the 'Naturblick' app. Germany was separated into four regions (North, East, South, West) which included quadrants with an area of 100×100 kms each. We divided the number of quadrants that contained nightingale song recordings by the number of total quadrants in Germany to determine the percentages for each year and whether the spatial distribution increased with time.

## Underlying patterns of species misidentifications

Species identification is challenging for citizen scientists (Crall et al. 2011) and one of the major data quality issues. We aimed to elucidate and better understand the development of species misidentifications over the course of the

project and underlying patterns. MP3 and m4a files were converted into the WAV format using the WaveLab 7 program to analyse them visually and acoustically with Avisoft SASLab Pro 5.2 (R. Specht, Berlin, Germany, sampling rate = 22.050 Hz, FFT = 1024 points, Hamming-Window, overlap 93.75%). Recordings were sorted into four types (nightingale songs, nightingale calls, other bird species, i.e., species misidentifications of the nightingale and no bird song). We examined the effect of dissemination activities in all three nightingale breeding seasons and of participation (i.e., anonymity, experience) on the recording types and descriptively analysed the respective percentages (number of recordings per classified type/all recordings).

To identify temporal factors for underlying patterns (calendar week and time of day), we descriptively compared the number of nightingale songs and other bird species recorded across all study years. The peak singing time of nightingales (23:00–1:00 h) as reported in the literature (see Jäckel et al. 2021) and other species of birds (5:00–7:00 h and 19:00–21:00 h) were accounted for to see if they have an influence on the rate of species misidentifications. To understand if and when certain species are most often misidentified as nightingales, we compared recordings from day (04:00–22:00 h) and night (22:00–04:00 h). Not all of the other bird species could be identified (too short, disturbing background noise).

As there were few species misidentifications in total (about 1–10% of submissions), we decided to work on the taxonomic group level. We analysed underlying patterns in terms of similarities (vocal, visual and species abundance) between the taxonomic group of the other bird species and the nightingale. Vocal similarities were determined by the following parameters: melodic (yes/no), complex (yes/no), nightingale song elements (whistle/trill/buzz) and usual time of singing (night/dawn/day). Nightingales are visually recognisable on their song post. Thus, we determined for the evaluation of visual similarities, whether the plumage

Fig. 2 Overview of the project's data scope development with regard to the total number of dissemination activities (**a**), number of participants (**b**), cumulative duration of recordings (**c**), and number of recordings from citizen scientist who participated anonymously or non-anonymously in 1, 2, or 3 years (**d**). The dashed line separates the years before (2018, 2019) and during (2020) the COVID-19 pandemic
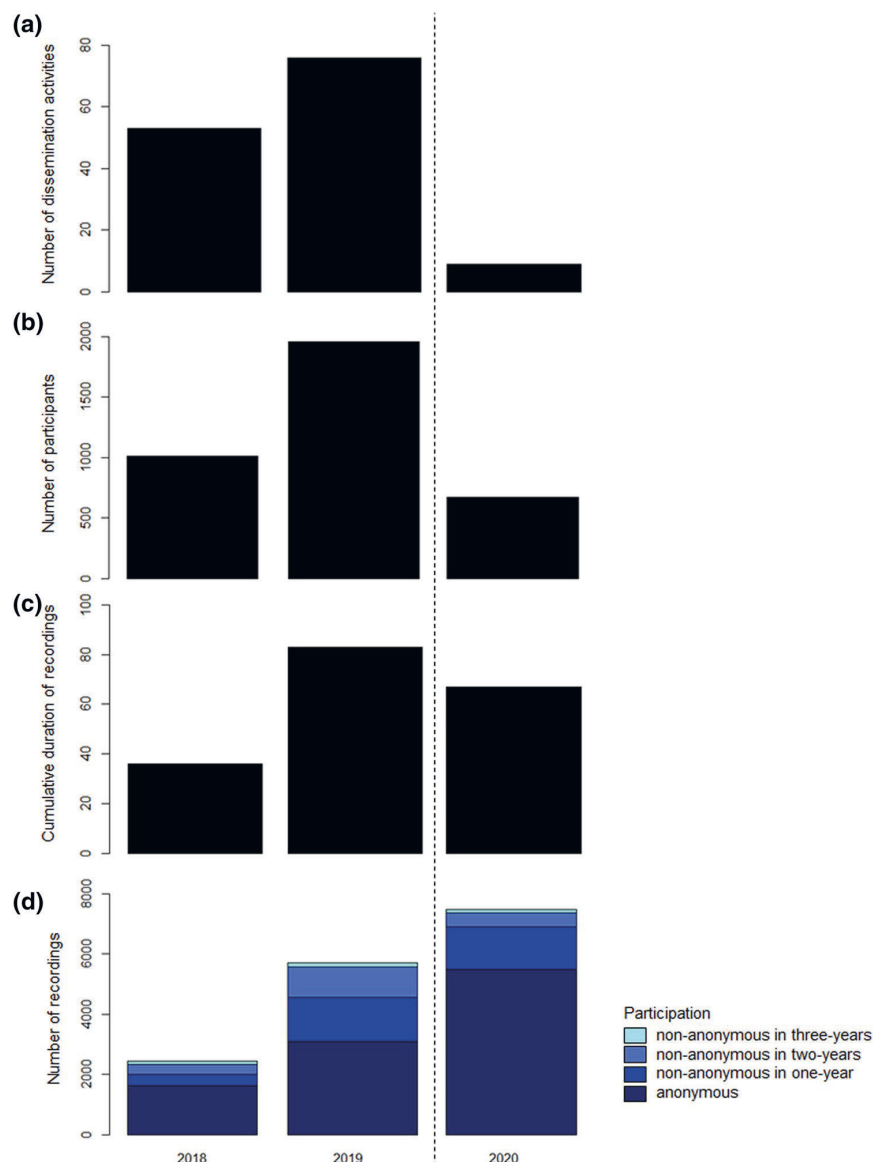
was identical to the nightingale (brown), similar (black) or different. As frequently occurring species are easier to identify than rarely occurring ones (Falk et al. 2019), we determined the species abundance in Germany from the 'Berliner Ornithologische Arbeitsgemeinschaft e.V.' (Berlin Ornithological Society) breeding bird monitoring (http://www.orniberlin.de/) and divided it into six groups (group 1 = ≤ 1,000, group 2 = ≤ 10,000, group 3 = ≤ 100,000, group 4 = ≤ 500,000, group 5 = ≤ 1,000,000, group 6 = > 1,000,000).

Statistical analyses

We performed all statistical analyses with R version 4.1.2 (R-Team 2021). A possible correlation between (i) the dissemination activities and the number of participants, (ii) cumulative duration and the recording type (iii) or number of recordings and the recording type was determined using Spearman's rank correlation test. We used a principal component analysis (PCA) and a general linear model (GLM) to compare similarities (vocal, visual and species abundance) between taxonomic groups and nightingales.
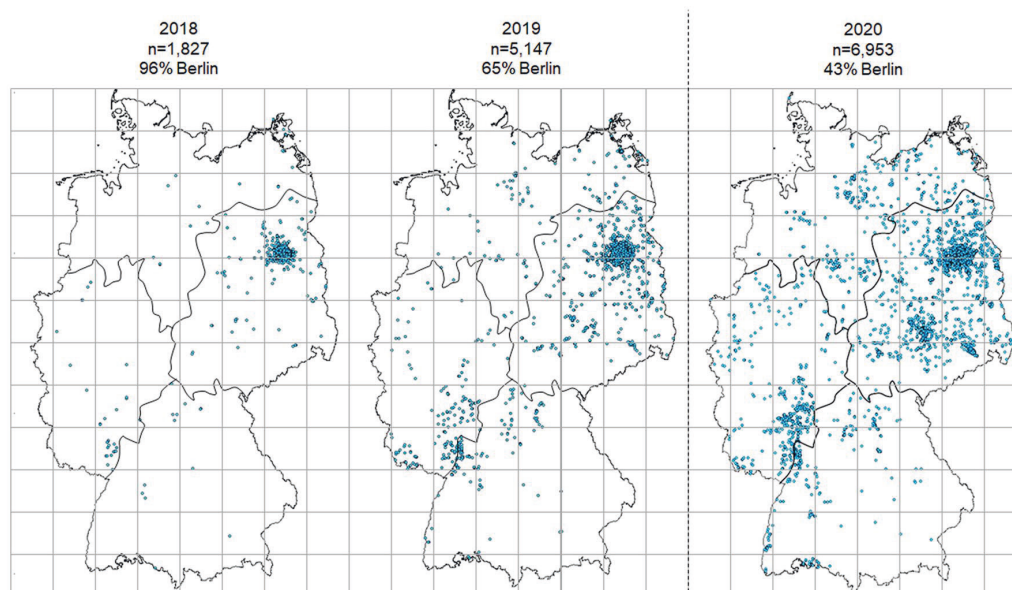
Fig. 3 The spatial distribution of nightingale song recordings in the CS project from Germany. The dashed black line separates the years before (2018–2019) and during the COVID-19 pandemic (2020). Black continuous lines show the four regions (North, East, South, West). Grey solid lines represent quadrants (100×100 kms)

Results

Development of the data scope over the study period

The project showed a positive community development over the project duration in terms of the data scope. The number of dissemination activities (Fig. 2a, Online Resource 1), cumulative duration (Fig. 2b) and number of participants (Fig. 2c) increased from 2018 to 2019 and decreased in 2020. Number of recordings increased continuously over the nightingale breeding seasons (Fig. 2d), this was due to a higher number of individual engagement especially in 2020. The majority of data were recorded anonymously, followed by citizen scientists who participated non-anonymously in 1, 2 and 3 years. Dissemination activities did not correlate with the number of participants, cumulative duration or number of recordings (Spearman's rank correlation test: $p > 0.05$).

Development of spatial distribution of CS recordings in Germany

Most recordings were obtained from the East of Germany, followed by the West, North and South (Fig. 3). This matches the natural distribution of the nightingale (reported by the German national bird breeding count by the 'Dachverband Deutscher Avifaunisten; Gedeon et al. 2014). During 2018, we had the lowest spatial distribution in terms of the percentage of quadrants which contained nightingale song recordings (42%). In 2020, during the COVID-19 pandemic, the spatial distribution for all regions was larger (90%) than in 2019 (74%). Over the years, the data were more geographically spread and no longer came mainly from Berlin as in the first year.

Underlying patterns of species misidentifications

The data quality of the recording types showed a positive development over the 3 years. The percentage of nightingale songs and calls (Fig. 4a) increased steadily over the study period. At the same time, the percentage of other bird species and no birds decreased continuously. In 2020, during the COVID-19 pandemic, the number of species misidentification was the lowest. Dissemination activities did not correlate with the recording types (Spearman's rank correlation test: $p > 0.05$). The percentages were similar between anonymous and non-anonymous participants (Fig. 4b, Online Resource 2). The highest percentage of recordings with nightingale songs and the fewest with other bird species were generated by non-anonymous citizen scientists who

took part in all 3 years, followed by those who attended 2 years, 1 year and anonymously.

A descriptive analysis of recordings from all years revealed that species misidentification was not affected by the song timing across calendar weeks or time of day. Most nightingale songs were recorded in the 16th and 17th calendar week (Fig. 5a). Recordings with other bird songs were made primarily between the 18th and 20th calendar week (see Jäckel et al. 2021). After the high nightingale breeding season when song is most common, slightly more (26th week) or the same amount (27th week) of other bird songs were recorded. Most recordings of nightingale song were made between 23:00 and 0:00 h and of other bird species at 21:00 and 4:00 h (Fig. 5b). There were more recordings with species misidentifications of other bird species recorded at hours outside the times when the nightingale song is regularly found (23:00 and 1:00 h).

During the day, European Blackbirds (*Turdus merula*, $n = 295$), Song Thrush (*Turdus philomelos*, $n = 35$) and House Sparrow (*Passer domesticus*, $n = 31$) were the species that were mainly misidentified as nightingales by participants (Table 1). European Blackbirds ($n = 30$), song thrush ($n = 12$) and European Robin (*Erithacus rubecula*, $n = 5$) were the species mainly misidentified at night. The Thrush

Nightingale (*Luscinia luscinia*), the common nightingale's sister species (Vokurková, et al. 2013) was recorded fourteen times.

A PCA of three similarities traits (vocal, visual and species abundance) showed patterns of species misidentifications. Two principal components explained 71% of the species misidentifications with other species for vocal and visual similarities (Fig. 6, PC1 = 44%) and species abundance (PC2 = 28%) between the taxonomic groups and the nightingale. Vocal similarities mainly in the Muscicapoidea and Sylvioidea and partly in Certhiodea may have had an influence (for an overview of the species see Table 1). Visual similarities of Passeroidea and Certhiodea may have had an effect. Species abundance may be a factor for Passeroidea. A general linear model of the first principal component showed that species misidentifications on the level of the taxonomic groups were influenced significantly by the vocal similarity (GLM: $df = 24$, p-value < 0.005; Tab. 2).
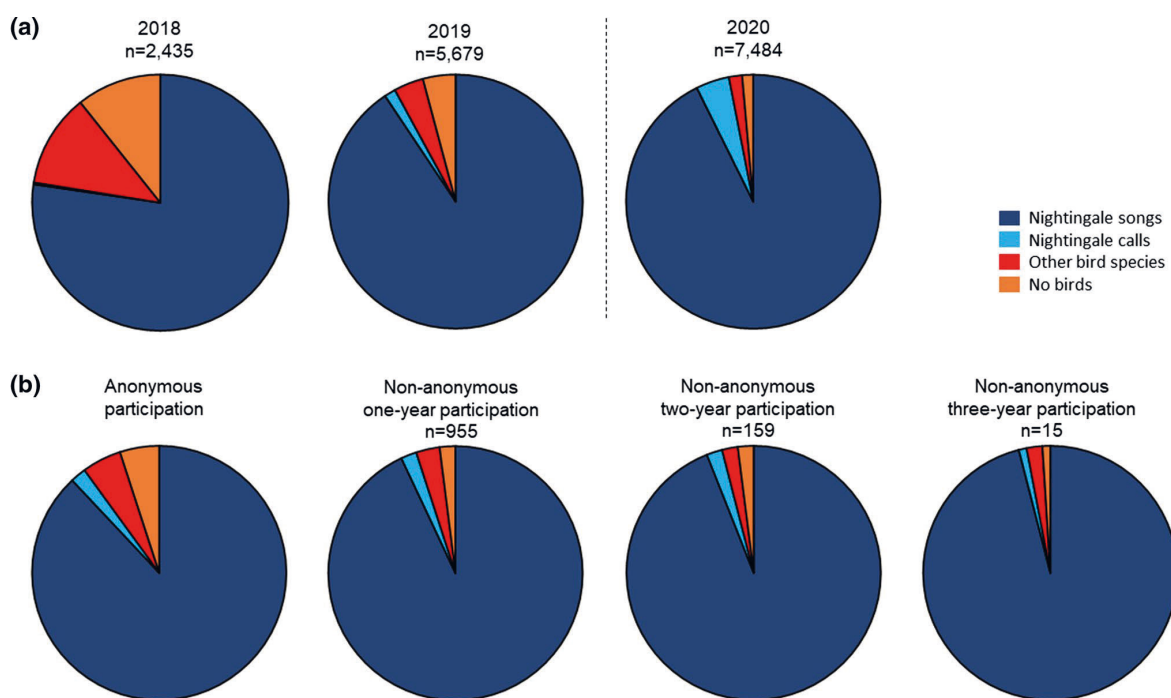
Fig. 4 Comparison of CS recordings with regard to percentages of recording types (**a**) and from citizen scientists who participated anonymously or non-anonymously (**b**). The dashed line separates the years before (2018, 2019) and during (2020) the COVID-19 pandemic
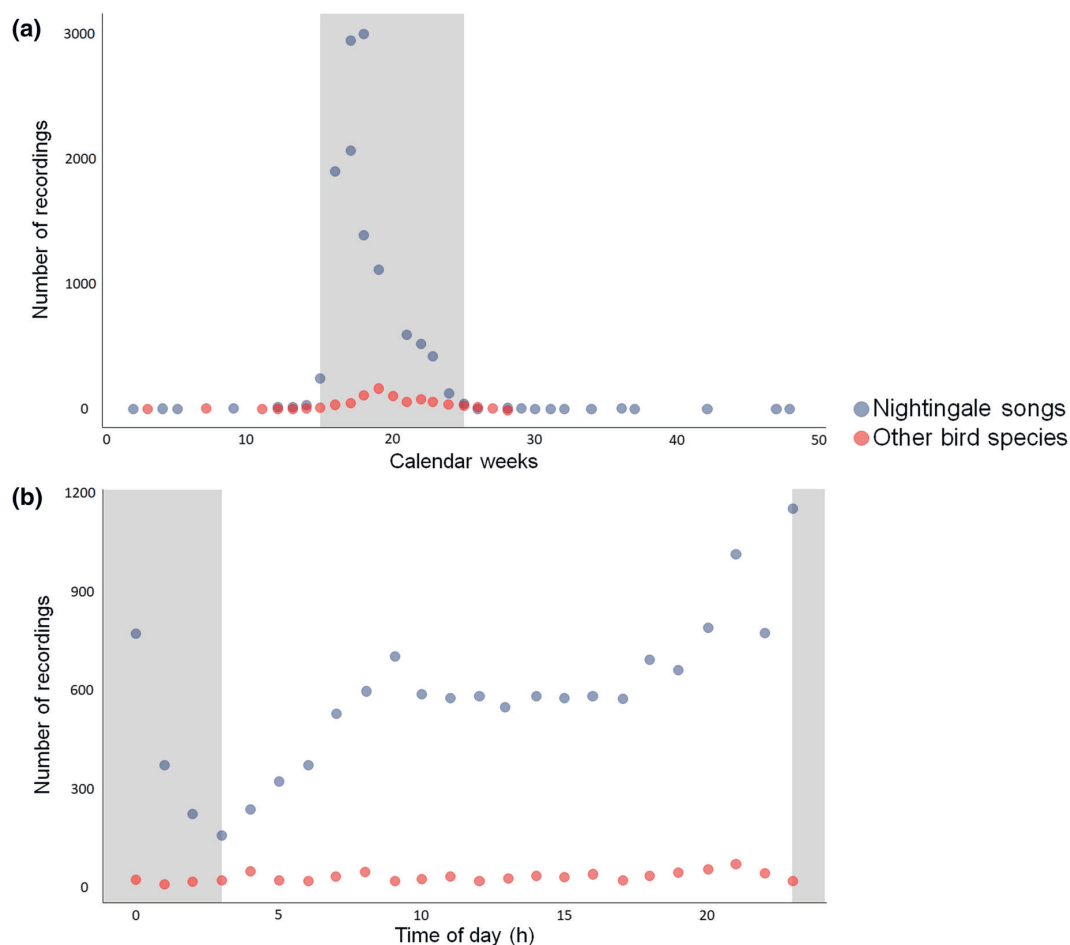
Fig. 5 The temporal distributions of CS recordings for 13,991 nightingales and 601 other bird species (i.e., species misidentifications) during the whole study period (2018–2020). In grey, the periods of nightingale song as reported in the literature is highlighted. **a** The weekly number of recordings in the course of the breeding season. **b** The number of recordings in relation to the time of day (h)

Discussion

In this paper, we investigated the development of community engagement and data quality of a 3-year CS project based on the song of the nightingale before and during the COVID-19 pandemic. We found an increase in the dynamics of community engagement over the course of the project in both the data scope and spatial distribution. The number of species misidentifications decreased during the study period and species misidentifications were mostly affected by vocal similarities of other species. In the following, we will discuss our findings in more detail with regard to two aspects: best practices and lessons learned that we believe could be useful to grow and improve (future) avian-based CS projects.

Best practises: success factors in citizen science

The nightingale CS project steadily grew in numbers of recordings collected, despite all the challenges that CS brings (Bonney et al. 2014; Wittman et al. 2019) and the COVID-19 pandemic brought in 2020. In total, 13,991 nightingale song recordings were submitted to the project by anonymous (64%) and non-anonymous participants (36%). We achieved the goal of high community engagement, even though only a few non-anonymous citizen scientists participated in all three nightingale breeding seasons. Similar to other CS projects (e.g., Segal et al. 2015; Seymour and Haklay 2017), most of the non-anonymous participants took part in only a single nightingale breeding season. However, proportionately, these 1-year citizen scientists produced many and anonymous ones produced

Table 1 The number of species misidentifications on the species level for 601 recordings of other bird species during the whole study period (2018–2020)

Species		Taxonomic group	Number of recordings	
English name	Latin name		Day	Night
European blackbird	*Turdus merula*	Muscicapoidea	295	30
Song thrush	*Turdus philomelos*	Muscicapoidea	23	12
European robin	*Erithacus rubecula*	Muscicapoidea	22	5
Common redstart	*Phoenicurus phoenicurus*	Muscicapoidea	10	1
Thrush nightingale	*Luscinia luscinia*	Muscicapoidea	8	6
European starling	*Sturnus vulgaris*	Muscicapoidea	4	0
Northern mockingbird	*Mimus polyglottos*	Muscicapoidea	2	0
Redwing	*Turdus iliacus*	Muscicapoidea	1	0
Bicknell's thrush	*Catharus bicknelli*	Muscicapoidea	1	0
Black redstart	*Phoenicurus ochruros*	Muscicapoidea	1	1
House sparrow	*Passer domesticus*	Passeroidea	31	0
Common chaffinch	*Fringilla coelebs*	Passeroidea	13	2
European greenfinch	*Chloris chloris*	Passeroidea	4	0
European goldfinch	*Carduelis carduelis*	Passeroidea	2	0
House finch	*Haemorhous mexicanus*	Passeroidea	1	1
European serin	*Serinus serinus*	Passeroidea	1	0
Grey wagtail	*Motacilla cinerea*	Passeroidea	1	0
Icterine warbler	*Hippolais icterina*	Passeroidea	1	0
Pine grosbeak	*Pinicola enucleator*	Passeroidea	1	0
Eurasian blackcap	*Sylvia atricapilla*	Sylvioidea	26	1
Great tit	*Parus major*	Sylvioidea	10	0
Great reed warbler	*Acrocephalus arundinaceus*	Sylvioidea	4	2
Willow warbler	*Phylloscopus trochilus*	Sylvioidea	3	0
Eurasian blue tit	*Cyanistes caeruleus*	Sylvioidea	2	1
Western bonelli's warbler	*Phylloscopus bonelli*	Sylvioidea	1	0
Arctic warbler	*Phylloscopus borealis*	Sylvioidea	1	0
Garden warbler	*Sylvia borin*	Sylvioidea	1	1
Eurasian wren	*Troglodytes troglodytes*	Certhioidea	11	0
Eurasian treecreeper	*Certhia familiaris*	Certhioidea	1	0
Carrion crow	*Corvus corone*	Corvoidea	4	0
Eurasian magpie	*Pica pica*	Corvoidea	1	0
Western jackdaw	*Corvus monedula*	Corvoidea	1	0
Eurasian scops owl	*Otus scops*	Strigidae	1	0
Common wood pigeon	*Columba palumbus*	Columboidea	2	0
European green woodpecker	*Picus viridis*	Picoidea Vigor	2	0

the majority of recordings. Anonymity may have been a success factor, without which two thirds of the data would not have been obtained. Bryant and Oliver (2009) demonstrated that dissemination activities do not seem to be a contributor to success for the duration or community engagement of participation in CS projects. Similarly, we assume that some citizen scientists liked joining our diverse dissemination activities but rarely generated recordings, while others recorded but did not attend. Even though press coverage and events did not seem to have an influence on the data scope (i.e., number of citizen scientists, duration and number of recordings), their reduction during the COVID-19 outbreak led to fewer participants and possibly shorter recording duration than in the previous years. These fewer citizen scientists in 2020 were individually more engaged and produced more recordings. The enhanced use as well as the improvement of the PRA may have led to an increase in the number and a decrease in the duration of recordings in the last year. Our findings are also consistent with other CS projects that reported higher individual engagement during the pandemic (e.g., Sánchez-Clavijo et al. 2021; Hochachka et al. 2021) based presumably on an increased desire to experience nature (Flaccus 2020) or outdoor activities (Venter et al.
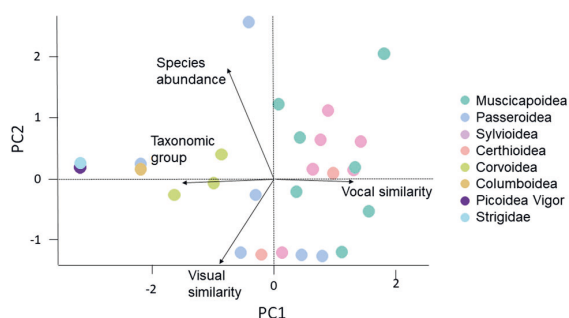
Fig. 6 A principal component analysis (PCA) for 601 other bird species recordings during the whole study period (2018–2020). We compared similarities (vocal, visual and species abundance) between taxonomic groups and nightingales. The arrow length indicates the degree of similarities, the arrow direction indicates the association of the factors with the principal components PC1 and PC2

Table 2 Results of a general linear model testing whether the first principal component (PC1) showed relationships between the traits and the taxonomic group

	Estimate	SEM	t	p
Taxonomic group	4.3249	1.6256	2.660	0.137*
Vocal similarity	− 0.613	0.1769	− 3.467	0.0020**
Visual similarity	0.5638	0.3200	1.762	0.0908
Species abundance	− 0.2869	0.1998	1.436	0.1639

2021). The geographical spread of our opportunistic data increased over the course of the project and was similar to the nightingale distribution in Germany (Gedeon et al. 2014). Moreover, our project was successful enough to expand the initial geographical focus from Berlin to a nationwide data coverage.

The low-threshold approach which engaged a wide audience of citizen scientists was also successful in this project. Contrary to Falk et al. (2019), participants as a mass may have improved, resulting in higher data accuracy than in other CS projects (e.g., Kosmala et al. 2016). In fact, the data quality of anonymous users was similar to those who participated for multiple years non-anonymously. This revealed that quality was not negatively affected neither by reduction of our dissemination activities (Bryant and Oliver, 2009) nor by anonymity of participants (Dickinson et al. 2010) but positively influenced by the project duration and the improved PRA.

Lessons learned: suggestions for future CS projects on birdsong

As there is evidence in the literature that in many cases, data quality is linked to avoidable errors in the study design (e.g.,

Bowser et al. 2020), we suggest the following measures for future CS projects:

Most species misidentifications did not occur at night, as expected, but were affected by vocal similarities of other species. We should, for example, have provided more information through the use of the already integrated 'Species Portrait' and 'Trait Selection' features of the app, which could have reduced confusions with other species based on vocal similarities. Citizen scientists should have been supported with more specific instructions for the recordings, i.e., how (smartphone orientation), where (regions), and when (during the night). Such a task-lead approach facilitates data collection for citizen scientists (Moczek et al. 2021) and enables them to get a common understanding of data quality (Land-Zandstra et al. 2021). Positive feedback and confirmation of the number of successfully recorded nightingale songs likewise might have helped to increase quality (Peltola and Arpin 2018) and motivate citizen scientists to participate long-term (Pandya and Dibner 2018). It has been suggested that rewards (Reeves et al. 2017), greater involvement in scientific processes and valuing individual contributions (Dowthwaite and Sprinks 2019) could lead to data generated not by a few participants, (90-9-1 rule: Gasparini et al. 2020), but by many and further over the whole project duration. In future projects, the aspects that influence data quality could be validated and specified by additional demographic data (e.g., leisure activities; Lee and Scott 2004). Equally informative would be a self-assessment item or a scale of skills and knowledge to compare the recordings (type, number and duration).

Overall, the project was very successful and has positively evolved over the years. As individual engagement, data scope, spatial distribution, and data quality have largely increased over the nightingale breeding seasons, this strongly suggests that CS projects need to be implemented over the long term. It also became apparent that it was indeed a good idea to focus on a specific species like the nightingale, since species identification is a difficult task (Crall et al. 2011) and the nightingale proved to be a charismatic focal bird.

Conclusion

The nightingale citizen science project has demonstrated that CS is a research approach that can contribute large datasets with data quality valid to science through increasing community development and engagement over time. For dialect research, many and diverse recordings from various locations are necessary, making our CS data highly valuable. The findings from our study may also offer great value for other CS projects to gain insights into best practices and to avoid systematic species misidentifications in

the future, which can lead to biased ecological estimates (Dickinson et al. 2010). In sum, our study may inspire other existing and evolving CS projects in ornithology to adapt their study design with regard to their modes of community engagement and ways to ensure data quality.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest This work was supported by BMBF (Förderkennzeichen: 01BF1709). The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

Ethics approval This study investigates the development of a citizen science project based on the nightingale song. The data of the citizen scientists were shared with our project with their approval via the 'Naturblick' app. For the recordings, we obtained the consent of participants to analyse their data. In Germany, the approval of an ethics committee is not required for such research questions and was, therefore, not obtained. We have, therefore, received all the necessary permissions required in Germany.

References

Amrhein V, Kunc HP, Naguib M (2004) Seasonal patterns of singing activity vary with time of day in the nightingale (*Luscinia megarhynchos*). Auk 121:110–117. https://doi.org/10.1093/auk/121.1.110

Bonney R, Shirk JL, Phillips T, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK (2014) Next steps for citizen science. Science 343:1436–1437. https://doi.org/10.1126/science.1251554

Bowser A, Cooper C, De Sherbinin A, Wiggins A, Brenton P, Chuang TR et al (2020) Still in need of norms: the state of the data in citizen science. Citiz Sci Theory Pract 5:1–16. https://doi.org/10.5334/CSTP.303

Bruckermann T, Greving H, Brandt M, Harms U (2021) Daten sammeln reicht mir! Bürgerwissenschaften: Welche Aktivitäten interessieren Beteiligte? IPN Journal 8:24–29. www.ipn.uni-kiel.de/de/publikationen/ipn-journal/ipn-journal-no-8

Bryant J, Oliver MB (2009) Media effects: advances in theory and research. Routledge, London

Catchpole CK, Slater PJB (2008) Bird song. Biological themes and variations. Cambridge Univ, Press

Crall AW, Newman GJ, Stohlgren TJ, Holfelder KA, Graham J, Waller DM (2011) Assessing citizen science data quality: an invasive species case study. Conserv Lett 4:433–442. https://doi.org/10.1111/j.1755-263X.2011.00196.x

Diblíková L, Pipek P, Petrusek A, Svoboda J, Bílková J, Vermouzek Z et al (2019) Detailed Large-scale Mapping of Geographical Variation of Yellowhammer *Emberiza citrinella* Song Dialects in a Citizen Science Project. Ibis 161:401–414. https://doi.org/10.1111/ibi.12621

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst 41:149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Dowthwaite L, Sprinks J (2019) Citizen science and the professional-amateur divide: lessons from differing online practices. J Sci Commun 18:1–16. https://doi.org/10.22323/2.18010206

Falk S, Foster G, Comont R, Conroy J, Bostock H, Salisbury A, Kilbey D, Bennett J, Smith B (2019) Evaluating the ability of citizen scientists to identify Bumblebee (Bombus) species. PLoS ONE 14(1–13):e0218614. https://doi.org/10.1371/journal.Pone.0218614

Frick S, Jaehne S (2013) Das Internetportal „ornitho. de " - Instrument zum Artenmonitoring. Landschaftspflege Und Naturschutz in Thüringen 50:174–175

Flaccus G (2020) Bird-watching Soars Amid COVID-19 as Americans Head Outdoors. Associated Press. https://apnews.com/article/94a1ea5938943d8a70fe794e9f629b13

Gasparini M, Clarisó R, Brambilla M, Cabot J (2020) Participation inequality and the 90–9–1 principle in open source. In: Proceedings of the 16th International Symposium on Open Collaboration 6:1–7. https://doi.org/10.1145/3412569.3412582

Gedeon K, Eikhorst W, Fischer S, Flade M, Frick S, Geiersberger I et al (2014) Atlas Deutscher Brutvogelarten. Stiftung Vogelmonitoring Deutschland

Glutz von Blotzheim UN (1988) Handbuch der Vögel Mitteleuropas. Aula, Wiesbaden

Hochachka WM, Alonso H, Gutiérrez-Expósito C, Miller E, Johnston A (2021) Regional variation in the impacts of the COVID-19 pandemic on the quantity and quality of data collected by the project eBird. Biol Cons 254:1–41. https://doi.org/10.1016/j.biocon.2021.108974

Jäckel D, Mortega KG, Sturm U, Brockmeyer U, Khorramshahi O, Voigt-Heucke SL (2021) Opportunities and limitations: a comparative analysis of citizen science and expert recordings for

bioacoustic research. PLoS ONE 16:1–25. https://doi.org/10.1371/journal.pone.0253763

Jäckel D, Mortega KG, Brockmeyer U, Lehmann GUC, Voigt-Heucke SL (2022) Unravelling the stability of nightingale song over time and space using open, citizen science and shared data. Front Ecol Evol 10:1–13. https://doi.org/10.3389/fevo.2022.778610

Kelling S, Johnston A, Hochachka WM, Iliff M, Fink D, Gerbracht J et al (2015) Can observation skills of citizen scientists be estimated using species accumulation curves? PLoS ONE 10:1–20. https://doi.org/10.1371/journal.pone.0139600

Kosmala M, Wiggins A, Swanson A, Simmons B (2016) Assessing data quality in citizen science. Front Ecol Environ 14:551–560. https://doi.org/10.1002/fee.1436

Land-Zandstra A, Agnello G, Gültekin YS (2021) Participants in citizen science, vol 13. Springer Nature, pp 243–259. https://doi.org/10.1007/978-3-030-58278-4_13

Lasseck M (2016) Improving bird identification using multiresolution template matching and feature selection during training. In CLEF (Working Notes), pp 490–501

Lasseck M (2018) Audio-based Bird species identification with deep convolutional neural networks. In: CEUR Workshop Proceedings, pp 1–11

Lepczyk CA (2005) Integrating published data and citizen science to describe bird diversity across a landscape. J Appl Ecol 42:672–677. https://doi.org/10.1111/j.1365-2664.2005.01059.x

Lee JH, Scott D (2004) Measuring birding specialization: a confirmatory factor analysis. Leis Sci 26:245–260. https://doi.org/10.1080/01490400490461387

Lewandowski E, Specht H (2015) Influence of volunteer and project characteristics on data quality of biological surveys. Conserv Biol 29:713–723. https://doi.org/10.1111/cobi.12481

Moczek N, Hecker S, Voigt-Heucke SL (2021) The known unknowns: what citizen science projects in Germany know about their volunteers—and what they don't know. Sustainability 13:1–10. https://doi.org/10.3390/su132011553

Pandya RE, Dibner KA (2018) Learning through citizen science. enhancing opportunities by design. National Academies Press, Washington, D.C.

Peltola T, Arpin I (2018) Science for everybody? Bridging the socio-economic gap in urban biodiversity monitoring. UCL Press

Phillips TB, Parker A, Bowser A, Haklay M (2021) Publicly generated data: the role of citizen science for knowledge production, action, and public engagement. In: Closing the knowledge-implementation gap in conservation science. Wildlife Research Monographs, vol 4, Springer, Cham. https://doi.org/10.1007/978-3-030-81085-6_4

Planillo A, Fiechter L, Sturm U, Voigt-Heucke S, Kramer-Schadt S (2021) Citizen science data for urban planning: comparing different sampling schemes for modelling urban bird distribution. Landsc Urban Plan 211:713–715. https://doi.org/10.1016/j.landurbplan.2021.104098

Randler C (2021) Users of a citizen science platform for bird data collection differ from other birdwatchers in knowledge and degree of specialization. Glob Ecol Conserv 27:1–10. https://doi.org/10.1016/j.gecco.2021.e01580

Reeves N, Tinati R, Zerr S, Van Kleek MG, Simperl E (2017) From crowd to community: a survey of online community features in citizen science projects. In: Proceedings of the 2017 ACM Conference on computer supported cooperative work and social computing, pp 2137–2152

Rögele A, Scheiter K, Randler C (2022) Can involvement induced by guidance foster scientific reasoning and knowledge of participants of a citizen science project? Int J Sci Educ 12:1–17. https://doi.org/10.1080/21548455.2022.2043567

Roy HE, Pocock MJ, Preston CD, Roy DB, Savage J, Tweddle JC, Robinson LD (2012) Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework. Wallingford, NERC/Centre for Ecology & Hydrology

Sánchez-Clavijo LM, Martínez-Callejas SJ, Acevedo-Charry O, Diaz-Pulido A, Gómez-Valencia B, Ocampo-Peñuela N et al (2021) Differential reporting of biodiversity in two citizen science platforms during COVID-19 lockdown in Colombia. Biol Cons 256:1–9. https://doi.org/10.1016/j.biocon.2021.109077

Searfoss AM, Liu WC, Creanza N (2020) geographically well-distributed citizen science data reveals range-wide variation in the Chipping Sparrow's simple song. Anim Behav 161:63–76. https://doi.org/10.1016/j.anbehav.2019.12.012

Segal A, Gal YA, Simpson RJ, Victoria Homsy V, Hartswood M, Page KR, Jirotka M (2015) Improving productivity in citizen science through controlled intervention. In: Proceedings of the 24th International Conference on World Wide Web 331–337 https://doi.org/10.1145/2740908.2743051

Seymour V, Haklay,M (2017) Exploring Engagement Characteristics and Behaviours of Environmental Volunteers. Citizen Science: Theory Pract 2:1–13. https://doi.org/10.5334/cstp.66

Stehle M, Lasseck M, Khorramshahi O, Sturm U (2020) Evaluation of acoustic pattern recognition of nightingale (*Luscinia megarhynchos*) recordings by citizens. Res Ideas Outcomes 6:1–9. https://doi.org/10.3897/rio.6.e50233

Sturm U, Tscholl M (2019) The role of digital user feedback in a user-centred development process in citizen science. J Sci Commun 18:1–19. https://doi.org/10.22323/2.18010203

Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S (2009) eBird: a citizen-based bird observation network in the biological sciences. Biol Cons 142:2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

Venter ZS, Barton DN, Gundersen V, Figari H, Nowell MS (2021) Back to nature: Norwegians sustain increased recreational use of urban green space months after the COVID-19 outbreak. Landsc Urban Plan 214:1–10. https://doi.org/10.1016/j.landurbplan.2021.104175

Vokurková J, Petrusková T, Reifová R, Kozman A, Mořkovský L, Kipper S et al (2013) The causes and evolutionary consequences of mixed singing in two hybridizing songbird species (*Luscinia* spp.). PLoS ONE 8:1–10. https://doi.org/10.1371/journal.pone.0060172

Wittmann J, Girman D, Crocker D (2019) Using iNaturalist in a coverboard protocol to measure data quality: suggestions for project design. Citiz Sci Theory Pract 4:1–12. https://doi.org/10.5334/cstp.131

https://support.rstudio.com/hc/en-us/articles/206212048-Citing-RStudio

Chapter V

Synthesis

1. Summary

Previous CS birdsong projects have been successfully conducted for species with a small, but not with a large, repertoire. Until now, there have not been any nightingale studies based on data from citizen scientists. The nightingale is a well-suited study object as its song has been extensively examined at the individual level since the 1980s (Hultsch, 1980). The aim of this thesis was to evaluate the method CS for bioacoustic research using dialect studies in the nightingale as a use case. In the following, I will evaluate the main findings of the individual chapters (**Chapters II - IV**), according to the respective research question introduced in the general introduction (**Chapter I**). I will give recommendations for future CS projects in the 'conclusion' section to improve data quality in other birdsong studies.

Are CS data as valid as those of academic researchers?

The first study (**Chapter II**) compared non-standardised CS data with standardised ones of academic researchers, which were generated under the same conditions (time, place). Such a comparison of quality under the same conditions was rare until now (e.g. Freitag, Meyer & Whiteman, 2016) and revealed different opportunities and limitations: In contrast to other studies (e.g. Paulos, 2009), my co-authors and I showed that the CS data scope in terms of the cumulative duration and number of recordings was considerably larger than the one of academic researchers. The technique used (smartphone actuality, brand) and not to the participants influenced how well songs could be visually distinguished and how accurate durations were measured. All user groups contributed differently to the data scope, which indicates that experience had no influence on the data quality. Unlike other studies (e.g. Dickinson, Zuckerberg & Bonter, 2010), this chapter demonstrated that the temporal distribution was better for CS than for academic researcher data. Although the citizen scientists did not receive a standardised protocol, their data followed the natural nightingale song behaviour quite closely. These recordings covered the whole breeding season in terms of calendar weeks and times of day for the first time. Limitations of CS data were also revealed as their mean duration was shorter and they were not completely accurate. Frequency measurements showed significant differences, which could be overcome with external calibrated microphones. The findings of this study served as a basis for a better understanding of the quantity and quality of the method CS for birdsong research. It was shown that experience and standardisation are not necessary for good quality.

Can dialect studies be conducted with CS data?

In the second study (**Chapter III**) non-standardised CS, open and standardised data from academic researchers were compared. This was the to date largest birdsong study over time and space which identified the following differences: In line with other studies (Couvet et al., 2008; Theobald et al., 2015), citizen scientist produced the largest data scope in terms of the total number of songs, song types and their diversity. CS and open data had a wider temporal and geographical coverage than the shared ones due to personal and financial constraints (Riesch, Potter & Davies, 2013). The datasets did not differ in the percentages of certain parts of songs, occurrence and relative frequency of song types. They all had a temporal and spatial bias. My co-authors and I demonstrated that birdsong studies could be conducted with CS data for both species with a small (Diblíková et al., 2019; Searfoss et al., 2020) and large repertoire as well as on the population level. The dataset combination revealed the greatest temporal and geographical stability in the nightingale, and indeed birds in general. We showed that standardised and non-standardised data complement each other positively and could be used perfectly together for bioacoustic research.

How does a CS project evolve over time?

Chapter IV compared the data from three consecutive years before (2018, 2019) and during the COVID-19 outbreak (2020). This was one of the few studies (e.g. Edgar & Stuart-Smith 2009) that analysed the project development over time and which revealed best practices and lessons learned: Nightingale song recordings increased continuously in total and per citizen scientist, with a peak in the last field season. Cumulative durations and number of participants first increased and then decreased over the course of the project. The spatial distribution largely expanded steady from Berlin to nationwide without a bias. Despite the lack of standardised protocols, the recordings followed the natural occurrence of the nightingale in Germany very closely. Misidentifications were lower (10% to 1%) than in other studies (Kosmala et al., 2016), reduced with time and were mainly influenced by vocal similarities between other bird species and the nightingale. Contrary to the assumption of Dickinson and colleagues (2010), anonymity was a success factor for the data scope and, like experience, had no negative influence on the misidentifications. In contrast, the project duration had an impact on the completeness, spatial distribution and accuracy, indicating that CS projects should be set up for the long term to enable good data quality.

2. Conclusion

This doctoral thesis demonstrated that CS is an appropriate research approach to generate data with a high quality. A great potential could be shown for a bird species with a large repertoire, indicating a likewise reliable use for species with a less complex song. Anonymity, inexperience and non-standardised protocols did not lead to incompleteness, temporal and spatial bias or inaccuracy, but to valid birdsong recordings. This is in line with other studies, which showed that CS data were as valuable as those from academic researchers (Coxen et al., 2017).

The chosen low-threshold approach of the CS nightingale project found a perfect balance between quantity and quality. The lack of standardisation led to a large amount of data (Jäckel et al., 2022). Smartphone apps with an in-built pattern recognition algorithm proved also successful to generate high quality data. The widespread use of smartphones will enlarge and enable new possibilities in future birdsong studies. They allow geo-located and time stamped data to be generated and shared at any location in real time (Newman et al., 2012). In the future, explanations about the correct position of the smartphone, optimal distance to the bird as well as the number, duration and spatial distribution of the recording would be helpful. Information about the research question and potential causes of misidentifications would increase data quality even further.

This shows how important an open evaluation is so that other CS projects can learn from it. My findings and research carried out in this doctoral thesis are of importance beyond the conducted studies. The developed song type catalogue has been shared with the Freie Universität Berlin for nightingale research in a behaviour biology master course. The knowledge gained provides essential insights for other bioacoustic studies, which collect or analyse data in equal ways. Birdsong projects have been growing steadily with several hundred involved participants worldwide, which contribute to open access databases (Vellinga, & Planqué, 2015; Hecker, Garbe, & Bonn, 2018). To ensure that this mass of available data is used more often and to increase confidence in the method CS, it is important to share knowledge about quality evaluations from other projects. The thesis should serve as a basis to inspire and guide further research questions.

Resulting novel research questions

As this thesis has shown that dialect studies can be conducted on a bird species with a large repertoire, I will outline in the following possible further research questions that could be examined in the future.

Song monitoring

The findings showed that CS recordings are well suited to capture all calendar weeks and times of day (**Chapter II**) which offers the potential to investigate temporal changes over the entire breeding season. Reduced song performance of buzz songs was found later in time (Weiss, Kiefer, Kipper, 2012), but whistles and trills changes have not yet been studied. Trills, like long buzz songs (Weiss, Kiefer, Kipper, 2012), are difficult to produce in other bird species (Podos, 1996; Kirschel et al., 2009). Thus, I would assume that the performance of trills also decreases towards the end of the breeding season. This should be investigated regionally to avoid spatial influence and could be assigned to citizen scientists from Berlin as this is where most of the participants in this project came from. The task would be to record the same individual or the same population over several calendar weeks. Likewise, recordings could be generated of the same male during day and night. In my bachelor thesis (Jäckel, 2013) I was able to show that there were no whistles that were only sung during the day or at night, but it is still unknown whether this also applies to trills and buzzes.

The good spatial coverage of the CS recordings of the whole of Germany (**Chapter IV**) could also be extended beyond the national borders across several countries up to the overwintering grounds in Africa. One study (Kipper, Sellar & Barlow, 2016) found differences between nightingale songs from Africa and Berlin. There exists no documentation of potential changes in the nightingale song during migration which takes several months, whereas nightingales feed enroute. Such a study would provide insights into whether the vocal variations found by Kipper and colleagues (2016) only occur in Africa, or whether they disappear during migration or on the breeding grounds. This would require the inclusion of citizen scientists from predefined regions so that the migration route is recorded completely.

Stabile song types

The identified stable song of nightingale offers the potential for further analysis. Of particular interest is the fact that the second study types (**Chapter III**) found stability in ten common and rare song and an earlier diploma theis (Schehka, 2004) identified 22 completely different ones. Future studies should investigate how many stable song types exist and how many of them belong to common or rare ones. For this, citizen scientists from many different regions could be invited to record different nightingales at night for one hour. Nocturnal nightingale recordings contain more songs and song types (Jäckel, 2013) and less anthropogenic or (a)biotic noise (Dominoni et al., 2020). One-hour recordings offer the potential to represent the entire repertoire of a nightingale (Kipper et al., 2004). This would require the 'Naturblick' app to increase the recording duration capacity to one hour.

Longer recordings would also provide the potential to investigate if certain song types remained stable over time and space because they had certain functions. This has already been suspected for cetaceans (Rekdahl et al., 2013) and there was also first evidence for this in the nightingale (Weiss et al., 2014). Weiss and colleagues showed that some song types were always followed by many (branches) or preceded by several others (bottleneck). The authors assumed that such characteristics are essential for the vocal sequence and code for dialects, as nightingales do not repeat the same song type immediately but only sing it again after many others. In future studies, it would be interesting to investigate whether the stable

song types all have the same or different characteristics (branches or bottleneck) and if this differs between the common or rare ones.

Fine structure measurements

The found absence of dialects in nightingale song (**Chapter III**) is consistent with other studies, which suggested that they are unlikely to emerge in species with a large vocal repertoire (Goodfellow & Slater, 1986; Lachlan et al., 2016; Tchernichovski et al., 2017). An extension of the analysed parameters, such as fine-structure measurements of frequencies and durations would allow this to be investigated in more detail. Although the same song type looks visually identical, it could differ in acoustic fine scale characteristics and thereby codes for dialects, as has been shown in other species (e.g. common yellowthroat; *Geothlypis trichas*: Bolus, 2014). While durations could have been measured using the CS recordings, frequency measurements were not possible due to the use of different devices (**Chapter II**). This would require citizen scientists to be equipped in advance with selected smartphones with calibrated external microphones in different regions. Future CS studies could then investigate the minimum and maximum frequency as well as durations of the entire song type or for and between elements, syllables and phrases. Individual differences would have to be determined in advance in order to distinguish them from potential dialects.

The Nightingale thus offers the potential for further CS projects. My findings are of great importance for an increased confidence in this method and serve as a basis for future birdsong studies.

References

Ahrends, A., Rahbek, C., Bulling, M. T., Burgess, N. D., Platts, P. J., Lovett, J. C., et al. (2011). Conservation and the botanist effect. Biological Conservation. 144, 131-140.

Amrhein, V., Kunc, H.P. & Naguib, M. (2004). Seasonal patterns of singing activity vary with time of day in the nightingale. The Auk. 121, 110–117.

Beecher, M. D. & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. Trends in Ecology & Evolution. 20, 143–149.

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V. & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience. 59, 977-984.

Bonney, R., Phillips, T. B., Ballard, H. L. & Enck, J. W. (2016). Can citizen science enhance public understanding of science? Public Understanding of Science. 25, 2–16.

Bolus, R. T. (2014). Geographic variation in songs of the common yellowthroat. The Auk: Ornithological Advances. 131, 175e185.

Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., et al. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. Biological Conservation. 208, 113-120.

Catchpole, C. K., & Slater, P. J. B. (2008). Bird Song: Biological Themes and Variations. Cambridge: Cambridge University Press.

Couvet, D., Jiguet, F., Julliard, R., Levrel, H. & Teyssedre, A. (2008). Enhancing citizen contributions to biodiversity science and public policy. Interdisciplinary science reviews. 33, 95-103.

Coxen, C. L., Frey, J. K., Carleton, S. A. & Collins, D. P. (2017). Species distribution models for a migratory bird based on citizen science and satellite tracking data. Global ecology and conservation. 11, 298-311.

Danielsen, F., Enghoff, M., Poulsen, M. K., Funder, M., Jensen, P. M. & Burgess, N. D. (2021). The concept, practice, application, and results of locally based monitoring of the environment. BioScience. 71, 484-502.

Diblíková, L., Pipek, P., Petrusek, A., Svoboda, J., Bílková, J., Vermouzek,Z., et al. (2019). Detailed large-scale mapping of geographical variation of yellowhammer *Emberiza citrinella* song dialects in a citizen science project. Ibis. 161, 401–414.

Dickinson, J. L., Zuckerberg, B. & Bonter, D. N. (2010). Citizen science as an ecological research tool: challenges and benefits. Annual review of ecology, evolution, and systematics. 41, 149-172.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., et al. (2012). The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment. 10, 291–297.

Dominoni, D., Smit, J. A., Visser, M. E. & Halfwerk, W. (2020). Multisensory pollution: Artificial light at night and anthropogenic noise have interactive effects on activity patterns of great tits (*Parus major*). Environmental Pollution. 256, 113314.

Edgar, G. J. & Stuart-Smith, R. D. (2009). Ecological effects of marine protected areas on rocky reef communities—a continental-scale analysis. Marine Ecology Progress Series. 388, 51-62.

Fischer, H. A., Gerber, L. R. & Wentz, E. A. (2021). Evaluating the Fitness for Use of Citizen Science Data for Wildlife Monitoring. Frontiers in Ecology and Evolution. 705.

Flesch, E. P. & Belt, J. J. (2017). Comparing citizen science and professional data to evaluate extrapolated mountain goat distribution models. Ecosphere. 8(2), e01638.

Freitag, A., Meyer, R. & Whiteman, L. (2016). Strategies employed by citizen science programs to increase the credibility of their data. Citizen Science: Theory and Practice. 1(1).

Gil, D., & Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. Trends in Ecology & Evolution. 17, 133–141.

Glutz von Blotzheim, U. N. (1988). Handbuch der Vögel Mitteleuropas. Wiesbaden: Aula.

Goodfellow, D. & Slater, P. (1986). A model of bird song dialects. Animal Behaviour. 34, 1579–1580.

Goller, M., & Shizuka, D. (2018). Evolutionary origins of vocal mimicry in songbirds. Evolution letters. 2, 417-426.

Greenwood, J. J. D. (2007). Citizens, science and bird conservation. Journal of Ornithology. 148: 77–124.

Forrester, G., Baily, P., Conetta, D., Forrester, L., Kintzing, E. & Jarecki, L. (2015). Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs. Journal for Nature Conservation. 24, 1-9.

Haywood, B. K., Parrish, J. K. & Dolliver, J. (2016). Place-based and data-rich citizen science as a precursor for conservation action. Conservation Biology. 30(3), 476-486.

Hecker, S., Garbe, L. & Bonn, A. (2018). The European Citizen Science Landscape – a Snapshot. London: UCL Press.

Hultsch, H. (1980). Beziehungen zwischen Struktur, zeitlicher Variabilität und sozialem Einsatz des Gesangs der Nachtigall (*Luscinia megarhynchos B.*). Doctoral thesis, Frei Universität Berlin.

Hultsch, H. & Todt, D. (1981). Repertoire sharing and song-post distance in nightingales (*Luscinia Megarhynchos B.*). Behav. Ecol. Sociobiol. 8, 183–188.

Jäckel, D. (2013.) Unterschiede im Gesang der Nachtigall (*Luscinia megarhynchos*) in der Nacht und am Morgen. Bachelor thesis, Frei Universität Berlin.

Jäckel, D., Mortega, K. G., Brockmeyer, U., Lehmann, G. U. & Voigt-Heucke, S. L. (2022). Unravelling the Stability of Nightingale Song Over Time and Space Using Open, Citizen Science and Shared Data. Frontiers in Ecology and Evolution, 10, 778610.

Kiefer, S., Spiess, A., Kipper, S., Mundry, R., Sommer, C., Hultsch, H., et al. (2006). First-Year common nightingales (*Luscinia Megarhynchos*) have smaller songtype repertoire sizes than older males. Ethology. 112, 1217–1224.

Kiefer, S., Scharff, C., Hultsch, H. & Kipper, S. (2014). Learn it now, sing it later? field and laboratory studies on song repertoire acquisition and song use in nightingales. Naturwissenschaften. 101, 955–963.

Hultsch, H., Kipper, S., Mundry, R. & Todt, D. (2004). Long-term persistence of song performance rules in nightingales (*Luscinia megarhynchos*): a longitudinal field study on repertoire size and composition. Behaviour. 141, 371-390.

Kipper, S., Sellar, P. & Barlow, C. R. (2017). A comparison of the diurnal song of the Common Nightingale (*Luscinia megarhynchos*) between the non-breeding season in The Gambia, West Africa and the breeding season in Europe. Journal of Ornithology. 158, 223-231.

Kirschel, A. N., Blumstein, D. T., Cohen, R. E., Buermann, W., Smith, T. B. & Slabbekoorn, H. (2009). Birdsong tuned to the environment: green hylia song varies with elevation, tree cover, and noise. Behavioral Ecology. 20, 1089–1095.

Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. (2016). Assessing data quality in citizen science. Frontiers in Ecology and the Environment. 14:551-560.

Lachlan, R. F., van Heijningen, C. A. A., ter Haar, S. M. & ten Cate, C. (2016). Zebra finch song phonology and syntactical structure across populations and continents-a computational comparison. Frontiers in Psychology. 7:980.

Landgraf, C., Weiss, M., & Kipper, S. (2015). Multiple song features are related to paternal effort in common nightingales. BMC Evolutionary Biology. 15:115.

Lasseck M (2018). Audio-based Bird Species Identification with Deep Convolutional Neural Networks. In: CEUR Workshop Proceedings.

Lasseck M (2019). Bird Species Identification in Soundscapes. In: CEUR Workshop Proceedings.

Lewandowski, E., & Specht, H. (2015). Influence of volunteer and project characteristics on data quality of biological surveys. Conservation Biology. 29, 713-723.

Miller-Rushing, A., Primack, R. & Bonney, R. (2012). The history of public participation in ecological research. Frontiers in Ecology and the Environment. 10, 285-290.

Moyer-Horner, L., Smith, M. M. & Belt, J. (2012). Citizen science and observer variability during American pika surveys. The Journal of Wildlife Management. 76, 1472-1479.

Munzinger, S. T. E. F. A. N. (2015). Citizen Science: Qualitätssicherung durch Motivation. Entomologie heute. 27, 171-176.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S. & Crowston, K. (2012). The future of citizen science: emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment. 10, 298-304.

Paulos, E. (2009). Designing for Doubt Citizen Science and the Challenge of Change. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.187.5824.

Palacin-Silva, M.V. (2018). Understanding civic participation in environmental sensing: a values driven approach. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–4.

Phillips, T. B., Parker, A., Bowser, A. & Haklay, M. (2021). Publicly Generated Data: The Role of Citizen Science for Knowledge Production, Action, and Public Engagement. In Closing the Knowledge-Implementation Gap in Conservation Science. 83-107 Springer, Cham.

Planillo, A., Fiechter, L., Sturm, U., Voigt-Heucke, S. & Kramer-Schadt, S. (2021). Citizen science data for urban planning: Comparing different sampling schemes for modelling urban bird distribution. Landscape and Urban Planning. 211, 104098.

Podos, J. (1996). Motor constraints on vocal development in a songbird. Animal Behaviour. 51, 1061–1070.

Poulsen, H. (1958). The Calls of the Chaffinch (*Fringilla coelebs L.)* in Denmark. Dan. Ornithol. Foren. Tidsskr. 52, 89–105.

Ratcliffe, L. & Otter, K. (1996). Sex differences in song recognition. In: Ecology and Evolution of Acoustic Communication in Birds (Kroodsma, D. E. & Miller, E. H., eds). Comstock Publishing Assoc., Ithaca, NY, pp. 339—355.

Rekdahl, M. L., Dunlop, R. A., Noad, M. J. & Goldizen, A. W. (2013). Temporal stability and change in the social call repertoire of migrating humpback whales. The Journal of the Acoustical Society of America. 133, 1785–1795.

Riesch, H., Potter, C. & Davies, L. (2013). Combining citizen science and public engagement: the Open AirLaboratories Programme. Journal of Science Communication. 12, A03.

Roth, T. & Amrhein, V. (2010). Estimating individual survival using territory occupancy data on unmarked animals. Journal of Applied Ecology. 47, 386-392.

Roy, H. E., Pocock, M. J., Preston, C. D., Roy, D. B., Savage, J., Tweddle, J. C. & Robinson, L. D. (2012). Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework.

Scott, W. A. & Hallam, C. J. (2003). Assessing species misidentification rates through quality assurance of vegetation monitoring. Plant Ecology. 165, 101-115.

Searfoss, A. M., Liu, W. C. & Creanza, N. (2020). Geographically well distributed citizen science data reveals range-wide variation in the chipping sparrow's simple song. Animal Behaviour. 161, 63–76.

Schehka, S. (2004). Konstanz im Gesang der Nachtigall (*Luscinia megarhynchos*) unter Berücksichtigung geographischer und zeitlicher Aspekte. Diploma thesis, Frei Universität Berlin.

Silvertown, J. (2009). A new dawn for citizen science. Trends in ecology & evolution. 24, 467-471.

Smith, W. H. & Hamed, M. K. (2020). Trade-offs in moving citizen-based anuran call surveys online during the SARS-CoV-2 pandemic: Lessons from rural Appalachia, USA. Ecology and Evolution. 10, 12465-12471.

Sprau, P. & Mundry, R. (2010). Song type sharing in common nightingales, *Luscinia megarhynchos*, and its implications for cultural evolution. Animal Behaviour. 80, 427–434.

Stehle, M., Lasseck, M., Khorramshahi, O. & Sturm, U. (2020). Evaluation of acoustic pattern recognition of nightingale (*Luscinia megarhynchos*) recordings by citizens. Research Ideas and Outcomes. 6: e50233.

Tchernichovski, O., Feher, O., Fimiarz, D. & Conley, D. (2017). How social learning adds up to a culture: from birdsong to human public opinion. Journal of Experimental Biology. 220, 124–132.

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., et al. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation. 181, 236-244.

van der Wal, R., Sharma, N., Mellish, C., Robinson, A. & Siddharthan, A. (2016). The role of automated feedback in training and retaining biological recorders for citizen science. Conservation Biology. 30, 550-561.

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., et al. (2021). The science of citizen science (p. 529). Springer Nature.

Vellinga, W. P. & Planqué, R. (2015). "The Xeno-canto collection and its relation to sound recognition and classification," in Proceedings of the 2015 CLEF, (Toulouse: CLEF).

Weiss, M., Kiefer, S. & Kipper, S. (2012). Buzzwords in females' ears? The use of buzz songs in the communication of nightingales (*Luscinia megarhynchos*).

Weiss, M. (2012). From syntax to details: Organisational principles and information encoding in the singing of nightingales (*Luscinia megarhynchos*). Doctoral thesis, Frei Universität Berlin.

Acknowledgements

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die Dissertation selbständig und nur unter Zuhilfenahme der aufgeführten Hilfsmittel und Literaturquellen angefertigt habe.

Declaration:
I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids.

Berlin, den 23.05.2022