Studying individual vocal communication in group-living songbirds

Dissertation submitted for the degree of Doctor of Natural Sciences

Presented by

Lisa Franziska Gill

at the

Universität Konstanz

Faculty of Sciences
Department of Biology

Constance, 12/12/2016

First supervisor: Prof. Dr. Martin Wikelski
Second supervisor: Prof. Dr. Wolfgang Goymann
Third supervisor: Prof. Dr. Iain Couzin
Studying individual vocal communication in group-living songbirds

Lisa F. Gill
# Contents

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>Summary</td>
<td>xi</td>
</tr>
<tr>
<td>Zusammenfassung</td>
<td>xv</td>
</tr>
</tbody>
</table>

## PART I – INTRODUCTION

1. **Section 1. Vocal communication**
   1.1. Sources of variation 3
   1.2. Functions 3
   1.3. Songbird vocalisations 4
      1.3.1. Song 6
      1.3.2. The potential of calls 7

2. **Section 2. The challenge: individual recordings in context**
   2.1. Tools 7
   2.2. Species 12
      2.2.1. Zebra finches 14
      2.2.2. Jackdaws 14
   2.3. Contexts 15

3. **Section 3. Chapter overview** 17
### PART II – CHAPTERS

#### CHAPTER 1:

**A minimum-impact, flexible tool to study vocal communication of small animals with precise individual-level resolution**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Abstract</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Introduction</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>Materials and Methods</td>
<td>31</td>
</tr>
<tr>
<td>6.1</td>
<td>Backpacks</td>
<td>31</td>
</tr>
<tr>
<td>6.2</td>
<td>Workflow</td>
<td>31</td>
</tr>
<tr>
<td>6.3</td>
<td>Animals</td>
<td>31</td>
</tr>
<tr>
<td>6.4</td>
<td>Experiments</td>
<td>32</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Effects of backpack attachment and battery exchange</td>
<td>32</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Selective sound recordings</td>
<td>34</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Synchronous sound recordings</td>
<td>36</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Battery life</td>
<td>37</td>
</tr>
<tr>
<td>6.4.5</td>
<td>Transmission distance</td>
<td>37</td>
</tr>
<tr>
<td>6.5</td>
<td>Analyses</td>
<td>37</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Sound analyses</td>
<td>37</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Vocal interaction analyses</td>
<td>38</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Statistical analyses</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>Results</td>
<td>41</td>
</tr>
<tr>
<td>7.1</td>
<td>Effects of backpack attachment and battery exchange</td>
<td>41</td>
</tr>
<tr>
<td>7.2</td>
<td>Selective sound recordings in normal and noisy conditions</td>
<td>41</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Focal versus non-focal vocalisations</td>
<td>42</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Backpack versus external</td>
<td>42</td>
</tr>
<tr>
<td>7.3</td>
<td>Synchronous recordings, battery life and transmission distance</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>Discussion</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td>Ethics statement and Acknowledgments</td>
<td>49</td>
</tr>
<tr>
<td>10</td>
<td>Supporting Materials</td>
<td>51</td>
</tr>
</tbody>
</table>
### CHAPTER 2:

**Patterns of call communication between group-housed zebra finches**

*change during the breeding cycle*

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Abstract</td>
<td>57</td>
</tr>
<tr>
<td>12</td>
<td>Introduction</td>
<td>59</td>
</tr>
<tr>
<td>13</td>
<td>Results</td>
<td>61</td>
</tr>
<tr>
<td>13.1</td>
<td>Pair formation and reproductive stages</td>
<td>62</td>
</tr>
<tr>
<td>13.2</td>
<td>Vocalisation types and reproductive stages</td>
<td>62</td>
</tr>
<tr>
<td>13.3</td>
<td>Vocal interactions, reproductive stages and successful egg-laying</td>
<td>64</td>
</tr>
<tr>
<td>14</td>
<td>Discussion</td>
<td>71</td>
</tr>
<tr>
<td>14.1</td>
<td>Calling behaviour in social environments: new approach, new results</td>
<td>71</td>
</tr>
<tr>
<td>14.2</td>
<td>Opportunistic breeding and vocalisations</td>
<td>72</td>
</tr>
<tr>
<td>15</td>
<td>Materials and Methods</td>
<td>75</td>
</tr>
<tr>
<td>15.1</td>
<td>Study design, housing and nest material</td>
<td>75</td>
</tr>
<tr>
<td>15.2</td>
<td>Sampling</td>
<td>76</td>
</tr>
<tr>
<td>15.2.1</td>
<td>Behavioural observations, nest checks and blood sampling</td>
<td>76</td>
</tr>
<tr>
<td>15.2.2</td>
<td>Vocalisation recordings</td>
<td>77</td>
</tr>
<tr>
<td>15.3</td>
<td>Analyses</td>
<td>79</td>
</tr>
<tr>
<td>15.3.1</td>
<td>Reproductive stages</td>
<td>79</td>
</tr>
<tr>
<td>15.3.2</td>
<td>Sound analyses</td>
<td>80</td>
</tr>
<tr>
<td>15.3.3</td>
<td>Vocal interaction analyses</td>
<td>81</td>
</tr>
<tr>
<td>15.3.4</td>
<td>Statistical analyses</td>
<td>82</td>
</tr>
<tr>
<td>15.3.4.1</td>
<td>Hormones</td>
<td>83</td>
</tr>
<tr>
<td>15.3.4.2</td>
<td>Vocalisation types and breeding activity</td>
<td>83</td>
</tr>
<tr>
<td>15.3.4.3</td>
<td>Within-pair interactions, breeding activity and breeding success</td>
<td>84</td>
</tr>
<tr>
<td>16</td>
<td>Ethical note and Acknowledgments</td>
<td>85</td>
</tr>
<tr>
<td>17</td>
<td>Supplementary Materials</td>
<td>87</td>
</tr>
<tr>
<td>17.1</td>
<td>Appendix 1: Hormones in relation to reproductive stages</td>
<td>87</td>
</tr>
<tr>
<td>17.2</td>
<td>Appendix 2: Call types and more detailed breeding stages</td>
<td>88</td>
</tr>
</tbody>
</table>
CHAPTER 3:
On-bird sound recordings: automatic acoustic recognition of activities and contexts 97

Section 18. Abstract 99

Section 19. Introduction 101
  19.1. Automatic Acoustic Recognition 102
  19.2. Aims 103

Section 20. Data Collection 105
  20.1. Birds and microphone backpacks 105
  20.2. Video-validation of sounds 106
  20.3. Annotation of field data 106

Section 21. Automatic recognition 109
  21.2. Event Detection System 111
  21.3. Postprocessing 113
  21.4. Handling Missing Data 114

Section 22. Evaluation 117
  22.1. Metrics 117
  22.2. Evaluation Schemes 118
  22.3. Results 119

Section 23. Discussion 129

Section 24. Conclusions 133

Section 25. Acknowledgments and Funding information 135

CHAPTER 4:
Bio-logging reveals extra-pair copulations and male copulatory calls in "strictly monogamous" wild jackdaws 137

Section 26. Abstract 139

Section 27. Introduction 141

Section 28. Results 143

Section 29. Discussion 147

Section 30. Methods 149
<table>
<thead>
<tr>
<th>Section Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.1. Loggers and settings</td>
<td>149</td>
</tr>
<tr>
<td>30.2. Nest-boxes and cameras</td>
<td>150</td>
</tr>
<tr>
<td>30.3. Logger application</td>
<td>150</td>
</tr>
<tr>
<td>30.4. Sound analyses</td>
<td>150</td>
</tr>
<tr>
<td>30.4.1. Identifying the signalling individual</td>
<td>151</td>
</tr>
<tr>
<td>30.4.2. Context recognition</td>
<td>151</td>
</tr>
<tr>
<td>Section 31. Acknowledgments, Competing financial interests, Author contributions</td>
<td>153</td>
</tr>
<tr>
<td>PART III – GENERAL DISCUSSION</td>
<td>155</td>
</tr>
<tr>
<td>Section 32. Individual sound recordings from on-animal microphones</td>
<td>157</td>
</tr>
<tr>
<td>32.1. Transmitters versus loggers</td>
<td>157</td>
</tr>
<tr>
<td>32.2. Drawbacks</td>
<td>158</td>
</tr>
<tr>
<td>32.3. Advantage: context</td>
<td>159</td>
</tr>
<tr>
<td>32.3.1. Vocal behaviour in context</td>
<td>159</td>
</tr>
<tr>
<td>32.3.2. Acoustic context recognition</td>
<td>160</td>
</tr>
<tr>
<td>Section 33. Calling behaviour of group-living songbirds in naturalistic contexts</td>
<td>163</td>
</tr>
<tr>
<td>33.1. Building a jackdaw call repertoire?</td>
<td>163</td>
</tr>
<tr>
<td>33.2. Jackdaw copulatory calls and extra-pair behaviour</td>
<td>163</td>
</tr>
<tr>
<td>33.3. Defining calling and responding individuals in vocal communication</td>
<td>166</td>
</tr>
<tr>
<td>33.4. Call types, timing and vocal interactions of zebra finch pairs in groups</td>
<td>167</td>
</tr>
<tr>
<td>Section 34. Overall conclusions</td>
<td>171</td>
</tr>
<tr>
<td>Author contributions and Acknowledgments</td>
<td>1</td>
</tr>
<tr>
<td>Bibliography</td>
<td>V</td>
</tr>
<tr>
<td>Appendix</td>
<td>XXIII</td>
</tr>
<tr>
<td>Appendix 1:</td>
<td></td>
</tr>
<tr>
<td>Detailed temporal structure of communication networks in groups of songbirds</td>
<td></td>
</tr>
<tr>
<td>Appendix 2:</td>
<td></td>
</tr>
<tr>
<td>Individual identity in songbirds: signal representations and metric learning for locating the information in complex corvid calls</td>
<td></td>
</tr>
</tbody>
</table>
List of Figures

1. Variation of vocal signals 5
2. Vocal communication inside an environment 7
3. Synchronous audio recordings 32
4. Transmitter and leg loop harness 33
5. Calling behaviour (a) and movement (b) before and after backpack attachment 35
6. Spectral analysis of the focal and non-focal recording of a distance call 36
7. Vocal communication in noise 38
8. Amplitudes of different sound events recorded by backpack 52
9. Interaction details from Fig. 7 (PSTH) 53
10. Position of on-bird microphones on freely behaving zebra finches 61
11. Group reproductive stages change over time 63
12. Call types used in our study 63
13. Female and male call-type usage at different nest stages 65
14. Vocal interactions within groups across reproductive stages 68
15. Call-type combinations, nest stages and successful egg-laying 69
16. Timeline of trials 76
17. Synchronous external and on-bird recordings 79
18. Group housing 92
19. Steroid hormone concentrations at baseline levels and nest stages 93
20. Changes in call repertoire at more detailed breeding stages 94
21. Increasing specificity of within-pair vocal interactions 95
22. Overview of the processing workflows used for automatic recognition 110
List of Tables

1  Comparison of audio transmitters and loggers          13
2  Estimated parameters for calling activity            51
3  Estimated parameters for locomotor activity         53
4  Overview and short description of different agonistic, affiliative or sexual, and neutral behaviours  90
5  Overview and short description of breeding stages and nest stages  91
6  Labelling scheme for the actions/contexts in our recordings 107
7  Summary statistics on male copulatory vocalisations for all, within-pair and extra-pair copulations 144
Summary
Studying animal vocal communication (sending and receiving acoustic signals produced in the vocal tract) in groups is a challenge because the presence of multiple signallers may impede clean, non-overlapping recordings of individuals, and makes it difficult to identify exactly which animal produced which sound. Further, defining senders and receivers in vocal communication can be difficult even in simple vocal exchanges, which is even more challenging when a group of individuals is vocalising close by. But especially in group-living animals, vocal communication often plays a crucial role for survival and reproduction: vocal signals are used to facilitate contact and coordination with an entire group, for example to optimise foraging and anti-predator behaviour, but also to find and communicate with important members inside the group, for example with offspring or a mate. Songbirds may use two different types of signals to communicate vocally: songs and calls. Songs are intricate, learned vocal patterns, often involved in mate attraction and territory defence. Calls are simpler in structure, but more variable in terms of the contexts in which they occur, making them more flexible for the use in communicative vocal interactions between individuals – but also making them difficult to study.

To investigate group vocal communication from a functional perspective, it would be ideal to disentangle the roles, costs and benefits of all individuals involved in vocal interactions, while providing complex contexts in which animals may exhibit their normal behaviours. One approach to this is to place a microphone directly on the sound-producing animal, which opens up the possibility to study different, even very “private”, communication channels of individuals inside groups. Due to technical advances (miniaturisation), this approach is becoming increasingly available, even for small animals, but so far, studies are rare that investigate such techniques, their potential effects, pros and cons, to make use of their full potential. The main aim of this dissertation was thus to investigate different methods for studying vocal communication of group-living songbirds in naturalistic contexts, in order to answer specific research questions.

I studied the effects of such methods on small songbirds under controlled conditions, and investigated their potential for communication studies (chapter 1). I also got involved in establishing a basis for acoustic context recognition and information yield of on-board microphone recordings (chapter 3), and used alternative approaches to define senders and receivers of vocal interactions (chapter 2, Appendix 1). Applying these methods, I was able to disentangle “who is who” in natural, i.e. spontaneously occurring, vocal interactions in group-housed zebra finches (*Taeniopygia guttata*) in the lab, and in wild, group-living jackdaws (*Corvus monedula*) in the field. In this way, I studied the importance of zebra finch call
communication of individuals and pairs inside groups over the course of changing life-history stages (chapter 2), and found that call types not only changed in terms of individual-level usage, but also in vocal interactions between individuals. I also found evidence that calling interactions between mates may be related to pair breeding success, thus highlighting the importance of calls in communication systems. In jackdaws, by revealing male copulation calls and extra-pair copulation behaviour (chapter 4), I was able to describe unknown aspects of vocal behaviour as a basis for future investigations, and challenge previous claims of absolute monogamy in this corvid species.
Zusammenfassung


Dazu untersuchte ich den Einfluss von “Mikrofonrucksäcken” auf kleine Singvögel unter standardisierten und kontrollierten Bedingungen, und diskutierte ihre Vor- und Nachteile im Hinblick auf Vokalisationsstudien (Kapitel 1). Ich beteiligte mich
auch daran, eine Basis für akustische Kontexterkennung und für einen sinnvollen Informationsgewinn von Mikrofonaufnahmen am Tier zu schaffen (Kapitel 3), sowie an alternativen Herangehensweisen zur Definition von Sendern und Empfängern vokaler Interaktionen (Kapitel 2 und Appendix 1). Anhand dieser Methoden konnte ich also herausfinden, "wer was gesagt" hatte in spontan auftretenden vokalen Interaktionen bei in Gruppen gehaltenen Zebra-finken \((Taeniopygia guttata)\) im Labor und bei in Gruppen lebenden Dohlen \((Corvus monedula)\) in freier Wildbahn. So untersuchte ich bei Zebra-finken während sich ändernder Paar- und Brutstadien die Rolle der Ruf-basierten Kommunikation von Individuen und Paaren innerhalb der Gruppe (Kapitel 2), und fand heraus, dass sich das Rufverhalten nicht nur auf der Ebene der Individuen veränderte, sondern auch in vokalen Interaktionen zwischen Tieren. Zusätzlich fand ich erste Hinweise darauf, dass diese Interaktionen einen Zusammenhang mit der erfolgreichen Eiablage eines Paares haben könnten, was die Rolle von Rufen in Kommunikationssystemen hervorhebt. Bei Dohlen fand ich heraus, dass die Männchen für spezifische Kopulationsrufe verantwortlich waren, und dass diese Rabenvögel entgegen vorheriger und viel-zitierter Literatur nicht absolut monogam sind (Kapitel 4). Diese neuen Ergebnisse stellen für zukünftige genetische und verhaltensorientierte Studien eine wichtige Basis dar.
PART I – INTRODUCTION
SECTION 1

Vocal communication

Cells, plants, fungi, animals – all such living “entities” have the potential to engage in communication, which means they can transmit and receive information from other entities. In nature, communicating entities may be of the same kind (Young, 2009; Heil and Karban, 2010; Bradbury and Vehrencamp, 2011; Rappel, 2016) or very different organisms (e.g. plants and fungi: Christensen and Kolomiets (2011); plants and animals: Schaefer et al. (2004)). Like these units themselves, the signals used for communication couldn’t be more diverse – ranging from electric current to simple proteins or odours, the seductive waving of a colourful limb, up to rhythmic drumming and spoken language – and may even have overlapping modalities (multimodal signalling). In acoustic animal communication, the pieces of information being transmitted and received are sounds. Such communicative sounds have highly diverse acoustic properties because they can serve different functions, occur in different media or habitats, and because they can be produced in very different ways: some result from a manipulation of external objects (e.g. hollow tree trunks) or external body parts (as in stridulation or wing-clapping), while others are produced internally, in specialised organs. In fact, many animals communicate via vocalisations, meaning sounds that are produced in the vocal tract. These sounds, too, are highly diverse both between and within communication systems, with variation occurring at different levels.

1.1. Sources of variation

Most obviously, vocalisations show variation in terms of their acoustic properties (e.g. duration, pitch, frequency modulation) which may result in different vocalisation types or in more subtle modifications within such vocal classes. But variation can also be found in the usage of vocalisation types, the timing of delivery both within (e.g. vocalisation rates) and between individuals (e.g. response latencies), or in a combination with other modalities (e.g. courtship dance). Such differences can be explained by stable factors such as size, sex or species. Some aspects are also heritable, and vocalisations often serve as species-specific signals, and may be linked with speciation (Slabbekoorn and Smith, 2002; Dereugnaucourt and Guyomarch, 2003; Lachlan and Servedio
But variation can also be driven by more immediate and variable internal or external influences, such as hunger state (Reers and Jacot 2011), or the social (Evans and Marler 1994; Townsend and Zuberbühler 2009; Menardy et al. 2014; Vignal et al. 2004) or acoustic properties of an environment (Brumm and Slabbekoorn 2005). For example in noise, animals make themselves heard by increasing the amplitude and/or changing the fundamental frequency of their vocalisations (Brumm and Todt 2002; Brumm 2006a; Kirschel et al. 2009; Slabbekoorn 2013). To ensure transmission of a signal, animals may also change temporal features of their vocalisations, e.g. by adjusting the timing, rate or duration of vocalisations (Brumm and Slabbekoorn 2005).

In my dissertation, I focus most on the variation lying in vocalisation types, their usage on an individual and on a between-individual level, as well as their relative timing in vocal interactions, with a specific focus on songbird calls (see below). Because vocal signalling involves a combination of these different sources of variation, it has the potential to carry high levels of information content. But as with other biological traits, analysing the variation of vocal parameters by themselves as isolated variables is not enough to gain a thorough understanding of the underlying mechanisms and functions. Instead, it is necessary to investigate specific features of vocal signals, set in an appropriate context. In the following, I will explain in more detail why this is a goal necessary to pursue, why this was and still can be a difficult task, and how this challenge can be overcome.

1.2. Functions

It has been demonstrated that animal vocalisations can convey an emotional or urgency state, up to a concrete semantic meaning by exhibiting differences in the acoustic structure on a gradual scale or as discrete acoustic classes (Bradbury and Vehrencamp 2011). A classic example for the importance of discrete semantic signals is the vervet monkey alarm system in which acoustically distinct vocalisation types given by conspecífics towards different predator threats elicit adaptive but conflicting behavioural response strategies (Seyfarth et al. 1980). But, as mentioned above, vocalisations may carry additional information, for example on species, age class or individual identity. The ability to recognise and being recognised by other individuals is a prerequisite for behavioural adaptations facilitating future encounters (Tibbetts and Dale 2007). Therefore, it underlies many animal societies (Beecher 1989; Ward and Webster 2016), and is also key to preferential investment in specific, important individuals (e.g. partner or offspring: Jouventin et al. 1999a,b). Such features may thus allow receivers to optimise their response to a given signal even further (Cheney and Seyfarth 1988).
If the production and maintenance (i.e., factors involved in transmitting, receiving, and processing) of such signals are less costly than a wrong decision—for example if a wrong flight response is likely to result in imminent death or injury—it is plausible that such a vocal communication system could become evolutionarily adaptive.

Further, vocalisations are often strongly linked with reproduction, the other important agent in evolution, because they can help finding potential mates in dense vegetation or over distance, choosing a suitable mate (Bradbury and Vehrencamp 2011), and may even be involved in driving physiological changes necessary for successful reproduction (Lehrman and Friedman 1969; Cohen and Cheng 1979; Cheng 2003).

Figure 1. Vocal signals can vary on different levels, each of which may carry information and thus in turn may influence potential receivers. For example, the same call type may have different meanings, depending on its acoustic modification, production rate, or the sender’s identity.

In the abovementioned example (Fig. 1), communication was unidirectional, and the sender and receiver of the signal easily identified (see Fig. 2A). But of course, reality is usually more complicated (see Fig. 2B). First, communication often goes both ways which means that animals may dynamically switch their roles as senders and receivers. Secondly, vocalisations are waves of sound that propagate through a given medium, which means the flow of information is prone to degradation and
is not necessarily confined to only one specific receiver (McGregor and Peake 2000; Bradbury and Vehrencamp 2011; Searcy and Yasukawa 2016). To obtain a more realistic picture of vocal communication, it is thus not only important to investigate vocal signals per se, but also the vocal interactions that they occur in. For instance, the timing and type of vocalisations involved in interactions have been shown to be socially important, not only to the animals directly involved (e.g. Brumm and Slater 2007; Searcy and Beecher 2009), but also by providing information for third-party individuals (Mennill 2002). This becomes even more interesting in group-living animals that rely heavily on vocal communication. In such species, vocalisations are often used to maintain contact and to coordinate within an entire group (Fichtel and Manser 2010), but they may just as well be used for locating and communicating with specific individuals inside such groups (Jouventin et al. 1999a; Balsby et al. 2012).

In sum, animal vocalisations are complex, multidimensional and to some extent heritable traits that can convey important information about their senders or environments, and have the potential to influence receivers’ responses. Given their diversity and importance for individual survival and reproduction, it is not surprising that vocalisations have become useful models for tackling mechanistic, behavioural, ecological and evolutionary questions in different fields of research.

1.3. Songbird vocalisations

Songbirds have become classic model species for studies involving vocal behaviour. They occupy a variety of ecological niches, encompassing different habitats, mating systems, feeding strategies etc. But above all, they produce an array of learned and unlearned vocalisations that, like other biological traits, can be quantified and compared between such niches, lineages or even experimental treatments (Catchpole and Slater 1995). The vocal repertoire of songbirds is usually separated into the eponymous “songs” and into “calls”. Although broad-scale differences might seem straightforward, such classifications are often challenging, especially at a finer scale, and depend to a large extent on the observed species, as well as on the choice of acoustic or context-driven definitions. In this dissertation, I do not attempt to engage in detail in this debate, but rather seek a pragmatic definition to delineate the major differences for the outline of my studies which focus mostly on calls (for reasons I will explain below). First, calls are usually defined as short, single vocal elements that may convey information in themselves (or in combination with other signals or contexts) while songs usually consist of multiple elements, often presented in rhythmic sequences, and repeated in a more or less standardised manner. Secondly, calls and song can be
Figure 2. Vocal communication inside an environment
A) A simplistic model involves a signal being transmitted by a sender through an environment, and picked up by a receiver.
B) A slightly more complex model visualises how quickly it becomes difficult to define senders and receivers, e.g. when the receiver responds vocally to the signal, vocal self-stimulation is involved, signals degrade or the information is received by a bystander, at a cost to the sender (eavesdropping).

1.3.1. Song. The song of birds has fascinated humans for centuries, becoming an inspiration for poetry and music, as well as the subject of different areas of research, including neuroanatomy, ethology, urbanisation studies, or human
linguistics. In the scientific literature, song has received the largest share of attention compared to calls, with most vocalisation studies revolving around a small number of songbird species with male-exclusive song (Riebel et al., 2005; Odom et al., 2014). In part, this can be explained by songs being the more conspicuous – and generally considered more pleasing – acoustic signals for human observers. Next, for a long time, this area of research has focused most on temperate-zone species, in which female song is less common than in the tropics (Slater and Mann, 2004). Further, in many species, females have higher investment during reproduction, and therefore choose their mate from competing males (Andersson and Iwasa, 1996; Wedell et al., 2006). Because songs are intricate, learned vocal patterns whose correct performance requires high amounts of sensory-motor training and control, it has been suggested that they can honestly reflect different aspects of male quality, similar to more physically graspable traits such as elaborate plumage or nest structures (Kroodsma and Byers, 1991; Catchpole and Slater, 1995; Bradbury and Vehrencamp, 2011).

In many species, males provide more than genetic material. Many bird species are at least socially monogamous, and often females and males cooperate in raising their brood (Cockburn, 2006). This means that although not all offspring might be genetically sired by both parents, both members of the social pair show high levels of investment (the claim of strict monogamy in birds has been disproven during the rise of genetic studies, reviewed by Petrie and Kempenaers, 1998; Griffith et al., 2002; Westneat and Stewart, 2003). Also, pair members need to cooperate and to synchronise in terms of physiology (reproductive state) (Hirschenhauser et al., 2008) as well as behaviour (nest visits) (e.g. Sanchez-Macouzet et al., 2014; Mariette and Griffith, 2015). In this case, the importance of male genetic quality alone recedes into the background, giving way to more pair-related aspects. Thus, if life-history stages associated with a pair's successful reproduction, such as mate choice, pair formation, pair maintenance or breeding, were to be facilitated by vocal communication between partners, both the female and male should be able to contribute vocal signals to form an interactive social behaviour, as for example in male-female song duets (e.g. Thorpe et al., 1972; Wickler and Seibt, 1980; Baldassarre et al., 2016).

### 1.3.2. The potential of calls.

As mentioned above, songbird calls are produced by both sexes. They occur in a variety of contexts, and the majority of them are presented in a less “formal” (Marler, 2004) manner than songs (Catchpole and Slater, 1995; Marler, 2004). They can be loud and conspicuous, like distance, alarm or mob calls, but may also be quiet and highly directional (Marler, 2004; Marler and Evans, 2008). By using quiet vocalisations, animals reduce the range...
of potential receivers, which has been associated either with aggressive (Dabelsteen et al., 1998; Templeton et al., 2012) or mating-related contexts (Gorissen and Eens, 2004; Elie et al., 2010; Ter Maat et al., 2014). There is even evidence suggesting pair-related interactive usage of calls (Blaich et al., 1996a,b; Gorissen and Eens, 2004; Ter Maat et al., 2014). Such “private communication channels” could thus provide insight into further aspects of vocal interactions, and the potential of bird calls for studying basic underlying mechanisms and evolutionary aspects of communication has been theoretically discussed (Marler, 2004; Marler and Evans, 2008).

So far, surprisingly few studies addressed in detail the interactive calling behaviour of songbirds. The bias of 20th century vocalisation research towards male song aside, there is also a very practical explanation for this: recording calls at the individual level without disrupting an animal's normal behaviour can be technically challenging because calls are often short, inconspicuous, and difficult to assign to specific individuals, especially if produced in naturalistic contexts, e.g. in close proximity to other sound-producing individuals.
SECTION 2

The challenge: individual recordings in context

In an ideal setting, we could observe, record and manipulate the behaviour of animals on extremely fine-tuned scales, under the most naturalistic conditions, to thoroughly examine underlying mechanisms and functions. In reality, studying how and why vocal signals are produced is not always an easy task. First, animals should be provided with an appropriate context in which they may exhibit the behaviour in question. However, obtaining adequate animal vocalisation recordings in undisturbed and naturalistic settings can be very challenging: vocalisations may be degraded or lost if recorded at a distance, and may overlap with abiotic (e.g. streams), biotic (conspecifics or heterospecifics) or anthropogenic background sounds. The impressive sounds of a dawn chorus or of a duetting pair of birds can be fascinating traits in themselves, but to investigate the role of the individuals taking part in vocal communication, as well as the importance of vocal communication for these individuals, it is crucial to obtain acoustic recordings that allow identifying the vocal signalers involved. To do so, some field studies focused on the behaviour of solitary birds vocalising at a distance. However, studies focusing on more social animals often had to choose: some obtained acoustic recordings of single individuals, but housed animals in isolation (Blaich et al., 1996a; Vignal et al., 2004) or in otherwise strongly simplified environments (Anisimov et al., 2014). Other studies were carried out in more natural settings or even in the wild, but could not easily distinguish the vocalising individuals (review by Thorpe et al., 1972; Elie et al., 2010), or only managed to assign the vocalisations of single individuals at a time (Mann et al., 2006).

Disentangling different forms of vocal communication in group-living animals to investigate different aspects of songbird calls was thus the main aim of this dissertation. A prerequisite for this was the evaluation and application of adequate approaches that allowed recording and analysing individual vocalisations (calls) of animals behaving freely inside groups, as well as relating them with the contexts in which they were produced. In the following paragraphs, I will describe the different tools and systems I eventually used to approach this challenge.
2.1. Tools

To investigate vocal behaviour of individual animals behaving freely in the presence of vocalising conspecifics or in other difficult contexts, a couple of technical approaches have become useful. In some settings, vocalisation partners produce loud and clearly different vocalisations, thus making it possible to tease apart “who is saying what”, by observing and recording the focal individuals (e.g. Brumm [2006b]). Microphone arrays can help disentangling more difficult cases by triangulation of sound (Blumstein et al. [2011]). However, if animals vocalise without conspicuous body movements, very softly and in very close proximity (Elie et al., 2010), or outside of the range of an observer (Couchoux et al. [2015]), external microphones may no longer record the vocalisations at all, or may no longer allow a reliable distinction between the involved signalers. In such cases, it has proven useful to place a microphone directly on the animal (Cvikel et al. [2014]; Ter Maat et al. [2014]; Couchoux et al. [2015]; Rekdahl et al. [2015]; Thiebault et al. [2016]). Two different methodological approaches are available for this (telemetry and bio-logging), both of which carry specific benefits and drawbacks which I summarise in Table 1. The largest difference between the two methods stems from the recording procedure itself. While loggers store data locally, transmitters send the acoustic signals to be recorded elsewhere. This seemingly small difference results in a number of implications, such as reducing the weight and increasing recording times for transmitters, due to lower battery and storage demands, as well as allowing perfectly aligned recordings of multiple individuals without clock-drift issues (if multi-channel recording is used). On the other hand, loggers have the advantage of producing continuous, stable vocalisation recordings, and are not affected by the orientation towards a receiving antenna, or any specific materials or distances in between. In my dissertation, I used both of these systems, depending on the specific requirements that each question and study species involved, which I explain below.

2.2. Species

To investigate the role of vocalisations in group-living songbirds, I chose two different study organisms: zebra finches (Taeniopygia guttata) and western jack-daws (Corvus monedula). Both species are highly gregarious but form a lifelong (socially monogamous) pair bond, respectively, (Zann [1996]; Dwenger [1989], thus making mating partners the most important subunit inside a group. These birds are also highly vocal, with both sexes producing a repertoire of calls (respectively: Zann [1996]; Dwenger [1989]). Because of the importance of pairs, it has been suggested that pair communication could be different compared to communication
Table 1. Feature description and comparison of the two on-board microphone techniques: audio transmitters and audio loggers.

<table>
<thead>
<tr>
<th>Features</th>
<th>Audio transmitters</th>
<th>Audio loggers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recordings made</td>
<td>Remotely</td>
<td>On animal</td>
</tr>
<tr>
<td>Weight</td>
<td>Light</td>
<td>Heavier</td>
</tr>
<tr>
<td>Recording duration (battery, storage)</td>
<td>Prolonged</td>
<td>Shorter</td>
</tr>
<tr>
<td>Synchronisation</td>
<td>Multi-channel, 100%</td>
<td>Addit. technique required, difficult</td>
</tr>
<tr>
<td>Device required for data retrieval</td>
<td>No</td>
<td>Yes (future remote download?)</td>
</tr>
<tr>
<td>Range limitation</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Acoustic distortions, degradation</td>
<td>Common</td>
<td>Rare (possible if damaged)</td>
</tr>
<tr>
<td>Stable signal-to-noise ratio</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of channels</td>
<td>Limited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Technical effort on-site</td>
<td>High (devices, cables etc.)</td>
<td>Low (logger only)</td>
</tr>
</tbody>
</table>

**SUMMARY:**
- Individual recordings: Yes (Yes)
- Recordings in complex habitats: Yes (Yes)
- Suited for: Small species (Yes), Aviaries/field (small home range) (Yes), Timing studies (Yes), Larger species, Aviaries or field, Vocal. types, within-indiv. differences
with the remaining colony (zebra finches: Zann (1996), jackdaws: Lorenz (1931), Dwenger (1989)).

2.2.1. Zebra finches. Because of its readiness to breed in captivity, and its male-exclusive song production, the zebra finch has become an avian model species for diverse scientific questions, ranging from neuroanatomy to the evolution of language (Griffin and Buchanan 2010). Much research has focused on song, but less is known about calling behaviour, which is exhibited by both sexes. Zebra finches produce thousands of calls per day when housed in social context (Beckers and Gahr 2010), but many of these calls are relatively soft, and not easily assigned to specific individuals (Elie et al. 2010, 2011). Thus, this species presented itself as a suitable model to examine – under close monitoring and controlled conditions in captivity, without risk for the animals (e.g. predation) – the use, performance and effects of small on-animal microphones for future individual-level investigations of calls in vocal communication (chapter 1). Further, because these animals come from unpredictable habitats in Australia, breeding can easily be induced throughout the year by providing suitable environmental conditions (Perfetto et al. 2007). Thus, these opportunistic breeders also allow observations during transitions in life-history stages. Therefore, we chose this model species to investigate different aspects of call communication of pairs and other group members by following individual-level calling behaviour over induced changes in reproductive states (chapter 2).

2.2.2. Jackdaws. Although the jackdaw only breeds once per year, it has also served a study species for different investigations, in captivity (e.g. Lorenz 1931, Tamm 1977, Katzir 1982, Wechsler 1988, von Bayern et al. 2007) as well as in the field (e.g. Roell 1978, Henderson and Hart 1993, Arnold and Griffiths 2003, Fulgione et al. 2003, Davidson et al. 2015). Due to a much discussed link between complex sociality and vocal communication (McComb and Semple 2005, Freeberg 2006), the jackdaw, with its ascribed high levels of cognition and complex social structures (Lorenz 1931, Roell 1978, Glutz von Blotzheim and Bauer 1993, Cramp and Perrins 1994) presents itself as a reasonable choice for studying complex vocal behaviour. There is also growing evidence that jackdaws can recognise conspecifics based on their vocalisations alone (Lorenz 1931, Dwenger 1989, Zandberg et al. 2014), like a variety of other animal species (Jouventin et al. 1999b, McComb et al. 2000, Koren and Geffen 2010, Miller and Wren 2012, Bergman and Sheehan 2013, Janik and Sayigh 2013, Mates et al. 2015). But most knowledge on jackdaw vocalisations is based on rather old textual descriptions, lacking state-of-the-art recordings and modern classification techniques.
These birds were particularly useful for my interests because they have a number of features that make them a suitable candidate for on-board microphones in the field. First, they are reasonably common in Europe, and, being a synantropic species, do not generally avoid humans (but: neophobia, [Lorenz, 1931]). Secondly, they nest in cavities and readily accept nest-boxes which they defend vigorously for at least throughout the (extended) breeding season. Nest sites are often limited, and these birds only breed once per year. Also, males and females show high levels of investment to raise their brood ([Lorenz, 1931], [Dwenger, 1989], [Henderson and Hart, 1993], [Cramp and Perrins, 1994]). Therefore, it seemed likely this species should be robust towards disturbances at the nest, which meant it would be feasible to perform nest checks, and also to catch adults inside their nests at low risk of nest abandonment. Lastly, jackdaws belong to the corvid family, and are thus rather large and sturdy, compared to other songbird species. Thus, it should be possible to deploy audio data loggers to record their vocal behaviour wherever they go, be it at the nest, at a feeding site or at the night roost which is often many kilometres away ([Cramp and Perrins, 1994]).

### 2.3. Contexts

In my dissertation, I use the word “context” as a proxy for some of the physical and/or social conditions in which vocalisations occurred, and which I will explain in more detail for the different chapters. As exemplified above, vocalisations are often associated with specific situations, for example during courtship or predator attacks. Therefore, providing overly simplified conditions in vocalisation studies may potentially not yield useful recordings, or may lead to misinterpretation of results. However, some questions require simplified, standardised and controlled experimental conditions to disentangle specific treatment effects without confounding factors (see chapter 1, Appendix 1 & 2). Also, to investigate social networks, it is often crucial to obtain interaction data from all individuals involved ([Kossinets, 2006], [Silk et al., 2015], [Farine and Strandburg-Peshkin, 2015]). For instance in chapter 2 (and Appendix 1), I aimed to define senders and receivers of vocal signals in a group statistically, by using the temporal information encoded in call occurrences between individuals (relative timing). Implementing this information to construct group “vocal networks” thus profited from a closed-system investigation, highly unlikely to achieve in nature (at least in this species). In this “intermediate-level context” study (chapter 2), I also provided changing environmental conditions (sudden onset of nest material availability), embedded in changing socio-sexual contexts (animals unknown to each other forming group, establishing pairs and passing through different breeding stages).
In chapter 3, I explore the notion of “acoustic context” of on-animal sound recordings in the wild. To do so, a validation of different sounds (animal movements and background sounds) made captive recordings indispensable to gain an understanding of sounds which could eventually be extrapolated to the field. Based on some of this information, chapter 4 investigates the vocalisations of a free-living, wild songbird, given in a specific example of natural context: copulation. In sum, this dissertation combines different levels of context complexity, ranging from experiments in simplified environments (chapters 1&3, Appendix 1&2), to studies involving “intermediate”, i.e. reduced but biologically meaningful and changing settings (chapter 2, Appendix 1), to complex, specific contexts in the field (chapters 3&4).
SECTION 3

Chapter overview

My dissertation contains four chapters in which I describe, evaluate and apply two different on-animal microphone methods as well as appropriate analytic tools for studying call communication in group-living songbirds, using zebra finches and jackdaws as model organisms. In the Appendix, I included the abstracts of two further manuscripts which I was involved in during the course of my PhD studies. They provide more detail on specific aspects – timing of vocal events (Appendix 1) and individuality of calls (Appendix 2) – that are related to the topic, but slightly beyond the main scope of my dissertation.

Chapter 1

In the first chapter of my dissertation [Gill et al., 2016], I describe a tool that allows studying vocal communication with individual-level resolution in small, group-living birds, and discuss its potential and limitations, both for the animals and for the scientific questions that can be addressed. In this joint authorship study, my co-authors and I examined the performance and the effects of 0.6g microphone telemetric backpacks on zebra finches under controlled conditions. By placing the backpacks directly on the animal, it was possible to record and reliably assign the vocalisations of specific individuals, even in different noisy environments. Transmitting the audio signal to a receiving base station and performing multi-channel recordings instead of recording directly on the animal (see Table 1) not only resulted in strongly reduced device weight, but also allowed long-term and perfectly synchronous recordings of multiple individuals behaving freely in structured environments and socially complex settings, without post-hoc synchronisation. In this manuscript, we also explored the backpacks’ effects on animal locomotor activity, as well as on the target behaviour of such methods: vocalisations. Compared to previous studies, we found that this methodology resulted in shorter habituation times (less than 3 days). However, we also highlight the importance of extended recordings that exceed potential initial alterations in behaviour, which was not possible in a previous study using a bio-logging approach in this species.
Chapter 2

In chapter 2 (Gill et al., 2015), I apply this technical approach (using an older, slightly heavier version) to gain an understanding of call communication in a small, group-living bird (zebra finch) over a changing environmental and socio-sexual context – which required a lightweight device allowing long-term individual-level recordings of loud and soft vocalisations from freely behaving birds. As argued above, in species with male-exclusive song, calls have the potential to be used by both sexes to form interactive social behaviours that may facilitate important life-history stages. Using microphone telemetry (after an appropriate habituation period), we continuously and simultaneously recorded all individuals present inside a closed-system group setting. In this way, I followed the individual vocalisations of male and female zebra finches previously unknown to each other, and analysed their vocal behaviour while the social group formed, pairs emerged and the birds passed through different breeding stages. We found that calling behaviour not only changed with respect to call-type usage on the individual level, but also during naturally occurring vocal interactions in the group. To objectively define signallers and responders in vocal interactions, we used the temporal information encoded in between-individual call timings. Investigating the timing and call types involved in such vocal exchanges between mates and other members of the group suggested that temporally fine-tuned calling interactions were associated with successful breeding of a pair. This means that call communication may not only play an important role during pair formation and early phases of the breeding cycle, but might be more directly linked to reproductive success than previously assumed, thus highlighting its evolutionary importance in communication systems.

Chapter 3

In chapter 3, I bring up the notion that microphone backpacks have the potential of providing us with more than vocalisation recordings, namely acoustic context. As described above, studying the vocal behaviour of individual animals moving naturally in their habitat has become more feasible through the use of on-board microphones. However, because sound recordings have high energy and storage demands, field studies using such technological approaches for investigating vocal behaviour of small animals are as yet extremely rare. Also, as discussed

1 Appendix 1 provides a more thorough discourse on the analytic tools for investigating the timing of calls in vocal interactions, by analysing call data from zebra finches under standardised conditions as well as reanalysing the more complex dataset described in chapter 2.

2 But not all topics are equally well suited for this technical approach. For example, confounding factors call for caution, making external microphones and standardised acoustic backgrounds attractive for studies on individual identity and vocal recognition (see Discussion, Appendix 2).
above, it is necessary to set the individual vocal behaviour in relation to other contextual data. But simultaneous observations can be limited in the field, and, especially in birds, severe weight limitations restrict the amount of additional data that can be collected on-board (such as accelerometry or GPS). Because backpack microphones designed for continuous vocalisation recordings move through the environment with their carrier animal, their recorded sounds often contain different acoustic events that occurred in the background, such as anthropogenic sounds or the vocalisations of conspecifics or heterospecifics. Additionally, some of the animal’s movements picked up by the microphones are reflected by characteristic sound patterns. Making this additional source of information available could thus be highly valuable for studying the vocal behaviour of animals, embedded in different contexts, in the field.

To investigate this, I applied audio loggers on freely behaving captive and on free-living wild jackdaws (*Corvus monedula*), and conducted a video-validation study. I started a collaboration with a machine-learning expert who used my video-validated and human-coded annotations to test different machine learning approaches on the data containing bird movements and background sounds. Although fully automated event detection and classification require further efforts, the soundscape analysis approach did succeed in identifying focal and non-focal vocalisations, and in extracting behavioural and contextual information that other technical approaches cannot provide on their own (GPS, accelerometry, etc.). Given the fact that on-animal sound recordings are likely to become more efficient and available with advancing technological developments, analysing the full information spectrum of these streams of sounds seems promising and an effort worth taking to understand when, how and why animals produce specific vocalisations.
Chapter 4
In this chapter, I applied some of the methods explained in chapter 3 to investigate a specific call type associated with a specific context: copulatory calls in jackdaws. These calls are produced during copulation, are very loud and conspicuous, and can be heard from dozens of metres away. But because copulations take place inside the dark nest cavities in this species, it was not yet possible to identify which sex produced them, and thus, their function was completely unknown. Although not all animals produce such sounds, in some vertebrate species, including humans, such vocalisations have been subject to intense study, and have revealed very different functions, ranging from mate attraction to the source of complex social information (Hauser 1998; Bradbury and Vehrencamp 2011).

Using the abovementioned audio loggers on free-living wild jackdaws, as well as nest-box video footage and sound analyses, I found that males were the source of these conspicuous vocalisations. In addition, I discovered not unsubstantial amounts of extra-pair copulations which, too, were supposedly accompanied by the loud copulatory calls. This information contradicts the commonly accepted notion of jackdaws being exclusively monogamous, and calls for modern, state-of-the-art genetic and behavioural investigations. Although I can only speculate on the function of these vocalisations so far, this study demonstrates that on-animal sound recordings may help describing and revealing unknown vocalisations and exciting non-vocal behaviours, to facilitate future detailed investigations of vocal communication in natural contexts.
PART II – CHAPTERS
CHAPTER 1

A minimum-impact, flexible tool to study vocal communication of small animals with precise individual-level resolution

Lisa F. Gill*; Pietro B. D’Amelio*; Nicolas M. Adreani*; Hannes Sagunsky; Manfred L. Gahr; Andries ter Maat

SECTION 4

Abstract

To understand both proximate and ultimate factors shaping vocal communication, it is fundamental to obtain reliable information of participating individuals on different levels: Firstly, it is necessary to separate and assign the individuals’ vocalisations. Secondly, the precise timing of vocal events needs to be retained. Thirdly, vocal behaviour should be recorded from undisturbed animals in meaningful settings. A growing number of studies used animal-attached microphones to tackle these issues, but the implications for the study species and the research question often receded into the background.

Here we aim to initiate a discussion about the limitations, possible applications and the broader potential of such methods. Using lightweight wireless microphone backpacks (0.75 g including customised leg-loop harness) combined with multi-channel recording equipment we captured vocal behaviour of small songbirds. We evaluated the effect of the devices at various levels, including an assessment of how vocal and locomotor activities were affected by initial device attachment and battery exchange. We compared our approach to existing studies and identified suitable research examples. We acquired continuous vocalisation recordings of zebra finches, and unequivocally assigned them to interacting individuals, with system-based synchrony, irrespective of background noise. We found effects of initial backpack attachment and of battery replacement on vocal and locomotor activity, but they were minimised through the extended recording duration (ca. 16 days) that outlasted habituation effects (ca. 3 days).

This method provides the tools to integrate individual vocal communications into a group setting, while enabling animals to behave freely in undisturbed, structured and acoustically complex environments. By minimising the effects on the animals, the behaviour under study, and ultimately on the research question, this approach will revolutionize the ability to capture individual-level vocalisations in a variety of communication contexts, opening up many new opportunities to address novel research questions.
SECTION 5

Introduction

Due to technological limitations in obtaining good quality vocalisation recordings from individual animals, many open questions remain about animal vocal communication, especially in naturalistic settings. A growing body of studies has recently come to use animal-borne devices on species ranging from whales (Johnson et al., 2009) to chipmunks (Couchoux et al., 2015), opening up new fields of communication research. Before this, it was challenging to record and assign vocalisations of focal animals that were behaving freely inside their natural habitats, or in the presence of multiple sound sources, such as vocalising conspecifics or other background noises (Otter et al., 1997; Brumm, 2004; Furrer and Manser, 2009; Clemmons and Howitz, 2010; Bousquet et al., 2011). For example, when studying highly synchronised vocal interactions in social contexts, like in duetting birds, the role of individuals has been very difficult (Thorpe et al., 1972; Mann et al., 2006) or even impossible (Elie et al., 2010) to investigate. Using modern remote sensing technology, such as microphone transmitters (Ter Maat et al., 2014; Gill et al., 2015) or audio loggers (Johnson et al., 2009; Ilany et al., 2013; Goldbogen et al., 2014; Anisimov et al., 2014; Couchoux et al., 2015), vocalisations are recorded directly from the sound-producing animal, and can be assigned unequivocally. It is however just as fundamental to retain the precise temporal information of vocal events (Brumm and Slater, 2007). This has been achieved through post-hoc synchronisation (Anisimov et al., 2014) or multi-channel recordings (Ter Maat et al., 2014; Gill et al., 2015). Now that it is becoming increasingly feasible to obtain data with individual information and temporal precision, mechanisms and function of vocal interactions can be investigated with individual-level resolution. In some cases, vocalisations could even be recorded while the animals covered large distances or moved freely in otherwise inaccessible natural habitats (Johnson et al., 2009; Cvikel et al., 2014; Couchoux et al., 2015). However, it has been shown that animal-attached devices may not only lead to short-term changes in movement patterns (Hooge, 1991; Schregardus et al., 2006; Chipman et al., 2007; Anisimov et al., 2014), but may also have longer-term fitness consequences (Phillips et al., 2009; Barron et al., 2010; Arlt et al., 2013). Such results, often coming from the field of movement ecology, suggest that the devices have the potential to influence...
exactly those behaviours that were studied (Ropert-Coudert and Wilson 2005). In vocal communication research, it would therefore be important to assess the effect of on-board devices on vocal behaviour (Anisimov et al. 2014), in addition to more general behavioural patterns such as locomotion. But due to the novelty of the field itself (Wilmers et al. 2015), and in view of new exciting discoveries, this and other methodological aspects often receded into the background. So far, most studies did not quantify any direct effects of device attachment on normal behaviour (Hiryu et al. 2008; Cvikel et al. 2014; Ter Maat et al. 2014; Gill et al. 2015), or neglected to investigate related aspects, such as handling before each recording period (Anisimov et al. 2014). In addition, on-board devices may have specific technical requirements that impose further constraints on animal wellbeing (e.g. frequent handling for data retrieval and/or battery exchange; unstructured environments necessary for infrared-based logger synchronisation (Anisimov et al. 2014). Because most investigations did not address in detail the implications of a specific approach for the study animals (Ropert-Coudert and Wilson 2005; McIntyre 2015), it is not clear how this could affect the design and outcome of the respective study. Therefore, the current literature is also lacking a thorough discussion of the limitations, of possible applications and of the broader potential of such methods.

Here, we propose a flexible method that could overcome many of such limitations, and describe in detail its technical specifications, setup and workflow. By focusing on the method’s direct and indirect consequences on the animals and the research questions involved, we discuss possible applications and the potential for understanding mechanisms and function of vocal communication. In vocal communication research, songbirds have often been used as model organisms. Recording birds with on-board devices represents an especial challenge, because flight and other behaviours can easily be impaired (Pennycuick et al. 2012; Vandenabeele et al. 2012; Anisimov et al. 2014). Therefore, we validate the applicability of on-board microphones for detailed investigations of animal vocal communication, using the example of small songbirds and an improved version of a previously published microphone transmitter (Ter Maat et al. 2014). This latest version is currently the lightest on-board device for sound recordings (0.56g, 0.75g including harness). In our study, we evaluate i) transient effects on vocal and locomotor behaviour of the initial backpack attachment and of handling associated with battery exchange on zebra finches (Taeniopygia guttata); ii) the system’s performance at accurately and selectively recording individual vocalisations, even in loud noise; iii) as well as its system-based between-transmitter
synchrony. In combination with an extended recording duration that outlasts habituation, these features allow tracking individual vocal behaviour, even of small animals, in acoustically and physically complex environments.
SECTION 6

Materials and Methods

6.1. Backpacks

The backpacks (Fig. 4) consisted of a circuit board including the transmitter (Ter Maat et al., 2014) (12.3 x 5.5 x 4.5 mm), a miniature condenser microphone (FB-23359, Knowles, USA), as well as a battery (Zn Air, p10, Power One, Germany) and an established leg-loop harness (Rappole and Tipton, 1991). In total, backpack weight was 0.56g without and 0.75g including the harness (4.78% of an average 15.7 g zebra finch in our colony). To build the harness, we attached an elastic cord (1mm diameter, Kordel elastic, Veno Hermann Veddeler GmbH, Germany) to a custom-built silicon casing (Fig. 4b). During attachment, harness length was customised for each bird to decrease the impact (Barron et al., 2010). An earlier study on an electrophysiology transmitter showed that the effects of anaesthesia on singing behaviour and locomotion exceeded the effects of handling (Schregardus et al., 2006). Therefore, and because backpack attachment was easy and fast (see Results, supplementary Video 1) birds were not anaesthetized.

6.2. Workflow

Each transmitter had a specific non-overlapping transmission frequency (270 MHz to 320 MHz) which was received by its own antenna (crossed Yagi directional antenna for 300 MHz, Winkler Spezialantennen, Germany) and radio receiver (AOR8600, AOR Ltd., Japan, modified to have an audio bandwidth of 12 kHz), and the signal was tracked by an oscilloscope (DSO-X 2004A, Agilent Technologies, USA). This allowed stable recordings of single channels. To digitize the acoustic signals, we used a multi-channel A/D converter (Fast Track Ultra 8R, M-Audio, USA) connected to a PC. Each of the resulting digitized signals were recorded in parallel using multi-channel software (16-bit, 44100 Hz; ASIO Rec, adapted by Markus Abels, MPIO Seewiesen). For a flow diagram, see Fig. 3.

6.3. Animals

We used a total of 10 adult, parent-raised, domesticated zebra finches (9 males, 1 female) that were kept on a 12/12-h light/dark cycle with food and water ad
Figure 3. Synchronous audio recording: a) Workflow, b) 4-hour and c) 24-hour test recording (a) Workflow of recording process: Each group member carries a backpack that transmits acoustic signals via a unique transmission frequency. Signals are received by one antenna and receiver per individual. After digitization ("A-D converter"), multi-channel software records all audio tracks simultaneously, in real time. External microphones (blue) may be connected to the system, e.g. as reference. (b+c) Time [hours] of 2 transmitters plotted against each other (X-axis: transmitter A; y-axis: transmitter B). Dashed lines indicate theoretical examples of clock drift (black upper: B faster than A; black lower: B slower than A; white: zero clock drift). Light blue dots show real values from a few hours at the end (b) of a 24-hour test recording (c). Data available at http://datadryad.org/review?doi=doi:10.5061/dryad.h8h35.

6.4. Experiments

6.4.1. Effects of backpack attachment and battery exchange. To evaluate direct behavioural effects of backpack attachment on zebra finches, we measured individual calling and locomotor activity in a standardized environment. Because handling is often necessary during an ongoing experiment (e.g. to change
batteries or download data) we also quantified changes in these behaviours due to battery exchange.

For this, 7 male zebra finches were housed in individual cages (54x40x28 cm) inside custom-built sound-proof chambers (70x50x50 cm) equipped with a microphone (TC20, Earthworks, USA) connected to the sound recording system mentioned above, and a video camera (Handy kam, UK). Sound and video recordings began as soon as the birds were placed in the sound chambers, and continued throughout the day, for 20 days in total. The first 4 hours after “lights on” were used to analyse and track vocal and locomotor activity (data available at http://datadryad.org/review?doi=doi:10.5061/dryad.h8h35). As a proxy for the birds’ vocal activity, we used the number of calls recorded by the external microphones during the 4-hour periods. Locomotor activity of the birds was measured through
a quantification of automated motion-detected multi-channel videos (Surveillance System V8.5.0.0, Geovision). Whenever a bird changed its location in the cage, continuous recording was triggered. If there was no further movement of the bird within 5 seconds, the recording stopped. The duration of all the videos was extracted (MediaInfo, 0.7.71) and summed up (using R (R Core Team 2014)) for each bird on each morning, and was subsequently used as a proxy for locomotor activity.

Isolating zebra finches impacts vocal activity (Perez et al. 2012), but on day 7, the calling activity of all birds reached a stable plateau (<15% day-to-day change for minimum 4 days; mean of all birds) which means this value could be used as baseline (i.e. habituation to cage). On the next morning (day 8), the birds were caught about 20 minutes before “lights on” and were equipped with a microphone backpack dummy (same weight and external case). Subsequently, the birds were released back inside their cages, and the above-mentioned sound and video recording scheme was continued. Once vocal activity had again reached a stable plateau (day 15, see above), with no statistical difference to baseline levels (day 7, Fig. 5), we proceeded with the next step of the experiment. To investigate the effect of battery exchange, we caught the birds on day 16 about 20 minutes before “lights on”, and exchanged the backpack batteries. Subsequently, the birds were returned to their cages, and the recording procedures described above were resumed. The experiment ended on day 20 because bird activity levels had returned to baseline values and had not changed significantly during the 4 previous days.

6.4.2. Selective sound recordings. The aim of the next experiment was to evaluate the performance and selectivity of the backpack recordings in challenging acoustic environments, such as loud noise or the presence of conspecifics. We thus fitted 3 zebra finches (2 males; 1 female for vocal stimulation) with microphone backpacks, and subsequently housed them together in a cage (1x1x1 m) equipped with a microphone (TC20, Earthworks, USA) and a loudspeaker (KENWOOD KFC-1761S, Kenwood Electronics, London, UK). After a 5-day habituation period, vocalisations were recorded via the backpack microphones as well as by the external microphone for the first 4 hours after “lights on”, on 2 subsequent days, during different conditions. On day 1, we recorded the vocalisations without changing the setup (“no-noise” condition). On day 2, we recorded the vocalisations during playback of loud, constant white noise (“noise” condition; 80dB, measured at 1 m from the speaker using an SPL meter; HD600, Extech, USA).

Assigning focal versus non-focal vocalisations
First, we evaluated the system’s performance at assigning the vocalisations to the
bird carrying the backpack (“focal bird”) in the presence of vocalising conspecifics, in both noise conditions. For each male (♂1, ♂2), we identified the loudest vocalisation type by comparing waveform amplitudes. Due to the similarity of some song syllables and distance calls, we combined these two categories as “loudest syllable type”. We counted how many syllables of the “focal” and of the “non-focal” individuals had been recorded in the “focal” individual’s backpack. Within each
Figure 6. Spectral analysis of the focal and non-focal recording of a distance call. Power spectra (left) and sonograms (right) of the same male distance call, recorded separately by the “focal” (top row, red) and by the “non-focal” bird’s backpack (bottom row, blue). Both recordings were normalised for spectral analysis. Note that the maximum-power frequency differed greatly between “focal” and “non-focal” recordings which allowed an unequivocal identification of the sound-emitting individual. Grey dashed lines indicate frequency of maximum power (peak). Also see Audio 2.

male’s recording, we used spectral differences to assign syllables to the “focal” or the “non-focal” bird. Backpack vocalisation recordings of the “focal” bird showed higher amplitudes (Fig. 3) and contained higher power in the low frequency bands than the “non-focal” birds (Ter Maat et al., 2014; Gill et al., 2015) (Fig. 6). These vocalisations are individual-specific (each bird has a unique structure, Zann, 1984), therefore we were able to validate the above described assignment based on spectral differences alone.

Performance of backpack versus external microphone recordings
Next, to assess the system’s efficiency at recording individual vocalisations despite loud external noise we counted the song events recorded by the backpack and by an external microphone, in the two noise conditions. We assigned the songs to either of the 2 males based on spectro-temporal differences, and used the most characteristic syllable as a proxy for song. Subsequently, we counted the number of song occurrences recorded by the external microphone and by the backpack microphone, for both males.

6.4.3. Synchronous sound recordings. Although the multichannel recording procedure results in synchronised audio recordings (system-based), we performed
a simple playback experiment to demonstrate this empirically. 6 audio trans-
mitters were distributed equidistantly on a wooden platform, at 80 cm from a
loudspeaker (KENWOOD KFC-1761S, Kenwood Electronics, London, UK). During a
period of 24 hours, we automatically broadcasted a zebra finch call from the loud-
speaker 3 times per hour at 75 dB (measured at 80 cm from the loudspeaker using
an SPL-meter, see above) while continuously recording sound via the backpacks.
The 6 resulting sound files were analysed using custom software, following the
paradigm described below (see 6.5.1). The timings of call onsets, as recorded by
the different transmitters, were plotted against each other (example in Fig. 3b
h8h35).

6.4.4. Battery life. Theoretical battery life of the transmitters was calculated
as the battery’s rated capacity divided by the device’s consumption of current
(100mAh/0.22mA = 454h, 18.9 days). For an empirical test, 5 transmitters were
supplied with a new battery (1.45V, see above) and placed in a room with constant
humidity and temperature (45%, 24°C). Their signal reception was checked once
a day. Battery life was noted as expired when there was no signal.

6.4.5. Transmission distance. Transmission distance was empirically tested
using 3 devices. It was defined as the furthest point that allowed signal reception
and was assessed in three different conditions: outside a building, without (i) and
with (ii) an amplifier (Rohde and Schwartz, Germany, 20 dB rf amplifier), and inside
a building (without amplification) (iii).

6.5. Analyses

6.5.1. Sound analyses. All sound files were processed and analysed semi-
automatically using custom software, to “cluster” (k-means) the individual vocal-
isations into calls and song (Ter Maat et al., 2014; Gill et al., 2015). In the habitu-
ation experiment, we focused on the number of calls per individual (all call types),
because, unlike song, calls were produced by the isolated birds in sufficient num-
bers for a sound statistical evaluation. In the backpack recordings, we measured
the mean amplitudes of the background noise and of the loudest syllables coming
from the focal and non-focal bird. All amplitude values were obtained by calcu-
ling the mean amplitude of the Hilbert transform of a given sound type using
the package “seewave” (Sueur et al., 2008) in R (Development Core Team) (Fig. 8).
We used 10 samples for noise and 15 each for focal and non-focal birds’ vocalisa-
tions. All power spectra and sonograms (Figs. 6, 7) were drawn in R (Development
Figure 7. Vocal communication in noise. (a) Sonograms of a loud syllable in the “no noise” (left) and “noise” (right, 80 dB) condition. Exact same syllables were recorded by the backpack (top) and external microphone (bottom). Note that the noise completely masks the syllable in the external microphone (bottom right), but does not appear in the backpack recording (top right). (b) Vocal network of 3 birds (1 female, 2 males) recorded via backpack microphones in the “no noise” (left) and “noise” (right) condition. Arrows indicate the direction and line thickness the amount of calls used to answer each individual (see Methods). White rectangles represent peri-stimulus time histograms (see Methods, Fig. 9, Audio 1).

Core Team) using the package “seewave” (Sueur et al., 2008) (Spectrum size 4096 points, FFT size 512, “hanning” window).

6.5.2. Vocal interaction analyses. For vocal interaction analyses, we used vocalisation recordings from the 2 male and 1 female in both noise conditions (see 6.4.2 Selective sound recordings, 4h per bird per day). We focused on the most frequent call, the “stack” (Ter Maat et al., 2014). As previously described (Anisimov et al., 2014; Ter Maat et al., 2014; Gill et al., 2015), zebra finches are likely to “reply” to another within 0.5 seconds. Therefore, we used the number of vocalisations emitted within this narrow time window as the number of “answers”. Based on this, we drew vocal networks between 3 interacting individuals (Fig. 7B), with the direction and thickness of the arrows reflecting the direction and the number of answers (ranging from 76 to 645). To visualise in more detail the timing of vocalisations between each set of 2 birds (data available at [http://datadryad.org/review?doi=doi:10.5061/dryad.h8h35](http://datadryad.org/review?doi=doi:10.5061/dryad.h8h35)), we plotted peri-stimulus time histograms (Abeles, 1982) (PSTHs, Fig. 9) in addition to the arrows. In these PSTHs, vocalisation onsets of 2 birds were aligned and summed up in 20ms-bins. For plotting, all bin values were normalised by the overall maximum value of a single bin (150).
The confidence intervals, indicated by horizontal red lines, are the result of a simulation of 1000 calls placed at random times in the recording (Ter Maat et al. 2014).

### 6.5.3. Statistical analyses.

All statistical analyses were performed using R (R Core Team 2014). We modelled the number of calls and of video duration (proxy for locomotor activity, see above) using linear mixed models, specifically applying the function `lmer` (package “`lme4`” (Bates et al. 2015). Both outcome variables, the number of calls and the video duration, were square root-transformed to approach a normal distribution of residuals. We used the day of the experiment as a categorical explanatory variable. For modelling the number of calls, we used individual id, the centred ($x_{centred} = x - \text{mean}(x)$) tarsus length and scorer id as random factors to account for between-individual differences in calling activity and size, as well as a possible observer bias, respectively. For the locomotor activity model we used the same random factors, excluding scorer id (automated scoring). Models including all combinations of random factors were compared to each other using AIC scores (Akaike 1970). Highest ranking models were the ones that included all random factors. These models were used for subsequent analyses of calling and locomotor activity. For both linear models, standard diagnostic plots were used to assess whether assumptions were met. For the model parameters we used flat prior distributions. This means the credible intervals (CrI) were equal to confidence intervals obtained by frequentist methods, and sensitivity analyses of prior distributions are not required. To obtain the new set of parameters we simulated 10000 values from the joint posterior distribution of the model parameters using the function “`sim`” of the package “`arm`”. The 95% CrIs shown in Fig. 5 were calculated from the 2.5% and 97.5% quantiles of the simulated values as lower and upper limits, and we used the means as estimates (fitted values). We used the derived parameters from the posterior distributions to test specific hypotheses (e.g. differences from the baseline). To calculate the probability of a difference between activity levels on different days, we calculated the proportion of differing simulated values (for example number of calls on different days). We considered them to be statistically different from each other (significant) if the posterior probability was larger than 0.95. See Tables 2, 3 for exact fitted values and credible intervals plotted in Fig. 5 (negative values set to 0).
Results

7.1. Effects of backpack attachment and battery exchange

Attaching the established and customized leg-loop harness (Rappole and Tipton, 1991) lasted 53.7 ± 18.2 seconds (mean ± standard deviation, n=7), and birds were able to fly immediately afterwards (Video 1). Handling duration while fitting the backpacks was thus comparable to, or even shorter than, regular bird banding – a standard procedure in ornithology. Calling behaviour (Fig. 5A, Table 2) strongly decreased immediately after backpack attachment ((fitday – fitbaseline)/ fitbaseline: by 99.97%; p<0.001), and gradually increased from the next day onwards. After 72 hours, there was no statistical difference (see Methods for definition) between pre- and post-backpack calling activity (p=0.102; but remaining 42.5% decrease). Locomotor activity (Fig. 5B, Table 3) decreased immediately (by 70.16%; p<0.001), but returned to pre-backpack values after 24 hours (27.96% decrease; p= 0.202). Capturing the birds and replacing the transmitter batteries (battery exchange: 5-10 seconds, total handling time including catching: <1 min) after birds had habituated to the backpacks significantly decreased calling only on the day of manipulation (by 61.05%; p=0.024, Fig. 5A; next day: 3.34% increase; p=0.532; all comparisons with baseline, see Fig. 5A, Table 2). Locomotor activity was not affected statistically but decreased by 45.31% (p=0.079, see Fig. 5B, Table 3). Thus, this disturbance was weaker and birds recovered faster than after initial backpack attachment. Additionally, they did not impede successful reproduction. Copulations, egg-laying and parental care were frequently observed (unpublished data). Also, using an older version of the backpacks on previously unmated birds resulted in a mean clutch size of 4.55 eggs (unpublished data from Gill et al. (2015)) which was similar to the mean in one of the institute’s breeding colonies (4.4, personal communication W. Forstmeier).

7.2. Selective sound recordings in normal and noisy conditions

Bird song can be strongly directional (Brumm and Todt, 2003; Brumm et al. 2011), and even without noise, the external microphone picked up fewer song events
than the backpack microphones. As explained above (see Methods 6.4.2 Selective sound recordings), this was not due to an incorrect assignment of "focal" birds’ vocalisations (Figs. 6 and 8). Instead, this means that backpack recordings give a more accurate representation of vocalisation events than external microphones.

7.2.1. Focal versus non-focal vocalisations. Assuming that ‘focal’ birds’ syllables had higher overall amplitudes (Fig. 5) and contained more power in the low frequency bands than the recordings of "non-focal" birds (Ter Maat et al., 2014; Gill et al., 2015) (Fig. 6, supplementary Audio 2, see Methods) resulted in 0-10% of “non-focal” syllables recorded by the “focal” backpacks during both noise conditions. Specifically, in the “no noise” condition we counted 936 (♂ 1) and 777 syllables (♂ 2) in the “focal”, and 92 (♂ 1) and 4 (♂ 2) in the “non-focal” backpack recordings (i.e. 9.8 and 0.5%, respectively). In the “noise” condition, we found 1353 (♂ 1) and 1244 (♂ 2) syllables in the “focal” and 88 (♂ 1) and 0 (♂ 2) in the “non-focal” backpack recordings (9.3 and 0%, respectively). Using the individually distinct syllables for validation gave identical results. The differences in spectral features alone were enough to accurately assign the signaller: none of these syllables were assigned to an incorrect signaller. Therefore, there was no need for any additional device, e.g. accelerometers (Anisimov et al., 2014), to reliably assign the vocalisations.

7.2.2. Backpack versus external. In the “no noise” condition, fewer events were detected by the external microphone than by the backpacks. Specifically, we found 492 (♂ 1) and 607 songs (♂ 2) in the external and 525 (♂ 1) and 655 (♂ 2) in the backpack recordings. During “noise”, songs were completely masked in the external recordings and were only detectable in the backpack recordings (950 (♂ 1) and 247 (♂ 2) songs, Fig. 7A, supplementary Audio 1). Thus, the on-board microphones were more accurate than conventional ones at collecting data from interacting birds across different environmental conditions. Therefore, it is possible to draw vocal networks and investigate vocal interactions with individual-level resolution, even in acoustically challenging conditions (Fig. 7B, Audio 1).

7.3. Synchronous recordings, battery life and transmission distance

The setup we describe here involved a single, multichannel recording device. This resulted in identical stimulus onsets in different recordings (n=6) (Fig. 3, http://datadryad.org/review?doi=doi:10.5061/dryad.h8h35), i.e. precise system-based temporal alignment of multiple audio tracks. Transmission distance (n=3) without using an amplifier ranged from 11.3 to 15.8m inside, and from 11.3 to 13.72m
outside. With amplification, higher values were obtained (25.1 – 26.7m). Battery life (n=5) amounted to 16.2 ± 0.83 days in standardised conditions (see Methods).
Using our transmitters, we were able to record multiple vocalising individuals unequivocally, even in noise, without any additional identification technique (e.g. accelerometry in Anisimov et al. 2014). For many questions in vocal communication, it is also important to reconstruct the precise timing of vocal events (Brumm and Slater 2007; Ter Maat et al. 2014; Gill et al. 2015). Our approach involved system-based synchrony of multiple audio tracks, thus eliminating clock-drift. Also, our microphone backpacks were fast and easy to apply, and allowed the birds to fly immediately afterwards (supplementary Video 1). From quantifying activity levels, we found that attaching backpacks affected not only locomotor, but also vocal behaviour, i.e. the behaviour which this methodology was intended to capture. Moreover, vocal activity was also affected by handling necessary for battery replacement or data retrieval (Anisimov et al. 2014; Couchoux et al. 2015) which, to our knowledge, has not been evaluated before. All measured behavioural effects were only temporary, with vocal and locomotor activity recovering after 72 and 24 hours after backpack attachment, and vocal activity recovering within 24 hours after battery exchange. Because stress levels are lower in non-isolated zebra finches (Perez et al. 2012), it is possible that more naturalistic experimental settings could result in even faster recovery rates.

An ideal way to study the ecology and evolution of vocal communication would be to capture all vocalisations and their drivers in natural settings. But complex social and physical contexts can be acoustically challenging (Brumm 2004; Clemmons and Howitz 2010; Bousquet et al. 2011), especially when individual-level information is important. Compared to most data commonly collected via remote sensing, acoustic recordings require high sampling rates, and thus have high storage and energy demands. Therefore, vocal communication has not been frequently studied using backpacks (Wilmers et al. 2015). Recent technology has made previously unexplored questions more accessible by providing data from individuals behaving in more biologically meaningful settings than ever before. Animal-borne recorders (loggers) have proven useful in gaining normalised sound recordings, irrespective of the distance towards any receiving base station (Johnson et al. 2009; Cvikel et al. 2014; Couchoux et al. 2015). However, they are...
subject to weight, storage and battery constraints. In addition, whenever multiple devices have internal clocks, their timing eventually drifts apart (clock drift). Post-hoc synchronisation of audio tracks is possible by using additional signals, such as a master device (Levin et al., 2015), GPS (Walker et al., 2015) or infrared pulses (Anisimov et al., 2014) as an external reference (example in Fig. 3B). However, this requires a substantial amount of data processing, and comes at a cost of precision. Moreover, such devices could affect potential research questions by increasing animal weight load or by imposing specific technical requirements, for example unstructured environments (necessary for infrared-based synchronisation (Anisimov et al., 2014)). Telemetric devices (transmitters), on the other hand, do not store information on the animals, and thus allow extended, continuous recording periods and system-based synchrony while maintaining small size and light weight (Ter Maat et al., 2014; Gill et al., 2015). Therefore, they may not only have smaller direct effects on the animals (Wilson and McMahon, 2006), but also allow longer continuous recordings that can bypass handling-related effects. However, their application can be restricted by transmission range. In our setup, the signals were received up to a distance of 26 m which allows recording animals in large enclosures and even in the field under certain conditions, if an adequate power supply is provided. Specifically, if the position of an animal can be anticipated it is possible to capture its vocalisations by placing the receiving antenna accordingly, for example close to the nest, roost, lek or colony. Or, if the species allows being approached (Radford and Ridley, 2008; Bousquet et al., 2011), it could even be possible to follow a set of vocalising individuals. Additionally, transmission distance and recording time could be increased if the study species could cope with an augmented weight.

In small animals, particularly in birds, devices need to be light-weight and ergonomic to prevent constraints on locomotion and other behaviours (Hooge, 1991; Barron et al., 2010; Vandenabeele et al., 2012, 2014; Blackburn et al., 2016). So far, the effects on vocal behaviour in addition to locomotor activity were reported only once (Anisimov et al., 2014). and, compared to in our study, ceased after a longer period of time. This was probably due to a device with higher proportional weight and less ergonomic harness whose attachment required anaesthesia. Especially in the field, it is important to ensure that such devices do not hamper normal locomotion (Barron et al., 2010; Blackburn et al., 2016). Our device is currently the lightest on-board microphone method and can be applied to animals as small as hummingbirds (ongoing study on black jacobins, Florisuga fusca, personal communications Monte & Gahr, 2016). However, deploying observation devices on living animals may influence various aspects of their lives (Hooge,
and weight is not the only property of a backpack that may affect them. Therefore, it is important to pay attention to unexpected complications that may arise. For example, during earlier stages of the device, we noted strong behavioural effects of the protruding antenna. In fact, in a previous study, the antenna had to be incorporated into the harness for females to ensure normal reproductive behaviours (Gill et al. 2015), which came at a cost of transmission quality. In the device’s current version, this has been overcome by incorporating the antenna wire in the backpack (coil). Thus, our backpacks allow successful reproduction (personal observations, see 6.4.1), and can be used in various complex physical and social settings.

In general, if recordings are made before animals have recovered from backpack attachment, there can be confounding effects in the outcome of an experiment. One way of overcoming this is through a thorough assessment and subsequent exclusion of any affected recordings, which is possible only when recording duration (up to 16 days in our study) exceeds the habituation period (up to 3 days in our study). Such long-term continuous recordings thus further extend the scope of scientific questions that can be addressed by enabling us to track within-individual changes in vocal behaviour of undisturbed animals, for example in response to changing contexts. For instance, it would be possible to investigate with minimal impact the vocal ontogeny of individuals, by following the development of all their vocalisations, even in complex social environments.

Because the multi-channel recording procedure results in perfectly synchronised tracks, it is not only possible to study spectro-temporal features of vocalisations, but also the temporal relationships between vocal events. This method would therefore be suitable for investigating vocal networks of highly vocal, group-living animals (see Fig. 7B). For example, the white-browed sparrow weaver (Plocepasser mahali) produces complex duets and chorus song that seem to be linked with social status and pair bond (Voigt et al. 2006). However, their vocalisations are difficult to assign to specific individuals. Because the singing behaviour is predictable in time and place, this species seems a promising candidate for field studies (ongoing pilot project, personal communications Leitner & Voigt, 2016). But fine-scale temporal precision and synchrony are not only important for studying vocal interactions. They are also a prerequisite for other subjects of investigation, such as the coordination of high-speed multimodal courtship behaviours (Bostwick 2003; Ota et al. 2015), echolocation (Cvikel et al. 2014), or electrophysiology (Ter Maat et al. 2014). Because the recording process allows connecting and recording further channels in parallel (e.g. spiking neurons, heart rate), it is possible to investigate the meaning of vocal signals by integrating them into a variety of contexts.
In conclusion, it is important to evaluate potential effects of a given methodology in order to minimise an influence on the study species and ultimately on the research question. Using our method, it is possible to quantify vocal behaviour with individual-level resolution in largely undisturbed subjects. Therefore, challenging aspects of vocal communication can be investigated in small animals, even in small birds, while they behave freely in structured, acoustically complex or changing environments, even in the field. Animal-attached microphones thus allow collecting unbiased vocal information from individuals in previously unexplored situations ranging from large groups to loud anthropogenic noise, which in turn will advance our understanding of mechanisms and evolution of animal communication.
SECTION 9

Ethics statement and Acknowledgments

Ethics statement
All housing and experimental procedures were approved by the Government of Upper Bavaria (55.2-1-54-2532-25-09 and 55.2-1-2532-181-12).

Acknowledgments
We are grateful to Lisa Trost and Markus Abels for technical support and advice. We thank Damien Farine and Rudolf Alexander Steinbrecht for useful comments and suggestions. We also thank the IMPRS research school for training and support.
SECTION 10

Supporting Materials

**Supplementary Video 1: Backpack attachment (real time)**
Demonstration of how to attach a transmitter backpack on a zebra finch (real time).

**Supplementary Audio 1: Sound files corresponding to Fig. 7**
Two calls recorded by backpack and external microphones in different conditions in the following order: Call 1 recorded in “no noise” condition via backpack (i) and via external microphone (ii). Call 2 recorded in “no noise” condition via backpack (iii) and via external microphone (iv).

**Supplementary Audio 2: Sound files corresponding to Fig. 6**
The identical call recorded by the caller backpack and by the backpack of the other individual (i.e. as “Focal” versus “Non-focal” call).

<table>
<thead>
<tr>
<th>Day</th>
<th>Lower</th>
<th>Upper</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>273.8517</td>
<td>1.047623</td>
</tr>
<tr>
<td>2</td>
<td>11.27283</td>
<td>1211.867</td>
<td>368.2638</td>
</tr>
<tr>
<td>3</td>
<td>271.0442</td>
<td>2301.281</td>
<td>1036.209</td>
</tr>
<tr>
<td>4</td>
<td>591.1981</td>
<td>3146.983</td>
<td>1628.205</td>
</tr>
<tr>
<td>5</td>
<td>685.9842</td>
<td>3406.461</td>
<td>1788.999</td>
</tr>
<tr>
<td>6</td>
<td>768.2764</td>
<td>3552.053</td>
<td>1930.71</td>
</tr>
<tr>
<td>7</td>
<td>724.7044</td>
<td>3452.677</td>
<td>1846.059</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>265.655</td>
<td>0.484488</td>
</tr>
<tr>
<td>9</td>
<td>36.4427</td>
<td>1443.643</td>
<td>490.1048</td>
</tr>
<tr>
<td>10</td>
<td>78.97351</td>
<td>1662.567</td>
<td>628.777</td>
</tr>
<tr>
<td>11</td>
<td>260.5525</td>
<td>2349.662</td>
<td>1061.494</td>
</tr>
<tr>
<td>12</td>
<td>475.0668</td>
<td>2927.306</td>
<td>1463.353</td>
</tr>
<tr>
<td>13</td>
<td>558.8101</td>
<td>3026.157</td>
<td>1553.253</td>
</tr>
<tr>
<td>14</td>
<td>498.0024</td>
<td>2952.411</td>
<td>1488.713</td>
</tr>
<tr>
<td>15</td>
<td>599.7624</td>
<td>3164.497</td>
<td>1656.169</td>
</tr>
<tr>
<td>16</td>
<td>119.5749</td>
<td>1832.445</td>
<td>719.002</td>
</tr>
<tr>
<td>17</td>
<td>751.116</td>
<td>3583.635</td>
<td>1907.76</td>
</tr>
<tr>
<td>18</td>
<td>700.0914</td>
<td>3393.852</td>
<td>1811.839</td>
</tr>
<tr>
<td>19</td>
<td>853.4044</td>
<td>3746.917</td>
<td>2053.378</td>
</tr>
<tr>
<td>20</td>
<td>1020.361</td>
<td>4159.362</td>
<td>2328.865</td>
</tr>
</tbody>
</table>

Table 2. Estimated parameters for calling activity
Figure 8. Mean normalised amplitudes of backpack recordings during “no noise” (light grey bars) and “noise” condition (dark grey bars): Baseline background levels (n = 10) and loudest syllables recorded by the Non-focal (n = 15) and Focal bird’s backpack (n = 15). In both noise conditions, the syllables recorded by the Focal bird’s backpack were much higher in amplitude than those recorded by the Non-focal ones.
Figure 9. PSTHs of the stack-call interactions of three birds in the “no noise” (left) and “noise” (right) condition. In each PSTH, the vertical dashed line represents the onset of the call given by the reference bird (symbol inside white square), and the coloured area represents normalised binned vocalisation onsets of the other bird (symbol inside corresponding coloured square). Red horizontal lines represent the upper and lower confidence intervals (see Methods).

<table>
<thead>
<tr>
<th>Day</th>
<th>Lower</th>
<th>Upper</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.866373</td>
<td>161.356</td>
<td>26.50014</td>
</tr>
<tr>
<td>2</td>
<td>42.61392</td>
<td>430.3366</td>
<td>186.7778</td>
</tr>
<tr>
<td>3</td>
<td>40.98511</td>
<td>432.8327</td>
<td>186.7778</td>
</tr>
<tr>
<td>4</td>
<td>122.9156</td>
<td>641.5026</td>
<td>330.0278</td>
</tr>
<tr>
<td>5</td>
<td>94.61405</td>
<td>570.4926</td>
<td>283.3611</td>
</tr>
<tr>
<td>6</td>
<td>107.7157</td>
<td>603.1087</td>
<td>306.25</td>
</tr>
<tr>
<td>7</td>
<td>159.1418</td>
<td>721.5347</td>
<td>393.3611</td>
</tr>
<tr>
<td>8</td>
<td>13.499</td>
<td>324.9939</td>
<td>117.3611</td>
</tr>
<tr>
<td>9</td>
<td>94.55955</td>
<td>567.7036</td>
<td>283.3611</td>
</tr>
<tr>
<td>10</td>
<td>148.5686</td>
<td>691.938</td>
<td>373.7778</td>
</tr>
<tr>
<td>11</td>
<td>175.5833</td>
<td>742.969</td>
<td>413.4444</td>
</tr>
<tr>
<td>12</td>
<td>199.6481</td>
<td>792.922</td>
<td>448.0278</td>
</tr>
<tr>
<td>13</td>
<td>209.6277</td>
<td>833.1265</td>
<td>469.4444</td>
</tr>
<tr>
<td>14</td>
<td>210.0272</td>
<td>829.9171</td>
<td>469.4444</td>
</tr>
<tr>
<td>15</td>
<td>197.6279</td>
<td>791.9319</td>
<td>448.0278</td>
</tr>
<tr>
<td>16</td>
<td>56.94039</td>
<td>469.3772</td>
<td>215.1111</td>
</tr>
<tr>
<td>17</td>
<td>237.8173</td>
<td>879.1853</td>
<td>506.25</td>
</tr>
<tr>
<td>18</td>
<td>178.1898</td>
<td>763.7226</td>
<td>420.25</td>
</tr>
<tr>
<td>19</td>
<td>223.261</td>
<td>852.2261</td>
<td>484</td>
</tr>
<tr>
<td>20</td>
<td>242.4351</td>
<td>882.3769</td>
<td>513.7778</td>
</tr>
</tbody>
</table>

Table 3. Estimated parameters for locomotor activity
CHAPTER 2

Patterns of call communication between group-housed zebra finches change during the breeding cycle

Lisa F. Gill*, Wolfgang Goymann, Andries ter Maat, Manfred Gahr

2 Gill et al., 2015. Patterns of call communication between group-housed zebra finches change during the breeding cycle. eLife 4. doi:10.7554/eLife.07770
SECTION 11

Abstract

Vocal signals such as calls play a crucial role for survival and successful reproduction, especially in group-living animals. However, call interactions and dynamics within groups remain largely unexplored because their relation to relevant contexts or life-history stages could not be studied with individual-level resolution. Using on-bird microphone transmitters, we recorded the vocalisations of individual zebra finches (*Taeniopygia guttata*) behaving freely in social groups, while females and males previously unknown to each other passed through different breeding stages. As birds formed pairs and shifted their reproductive status, their call repertoire composition changed. The recordings revealed that calls occurred non-randomly in fine-tuned vocal interactions, and decreased within groups while pair-specific patterns emerged. Call-type combinations of vocal interactions changed within pairs and were associated with successful egg-laying, highlighting a potential fitness relevance of calling dynamics in communication systems.
SECTION 12

Introduction

Vocal communication plays an important role for a variety of social animals, because it is often directly linked with individual survival and successful reproduction. Vocal signals are especially important in group-living species, because they can be used to maintain group cohesion and coordinate common activities (reviewed by Fichtel and Manser (2010)), but also to recognise, locate and interact with specific individuals inside such groups (Jouventin et al. 1999a,b; Aubin and Jouventin 2002; Balsby et al. 2012; Ter Maat et al. 2014). However, in songbirds, most vocalisation studies have focused on male song and its relationship with hormones and reproduction (Nottebohm et al. 1987; Perez et al. 2012; Gahr 2014) in solitary, territorial, temperate-zone species (Marler 2004). But songbirds also produce calls in a variety of contexts (Marler 2004), sometimes in very large numbers throughout the day (Beckers and Gahr 2010). The usage and function of such calls is still unknown, mainly because it has been challenging to investigate with individual-level resolution naturally occurring vocal interactions between group members, in relevant contexts or different life-history stages.

It has been hypothesized that zebra finches (Taeniopygia guttata), which are group-living, socially monogamous and opportunistically breeding songbirds (Zann 1996; Perfito et al. 2007), share a different form of vocal communication with their lifelong partner compared to other members of their group (Zann 1996). In this species, both sexes produce diverse calls in large numbers, in various social contexts (Zann 1996; Beckers and Gahr 2010; Ter Maat et al. 2014), and depending on group structure (Elie et al. 2011). Some soft calls, i.e. low-amplitude vocalisations used in close-range signalling (Dabelsteen et al. 1998), have been suggested to play a role in pair communication (Zann 1996; Elie et al. 2010; Ter Maat et al. 2014). However, until recently, it was not possible to record and reliably assign all calls of individual zebra finches in the presence of their mates and within a group (Elie et al. 2011). Because calls are short and may be low in amplitude, especially when used at close range, earlier studies often resorted to strongly reduced social contexts or impoverished environments to investigate vocalisations at the individual level (Blaich et al. 1996a; Vignal et al. 2004; Anismov et al. 2014). Hence, these studies mainly addressed mechanistic questions of vocal behaviour,
but to understand the underlying meaning of calling, individual-based information in a socially relevant context is necessary. Our aim was thus to investigate with individual-level resolution the calling behaviour of zebra finches that were behaving freely in social groups, in a changing environment. We aimed at studying the dynamics of different call types and their usage – on the individual level as well as in interactions between mates and other group members – in relation to reproductive state and successful egg-laying. To do this, we housed groups of four females and four males previously unknown to each other together in large aviaries for about three weeks, provided them with nest material, and continuously recorded individual vocalisations of all group-members using microphone telemetry (= on-bird microphone transmitters). While the birds formed pairs and passed through different stages of their breeding cycle, we recorded vocalisations, performed behavioural observations, took blood samples for hormone determination, and monitored their nests to measure reproductive performance.
SECTION 13

Results

When housed together in social groups in large aviaries (Fig. 18), female and male zebra finches that were equipped with individual on-bird microphones (Fig. 10) and that were previously unknown to each other, formed pairs and started to breed after nesting material had been added (Fig. 11). At the same time, the call repertoire (Fig. 12) of individually recorded birds changed (Figs. 13, 20), and vocal interactions in the group became increasingly pair-specific (Figs. 14, 21). In within-pair calling interactions, the combination of call types involved changed (Fig. 14), and was related to successful egg-laying (Fig. 15).

Figure 10. Position of on-bird microphones on freely behaving zebra finches. Close-up pictures of backpacks fitted on zebra finches in an aviary. The visible white “backpack” contains a microphone (at the bottom), a radio transmitter and a battery. It is placed on the bird’s back (centre of mass) in a way that the microphone faces inwards, i.e. towards the bird’s body. Unlike the backpacks, harnesses disappear under the birds’ feathers (left and right panel), therefore, to demonstrate how they were fixed on the birds, the two silicon loops around the head and the abdomen are represented by dashed lines in the left panel. They were closed at the front of the bird with a small knot.
13.1. Pair formation and reproductive stages

30 out of 32 birds formed pairs (for definitions, see methods, Tables 1 and 2) within one to seven days, and began occupying nest boxes within 5.6 ± 3.3 days (mean ± standard deviation [SD]). The birds were not entirely synchronized in their reproductive stages and some nest-building began even before actual nest material was provided (using scraps of food or single threads from backpack material). But the addition of proper nest material triggered nesting activities in all birds but one, resulting in a mean onset of nest-building at 7.3 ± 2.4 days. Two categories of reproductive status, “nest stage” and the more detailed “breeding stage” resulted from our behavioural observations and nest checks (see methods) and are depicted in an overview of reproductive activity over the first 20 days (Fig. 11). Hereby, “nest stage” reflected large-scale changes of reproductive stages, and was confirmed by correlated changes in sex steroid levels (Fig. 19, see Appendix 1). Zebra finches reproduce as soon as environmental conditions permit to ensure successful breeding (see discussion). Therefore, birds that produced a clutch of eggs during the trials were termed “successful” and those that did not as “unsuccessful” at egg-laying. Pairs had eggs after 11.6 ± 5.8 days and began incubating them after 18.3 ± 5.1 days.

13.2. Vocalisation types and reproductive stages

Apart from calls related to parental behaviour (“thucks”), we found all vocalisation types described by Zann [1996]. The five most frequently used and distinct call types (“distance calls”, “stacks”, “tets”, “cackles” and “whines”, Fig. 12) were used for further analyses (182,752 calling events).

Within these five call types, we found differences in the number of vocalisations uttered per bird and recording in relation to reproductive stage and sex (Figs. 13, 20, Appendix 2, see methods for sample sizes). For distance calls, cackles and whines, the numbers of vocalisations changed for males and females in the same way (no interaction between “nest stage” and sex). Loud distance calls were produced most when birds were not yet paired or nesting, with highest levels during pre-nesting, that then decreased during the early and again during the later nest stages (Fig. 13). Fstage = 18.5, Fsex = 1.13, $R^2_{marginal} = 0.20$, $R^2_{conditional} = 0.37$; see methods for an explanation of $F$, $R^2_{marginal}$, $R^2_{conditional}$). Cackles and whines increased at the onset of reproductive activities, with cackles showing a peak in both sexes during the early nest stage (Fstage = 3.80, Fsex = 2.28, $R^2_{marginal} = 0.08$, $R^2_{conditional} = 0.33$; Fig. 13), and whines showing a peak during early and later nest stages (Fstage = 8.34, Fsex = 1.58, $R^2_{marginal} = 0.12$, $R^2_{conditional} = 0.24$; Fig. 13). Tets did not change throughout the three nest stages for either sex (Fstage = 0.49, Fsex =
Figure 11. Group reproductive stages change over time. Percentage of birds (nbirds = 32) assigned to the seven detailed breeding stages (coloured bars) and the three corresponding condensed nest stages over the first 20 days of the trials. Earliest onset of nest-material provisioning (day 7) is indicated by a green dashed vertical line.

Figure 12. Call types used in our study. Example spectrograms of female (A1) and male (A2) distance, tet (B), stack (C), cackle (D) and whine (E) calls. x-axis: time [ms], y-axis: frequency [Hz].

\[ R^2_{\text{marginal}} = 0.01, R^2_{\text{conditional}} = 0.34; \text{Fig. 13} \]. Stack calls showed an interaction between nest stage and sex: the amount did not change for males (Fstage = 0.06, \( R^2_{\text{marginal}} = 0.001, R^2_{\text{conditional}} = 0.42; \text{Fig. 13} \)), but in females, stacks were produced slightly more often during the early nest stage than during the pre-nesting stage (Fstage = 2.37, \( R^2_{\text{marginal}} = 0.05, R^2_{\text{conditional}} = 0.31; \text{Fig. 13} \)). In sum, this shows that call-type usage, i.e. the repertoire composition, changed...
at the individual level over the breeding cycle. Because the recordings were gained in temporal synchrony between all group members (see methods, Fig. 17), our study also allowed investigating how individual birds used calls to interact with other individuals of the group.

### 13.3. Vocal interactions, reproductive stages and successful egg-laying

These interaction-level data indicated that vocal networks were dynamic, differed in pair and group communication, and were related to the breeding stages. Peri-stimulus time histograms (PSTHs, for details see methods and [Ber Maat et al. 2014](#)) that compare the onsets of the birds’ vocalisations relative to each other revealed that calls did not occur randomly. Instead, in many cases, the calling behaviour of a specific bird elicited significant changes in the calling behaviour of another bird within a time frame of 0.5 seconds (relative to each dyad’s baseline). The resulting correlation indices (see methods for calculation and sample sizes) were plotted in confusion matrices showing all possible combinations of birds and call types (Fig. 14A). These matrices demonstrated dynamic interaction patterns between the birds: while birds shared many “significant interactions” with other birds in various call types at the beginning of the group trial, interactions decreased and became more and more specific with progressing reproductive activities (Fig. 14A). On the day that nest material was provided, the diagonal between the top left and the bottom right lit up (Fig. 14A), suggesting that an interaction pattern emerged that was pair-specific and synchronous within each group. Two of the 10 pairs did not show this vocal interaction pattern (pairs L and F in Fig. 14A), and also did not lay eggs.

In total, out of all possible combinations between all call types from all individuals (within-bird interactions excluded) only 6.5% resulted in significant interactions (see methods for definition). 4.8% showed positive values, that is the calls of one bird led to an increase in the calls of another bird, and 1.7% showed negative values, that is the calls of one bird led to a decrease in the calls of another bird. Within pairs (n=10), 9.2% of the possible interactions were significant, with 8.3% being positive and 0.9% negative. Further, the data suggest that vocal activity and the amount of vocal interactions decreased with progressing breeding stages, but the ratio of vocal interactions with the partner compared to those with other group members showed a 5-6 fold increase (from 0.74 when unpaired to 4.24 during incubation; Fig. 21).

Within pairs, not all possible call-type combinations were used in significant positive vocal interactions, for example, distance calls were never used in combination with whines. The highest percentages of within-pair calling interactions...
Figure 13. Female and male call-type usage at different nest stages. Boxplots of the number of vocalisations (natural log- or square-root transformed) per four hours of recordings for the different vocalisation types in relation to the three Nest stages, analysed separately for females (red, nfemales = 12) and males (blue, nmales = 10) from the three trials. They show that call types change differently over Nest stages: distance calls decrease, and cackles and whines increase. Thick black horizontal line = median of observations, box = 25% - 75% quantile of the observations (length = interquartile range), whiskers = last observation within 1.5 times the interquartile range, circles = observations farther than 1.5 time the interquartile range from the edge of the box, coloured point = fitted value (Bayesian estimate), coloured vertical bar = 95% credible intervals (CrI) of the fitted value. If Bayesian estimates (coloured points) and CrI (vertical coloured lines) do not overlap inside single plots, there is a difference in the number of vocalisations used in relation to Nest stage. Such differences are indicated by different letters at the top of each box. Sample sizes during Pre-, Early and Later Nest stage were: 24, 20 and 35 data points coming from 9, 10 and 12 females, and 23, 18 and 26 data points coming from 8, 8 and 10 males.
took place in tets, stacks and cackles (Fig. 14B). Same-call interactions (bottom left to top right diagonal in Fig. 14B) were not more common than interactions between different call types. However, same-call interactions were more symmetrical between females and males, and changed over the breeding stages almost in the same way for both sexes. In contrast, different call-type interactions were less symmetrical between the sexes. For example, tets were more likely to be answered by stacks when the responding bird was a female, and stacks were more likely to be answered by tets or cackles when the responding bird was a male, especially at the onset of nesting activities (Fig. 14B). In this case, the asymmetries thus changed over the breeding stages, with a peak at “nest inspection”. Breeding stage thus had an effect on different call-type combinations.

The number of positive within-pair calling interactions was not only related to reproductive state, but also to whether or not a pair succeeded in producing a clutch of eggs (“successful egg-laying”). The number of call-type combinations with significant interactions increased over reproductive stages for pairs that laid eggs, but failed to do so for pairs that did not lay eggs (Fig. 15). This means that pairs involving in more call-type interactions at certain stages were more likely to produce a clutch of eggs. Successful pairs only shared significant interactions in few call-type combinations before nesting (1.4 ± 0.83 call-type combinations ± SD) and increased these interactions during the later nest stage (8.92 ± 4.41). Unsuccessful pairs, on the other hand, showed decreasing numbers and also higher levels of variation throughout the breeding stages (from 4.48 ± 3.35 to 1.25 ± 1.5, respectively), suggesting a less specific usage of call types in interactions.
Figure 14. Vocal interactions within groups across reproductive stages. **A**. Vocal interaction matrices Examples of vocal correlation indices (from -1 to 1, see colour scale) resulting from PSTHs for all bird and call-type combinations during different phases of the trials (different days indicated above each box), for trials I, II and III (nbirds = 6, 8 and 8). All initiating birds (x-axis) and responding birds (y-axis) are represented by capital letters (pink: females, blue: males) and are subdivided into the five call types. Note that grey squares (= zero values) indicate there was no significant interaction in the respective dyad, and does not mean there were no vocalisations (see methods). Same capital letters indicate members of a pair, and within-pair interactions can be found in the diagonal from top left to bottom right. Note an increase in within-pair interactions and a decrease in overall group interactions with progressing reproductive stages (left to right). Inserts in Fig. 14A (lower right corner) explain the different interaction levels in the group (highlighted in pale blue) and the call-type interactions (highlighted in pale yellow). The dark grey diagonal from bottom left to top right represents within-bird interactions which were excluded from the analyses. Same-sex interactions are emphasized by pink (female-female) or blue (male-male) outer lines. In trial I, white squares represent missing values. The dataset is available at http://datadryad.org/review?doi=10.5061/dryad.vt69s. **B**: Within-pair vocal interactions at different breeding stages Summary graph of positive within-pair calling interactions in relation to different call-type combinations, sex and the detailed breeding stages (npairs = 10). Initiating birds' call types are plotted on the x-axis and percentages of positive responses (pink: females, blue: males) are plotted on the y-axis, in the corresponding call types. Note that both females and males were initiating and responding birds.
Figure 15. Call-type combinations associated with nest stages and successful egg-laying. Boxplot of within-pair number of combinations with significant interaction (positive) over Pre- (nPre = 8), Early (nEarly = 8) and Later Nesting (nLater = 10) for pairs that were successful (orange, n = 6) and unsuccessful (grey, n = 4) at producing a clutch of eggs within the three-week trials (Later Nest stage here refers only to Nest-building). Note the increase in call-type interactions of successful pairs across the Nest stages. Thick black horizontal line = median of observations, box = 25% - 75% quantile of observations (length = interquartile range), whiskers = last observation within 1.5 times the interquartile range the edge of the box, black dots = observations farther than 1.5 time the interquartile range from the edge of the box.
Discussion

Our study showed with individual-level resolution that call-based vocal communication of group-living zebra finches changed across reproductive stages. Using the temporal information encoded in call onsets, we found that the timing of calls was not random but instead occurred in significant vocal interactions between individuals of social groups. Both individual-level call-type usage and calling interactions between mates and other group members changed with breeding stages. The quality of calling interactions between pair members during several reproductive stages was correlated with successful egg-laying.

14.1. Calling behaviour in social environments: new approach, new results

In songbirds, there is increasing evidence that not only song, but also calls can play a role in reproduction (Groth, 1993; Marler, 2004; Elie et al., 2010; Ter Maat et al., 2014). However, the usage and function of call repertoires, especially in group-living species, has been unknown, so far. The zebra finch is one of the prime model organisms for studies on vocal communication, especially with regard to song. Its vocal repertoire, including song and different call types, has been described in most detail by Zann’s laboratory and field observations (Zann, 1996) to which we found various parallels in our data. For example, as suggested (Zann, 1996), loud distance calls occurred most before birds were paired or nesting, and cackles and whines increased at the onset of breeding activities (Figs. 13, 20). However, in most previous analyses of zebra finch vocal behaviour, technical limitations constrained a reliable separation between individual sound sources when birds behaved in social contexts involving direct contact with multiple individuals (Zann, 1996; Elie et al., 2010, 2011). This was especially relevant when birds vocalised quietly and in close proximity to each other (Elie et al., 2010). Although it has been suggested that the vocal output of entire zebra finch groups depends on group structure (Elie et al., 2011) and that quiet calls may play a role in pair formation (Zann, 1996; Elie et al., 2010; Ter Maat et al., 2014), these previous studies did not investigate calling interactions inside groups with individual-level resolution. For instance, it had been stated that quiet tet calls are produced...
almost at all times, are not directed at specific individuals, and therefore do not stimulate specific replies (Zann, 1996). Instead, we found that tet calls did not occur randomly, but in meaningful interactions between individual birds. Other studies recorded individual vocalisations, but resorted to strongly reduced social contexts and environmental enrichment, often coupled with frequent disturbances through bird handling (e.g. Blaich et al., 1996a; Vignal et al., 2004; Anisimov et al., 2014). Therefore, despite slightly different vocalisation classification paradigms (Vignal et al., 2004; Elie et al., 2011; Anisimov et al., 2014; Ter Maat et al., 2014), between-study differences in the birds’ vocal repertoire contents are most likely due to social context, as this can impact a multitude of physiological and behavioural aspects which in turn may be linked to vocalisations, e.g. high rates of stack-call production in isolated birds (Zann, 1996). Our setup allowed combining a species-relevant context with recording techniques that ensured longer-term individual recordings of freely behaving birds with infrequent bird handling. This enabled us to explore new aspects of vocal communication, including functional aspects of vocal interactions, as discussed below.

14.2. Opportunistic breeding and vocalisations

Zebra finches are opportunistic breeders, and in arid habitats, they rely on short and unpredictable periods of rainfall to successfully rear their young. Therefore, birds need to start breeding immediately when environmental conditions permit (Zann et al., 1995; Prior et al., 2013), and may do so throughout the year, in the wild as well as in captivity (Zann, 1996; Perfito et al., 2007; Perfito, 2010). In our study, newly joined birds quickly formed pairs and increasingly engaged in reproductive activities which were correlated with increased concentrations of gonadal hormones (see Appendix 1). In parallel, shifts in reproductive stages were associated with changes in calling behaviour. First of all, at the individual level, the call repertoire changed, sometimes showing sex-specific patterns. Second, vocal networks were dynamic, and showed increasing differences between pair and group communication. The most synchronous and therefore most apparent changes in reproductive stages occurred at the sudden onset of nest material which was accompanied by nesting behaviours in most birds, and was reflected by a pair-specific pattern lighting up in the vocal interaction matrices. Intriguingly, the quality of within-pair vocal interactions was associated with successful egg-laying. Although all pairs engaged in nesting behaviours at some point, those sharing interactions in more call-types during later nest stages were more likely to succeed in producing a clutch of eggs during the three-week trials. In addition, the different levels of variation, especially before the onset of breeding activities (pre-nesting), suggest that successful pairs were more specific in their call-type usage.
in interactions, demonstrating the importance of call types in pair communication during the breeding cycle. Because in the wild, birds need to start breeding immediately with the unpredictable onset of rain (see above), such rapid changes in fine-tuned pair communication could be essential for successful reproduction. Our findings thus offer an additional aspect of opportunistic breeding behaviour by showing that changes in the environment, leading to changes in reproductive state, were accompanied by transient changes in the calling behaviour inside groups, involving a shift towards dyadic pair communication.

In zebra finches, song has been shown to be important for mate choice and pair formation but to lose significance once a stable pair bond is established (Adkins-Regan and Tomaszycki, 2008). Our findings not only support that calls and calling interactions between mates are important for pair formation (Elie et al., 2010; Ter Maat et al., 2014), but also suggest an important role for successful reproduction (egg-laying). In this species, song is produced only by males, and thus constitutes a unilateral signal. Calls, on the other hand, are produced by both sexes and can be exchanged bilaterally. Therefore, they have the potential to be used in mutual and more complex behavioural interactions supporting pair formation and synchronisation as well as pair-bond maintenance, as suggested for other mutual behavioural displays (reviewed by Bradbury and Vehrencamp, 2011). Hereby, some call types are used more frequently in vocal interactions than others. We therefore suggest that for rapid calling exchanges especially the soft and short tet, stack and cackle calls may be more suitable than longer and more variable calls, such as whines.

It remains to be explored whether or how gonadal hormones affect calling behaviour across changing reproductive stages (see Appendix 1). Due to large differences in the temporal resolution of sampling methods for hormones and behaviour, a direct comparison between concentrations of circulating hormones and of call communication is difficult (see Appendix 1 for a more detailed discussion of hormones and calling dynamics in groups). Next to hormones, the decision of an initiating bird to produce a certain call and of a responding bird to answer with a certain call may depend on the specific context, on the behaviour of the other group members, or on the responder’s previous experience (as suggested for social bonding in other bird species, Vignal et al., 2004; Emery et al., 2007). To date, we can only speculate on how such vocal interaction patterns are established in this songbird species. Learning might play a role in females and males, because both sexes initiated calling interactions and responded to calls in an increasingly synchronised pattern during progressing reproductive stages. Also, it remains to be investigated whether call-type usage between potential mates supports assortative mating (Moravec et al., 2006), or has a direct effect on fertility,
as suggested for specific vocalisations in budgerigars (Melopsittacus undulatus) (Brockway 1965, 1967). Physiological and behavioural synchronisation of pair members have been suggested to play an important role in successful reproduction (Wickler and Seibt 1980, Cheng 2003, Hirschenhauser et al. 2008, Ouyang et al. 2014). To our knowledge, ours is the first study to demonstrate an association between intra-pair calling dynamics and successful egg-laying, thus highlighting the potential fitness relevance of calls and calling interactions in communication systems.
SECTION 15

Materials and Methods

15.1. Study design, housing and nest material

Between November 2011 and August 2012, four group-trials of about three weeks each were carried out in succession on a total of 32 adult zebra finches (4 females and 4 males each) which were fully adult offspring from our breeding colony. Future group members of the opposite sex had never previously met, i.e. had been raised in separate rooms, and siblings were not included in the same trials. We also ensured that birds had not bred before. Birds were individually recognised by one numbered and two coloured leg bands, and were kept in a 14:10 hours light:dark cycle with ad-libitum access to water and seeds and additional greens and egg-food.

The birds were caught from large same-sex aviaries and equipped with on-bird audio transmitter “backpacks” (Fig. 10). They were subsequently held in smaller same-sex aviaries (170 x 165 x 80 cm) for habituation to the backpacks before the beginning of the trials. Different habituation phases have been reported in similar studies, which is likely due to differences in the weight of the backpacks applied to the birds. Reported backpack weights on captive zebra finches ranged from 0.6g (Gill et al., 2016) to 3g (Anisimov et al., 2014), and resulted in habituation phases between a few days (one day for movement, four days for call rates, (Gill et al., 2016) up to about two weeks (Anisimov et al., 2014). Our backpacks weighed 1g, and based on prior observation (unpublished data), we chose a one-week habituation period to ensure birds had fully recovered from any behavioural effects of the backpacks and any associated handling.

Each trial began by placing the four males and the four females inside a large aviary (Fig. 18 for timeline see Fig. 8). This mixed-sex aviary (2m x 2m x 2.5m) contained four large perches, four empty nest boxes and eight antennas protruding vertically into the top area of the cage, thus offering additional perching opportunities (Fig. 18). One week after the beginning of the recordings, nest material (coconut fibres and soft white lining) was provided and recurrently refilled. While the birds went through different breeding stages, they were blood-sampled for hormone analyses (see Appendix 1), and non-vocal behaviours were recorded during regular observations. Vocal behaviour was recorded almost continuously.
throughout the day by means of microphone telemetry, and exemplary morning sound recordings were analysed (see methods on sound analyses). Behavioural observations and handling of technical equipment were carried out from behind a large, green curtain inside the experimental room, and to control for human disturbance, the observer quietly entered the room at least 10 minutes before each observation period. Animal care, nest checks as well as any handling of birds, took place outside of recording periods (see below).

For behaviour and hormone analyses, the data from all four trials (0, I, II and III) were used (n = 32). For sound analyses, only trials I, II and III were used, because trial 0 served as a test run for the sound recordings and experimental design.

![Timeline of trials](image)

**Figure 16.** Timeline of trials. Timeline indicating housing conditions and approximate timing of: backpack application, habituation phase (minimum seven days), beginning of trial and sound recordings (ca. three weeks, see methods), onset of nest material availability, end of trial. Of the continuous sound recordings, mornings (220 ± 20 minutes) of different days (differed between trials, not indicated in graph) equally representing birds’ breeding stages were analysed. Blood sampling occurred three times (indicated by red arrows): before the beginning of the trials, one day after males and females were joined and one day after nest material became available.

### 15.2. Sampling

**15.2.1. Behavioural observations, nest checks and blood sampling.** To establish which individuals formed pairs, to define their breeding stages, and to record agonistic and prosocial behaviour within the groups, behavioural observations were carried out for 30 minutes at least twice a day from day 1 to 4, at least once per day until day 10, and at least once every two days until birds had laid eggs (days 16 – 23). Prolonged close proximity and tactile contact, such as
clumping and allopreening, between zebra finches has been shown to be indicators of pair formation (Butterfield 1970; Silcox and Evans 1982; Tomaszycki and Adkins-Regan 2005). Therefore, the observer noted the behaviours listed in Table 1 by coding them directly into a prepared spreadsheet on a laptop computer. Behaviours of short duration, e.g. displacements, copulation solicitation or entering the same nest box, were recorded as frequencies, while longer lasting behaviours, e.g. clumping or incubating, were sampled once every two minutes and processed as durations. The given behaviour was recorded along with its time and location, and, if applicable, the identity of sender and receiver were noted (e.g. for displacements or allopreening).

Nest checks were carried out at least every second day between 7 and 9 hours from lights on when birds usually were relatively inactive. Nest checks included visual inspection of the nest material inside each nest box, counting the number of eggs and hatched chicks inside a nest, and individually marking them with a coloured pen. Sound recordings were paused during this time, and only those preceding nest checks were included in this manuscript. Behavioural observations were resumed a few hours after nest checks.

Blood samples to determine baseline hormone concentrations were taken at least one week before the beginning of the trial when the birds were still held in same-sex aviaries (see timeline in Fig. 10). They were bled again one day after the beginning of each trial, i.e. one day after being joined in the large mixed-sex aviary and one day after nest material had been added, a week later. To minimise any effects of blood sampling, we only included sound recordings from the same day as blood sampling, if they had been recorded before the procedure.

Birds were caught from the aviaries with hand-nets by two people and passed on to four other people who bled them. Hereby, the brachial vein was punctured using sterile syringes, and a small amount of blood (ca. 70 μl) was taken using a heparinised capillary tube, and subsequently transferred to a tube on ice. All blood samples were collected and stored on ice within 5 – 10 minutes of the initial disturbance of opening the door to the experimental room (245 ± 27 s). As soon as all birds had been processed, they were released back inside the aviary. Within 15 minutes, all tubes were brought to the lab where they were centrifuged at 3000 rpm at room temperature for 10 minutes. Plasma was separated from the blood cells and stored at -80°C.

15.2.2. Vocalisation recordings. To capture individual vocalisations in a group setting, we used microphones mounted directly on the birds (on-bird microphones), and hereby chose microphone transmitters (Sparrow System, Fisher, III, USA). This provided a number of practical advantages in this setting, compared to
using loggers (Anisimov et al., 2014), some of which we explain in the following. Due to the unified external recording process when using transmitters (see below), it was possible to gain recordings from all individuals simultaneously and in temporal synchrony (Fig. 17) within each group. Thus, birds could be housed in a relatively large three-dimensional environment (2 x 2 x 2.5 metres) including nest boxes and perches at different heights (Fig. 13), because transmitters do not require synchronisation to an external, e.g. visual signal (Anisimov et al., 2014). In addition, our devices – containing a miniature microphone, a radio transmitter and a button cell battery – were light (1g), and batteries were replaced only after 6 – 10 days, which strongly reduced the amount of disturbances associated with bird handling, compared to previous studies (e.g. Vignal et al., 2004; Anisimov et al., 2014).

Devices were fixed on the back of the birds by a thin silicon harness, with straps around the neck and abdomen (Fig. 10). In a different study (Anisimov et al., 2014), additional accelerometers were used to distinguish between vocalising individuals. However, as previously described (Ter Maat et al., 2014), directing the miniature microphone towards the bird’s body, i.e. facing inwards (Fig. 10) instead of facing outwards, strongly reduces the probability of recording vocalisations of non-focal individuals, because microphones are attached as closely as possible to the sound source, and external sounds are dampened by the backpack. In addition, due to differences in spectral properties (Fig. 17; also see methods section on sound analyses), any (rarely) recorded non-focal birds’ calls were easily detected and excluded during the sound analyses.

Backpacks were modified for females in such a way that the originally dorso-caudally protruding antenna was incorporated into the silicon loop to ensure that copulation remained possible. As this procedure tended to decrease radio signal strength, it was not done for males who exhibited normal copulation behaviour despite a protruding antenna. Eight communications receivers (AR8600, AOR, USA) set up outside the aviary were connected to the eight antennas inside the aviary and received the respective AM-modulated radio signals (transmission frequency: 375 – 380 mHz) coming from the microphone transmitters on the backs of the eight birds. A 16-channel AD-converter (Sonic Core A16 Ultra) was used to digitise the analogous signals. Using a custom multi-channel programme (ASIO Rec, Markus Abels, MPIO Seewiesen), we recorded all eight channels at a rate of 44100 Hz and stored them as uncompressed files (.wav) on a single computer to ensure temporal synchrony between them. Recordings were automatically started and stopped for all eight birds simultaneously and were made for at least four hours in the mornings and evenings every day. An exemplary subset of the recordings was analysed for this manuscript (see methods on sound analyses).
Figure 17. Synchronous external and on-bird recordings. Example spectrograms (x-axis: time [ms], y-axis: frequency [kHz]) of synchronous recordings with A: group recording from external microphone without individual information and B – D: individual recordings from three (out of eight) on-bird microphones. Dark vertical lines (in C and D) represent wing beats that hardly show up in the noisy external recording. Note the higher power in the low frequencies in the on-bird microphone recordings, compared to the external recording.

15.3. Analyses

15.3.1. Reproductive stages. To objectively describe pairs’ breeding stages, two sets of categories were created from the 30-minute observations and nest checks (Table 2, Fig. 11). The first one ("breeding stages") was in greater detail and was summarised to form the second one ("nest stages") which was required for statistical analyses.

For the evaluation of hormones at different reproductive stages, we used the results from 95 blood samples collected at the three time points (baseline n=31, one day after joining of birds n=32, one day after adding of nest material n=32) from the 16 male and 16 female zebra finches. Using a modified version (Goymann et al., 2008) of the radioimmunoassay method established by Wingfield and Farner (1975), we determined concentrations of testosterone (T), dihydrotestosterone (DHT), oestradiol (E2) and progesterone (P4). Samples were run in duplicates for each individual and measured in a single assay. The intra-assay variation of extracted chicken pools was 12.4% for T, 17.8% for DHT, 29.7% for E2 and 15.5% for P4, respectively. Steroid extraction efficiency (mean percentage ± standard deviation [SD]) was 70.9 ± 2.7 (T), 73.7 ± 4.3 (DHT), 59.9 ±4.0 (E2), 55.9 ±11.0 (P4), and the detection limits measured in pg per tube were 0.39, 0.43, 0.2 and 2.38, respectively. As 77 out of the 95 E2 samples were below the detection limit, this hormone was excluded from further statistical analyses.
15.3.2. Sound analyses. Due to technical issues, we had to reject the recordings from two out of four males from trial I. Six (trials I and III) or seven (trial II) sound files per trial were chosen post hoc according to the following rationale for analysis: The respective recordings were started within 2 hours of lights on, and continued for a maximum of four hours (i.e. finished before birds’ midday). To minimise any external effects on the vocalisations, sound files qualifying for analysis also preceded any major disturbances (e.g. nest checks or blood sampling, see below and Appendix 1). Because breeding stages were not synchronous between birds and groups, sound files were chosen for analysis so as to give a relatively balanced representation of all pairs’ breeding stages. In total, we analysed 146 sound files (67 for males, 79 for females) coming from up to 10 male and 12 female zebra finches (10 pairs) in the three different trials, summing up to about 535 hours of sound recordings.

In all of the following steps, temporal information was retained between individual recordings. To remove low frequency noise, e.g. originating from external technical equipment, we used a high-pass filter (200 Hz cut-off frequency; apple Audio Unit) in Amadeus Pro (2.0.5, HairerSoft, UK) on all recordings. We used custom programs (available at [https://github.com/ornith](https://github.com/ornith)) to detect, extract and timestamp sound events, to identify and remove noise (e.g. wing-flapping or noise from degraded radio signal) and non-focals’ vocalisations, to automatically define onset and offset of syllables, and to classify them via a k-means clustering paradigm ([Ter Maat et al.] 2014). During this procedure, the (rare) vocalisations that had been recorded from non-focal individuals were easily detected and removed, because they show less power in the low frequencies, compared to the focal birds’ recordings (Fig. 17). If a syllable’s start or end had been defined inaccurately (e.g. due to overlapping noise) and its type was therefore not identifiable or its actual onset did not coincide with the onset of the noisy interval, it was manually deleted. Thus, we retained as many syllables as possible for the analysis but reduced potential confounding artefacts in the temporal information of syllable onsets. As zebra finch males may incorporate some of their calls into their song ([Zann] 1996), we arranged all syllables in their order of occurrence before clustering and were thus able to separate similar vocalisation types used in song bouts from those occurring as single calls. The program compares each sonogram in an automatic process with respect to the following nine temporal-spectral parameters: duration, mean frequency and standard deviation (SD), mode frequency and SD, Wiener entropy and SD, first peak and SD. The output was manually refined by visual inspection of the spectrograms so as to reach a classification of different call types according to the descriptions by [Zann] (1996). As is described there, cackles, arcs and whines are associated with the onset of nesting activities and
might blend into each other - with cackles being the shortest and the most in-
flexed ones and whines the longest and the least frequency-modulated versions. Arcs lie in between these two types in terms of occurrence, duration and fre-
quency modulation. Therefore, we decided to use only the calls at the ends of the continuum, namely “cackles” and “whines” and to discard the “arcs” from further analyses. If it was not possible to unambiguously assign single calls to a category, they were moved to a different class which was not included in further analyses.
As the aim of our study was to investigate the role of calls in zebra finch groups and the clustering paradigm was optimized for calls, we did not include song in our statistical analyses. Instead, we chose “distance calls”, “stacks”, “tets”, “cackles” and “whines” (Fig. 12) because they were the most frequently used and easily classified call types. We found the same call types for males and females despite potential sex-specific differences in spectral features, as for example in the loud distance calls that contain learned frequency modulations in males but not in fe-
males (Simpson and Vicario, 1990) (Fig. 12A).

The numbers of vocalisations were calculated for each vocalisation type for each bird for each recording. As the durations of recordings were not always identical (220 ± 20 min), these values were extrapolated to the longest recording (243 minutes), to avoid working with fractions. For the overall calling activity, we summed up these extrapolated numbers of vocalisations in the five call types for each bird and recording period. In addition to this absolute measure of call usage, we divided the numbers of calls per type by the total sum of calls to obtain the relative proportion of a bird’s call types for each recording period.

15.3.3. Vocal interaction analyses. Custom software written in C++ [avail-
able at https://github.com/orbith] was used to merge the up to eight single res-
ult files per recorded interval. The data was thus stored in a large text file con-
taining the starting point of each vocalisation within the recorded period, its syl-
labile type and the sender’s identity. In R, (3.1.1, R Development Core Team) these tables were concatenated and supplemented with information on the respective breeding stages and hormone data. Previous examinations (unpublished data) of spontaneous calling interactions in zebra finch groups revealed that a large pro-
portion of vocalisations occurring within a few seconds of another bird’s calls in fact occurred within half a second. As described in (Ter Maat et al., 2014), for each recording period, vocalisation starting points of two birds were aligned to each other in peri-stimulus time histograms (PSTHs) to gain cumulative sums from which a vocal association index can be calculated. We, too, chose a time window of interest of four seconds before and four seconds after call onset for our PSTHs with a binwidth of 50 ms (i.e. 160 bins in total). The number of calls in
the first 0.5-second bins (Nbase) served as baseline and those in the 0.5-second bins after call onset (Nresponse) as the response. The correlation index of a “response” for each set of two birds relative to each other was calculated as follows:

\[ R_{\text{response}} = \frac{N_{\text{response}} - N_{\text{base}}}{N_{\text{response}} + N_{\text{base}}} \]

Using a 95% confidence interval, we determined whether the occurrences of calls in a specific combination of individuals and call types could be considered random or as a “significant interaction”. This will be referred to as a non-zero correlation index (see below for statistical analyses). In this way, vocal activity was considered in calculating the strength of significant interactions. For example, a strong positive correlation in a dyad (turquoise square in Fig. 5A) is not a result of high vocal activity, but of accurate timing of vocalisations between two individuals. Likewise, a value of 0 (grey square in Fig. 14A) means that in this particular dyad, the significance criterion was not met (there was no significant vocal interaction), but does not mean that no vocalisations took place. 50 msec after call onset were removed from the analyses to avoid crosstalk (i.e. in case the same birds had been recorded by both backpacks). Data were considered valid only if each bin contained on average one or more events (i.e. total > 160 events).

From the correlation indices, we generated confusion matrices for all combinations of the five call types for all combinations of birds in a group (40 x 40 possible combinations for trials II and III and 30 x 30 matrices for trial I due to missing values from two males) (Fig. 14A). We did not find significant interactions for all groups on all days: on 2 out of the 20 analysed days, only within-bird interactions occurred frequently enough to exceed the confidence interval.

To further analyse these values for the different breeding stages and call type combinations, we only used within-pair interactions (trials I – III), because group members did not always have the same breeding stages on the same days. For within-pair interaction data, we calculated the percentage of positive to all possible call combinations for males and females of each pair in each breeding stage and for each call-type combination. We also calculated the percentage of positive interactions for pairs that laid eggs during the trials and those that did not (overall, when unpaired, during nest inspection and nest-building). To investigate pair specificity of vocal interactions, we calculated the percentage of possible interactions at different breeding stages with the partner and the other group members (n = 16 birds, from trials II and III, due to incomplete data in trial I).

**15.3.4. Statistical analyses.** Statistical analyses were performed using R version 3.1.1 ([R Core Team](https://www.r-project.org) 2014). To account for our study’s repeated measure design, we used mixed models to analyse hormones (see Appendix 1) and vocal
activities over the breeding cycle. An information theoretic approach was used to analyse the relationship between pair call-type combinations and successful egg-laying (see below). As frequentist analyses do not allow an exact calculation of the degrees of freedom (Bolker et al., 2009), we chose a Bayesian statistical approach (with uninformed priors). This allows drawing inferences as meaningful differences between groups by evaluating the ranges of their 95% credible intervals (CrI, range of group means, at 0.95 certainty). Thus, we ran linear mixed-effects models (lmer) using the R package "arm" (version 1.7-05) on the different dependent variables, with bird identity as a random factor. Posterior means and CrI were calculated using the function sim (running 5000 simulations), and were compared to find meaningful differences between groups. Therefore, if the mean estimate of one group did not overlap with the CrI of another group, it could be inferred that these groups differed from each other. Posterior means and CrI were plotted alongside boxplots of the raw data for vocal activities and hormones in relation to nest stages, and differences are indicated by different letters above each box (Figs. 13, 19). For reasons of clarity, we provide a table containing these values for the more detailed breeding stages (Fig. 20, Appendix 2).

Graphical methods (plotting model residuals) were used to evaluate model fit and whether model assumptions were met. To improve residual distributions, some of the numeric dependent variables were natural-log or square-root-transformed (indicated in graphs). To describe the variance explained by our models, we provide F values (ratio of between-group to within-group variance) and marginal and conditional R-square values that range from 0 to 1 and describe the proportion of variance explained by the fixed and by the fixed and random effects combined, respectively. They were calculated according to Nakagawa and Schielzeth (2013). As we expected to find marked differences both in the hormone concentrations and in the vocalisations of males and females, sexes were analysed separately to keep the models as simple and powerful as possible. Including trial ID as a further explanatory variable did not affect the outcome of any statistical test, and therefore, it was removed from all models.

15.3.4.1. Hormones. To find out about the relationship between hormone levels and nest stages, we ran a model for each of the three hormones (log-transformed) with a four-factor variable including baseline levels and the three nest stages (pre-nesting, early, and later nest stages) as fixed effect and bird identity as random factor.

15.3.4.2. Vocalisation types and breeding activity. To get an overview over the usage of different call types during different phases of the breeding stage, we ran a model for each vocalisation type on the frequency of vocalisations per
recording period (extrapolated to the longest recording duration) in relation to the three-factor variable nest stage (pre-nesting, early, and later nest stages), again with bird identity as random factor. For a more detailed impression of the relative changes in call usage according to breeding stages, we ran another set of models on the proportion of vocalisations of one call type in relation to all call types over the more detailed variable ‘breeding stage’ as fixed effect (Appendix 2). This was done on all vocalisation types separately and always included bird identity as a random factor.

15.3.4.3. Within-pair interactions, breeding activity and breeding success. As data points were few (10 pairs, subdivided into groups), we decided on a compromise between pure description of the data and complex models accounting for repeated measures to get an idea of whether calling interactions between pair members at different stages of the breeding cycle were related with whether or not pairs would lay eggs within the three weeks of the trial (pair success). To be able to compare successful and unsuccessful pairs (Fig. 15), we only included breeding stages before the egg-laying phase, i.e. until “nest building”, as “egg-laying” and “incubation”, by definition, only occurred in “successful” pairs. If the same pairs were sampled multiple times per breeding stage, mean values were calculated. We then performed model selection according to Anderson (2008) on the number of positive interactions in relation to nest stage and success, assuming a poisson distribution. After checking model assumptions, we calculated corrected AIC values (cAIC) to select the best fitting model (lowest cAIC). This turned out to be the model including an interaction between the two explanatory variables.
SECTION 16

Ethical note and Acknowledgments

Ethical note
Animal housing and welfare were in compliance with the European directives for the protection of animals used for scientific purposes (2010/63/EU). Protocols were approved by the Government of Upper Bavaria.

Acknowledgments
We would like to thank Hannes Sagunsky and Markus Abels, Susanne Seltmann, Ingrid Schwabl and especially Lisa Trost for technical support, and Monika Trappschuh for helping with hormone analyses. Further, we thank Nicole Geberzahn and Wolfgang Wickler for their helpful comments on previous versions of the manuscript. We thank the International Max Planck Research School (IMPRS) for Organismal Biology for training and support.
Supplementary Materials

Source data for figures are available at http://datadryad.org/resource/doi:10.5061/dryad.vt69s

17.1. Appendix 1: Hormones in relation to reproductive stages

The steroid hormones testosterone, dihydrotestosterone (DHT), oestradiol and progesterone are known to be associated with different reproductive stages (Wingfield and Farner 1993). Therefore, we analysed plasma steroid concentrations of our birds at different time points to supplement behaviourally classified reproductive stages with information on the physiological status (for methods, see main manuscript). Reproductive status was correlated with hormone concentrations. Testosterone increased for females and males over the course of each trial: In females, testosterone concentrations were higher for the later nest stage compared to baseline levels ($F_{stage} = 1.6$, $R^2_{marginal} = 0.06$, $R^2_{conditional} = 0.43$). Males expressed higher testosterone concentrations in early and later nest stages compared to the baseline. Further, at the later nest stage, males had higher testosterone levels compared to the pre-nesting stage ($F_{stage} = 8.36$, $R^2_{marginal} = 0.21$, $R^2_{conditional} = 0.65$; Fig. 19). Dihydrotestosterone (DHT) did not change for females during the course of each trial ($F_{stage} = 0.28$, $R^2_{marginal} = 0.01$, $R^2_{conditional} = 0.37$; Fig. 19). For males, there were no differences in DHT concentrations between baseline and pre-nesting, or between early and late nest stages. DHT concentrations were, however, higher at the early and late nest stages than at baseline and pre-nesting ($F_{stage} = 5.46$, $R^2_{marginal} = 0.18$, $R^2_{conditional} = 0.57$; Fig. 19). Progesterone changed for both, females and males. In females, progesterone levels were higher during the later nest stage compared to all other stages ($F_{stage} = 9.39$, $R^2_{marginal} = 0.30$, $R^2_{conditional} = 0.52$; Fig. 19). In males, progesterone concentrations were higher during early nest stage compared to the baseline ($F_{stage} = 2.26$, $R^2_{marginal} = 0.13$, $R^2_{conditional} = 0.16$; Fig. 19). Newly joined birds quickly formed pairs and increasingly engaged in reproductive activities. These were correlated with increased concentrations of plasma testosterone in females and of androgens (DHT and testosterone) in males. Our results are
thus in line with studies showing that changes in the social and abiotic environment – for example the presence of receptive females or of nest material – are reflected by changes in plasma testosterone levels and the emergence of mating behaviours (Riters and Alger 2011). In addition, progesterone, associated with pair-bonding (Smiley et al. 2012) and reproductive behaviours in females, has been shown to be elevated especially around the egg-laying phase (Wingfield and Farner 1993; Sockman and Schwabl 1999), further corroborating our results. Whether the differences in calling behaviour over the changing reproductive stages reflect direct effects of gonadal hormones remains to be explored. Androgens are known to affect the abundance of vocalisations of many vertebrate taxa (review in Bass and Remage-Healey 2008), and in our study, changes in male androgen levels paralleled changes in particular call types across reproductive stages (decreasing distance calls and increasing cackles and whines). In females, there were no similar patterns in gonadal hormones and call usage, as only progesterone increased during the later breeding stage. Therefore, if gonadal hormones affected call usage in females and males, sex-specific hormonal profiles would mediate such dynamics. It is possible that gonadal hormones affected auditory preferences or auditory-motor interfaces, leading to a change in call usage via changed neural responsiveness (Metzdorf et al. 1999; Avey et al. 2008; Remage-Healey et al. 2010). However, it is unlikely that the slowly fluctuating gonadal hormones determine the closely timed vocal interactions. Rather, they may accompany changes in reproductive states at a larger temporal scale. Unfortunately, a direct comparison between concentrations of circulating hormones and of call communication cannot yet be drawn, due to extreme differences in the temporal resolution of reliable sampling methods available to date for hormonal data and for behavioural and acoustic information. Nevertheless, our findings show that hormones were correlated with changes in reproductive stages, which, in turn, were correlated with calling behaviour on various levels.

17.2. Appendix 2: Call types and more detailed breeding stages

To find out in more detail how the relative usage of vocalisation types might change during which part of the breeding cycle, we examined the effects of sex and of the seven detailed breeding stages on the proportion of vocalisations for each vocalisation type, i.e. the number of vocalisations of a specific vocalisation type during a recording divided by the number of all vocalisation types during the same recording period. Here, we found an interaction between sex and breeding stage for all vocalisation types except for distance calls (Fig. 20). In males and females, distance calls were used less in all other breeding stages than when birds were
‘unpaired’. In females, they also occurred more frequently when birds were during the ‘no nest’ stage (paired but without own nest) than during ‘nest-building’ and ‘female egg-laying’ (Fstage = 8.60, Fsex = 0.29, $R^2_{marginal} = 0.32$, $R^2_{conditional} = 0.40$). Tets showed two peaks during the breeding stages for females: one when birds were paired without own nest (‘no nest’) and one during ‘egg-laying’ (Fstage = 2.69, $R^2_{marginal} = 0.08$, $R^2_{conditional} = 0.20$). In males, tets occurred at intermediate levels during the “unpaired” and “no nest” stages, reached a peak during “nest inspection” and then decreased again (Fstage = 5.36, $R^2_{marginal} = 0.22$, $R^2_{conditional} = 0.39$). Female stacks occurred frequently during “unpaired”, “territorial” and “nest inspection” and were reduced during “no nest” and “nest-building” which was followed by another increase (Fstage = 2.27, $R^2_{marginal} = 0.09$, $R^2_{conditional} = 0.16$). For males, the proportion of stack calls was lower than for females and showed less prominent differences between breeding stages: there was a decrease over the course of the breeding stages. The highest peak, however, was at “female egg-laying” (Fstage = 2.09, $R^2_{marginal} = 0.08$, $R^2_{conditional} = 0.34$). Females used cackles least when they were “unpaired” and most during “nest-building” (mean estimate lay on the higher credible interval of “egg-laying”) (Fstage = 4.12, $R^2_{marginal} = 0.30$, $R^2_{conditional} = 0.41$). In males, the number of cackles gradually increased over the breeding stages until “territorial” and “nest-building” was reached, and then gradually decreased back to “unpaired” values (Fstage = 4.67, $R^2_{marginal} = 0.11$, $R^2_{conditional} = 0.52$). A similar pattern was found for whines in females (Fstage = 1.96, $R^2_{marginal} = 0.35$, $R^2_{conditional} = 0.11$) and in males (Fstage = 3.16, $R^2_{marginal} = 0.05$, $R^2_{conditional} = 0.63$): there was an increase in the proportion of whines over the breeding stages, reaching maximum values at “territorial” and “nest-building”, followed by a decrease back to “unpaired” levels (Fig. 20).

Video 1 Zebra finches behaving freely with on-bird microphones [http://dx.doi.org/10.7554/eLife.07770.017](http://dx.doi.org/10.7554/eLife.07770.017)

Example video and external audio recording of Zebra finches behaving freely inside aviary (partial view) on the day of nest material. In the video, note the small white objects on the birds’ backs (microphone backpacks) that allow normal behaviours, e.g. flight and collection of nest material. In the audio, note the soft, overlapping vocalisations, and wing beat sounds (see Fig. 17).
<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Description</th>
<th>Count/duration</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displacement</td>
<td>Focal bird arrives at another bird’s location forcing it to leave</td>
<td>count</td>
<td>Agonistic</td>
</tr>
<tr>
<td>Fighting</td>
<td>E.g., bill-fight, full body fight, chasing</td>
<td>count</td>
<td></td>
</tr>
<tr>
<td>Clumping</td>
<td>Birds sit in direct physical contact with each other</td>
<td>duration</td>
<td>Affiliative or sexual behaviour</td>
</tr>
<tr>
<td>Allopreening</td>
<td>One bird preens another bird</td>
<td>duration</td>
<td></td>
</tr>
<tr>
<td>Cop. solicitation</td>
<td>Female fans tail at male</td>
<td>count</td>
<td></td>
</tr>
<tr>
<td>Copulation</td>
<td>Male mounts female</td>
<td>count</td>
<td></td>
</tr>
<tr>
<td>Enter nest box</td>
<td>Birds enter the same nest box without fighting</td>
<td>count</td>
<td></td>
</tr>
<tr>
<td>Foraging</td>
<td>Bird is foraging on ground, feeding, drinking</td>
<td>duration</td>
<td>Neutral</td>
</tr>
<tr>
<td>Preening</td>
<td>Bird is self-preening</td>
<td>duration</td>
<td></td>
</tr>
<tr>
<td>Flying</td>
<td>Bird flies around in aviary</td>
<td>count</td>
<td></td>
</tr>
<tr>
<td>Incubating</td>
<td>Bird sitting inside nest box with eggs</td>
<td>duration</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Overview and short description of different agonistic, affiliative or sexual, and neutral behaviours and whether they were measured as occurrences (count) or every two minutes (duration).
### Table 5. Overview and short description of breeding stages and nest stages

<table>
<thead>
<tr>
<th>Breeding stage</th>
<th>Description</th>
<th>Nest stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaired</td>
<td>Bird does not show increased prosocial behaviour towards specific individual</td>
<td>Pre-nesting</td>
</tr>
<tr>
<td>No nest</td>
<td>Paired but without nest</td>
<td>Early nesting</td>
</tr>
<tr>
<td>Territorial</td>
<td>Pair defending nest site without nest material</td>
<td></td>
</tr>
<tr>
<td>Nest inspection</td>
<td>Pair inspecting different nest boxes</td>
<td></td>
</tr>
<tr>
<td>Nest building</td>
<td>Pair bringing nest material to nest box</td>
<td>Later nesting</td>
</tr>
<tr>
<td>Laying</td>
<td>Pair’s female laying eggs</td>
<td></td>
</tr>
<tr>
<td>Incubation</td>
<td>Pair members incubating</td>
<td></td>
</tr>
</tbody>
</table>
Figure 18. Group housing. Overview of experimental room including aviary and technical set-up.
Figure 19. Differences in steroid hormone concentrations at baseline levels and the three nest stages. Boxplots of testosterone (TESTO), dihydrotestosterone (DHT) and progesterone (P4) concentrations with natural log-transformed values at baseline (Base) and the three nest stages Pre-, Early and Later Nesting, analysed separately for the 16 females (red, nBase = 16, nPre = 11, nEarly = 5, nLater = 13) and 16 males (blue, nBase = 15, nPre = 12, nEarly = 5, nLater = 12) from four trials. Bayesian estimates are plotted as coloured points, and 95% credible intervals as coloured lines. For an explanation of abbreviations, lines and symbols see Fig. 4. For methods, detailed results and discussion of hormones in relation to nest stages see Appendix 1.
Figure 20. Changes in call repertoire at more detailed breeding stages. Boxplots of the proportion of call types (y-axis) for females (top, nfemales = 12) and males (bottom, nmales = 10) over the detailed breeding stages (x-axis, UP: Unpaired, NN: No nest, Inspec: Nest inspection, Terr: Nest defence, Build: Nest building, Lay: Egg-laying, Incub: Incubation). Sample sizes during different breeding stages were 10, 14, 8, 12, 18, 10 and 7 data points coming from 7, 7, 6, 8, 12, 7 and 5 different females, and 10, 13, 6, 12, 15, 6 and 5 data points coming from 7, 6, 4, 8, 10, 5 and 3 different males. Bayesian estimates and CrI are plotted as coloured points and vertical bars (red: females, blue: males). For further explanations and abbreviations see Fig. 4. Note the different scale and missing Bayesian statistics for “whines” (right), as these consisted a comparably small part of the repertoire and included many zero values, especially at the earlier breeding stages. A table showing all Bayesian estimates and credible intervals can be found in Appendix 2.
Figure 21. Increasing specificity of within-pair vocal interactions. Vocal activity (number of distance, tet, stack, cackle and whine calls, green points = means, green lines = standard errors) and specificity of vocal interactions within pairs (red bars) and with non-pair members (grey bars) over the breeding stages (UP: Unpaired, NN: No nest, Insp: Nest inspection, Terr: Nest defence, Build: Nest-building, Lay: Egg-laying, Incub: Incubating). Note that the pair specificity of vocal interactions increased across breeding stages, and after birds had paired, they always shared more interactions with their partner than with other group members. Also note that for this graph, only trials II and III (n = 16 birds) were included due to the missing values in trial I (would allow investigation of within-pair interactions but not of vocal interactions from the entire colony). Sample sizes were thus 16, 25, 12, 14, 22, 12 and 10 observations coming from 10, 8, 8, 12, 16, 10 and 6 focal birds.
CHAPTER 3

On-bird sound recordings: automatic acoustic recognition of activities and contexts

Dan Stowell*, Emmanouil Benetos, Lisa F. Gill

3(submitted)
SECTION 18

Abstract

We introduce a novel approach to studying animal behaviour and the context in which it occurs, through the use of microphone backpacks carried on the backs of individual free-flying jackdaws (*Corvus monedula*). These sensors are increasingly used by animal behaviour researchers to study individual vocalisations of freely behaving animals, even in the field. However such devices may record more than an animal’s vocal behaviour, and have the potential to be used for investigating specific activities (movement) and context (background) within which vocalisations occur. To facilitate this approach, we investigate the automatic annotation of such recordings through two different sound scene analysis paradigms: a scene-classification method using feature learning, and an event-detection method using probabilistic latent component analysis (PLCA). Results are comparable with the state of the art in sound scene analysis; we find that the current recognition quality level enables scalable automatic annotation of audio logger data, given partial annotation, but also find that individual differences between animals and/or their backpacks limit the generalisation from one individual to another. we consider the interrelation of ‘scenes’ and ‘events’ in this particular task, and issues of temporal resolution.
SECTION 19

Introduction

Studying the behaviour of animals in real time and in their natural environments is becoming more and more feasible through the use of animal-borne loggers or other remote sensing technology (Wilmers et al., 2015). These technologies have provided insight into different aspects of physiology and behaviour, such as heart-beat (Laske et al., 2011) or migratory routes (Schofield et al., 2010; Newman et al., 2012), which in turn can help us understand basic mechanisms up to evolutionary drivers, as well as support decision-making processes in nature conservation or disease management.

To reconstruct daily activity patterns, many remote-sensing studies have used methods that provide information on the location of an animal in space (today most commonly GPS: Global Positioning System). To get more fine-scale information, spatial data have been combined with accelerometry which can shed more light on the actual activities of an animal (Shamoun-Baranes et al., 2012; Wilmers et al., 2015). However, the immediate causes or related contexts of specific animal behaviours were often not identifiable through these technologies, and required additional information sources.

Recently, microphone backpacks have become useful tools to investigate different aspects of vocal behaviour in naturalistic contexts, even in small animals (Hiryu et al., 2008; Ilany et al., 2013; Couchoux et al., 2015; Gill et al., 2016). By picking up the vocal sounds close to their production origin, researchers are now able to record and identify vocalisations from the signal-emitting individuals, even in physically or acoustically challenging environments. But in small animals, unlike for example in whales (Stimpert et al., 2015), it is often not (yet) possible to apply tags that provide multiple channels of information simultaneously, due to weight limitations—especially in birds. Thus, placing vocal behaviour into relevant context can be limited to specific situations in which a simultaneous collection of further data is possible.

Because an on-board microphone moves along with its bearer, most microphone backpacks do not exclusively record vocalisations, but also other sounds. Firstly, depending on their sensitivity, the microphones have the potential to pick up a
variety of background sounds. Secondly, specific movement patterns of the animal resulting in characteristic sound patterns might reveal aspects of the animal’s behaviour, e.g. “running” or “self-scratching” (noted by Ilany et al., 2013; Couchoux et al., 2015). But, to date, this has not been investigated in detail.

19.1. Automatic Acoustic Recognition

Successful identification of animal-related sounds could provide a unique opportunity because it may allow investigating not only the behaviour of the animal itself, but also different aspects of its abiotic and biotic environment—which is currently not possible by recording the spatial position or movement of single individuals, without further data collection. This in turn could be useful for various purposes (as above: from basic research to conservation, e.g. effects of anthropogenic noise), but analysing such signals/soundscapes remains a challenge to date. Manual annotation is possible for small datasets, though hard to scale up; further, for free-flying birds there will usually be no visual/video support for manual annotation. Hence there is strong potential for microphone backpack methodologies to be augmented by automatic acoustic recognition of bird activities and their contexts.

The problem of automatic animal context recognition from audio is directly related to the emerging field of sound scene analysis (also termed acoustic scene analysis), and more specifically to the two core problems in the field, namely sound scene analysis and sound event detection (Stowell et al., 2015). Since the context in question can refer either to an animal’s current activity or background sounds, the problem can be viewed as either or both of searching for specific acoustic events (e.g. related to flapping wings in the context of flying) or evaluating the overall properties of a continuous sound scene (e.g. background sounds indicating that an individual is based in a nest).

The vast majority of approaches in the field of sound scene analysis either fall directly into the problem of sound scene recognition (which typically refers to identifying scenes based on location-specific characteristics, e.g. park, car, kitchen) or the problem of sound event detection (which refers to identifying instances of sound events with a start and end time, e.g. door slam, scream) (Stowell et al., 2015). An approach that is closer to the present work is proposed by Eronen et al. (Eronen et al., 2006), who developed a computationally efficient classification-based system for audio-based context recognition in urban environments, where ‘context’ referred to both locations (e.g. train, street) but also to specific activities (e.g. construction, meeting). In Heittola et al. (2013), Heittola et al. proposed a system for sound event detection, which is however dependent on the context of
each sound scene. A system based on hidden Markov models (HMMs) with multiple Viterbi decoding was proposed, which was able to identify to a relative degree of success 60 types of sound events, being present in 10 different types of location-dependent audio-related contexts.

Another related strand of research is speaker diarisation, in which multi-party speech recordings are analysed such as discussions in meetings, and the primary goal is to recover a transcript of which party spoke when (Tranter and Reynolds 2008; Anguera et al. 2012). In speaker diarisation, the emphasis is primarily on speech and so the range of sound types considered is often highly constrained. Also the targets of transcription are individual speaking sources rather than aggregate contextual categories. Much work in speaker diarisation treats the transcription task as monophonic (only one speaker at a time), although recent directions are beginning to address overlapping speech (Anguera et al. 2012). Generalisation across different domains (e.g. conference meetings versus broadcast news) is also an open topic, indicating the difficulty of these types of problem in general.

When placing the present study in context with related work in sound scene analysis, it is important to maintain a focus on the downstream use of the data, which must influence the way we design and evaluate systems. Typical applications in animal behaviour include: (a) aggregating timelines to produce an overall model of a species’ diurnal cycle of activity, or creating “time budgets”; (b) data-mining to search for one or many instances of a particular phenomenon. A transcript is rarely the end goal in itself. As an example consequence of this, for the applications just mentioned it may often be helpful to obtain a probabilistic or confidence-weighted output rather than merely a list of events, for optimal combination of information or best guidance of subsequent manual effort.

19.2. Aims

The aims of this study were thus to find out whether the recordings from microphone backpacks could be useful for investigating the immediate context in which individual vocalisations occur, such as an animal’s current activity (movement sound) or vocalising conspecifics (background sound), and to investigate the extent to which this could be facilitated by automatic acoustic recognition. To do so, we used video-validated and human-coded on-bird sound recordings from captive and free-flying jackdaws (Corvus monedula), to test the performance of different automatic recognition algorithms. We experimentally compared two different sound recognition paradigms (classification and event detection), as well as combinations and variants, and how they performed in terms of recognising
the various categories of activity and context that are of interest for measuring animal behaviour.

In the following we describe the data collection process (Section 20) before giving details of our two automatic recognition systems (Section 21). Our evaluation method and its results are presented in Section 22 and then in discussion (Section 23) we consider the implications of our study for the automatic annotation of animal-attached sound recordings.
Data Collection

20.1. Birds and microphone backpacks

For the current study, we used a subset of on-bird sound recordings obtained during a different study (Gill et al., in preparation). The analysed data were collected in the South of Germany, from 12 individual jackdaws (*Corvus monedula*, 7 captive-housed and 6 free-living), early in the years of 2014 and 2015. Backpack application was approved by the Government of Upper Bavaria and in compliance with the European directives for the protection of animals used for scientific purposes (2010/63/EU). The backpacks consisted of a commercially available digital voice recorder (Edic Mini Tiny A31, TS-Market Ltd., Russia), a rechargeable battery (ICP581323PA to ICP402035, Renata, Switzerland), a radio transmitter for relocation (BD-2 Holohil, Canada) and a shrinking tube casing. Loggers were charged, programmed and read out via PC connection and the according software (RecManager, version 2.11.19, Telesystems, Russia). They were set to record continuously for a few hours every morning, for a few days, beginning one day post capture (at 22050 Hz sampling rate, uncompressed .wav format). This provided coherent vocalisation data and acoustic background information, as opposed to using amplitude-based triggers (but at a cost of storage and battery). For backpack attachment, birds were either trained to fly inside a smaller compartment of the aviary where they were caught using bird nets (captivity), or trapped inside their nest boxes (wild). The backpacks were fitted using approved attachment methods (glue, or via a harness similar to Karl and Clout [1987], and following common recommendations (<< 5% of body weight Caccamise and Hedin [1985], close to centre of gravity [Vandenabeele et al., 2014]). Birds were individually identified by colour rings. After capture and backpack attachment (20 mins ± 4.1 SD), they were observed using binoculars and/or radio-telemetry, and all of them were immediately able to fly upon release. For further details on procedures and animal welfare, see Gill et al. (in preparation).
20.2. Video-validation of sounds

For a video-validation of on-bird sound data, video footage was collected from the captive birds during backpack recording hours. For this, an observer sat inside the aviary and video-recorded focal birds using a handheld camcorder (JVC Camcorder Everio GZ-MG77E, Japan). All sound files used for video validation were processed, played back, visualised (waveform or spectrograms: FFT window size 512, Hann, 0–10000 Hz, gain 20–35 dB, range 45 dB) and annotated in Audacity (Version 2.0.5) by LFG. Corresponding sound and video files were cut to match, and were then played back simultaneously, at normal speed (using Audacity, see above, and using VLC, Version 2.1.5). First, the sounds were annotated step-by-step with the corresponding visual information (see Table 6). If the focal bird was temporarily out of sight, this was labelled as missing data. Secondly, labels were added for acoustically distinct background sounds, such as vocalising jackdaws. Next, the annotation track (labels, start and end points) of each recording was exported as a text file. To balance between fine detail and sufficient sample size, the original labels were used to create slightly broader behavioural and contextual categories (Table 6).

An example clip of annotated data is visualised in Figure 23(a). In Supplementary Information we provide videos showing the studied birds in some example contexts, along with standard and backpack microphone recordings to illustrate the characteristics of the specific kind of sound recordings dealt with in this work.

20.3. Annotation of field data

Having worked with hours of sound and video recordings from jackdaw backpacks, we had learned a good deal about the acoustic representation of behaviours and were able to annotate the sounds in new files in almost as much detail as in combination with the according visual information (at least at the behavioural category level). Thus, the field recording subset was annotated by LFG based on aural and visual inspection of sounds, as learned from the captive dataset and from observations in the field, but also taking into account differences in the sounds due to different materials in the field (e.g. walking on different substrates), as well as different durations (e.g. prolonged flight). Two labels were added that had not been recorded in captivity: copulations; begging chicks inside the nest (Table 6).
<table>
<thead>
<tr>
<th>Sound type</th>
<th>Category</th>
<th>Label examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Movement</strong></td>
<td>Flying</td>
<td>Flying</td>
</tr>
<tr>
<td></td>
<td>Walking</td>
<td>Run, walk</td>
</tr>
<tr>
<td></td>
<td>Looking around</td>
<td>Look</td>
</tr>
<tr>
<td></td>
<td>Manipulation</td>
<td>Food, stick</td>
</tr>
<tr>
<td></td>
<td>Self-maintenance</td>
<td>Bill-wipe, preen</td>
</tr>
<tr>
<td></td>
<td>Small movement</td>
<td>Turn</td>
</tr>
<tr>
<td></td>
<td>Shaking</td>
<td>Body, head</td>
</tr>
<tr>
<td><strong>Vocalisations</strong></td>
<td>Focal call</td>
<td>Contact call</td>
</tr>
<tr>
<td></td>
<td>Non-focal call</td>
<td>Non-focal call</td>
</tr>
<tr>
<td></td>
<td>Allofeed vocalisation</td>
<td>Allofeed vocalisation</td>
</tr>
<tr>
<td></td>
<td>Background call</td>
<td>Bg mobbing</td>
</tr>
<tr>
<td><strong>Background</strong></td>
<td>Carrion crow</td>
<td>Carrion crow</td>
</tr>
<tr>
<td></td>
<td>Chickens</td>
<td>Hen, cock</td>
</tr>
<tr>
<td></td>
<td>Colony sounds</td>
<td>Church bells</td>
</tr>
<tr>
<td></td>
<td>Noise</td>
<td>Traffic noise</td>
</tr>
<tr>
<td><strong>Combination</strong></td>
<td>Allofeeding</td>
<td>Allofeeding</td>
</tr>
<tr>
<td></td>
<td>Copulation</td>
<td>Copulation</td>
</tr>
<tr>
<td></td>
<td>Nest</td>
<td>Entering nest</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Antenna</td>
<td>Antenna</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>Missing video</td>
</tr>
</tbody>
</table>

Table 6. Labelling scheme for the actions/contexts in our recordings. The ‘Category’ column gives the class labels used in the present study, with the other columns indicating the broader or more specific labelling used during manual transcription.
SECTION 21

Automatic recognition

To train recognition algorithms, we used a total of 8.4 hours of video-validated (captive: 43–100 minutes per bird) and 18.5 hours of human-coded (wild: 164–198 minutes per bird) sound recordings and their respective annotations. We next describe the automatic recognition systems that we evaluated, which are summarised in Figure 22.

21.1. Classifier-based System

The first system we used for activity and context recognition sits within the classification-based paradigm. We used our feature learning and classification method previously developed for bird species classification from vocalisations (Stowell and Plumbley, 2014). Importantly, this approach applies spherical k-means feature learning to Mel-spectrogram patches, in order to transform the input signal into a rich feature space suitable for applying a standard classifier. This particular feature learning algorithm is conceptually related to an unsupervised convolutional neural network, but its simplicity makes it eminently scalable to big data (Coates and Ng, 2012; Stowell and Plumbley, 2014). In this work, we segmented input audio into contiguous five-second clips, from which we calculated Mel spectrograms (FFT window size 1024 with 50% overlap), and applied median-clipping noise reduction to each frequency band. Unlike in the cited previous work, for these data we did not apply high-pass filtering, since we expected some classes to be indicated in part by lower-frequency or broadband components. During training we applied a single pass of the feature learning described in (Stowell and Plumbley, 2014) to these data, learning a high-dimensional projection onto 500 features. We then transformed the training and test data into this new feature space, before summarising each audio clip by the mean and standard deviation of each feature.

The summary features were used as input to a random forest classifier (Breiman, 2001) having 200 trees and trained using an entropy-based criterion for splitting branches. The data in this task is highly unbalanced, with some classes very sparsely represented. A random forest classifier is typically able to handle unbalanced (and high-dimensional) data well. However, an option available to us was to
Figure 22. Overview of the processing workflows used for automatic recognition.

Reweight the data to give equal prominence to positive and negative classes. This was particularly pertinent as the subsequent HMM postprocessing (see subsection 21.3) also makes use of the relative class balance. We therefore trained the classifier in both modes, equally weighted and balanced-reweighted, to inspect the effect of this choice.
21.2. Event Detection System

The second system used for activity and context recognition is adapted from the system of [Benetos et al., 2016], which was originally proposed for sound event detection in office environments. Thus, this approach attempts to recognize contexts as a collection of acoustic events related to each context, as opposed to the previous approach which was based on modelling the overall characteristics of an acoustic scene. The system extends probabilistic latent component analysis (PLCA) [Shashanka et al., 2008], a spectrogram factorisation technique which can be viewed as the probabilistic counterpart of non-negative matrix factorization (NMF) [Li and Seung, 1999]. The PLCA-based model assumes that an audio spectrogram can be decomposed as a series of sound activities or contexts, which can potentially overlap over time. Each activity is produced as a combination of sound exemplars, which have been pre-computed from training data.

For preprocessing, a time-frequency representation \( V_{f,t} \) (\( f \) is the frequency index and \( t \) is the time index) is computed by processing the input waveform with an equivalent rectangular bandwidth (ERB) filterbank [Moore, 1995], using the approach of [Vincent et al., 2010]. The filterbank uses 250 filters which are linearly spaced between 5 Hz and 10.8 kHz on the ERB scale, and has a 23ms time step.

Given that in the context of on-bird sound recordings several activities exhibit information in higher frequencies, a linear pre-emphasis filter is applied to \( V_{f,t} \) for boosting high frequency content. See Figure 23(b) for an ERB spectrogram of a recording from the captive subset, along with the respective context annotation.

The PLCA-based model takes as input \( V_{f,t} \) and approximates it as a bivariate probability distribution \( P(f, t) \), which is in turn decomposed into a series of spectral templates per sound activity/context and exemplar index, activations over time for each context class, as well as an auxiliary probability for the activation of each exemplar per context class over time. The model is formulated as:

\[
P(f, t) = P(t) \sum_{c,e} P(f|c, e)P(e|t)P(e|c, t)
\]

(1)

where \( c \in \{1, \ldots, C\} \) denotes the context class and \( e \in \{1, \ldots, E\} \) denotes the exemplar index. On model parameters, \( P(t) = \sum_f V_{f,t} \), which is a known quantity. Dictionary \( P(f|c, e) \), which in this system is pre-computed from training data, contains spectral templates per context class \( e \) and exemplar \( e \). The main output of the PLCA model is \( P(e|t) \), which is the probability of an active context per time frame \( t \). Finally, the model also contains the auxiliary probability \( P(e|c, t) \), which denotes the contribution of each exemplar \( e \) for producing a context \( c \) at time \( t \).
Figure 23. (a) Context annotations for a segment of recording from a captive bird. (b) The ERB spectrogram of the same recording corresponding to the region marked with vertical dashed lines in figure (a).
The unknown model parameters $P(c|t)$ and $P(e|c, t)$ can be iteratively estimated using the expectation-maximization (EM) algorithm [Dempster et al. 1977]. For the E-step, the following posterior is computed:

$$P(c, e|f, t) = \frac{P(f|c, e)P(c|t)P(e|c, t)}{\sum_{c,e} P(f|c, e)P(c|t)P(e|c, t)} \tag{2}$$

Using the above posterior, $P(c|t)$ and $P(e|c, t)$ can be estimated in the M-step as follows:

$$P(c|t) = \frac{\sum_{e,f} P(c, e|f, t)V_{f,t}}{\sum_{c,e,f} P(c, e|f, t)V_{f,t}} \tag{3}$$

$$P(e|c, t) = \frac{\sum_{f} P(c, e|f, t)V_{f,t}}{\sum_{e,f} P(c, e|f, t)V_{f,t}} \tag{4}$$

Parameters $P(c|t)$ and $P(e|c, t)$ are initialised in the EM updates with random values between 0 and 1 and are normalised accordingly. Eqs. (2) and (3)-(4) are iterated until convergence. In our experiments, we found 30 iterations to be sufficient.

In order to extract dictionary $P(f|c, e)$ from training data, first spectra $V^{(c)} \in \mathbb{R}^{F \times T_c}$ that correspond to an active context class are collected, where $T_c$ corresponds to the number of spectral frames that contain an active context class $c$. Then, for each context class a list of exemplars is created by performing clustering on $V^{(c)}$ using the k-means algorithm; here, the number of exemplars $E = 40$, following experiments on the training data.

The output of the PLCA model is given by $P(c, t) = P(t)P(c|t)$, i.e. the context activation probability, weighted by the energy of the spectrogram. Since $P(c, t)$ is a non-binary representation, it needs to be converted into a list of estimated contexts per time frame. The first option of post-processing $P(c, t)$ is by performing thresholding, where threshold values were estimated per context class using training data. Finally, active contexts with a small duration (shorter than 120ms) were removed. Additional post-processing options are discussed in the following subsection.

### 21.3. Postprocessing

Given the output from either the classifier or PLCA detector, we then optionally applied hidden Markov model (HMM) postprocessing to the estimated event sequences. See [Rabiner 1989] for an overview of HMMs. HMM-based postprocessing is a common procedure using knowledge about the temporal structure
of event sequences (gleaned from the training set) which knowledge may not otherwise be reflected. In particular, in our case the classifier treats each five-second segment as independent, neglecting information from neighbouring segments. Likewise, the PLCA event detection system considers each 23 msec output frame as independent.

Since our task was polyphonic, having multiple “channels” in parallel whose activation could be on or off, there was a combinatorially large number of possible states at any time \(2^k\), with \(k\) the number of classes). To deal with this large state space we applied the HMM in two alternative ways: (a) applying a single HMM to the entire system, whose set of possible states is the whole set of state combinations observed in the training data; or (b) independently applying a two-state, on/off HMM to the data of each class. Each approach has advantages and drawbacks. Treating channels as independent may lead to efficient training given a limited amount of data, but it neglects interaction effects which could help to resolve ambiguous situations. Therefore we tested both approaches.

We trained the HMMs generatively, using Laplacian smoothing of the transition tables—i.e. initialising each possible transition with a small uniform weight, which yields a prior equivalent to having observed one instance of each possible transition. The emission model for each HMM state was a Gaussian mixture model (GMM). To initialise and to select the number of GMM components, we applied the Dirichlet process GMM approach (Rasmussen, 1999) to the entire training dataset (sometimes called a universal background model or UBM), then for each HMM state we trained its emission model by variational inference initialised from the UBM. We used the GMM implementations provided by scikit-learn 0.17 (Pedregosa et al., 2011).

Having trained a HMM, there are multiple ways to apply it to new data. We explored the use of forward filtering—producing probabilistic “fuzzy” output which may then be thresholded if definite decisions are required—and Viterbi decoding—producing a single definite output, as the maximum likelihood state sequence given the observations. This then resulted in four kinds of HMM postprocessing: filtered or Viterbi-decoded output, from a jointly or independently-trained HMM.

### 21.4. Handling Missing Data

Occasional time-regions of the data were labelled as missing data (‘NA’), when birds were occasionally off-camera. These regions (around 17 minutes total, out of the 8.4 hours of captive audio) were excluded from the training of the classifiers and HMMs. For the PLCA-based system, the NA class was not used to create the pre-extracted dictionary \(P(f|c,e)\), and any spectral frames belonging to the NA
class were not used in the training data. In the test phase, any NA regions in the ground truth are set to be non-active, where any time frames $t$ in the PLCA model output that correspond to the NA regions are set so that $P(c, t) = 0$. ‘NA’ regions were excluded from the calculation of our evaluation statistics, due to the lack of ground truth for comparison.
SECTION 22

Evaluation

22.1. Metrics

As discussed in Section 19, the evaluation must be designed with regard to the planned or typical downstream use case—i.e. what tasks or analyses do we expect to follow on from such automatic annotation? For the present task, this bears upon the figures of merit which one calculates, as well as on issues such as the temporal granularity or temporal tolerance. It is desirable for an automatic system to recover exactly-timed transcriptions of every vocalisation, action and context given in the audio, but for some of the downstream tasks we consider the overriding aim does not require the highest resolution, for example when characterising time budgets across large datasets, or locating examples of certain activity. Hence our main evaluation measures were calculated at a five-second granularity (the same granularity as was used for the classifier). The output of the classifier-based system was itself at a five-second granularity; for the PLCA-based system, the output was sampled at 23ms steps, as in the input time-frequency representation \( V_{ft} \). We therefore grouped its outputs into five-second segments, and the output for each 5-sec segment was either the mean or the maximum of the 23 msec-step frames corresponding to that time segment.

Evaluation metrics for automatic transcription have been debated in music informatics and in sound scene analysis. Recently Mesaros et al. reviewed such measures for general sound event detection, discussing issues including the use of high-resolution versus segment-based metrics [Mesaros et al., 2016]. In their terminology our main metrics are segment-based, using five-second segments. However, Mesaros et al. consider only the evaluation of “definite” transcripts, not transcripts with probabilistic/ranked/fuzzy annotations, and as a result their review does not include statistics useful for evaluating the latter type of output. Foster et al., working with probabilistic outputs, use a four-second segment size and use the area under the ROC curve (“AUC”) as their figure of merit [Foster et al., 2015].

The AUC is widely used as an evaluation measure for detection and classification tasks, and has many desirable properties [Fawcett, 2006]: unlike raw accuracy, it is not impeded by “unbalanced” datasets having an uneven mixture of
true-positive and true-negative examples; and it has a standard probabilistic interpretation, in that the AUC statistic tells us the probability that the algorithm will rank a random positive instance higher than a random negative instance. This last feature makes it particularly suitable to evaluating with regard to downstream tasks in which the subsequent postprocessing will for example involve manually confirming/refining the separation of positive and negative instances. Hand criticises the AUC statistic (Hand and Anagnostopoulos 2013), but reluctantly confirms that its use is well-founded when the downstream makes use of the ranking information, for example to allocate a fixed budget of manual postprocessing time. An alternative widely-used evaluation measure is the “F score”: the harmonic mean of precision (robustness against false positives) and recall (robustness against false negatives) of a system (Mesaros et al. 2016). The F score is particularly suited to information-retrieval type applications, such as downstream tasks in which the user might for example wish to retrieve a subset of positive examples from a large database. The F score requires definite, binarised output; for fuzzy outputs, this requires postprocessing such as thresholding.

In the present work we calculated both the AUCs and the F scores for our systems, yielding slightly different perspectives on their relative performance. Both measures were calculated from the segment-wise output with five-second segment durations. AUCs were calculated separately for each class (our plots will show averages across classes). To use the F score with fuzzy outputs, we chose binarisation thresholds to optimise the score on the training data, before applying the same thresholds to the testing data in each case. This can be done with one threshold per class or with a single threshold; we tested both variants. To summarise the F score we calculated it across all classes, rather than averaging the per-class F scores, since the latter would be numerically unstable especially with sparse data (Mesaros et al. 2016).

### 22.2. Evaluation Schemes

Our data consisted of annotated long-duration audio from multiple individual birds, one set in captive conditions and one set in field conditions, with multiple recordings from each individual (3–8 per individual for captive; 2 per individual for field, of longer duration). We used this data to evaluate system performance in various crossvalidation scenarios:

**EachCap: Captive, strictly per-individual:**
A system was trained with one half of an individual’s recordings, and tested with the other. The converse was also done, and then results aggregated over all captive individuals (yielding 14 ‘folds’).
X-Y: Captive, pooled.:
A system was trained with examples from each individual—half of the recordings from each individual—and tested with the remainder. This gave 2 crossvalidation folds. Note that X-Y is constructed so that all the testing files come from birds also seen in the training data.

A-B: Captive, pooled and stratified.:
All recordings from each individual were allocated to one of two partitions. This is similar to X-Y except that no bird used for training is used for testing.

Cap-Field.:
In this case the captive data is used for training, and the field data used for testing. (Here we used only one crossvalidation fold.) It is the most challenging case: as well as the train and test sets having no birds in common, the recording situation is also different.

EachField: Field, strictly per-individual.:
As EachCap, but for the field data (12 folds).

Each of these scenarios relates not just to different degrees of generalisation, but to different downstream applications of automatic recognition technology. For example, a researcher may wish to annotate a fraction of a recording and then invoke automatic recognition for the remainder; or to use a fixed system trained on one set of birds, e.g. observed in captivity, and to apply it to new unknown recordings.

Finally, since the PLCA-based system produced its output at a higher resolution (i.e. for each 23ms frame), we used this opportunity to explore how the temporal resolution interacts with evaluation procedures and metrics. For this we repeated our evaluation using the segment-based F score, but using a much smaller segment size of 0.1 seconds, as compared with the 5 sec segment size used in the main experiments. In order to ensure a fair comparison, sets of class-specific thresholds were computed from training data for each evaluation segment size (i.e. 100 msec and 5 sec) separately. The F-measure was computed directly on the raw high-resolution output of the PLCA-based system.

22.3. Results
As intended, the choice of microphone placement led to high-amplitude recordings for sounds from the focal bird (calls, flying, and other movements) while other background sounds were quiet but still largely audible (see Supplementary Information for examples). The occurrence of the annotated actions and contexts
Figure 24. Total ground-truth durations of annotated regions of each category.

in the collected data was relatively sparse (Figure 24), with every class being active for less than 16% of the total time in both datasets.
We evaluated each of our systems in two configurations: the classifier-based system with unbalanced or balanced class-weighting for training; and the PLCA system with mean- or maximum-based temporal downsampling. In each case the differences between configurations were small, and so for clarity of presentation we will plot results from just one of each system (unbalanced classifier, mean-downsampling PLCA). We will refer to differences in outcomes from the system configurations where relevant.

Overall, the quality of automatic recognition showed a strong dependency on the choice of crossvalidation setup, i.e. on the relationship between the training data and the test data (Figure 26). As one clear example: the designs of the X-Y and A-B schemes were very similar except that the latter ensured that birds used for testing were not used for training; this change incurred a substantial penalty both in AUC and F score, implying that individual differences were highly pertinent. The X-Y scheme in turn was similar to the EachCap scheme except that it pooled the training data across individuals. Curiously, this pooling led to very similar F scores as EachCap, but to a marked difference in AUC: judged by AUC, the pooling of training data seems to have led to better generalisation properties, for both of the recognition algorithms tested. Judged by F score, both EachCap and Each-Field, using systems trained specifically for each individual, attained many of the strongest results. As expected, schemes involving generalising to unseen conditions had lower recognition scores—both A-B (generalising to new birds) and Cap-Field (generalising to new birds and to new recording environments).

As this task has not been evaluated before, there are no direct external comparisons for the overall recognition quality. The segment-wise F-measures are broadly comparable to those presented in [Mesaros et al. 2016] (for an indoor event-detection task with fewer categories and a different segment duration). In the present comparison of two different approaches, the classifier-based system generally outperformed the PLCA-based system: by an average of 5 percentage points on AUC, and 8 percentage points on F score. Figure 25 shows an example of the output from the classifier-based system overlaid with the groundtruth annotation, giving a rough visual indication of the kind of output that corresponds to the results obtained.

The effect of HMM postprocessing led to different results when considered via F score or AUC. The F score statistics (Figure 26 upper) often showed a mild improvement when HMM postprocessing is added, particularly for the classifier-based system; while the AUC statistics (Figure 26 lower) unanimously indicated worse results with HMM postprocessing (the leftmost result in each cluster, the unprocessed output, performing best).

To binarise continuous-valued output, we found that per-class thresholding was
not particularly better than a single threshold in general, except in the case of the raw PLCA output. This exception is because the raw PLCA output is expressed in terms of activation magnitude (i.e., related to the energy of each context class in the spectrogram), which does not have comparable meaning across classes, and so per-class thresholding is highly pertinent in that case. For the HMM-postprocessed outputs, a single threshold often slightly outperformed per-class thresholds, which is probably due to a slight reduction in overfitting the threshold choice.

The classes (categories) used in this study are highly diverse in kind, and so to drill further into system performance it is important to inspect performance on a per-class level (Figure 27 showing results for the X-Y and EachField scenarios). It is immediately clear that detection quality exhibits some correlation with the quantity of positive examples available for training (cf. Figure 24), although the focal call category is particularly well detected by the classifier system despite being relatively sparse in the training data. (Focal calls are behaviourally important; they are also the signal class for which our classifier was originally implemented.) Figure 27 also decomposes the F score into its components: precision and recall. When the classifier reaches a high F score it is often achieving strong precision, while when the PLCA does well it achieves strong recall.

The per-class results for the most difficult evaluation condition, Cap-Field, show that the generalisation to new individuals and new environments has a differential effect on recognition quality (Figure 28). Importantly, the classifier-based system is able to generalise well on one of the more important categories—focal call—as well as on self-maintenance, yet the performance on some other categories—walking, flying, bg jackdaws—drops off markedly. The performance of the PLCA-based system does generalise on some categories—looking around, self-maintenance—but exhibits lower performance in other categories, including focal calls.

Figure 29 shows a different view of the temporal nature of our data. For selected classes in a chosen recording, it summarises the true or inferred activity levels in broad (five-minute) time-steps. Both systems exhibit some mismatch with the ground-truth, though the output from the classifier-based system can be seen to better match the true contours of activity. In particular the classifier-based system shows a tendency to better match the true sparsity levels of class activations. A final comparative study was made using the higher-resolution 23 msec step raw output of the PLCA-based system, comparing this against the 5 sec mean-pooled segments. Using the X-Y crossvalidation scenario, the performance in terms of segment-based F-measure with 5 sec segment size was 39.07% when using the 23 msec output, and 38.03% when using the 5 sec mean-pooled output. When
Figure 25. An example of an automatic annotation from a relatively strongly-performing system (classifier; HMM filtering; Each-Cap condition). The black and white regions are correctly-identified as on and off respectively. Red are false-positive detections, and blue false-negatives. (Best viewed in colour).
however the high-resolution output was evaluated using the segment-based F-
measure with a 100 msec segment size, performance dropped to 22.19%. These
results indicate that the higher-resolution output can lead to a small improvement
over the pooled output, and that the numerical value of the chosen evaluation stat-
istic depends strongly on the temporal granularity of evaluation. The reduced
performance when evaluated at high resolution may be partly due to issues in the
temporal precision of the inferred and/or the ground-truth annotations.
Figure 27. F score, Precision and Recall (all in %) for each class separately, for 4 systems tested under the X-Y and EachField crossvalidation cases, using per-class thresholding.
Figure 28. Per-class results as in Figure 27 but for the Captive-Field condition.

Fig. 30 shows an example high-resolution output using the PLCA-based system for recording MohawkMOV00F_a from the captive set, which in this case reached a 100 msec segment-based F-measure of 54.1% using the X-Y crossvalidation scheme. A few observations can be made from Fig. 30: the system was able to successfully detect overlapping contexts, in this case background colony sounds and looking around movement. However, the output was often fragmented, as for example can be seen for detected flying events. Another notable issue is the high number of false alarms as compared to missed detections (which translates into high precision and low recall, as shown in Fig. 27). So for example, flight events present in the recording were correctly detected as flight, but at the same time the output produced false positives for the manipulation and self-maintenance classes.
Figure 29. Temporal activity profiles for one of the field recordings, for 8 selected classes. Each panel shows a bar chart plotting, for each subsequent five-minute interval, the proportion of time that the class was active. This was calculated as the proportion of 5-second segments in that interval that were labelled positive; for probabilistic outputs, the ‘fuzzy’ probabilistic decisions were summed. We compare an example of the manually-annotated ground truth (top row), the classifier inference (middle row), and the PLCA inference (bottom row). The two systems were in the Each-Field condition, with per-class HMM filtering as postprocessing.
Figure 30. The 23 msec step output of a recording from the captive set, using the PLCA-based system with the X-Y cross-validation scheme. The colour scheme is as in Fig. 25.
Our study has investigated a novel task in animal sound recognition, approaching it via two polyphonic sound recognition methodologies related to those previously studied in environmental and bird sound. Overall evaluation figures are comparable with the state of the art in these neighbouring tasks (Stowell et al., 2015; Mesaros et al., 2016). The details of the timelines recovered (Figures 25, 29, 30) show that across all conditions, further development is needed before this paradigm can be deployed for fully automatic analysis of animal behaviour patterns from audio data. Of the two recognition systems studied, the classifier-based system consistently led to stronger results, including a better match to the temporal characteristics of the true annotations (Figure 29), however, the PLCA-based system has an advantage of directly outputting a high-resolution (frame-by-frame) annotation, which may be particularly desirable in some applications, such as investigating the short-time vocal interactions between individuals.

Our sequence of crossvalidation tests demonstrated that generalising to new individuals and new environmental conditions remains a critical challenge for automatic sound recognition, certainly when judged by F score (Figure 26), especially when aiming at extrapolating from captive to field datasets. The present results suggest that to annotate field recordings, the best strategy could be to train a human annotator on the captive data to annotate a small subset of field recordings from individuals which in turn could be used to train the classifier for further field data analyses. Crucially, our study investigated the automatic recognition of a diverse set of classes, each of them pertinent for the study of animal communication and behaviour. The classes vary widely in their acoustic realisations, from single sound events such as calls, to behaviours such as walking heard as compound events or sound textures. Consequently, as expected there were wide variations in recognition performance across classes. The strongest-performing system achieved good F scores for focal calls, flying, self-maintenance and walking. In general, performance levels could be correlated with how well the class of interest was represented in the training data. The sound of flying is quite clear to a human annotator, especially in the field where birds may fly continuously for...
15 minutes or longer. Very short flights (less than 1–2 seconds) are more difficult, and require more attention, because they may be confused e.g. with feather ruffling. Especially the captive dataset was characterised by such short flights, which may explain why the relatively good scores for automatic detection of flying were still lower than anticipated. Suitable features and detectors for such noisy, loosely periodic sounds thus remain a topic for further development.

In manual inspection, we noted a tendency for systems to output detections for focal call and non-focal call at the same time. This can be attributed partly to acoustic similarities between the classes: the microphone placement was designed to assist with discriminating these categories, though in some instances it remained difficult even for a human annotator. Some acoustic differences included the effects of close-mic recording, giving increased low-frequency energy for the focal call over the non-focal call. We did not adapt our time-frequency representations specifically for this feature, and one future development could include such adaptation. A rival explanation for the confusion of focal and non-focal calls is that the two do tend to co-occur in close temporal proximity (< 1 seconds), and so the systems may be influenced more by the class co-activation (at the 5-second resolution) rather than acoustics. This highlights the tension inherent in selecting a time resolution for analysis; for studies such as this, in which the different categories operate with rather different temporal characteristics, an option may be for the system—and also the evaluation—to use a class-dependent time resolution.

In the present study we found relatively little benefit in HMM postprocessing of system output. Its purpose was to refine per-segment estimates by making use of temporal dependencies between segments. In some configurations it led to a mild improvement in results, though in some other configurations it led to deterioration. We did however find a consistent result that HMM filtering led to better results than Viterbi decoding, and that a per-class HMM was better than a unified HMM. The classifier-based system treated each segment entirely independently, and so should have benefited from some temporal smoothing. One interpretation is that simple Markovian dependency (at the 5-second timescale) does not reflect enough of the temporal structure present in the data, and that more sophisticated temporal models might be investigated.

Some of the differences in interpretation implied by the AUC and the F score might be attributed to the fact that F score requires fuzzy/probabilistic outputs to be binarised at a specific threshold, whereas the AUC uses the continuous data and thus generalises over all possible thresholds. In a typical practical application, the user will know the relative cost of false positives and false negatives—i.e. the relative importance of high precision and high recall—and can set a threshold
based on this balance. The standard F score weights the two equally. However, downstream applications might imply different priorities, such as high precision in the case of a user retrieving examples of specific behaviour. In those cases it would be desirable to use the generalised F score, sometimes referred to as $F_\beta$ where $\beta$ is the desired precision/recall ratio. This would be used not only for evaluation but for threshold-setting.

As already discussed, we consider that the current level of performance is not yet at level for blind application to new data. As with tasks in neighbouring disciplines—speaker diarisation and polyphonic music transcription—the task is difficult and the development of full automation will require refinement of methods adapted for the specific characteristics of the signals in question. This is particularly true for categories indirectly represented via clusters of related sound events. The present study with its diverse set of sound categories raises the possibility that a good detection system may benefit from using an entirely different system for each class, perhaps using different timescales. A further possible direction in relation to the timescale is the possibility of using dynamic time resolution. The appropriate time resolution at which to consider animal behaviour is a discussion well-rehearsed in ethology; if time resolutions could be dynamically inferred per-class from data, this might inform debate as well as improving system performance.

We investigated the performance of systems using segment-based evaluation measures. Our segment size of 5 seconds was chosen based on manual inspection of pilot data as well as on considerations of the target application. The classifier-based system was also configured to operate at this resolution; such a classifier-based system typically operates over segments of this size (not at ‘frame-wise’ resolution such as 23 ms) in order to make stable classification decisions. Segment-based evaluations aggregate higher-resolution data using a max-pooling approach (Mesaros et al., 2016), with the curious side-effect that a single positive item anywhere within the 5 sec segment leads to the whole segment considered active. To mitigate this effect, in future evaluations one might use a smaller (and data-driven) segment size for evaluation, even in the case that the system gives output at a larger segment size; perhaps more fundamentally, the max-pooling could be replaced with a parametric threshold (e.g. percentile-based) to reduce the effect of false-positive ‘blips’ on the evaluation outcome.

In the present work we considered interactions between the annotated categories via co-occurrence dependencies (positive or negative) implicitly learnt from the data: the classifier-based system used a single classifier predicting for all classes at once, the PLCA-based system had the opportunity to ‘explain away’ a portion of energy as belonging to one class rather than another, and the HMM
postprocessing was able to use a single HMM model across all classes (though this was not found to be better than per-class HMMs). Future work could consider alternative approaches to the relationships between categories. Hierarchical models such as the context-dependent sound event detection of [Heittola et al., 2013] may be suitable, or switching state-space models (switching SSMs), where the discrete “switch” would correspond to a context and the context-dependent SSMs would detect specific sound events or background sounds.
Conclusions

We have introduced an application of audio recognition specifically for sound recordings from animal-attached microphones, to enable analysis of the activity of a focal animal as well as the context of such activity, i.e. the environment around it as conveyed acoustically. This enables researchers to study the animal’s behaviour as well as the context of that behaviour, i.e. the environment around it as conveyed acoustically. We applied automatic recognition to data collected from lightweight backpack loggers carried by free-flying birds (jackdaws) in an aviary and in the field.

We directly compared a scene-classification and an event-detection approach to this task. The classification method made use of a feature learning method developed for bird vocalisations. For event detection, we introduced a modified PLCA method, improving on previously-published work in related domains. In evaluation, the classifier-based method performed most strongly.

We find that the current recognition quality level enables scalable automatic annotation of audio logger data, given partial annotation, but also find that individual differences between animals and/or their backpacks can reduce recognition rates when generalising to previously-unseen individuals. This approach to studying animal behaviour in single individuals requires further development for full automation and application to previously-unseen individuals. However, as on-animal microphones become increasingly common, this seems an effort worth taking to eventually extract meaning from such streams of sounds by facilitating the analyses of vocalisations, as well as some of their associated behaviours and acoustic contexts, without additional data collection and devices. Combining such results with an animal’s position in space or relative to its conspecifics, and with detailed acceleration data, would provide us with a more complete picture of what animals do and even provide hints why they do it, to tackle many remaining open questions in mechanistic, evolutionary and conservation-related areas of behavioural research.
Acknowledgments and Funding information

We would like to thank Tiffany Magdalena Pelayo van Buuren and Magdalena Mair for assistance in the field; Katrin Mayer for collecting video footage and helping with sound files; Auguste von Bayern, Wolfgang Goymann, Andries Ter Maat and Manfred Gahr for their support without which this work had not been possible; and to the von Bayern family for granting access to the premises and facilities. DS is supported by EPSRC Early Career research fellowship EP/L020505/1. EB is supported by a UK Royal Academy of Engineering Research Fellowship (grant no. RF/128). LFG is funded by the Max Planck Society.
CHAPTER 4

Bio-logging reveals extra-pair copulations and male copulatory calls in “strictly monogamous” wild jackdaws

Lisa F. Gill* and Manfred Gahr

4(submitted)
Abstract

Being noisy during copulation can be very costly: it makes animals vulnerable to predator attacks and intra-specific conflict. However, some species produce loud, context-specific vocalisations during copulation, sometimes revealing explicit information for conspecifics. To understand the role of such signals, it is important to know the context in which they were produced, as well as the identity of the sound-producing individuals. This can be very difficult when animals vocalise in close proximity and in a “private” context, as during copulation. To tackle this challenge, we deployed audio loggers for the first time on free-living wild songbirds, jackdaws (Corvus monedula), and fitted nest-boxes with video surveillance cameras. These colony breeders pair for life and often serve as a rare example of absolute monogamy in birds. Copulations take place inside dark nest cavities, and are accompanied by loud, context-specific vocalisations unknown from which sex. Our bio-logging approach identified the males as the source of copulatory calls, and even revealed extra-pair copulations. Interestingly, both intra-pair and extra-pair copulations were accompanied by these conspicuous vocalisations. Our results thus challenge the previous view of absolute monogamy in jackdaws, and call for more detailed, state-of-the-art investigations of copulatory vocalisations and extra-pair behaviour in these birds.
SECTION 27

Introduction

Loud, conspicuous vocalisations during copulation have been described for very different animal species, from large tortoises (Crawford and Awbrey [1978]) to small mammals (White et al. 1998; Liu et al. 2013). However, it seems plausible that conspicuous vocalisations combined with low vigilance during copulation may be costly. Indeed, it has been shown that reproductive behaviours in general (Gwynne 1989; Magnhagen 1991), and more specifically, being noisy during copulation (Siemers et al. 2012) may increase predation risk. Openly displaying engagement in copulation through vocalisations may also cause conflict with conspecifics, often resulting in mating disruption or aggression, especially towards subdominant individuals (Hauser 1998; Løvlie et al. 2014). But copulatory vocalisations have also been correlated with increased mating success (reviewed by Hauser 1998).

Investigating mechanisms, ecology and evolution of such vocal behaviours in detail often requires an exact identification of the sound-producing individuals, which can be very challenging – especially in the field, when visual confirmation is unavailable, or when animals vocalise in close proximity, as during copulation. Thus, although many vocal signals are known to be associated with sexual contexts (Hauser 1998; Bradbury and Vehrencamp 2011), most mating-related vocalisation studies focused on pre- or post-copulatory vocal behaviours (e.g. for mate attraction, courtship or mate-guarding) (Hauser 1998), and comparably little is known about vocalisations occurring during the act of copulation itself. In primates (reviewed by Hauser 1998) including humans (Brewer and Hendrie 2011), the role of vocalisations during copulation has been discussed, e.g. with respect to optimal fertilisation and as information source for third-party conspecifics. In some bird species, such calls have been shown to attract females (Anoop and Yorzinski 2013) or males (Løvlie et al. 2014) to a mating site, sometimes even “deceitfully” (Dakin and Montgomery 2014). In sum, copulatory calls have the potential to reveal very specific selection pressures, costs and benefits, as in underlying intra- and inter-sexual conflicts or predator-prey relationships.

Jackdaws (Corvus monedula) are gregarious songbirds of the corvid family that produce context-specific, very loud “räääh-räääh-räääh” calls during copulation...
These birds breed in colonies, but form strong life-long pair bonds (Lorenz 1931; Dwenger 1989; Glutz von Blotzheim and Bauer 1993; Cramp and Perrins 1994; Liebers and Peter 1998; Henderson et al. 2000). In diverse areas of research, they have been used frequently as an example of absolute monogamy which is rare among birds (e.g. Neodorf 2004; Spottiswoode 2004; Emery et al. 2007 citing “strict monogamy” reported by Henderson et al. 2000). However, direct observations of jackdaw copulations are challenging, because this behaviour usually take place inside the nest cavity (Dwenger 1989; Cramp and Perrins 1994; Liebers and Peter 1998). Despite this visual “secrecy”, the accompanying copulatory vocalisations can be heard from quite a distance, and seem to elicit strong behavioural responses in the remaining colony, especially at the onset of the breeding season (expressed by general agitation and activity, flying about and repeated nest-site inspections; LFG personal observations and personal communications with A. von Bayern). However, next to suggestive evidence these calls could be produced by the male (Dwenger 1989: hand-reared male attempting copulation with human hand), it is not known which sex produces these sounds (Dwenger 1989; Cramp and Perrins 1994). Consequently, this interesting vocal behaviour is hardly described, and its function is, to date, unknown.

To record calls with individual-level resolution, we applied individual on-bird microphones for the first time to wild songbirds in the field, and combined this method with nest-box video footage, to overcome the challenge of directly observing copulation behaviour of jackdaws in the wild, and of identifying the source of jackdaw copulatory vocalisations.
In total, we obtained on-bird audio recordings of 18 copulations (Fig. 31) performed by 4 males (11, 1, 4, 2 copulations each) during the recording period (42, 21, 21 and 31 hours each). Corresponding video footage was obtained for 3 of the 4 males (6, 0, 3, 2 copulation videos each). We found that in all 18 cases, the focal birds, i.e. the respective male, produced the loud copulatory vocalisations, while females often produced short calls (Fig. 32, Methods, Audio1).

All 18 copulations took place inside a nest-box, but apparently not all inside the nest-box of the recorded pair. In 2 out of 4 males, we detected extra-pair copulations (3 events: 2, 1, 0, 0 per male): simultaneous on-board sound and nest-box video recordings revealed the respective male was copulating inside a nest (on-board sound) while the respective female partner was sitting alone inside the pair’s nest-box (video). Unfortunately, we did not obtain video footage of the extra-pair copulations for these males, but did so for at least one extra-pair copulation in one female (without microphone) paired with one of the focal males, and also observed further extra-pair copulations in the colony.

For both within-pair and extra-pair copulations, males almost always (n=15) exited the nest-box immediately after copulation. Bouts of male copulatory vocalisations lasted 3.42 to 59.60 (19.45 mean ± 20.35 standard deviation [SD]) seconds, and single vocalisations 0.11 to 1.10 (0.51 ± 0.14 SD) seconds. It is worth noting that the male from which we obtained both within-pair and extra-pair copulation data, changed its behaviour during extra-pair copulation, increasing the mean bout duration and number of calls (from ca. 4 to 22 seconds, and from 8 to 37.5 calls, see 7).
Table 7. Summary statistics on male copulatory vocalisations for all, within-pair and extra-pair copulations (no separate column for the 4 events without video). Note bird 1 performed within-pair and extra-pair copulations.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Within pair</th>
<th>Extra pair</th>
<th>Bird</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bouts</td>
<td>18</td>
<td>11</td>
<td>3</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>bird1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>bird2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>bird3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>bird4</td>
</tr>
<tr>
<td>Total number of calls</td>
<td>578</td>
<td>359</td>
<td>115</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>147</td>
<td>48</td>
<td>24</td>
<td>bird1</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>0</td>
<td>37</td>
<td>bird2</td>
</tr>
<tr>
<td></td>
<td>344</td>
<td>264</td>
<td>0</td>
<td>bird3</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>47</td>
<td>0</td>
<td>bird4</td>
</tr>
<tr>
<td>Mean number of calls per bout ± SD</td>
<td>31.94 ± 32.46</td>
<td>32.64 ± 37.08</td>
<td>37.33 ± 4.51</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>13.36 ± 12.11</td>
<td>8 ± 0.63</td>
<td>37.50 ± 6.36</td>
<td>bird1</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>-</td>
<td>37</td>
<td>bird2</td>
</tr>
<tr>
<td></td>
<td>86.00 ± 14.72</td>
<td>88.00 ± 17.35</td>
<td>-</td>
<td>bird3</td>
</tr>
<tr>
<td></td>
<td>23.50 ± 12.02</td>
<td>23.50 ± 12.02</td>
<td>-</td>
<td>bird4</td>
</tr>
<tr>
<td>Mean bout duration ± SD [sec]</td>
<td>19.45 ± 20.35</td>
<td>19.33 ± 22.52</td>
<td>24.01 ± 3.76</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>7.06 ± 7.31</td>
<td>3.68 ± 0.28</td>
<td>21.84 ± 0.10</td>
<td>bird1</td>
</tr>
<tr>
<td></td>
<td>28.35</td>
<td>-</td>
<td>28.35</td>
<td>bird2</td>
</tr>
<tr>
<td></td>
<td>52.62 ± 11.51</td>
<td>52.32 ± 11.51</td>
<td>-</td>
<td>bird3</td>
</tr>
<tr>
<td></td>
<td>16.83 ± 7.65</td>
<td>16.83 ± 7.65</td>
<td>-</td>
<td>bird4</td>
</tr>
<tr>
<td>Mean call duration ± SD [sec]</td>
<td>0.508 ± 0.139</td>
<td>0.499 ± 0.127</td>
<td>0.495 ± 0.149</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>0.45 ± 0.14</td>
<td>0.45 ± 0.15</td>
<td>0.44 ± 0.11</td>
<td>bird1</td>
</tr>
<tr>
<td></td>
<td>0.60 ± 0.16</td>
<td>-</td>
<td>0.60 ± 0.16</td>
<td>bird2</td>
</tr>
<tr>
<td></td>
<td>0.51 ± 0.12</td>
<td>0.50 ± 0.10</td>
<td>-</td>
<td>bird3</td>
</tr>
<tr>
<td></td>
<td>0.57 ± 0.18</td>
<td>0.57 ± 0.18</td>
<td>-</td>
<td>bird4</td>
</tr>
</tbody>
</table>
Figure 31. Schematic of copulating pair. Female crouching inside nest cup, and mounted by the male wearing an audio logger on its back (white object with dashed black lines; in reality inside black shrinking tube). Presumably for balance, the male props its tail feathers against the nest floor and flaps its wings during copulation. The backpack, partly covered by feathers, touches the male’s body. The microphone is thus closer to the male than to the female vocal tract, and female vocalisations are dampened by both birds’ bodies (see Methods, Fig. 32).
Figure 32. On-board sound recordings during copulation of A) an entire copulatory call bout; B) two focal (carrying microphone) and one non-focal (without microphone) bird calls. Top: waveform, bottom: corresponding spectrogram. Thick black and red bars indicate vocalisations emitted by the male and female, respectively. Male long, harsh (fast amplitude-modulation, high energy over large frequency range) and loud copulatory vocalisations drown out female short calls when overlapping. Note additional sounds from movement such as male wing beats and exiting nest-box after copulation (see Methods and Audio1).
Discussion

Jackdaw copulatory vocalisations are a puzzling behaviour, as they occur during copulation inside the nest cavity, i.e. visually concealed from conspecifics, but at the same time are acoustically conspicuous, and can be heard from dozens of metres away. Previously, direct observations of jackdaw copulations were limited (Dwenger, 1989; Cramp and Perrins, 1994; Liebers and Peter, 1998). Further, it was impossible to find out which sex produced the copulatory calls, thus limiting the extent to which this behaviour and its function could be investigated. Using on-board microphones for the first time on free-living wild songbirds, we not only found that males were the source of these characteristic copulatory vocalisations in all 18 cases (n = 4 individuals), but, by combining this acoustic information with nest-box video footage, we also revealed instances of extra-pair copulations. Previous genetic and behavioural investigations of jackdaws suggested very low to zero levels of extra-pair fertilisations, but lacked actual observations of copulations (respectively: Liebers and Peter (1998); Henderson et al. (2000)). In our study, exact numbers of extra-pair mating behaviour and genetic parentage results are not available for the entire colony, but we managed to record a small subset of the observed extra-pair copulation events (minimum 2 out of our 4 audio-tagged males and one female partner, resulting in a minimum of 4 out of 19 copulations, see Table 7). If extra-pair copulations were as rare as previously suggested (Liebers and Peter (1998); Henderson et al. (2000)), it would be highly unlikely to obtain multiple recordings of this behaviour based on a small set of observations. Therefore, future research will need to re-evaluate the commonly cited “strict monogamy” in jackdaws (Henderson et al. (2000)) through state-of-the-art behavioural and molecular investigations of parentage in this species. Based on the information from individual-level recordings, we can now begin investigating the function of male jackdaw copulatory vocalisations in more detail. These calls are very loud and, to our knowledge, exclusively produced during copulation (Dwenger, 1989; Cramp and Perrins, 1994) which means they could not only attract predators, but may also entail conflict with conspecifics. Indeed, in other group-lived corvids, copulating birds are often interrupted or attacked by conspecifics (Coombs, 1960). In our study population, this receives some support.
by high arousal in the colony outside, but we did not obtain observation data on
the aftermath of copulations for the recorded birds. However, the latter were not
attacked or interrupted inside the nest-box, even during extra-pair copulations
which, too, were accompanied by loud “tell-tale” copulatory vocalisations. This
is surprising because extra-pair copulations may be “punished” in other species
(Valera, 2003; Wedell et al., 2006; Kempenaers and Schlicht, 2010). This obser-
vation also seems to rule out a role of these calls as strictly pair-related “mini-
croterritorial” signals (nest-site defence). Instead, given the growing evidence
that jackdaw vocalisations are individually distinct, and, in some cases, can be
recognised as such by conspecifics (Lorenz, 1931; Dwenger, 1989; Zandberg
et al., 2014; Stowell et al., accepted; Gill et al. in preparation), copulatory calls
may provide identity-associated information for conspecifics. As in other species
or with other vocalisations, they could also play a role in female stimulation and
synchronised copulation for optimal fertilisation (Brockway, 1965; Lehrman and
Friedman, 1969; Hauser, 1998), or even in post-copulatory sexual selection (Birk-
head, 2010). However, although copulatory vocalisations were associated with
male wing-flapping during mounting, it was not possible to identify the exact time-
point of vocalisations with respect to actual cloacal contact – this remains for
future study, involving advanced technical approaches.
In conclusion, our study shows that on-board microphone methods, in combin-
ation with other observation techniques, are useful in revealing previously un-
known aspects of vocal and non-vocal animal behaviour in the wild. This approach
generated unexpected findings which put previous claims of absolute monogamy
in jackdaws in perspective, thus calling for further detailed investigations of extra-
pair behaviour and parentage. So far, it is unknown why jackdaws produce con-
spicuous vocalisations during copulation that may reveal both intra-pair as well
as extra-pair copulations, and we can only speculate on the function of these
copulatory calls. But only by understanding the context and reliably knowing the
source of such sounds can we begin investigating in detail their associated po-
tential costs and benefits, to approach a better understanding of mechanisms as
well as the evolution of vocal communication.
Methods

Procedures were in accordance with the European directives for the protection of animals used for scientific purposes (2010/63/EU), and were granted approval by the Government of Upper Bavaria. Adult free-living jackdaws (Corvus monedula) breeding in a colony in Bavaria, Germany, were trapped inside nest-boxes, individually marked with one numbered aluminium and three colour rings, equipped with microphone backpacks (audio loggers), and subsequently released. The dataset presented here is a subset of on-board sound recordings collected by the first author in the course of a study on adult jackdaw vocalisations in the wild (PhD thesis), taking place during the breeding seasons 2013-2015. Birds were recorded at different breeding stages, from nest-building to incubation of eggs and during provisioning of chicks (aged at least 12 days post hatching) inside the nest. For the current study, we analysed on-board sound recordings and corresponding nest-box videos of 4 individuals (all males) that had performed copulations (n = 18) during the recording period (1 further individual did not copulate, 6 other individuals were recorded later during breeding). All recorded copulations were performed just before or during the female egg-laying period (2014 and 2015), i.e. during the female fertile phase.

30.1. Loggers and settings

To obtain individual sound recordings, we used audio loggers (Edic Mini Tiny A31, TS-Market Ltd., Russia) customised with a rechargeable battery (ICP581323PA to ICP402035, Renata, Switzerland) and lighter casing (shrinking tube). For locating animals and recovering the devices, we added small radio transmitters (BD-2 Holohil, Canada) that could be tracked using a 5-fold Yagi antenna (F150-151-5FB, Wildlife Materials Inc., USA) and a handheld receiver (AOR 8200, AOR, USA). The mean total weight of the backpacks was 8.5 g (min: 6.3, max: 9.5 g) which made up of a mean of 3.5% (± 0.5% standard deviation = SD) of the birds’ body mass (205-280g), by selecting lighter backpacks (lower battery performance resulting in shorter recording times) for lighter birds. Loggers were connected to a PC for charging, setup and data retrieval (supplemented software: RecManager, version 2.11.19, Telesystems, Russia). We obtained coherent vocalisation recordings
(22050 Hz, uncompressed .wav format) from multiple days by programming a few hours of continuous morning sound recordings for a couple of days. To maximise the yield of vocalisation recordings, loggers did not start until the morning after capture, thus reducing potential initial artefacts in vocalisation rates immediately following logger attachment (Gill et al. 2016).

30.2. Nest-boxes and cameras

The colony nested inside a large building inside wooden nest-boxes, chimneys and small crevices. The nest-boxes were accessible from inside the building, and before the breeding season, had been fitted with small doors for trapping the birds, large apertures for taking out birds, and small holes for infrared-supported nest-box video cameras (420 TVL, Handykam, UK). For habituation, functional cameras or dummies were installed at least four days before the capture of a target bird. The cameras were connected to a central multichannel motion-triggered video surveillance system (GV-1480, GeoVision, USA) to record copulations, to supervise normal pre- and post-backpack nesting behaviours, and in some cases to determine the sex of the birds. For the catching procedure, the camera was connected to a small portable screen instead, which allowed capturing the targeted bird, under controlled conditions, in particular when the partner was absent.

30.3. Logger application

Upon capture, a bird was immediately taken out of the nest, placed inside a cotton bag, and carried to a quiet working area. It was weighed using a spring balance, and fitted with a small bag over its head. Backpacks were applied close to the centre of gravity (Wilson and Vandenabeele 2012; Vandenabeele et al. 2014), using glue as a common temporary attachment method: we sowed a piece of cloth to the backpack, trimmed the bird’s dorsal feathers (< 5 mm), covered both surfaces with a small amount of flexible glue (Pattex Gel, Germany), and held them together for ca. 30 seconds. Birds were immediately released (20 minutes ± 4.1 standard deviation = SD). Because backpacks fell off after about a week, recapture or any risk of injury related with alternative harness methods were avoided.

30.4. Sound analyses

Backpack sound recordings were acoustically and visually inspected in Audacity (Version 2.0.5) using waveforms and spectrograms (FFT window size 512, Hann- ning, 0 - 10000 Hz, gain 20 - 35 dB, range 45 dB). Spectrograms included in the manuscript were created using Raven (Lite 1.0, Cornell Lab of Ornithology).
30.4.1. Identifying the signalling individual. To identify the vocalisations of focal (wearing microphone) and non-focal bird (copulation partner without microphone), we made use of the microphone backpack’s position (Fig. 31), and of characteristic sound properties of on-board sound recordings (Fig. 32). First, distant background vocalisations were easily separated from foreground vocalisations by pronounced amplitude differences. Second, the vocalisations of focal birds also carried more power in low frequency bands, compared to those of non-focal animals (similar to Ter Maat et al. (2014); Gill et al. (2016)). Third, because vocalisations are coupled with body movements, focal birds’ calls contained a pronounced acoustic onset or “prefix”, compared to those of non-focal individuals. Fourth, focal bird calls were always recorded at the same distance from the carrier bird, while non-focal bird calls, for example during a continuous sequence of calling, could reversibly change its “acoustic depth” (background versus foreground). Thus, identification was facilitated by males and females entering the nest-box at different times while calling continuously. Lastly, because jackdaw males mount females for copulation, the microphone on the male’s back was at the largest possible distance from the vocalising female (in this specific context), and the female’s vocalisations were thus attenuated by the bodies of both mating birds (Figs. 31, 32).

30.4.2. Context recognition. Because these microphones move through the same acoustic environment as their carriers, they do not exclusively record vocalisations, but also characteristic movement patterns (e.g. flight, self-maintenance) and background sounds (e.g. anthropogenic sounds, vocalising conspecifics). Some of these sounds were annotated following video-validation to acoustically detect copulations, and whether a focal bird was inside or outside a nest-box (Fig. 32). For this, a human annotator (LFG) inspected these recordings, and was trained as follows to acoustically recognise the sounds of copulations (via field observations and video-validation from nest-box cameras) and of the focal bird being inside versus outside a nest-box (from nest-box cameras, and in a captive validation study (Stowell et al., under review). Copulations were detected as a combination of the abovementioned loud copulatory vocalisations, wing-flapping and other movement sounds, and were sometimes accompanied by shorter simultaneous vocalisations. The acoustic features of a focal bird being inside as opposed to outside a nest-box were 1) a combination of lower overall levels of ambient sounds, slight reverberation of vocalisations, and “wooden” sounds (beak, claws or body touching wooden nest-box walls) when inside the nest-box, and 2) found in
between characteristic loud “scraping” sounds caused by the microphone touching the entrance hole during entrance and exit (Audio 1).

Audio1 3 male on-board sound recordings separated by silence. i) a vocal event involving one call by a non-focal but close individual, one call by the focal and 3 calls by distant individuals, ii) 1 female short call in between 2 a different male’s copulatory calls (from Fig. 32B), iii) entire copulation call bout (from Fig. 32A). Note scraping sound as male exits nest-box at the end.
Acknowledgments

We thank Tiffany Pelayo van Buuren for field assistance, Ingrid Schwabl for video-coding, Wolfgang Goymann for comments and support, Auguste von Bayern for discussions and access to the jackdaw colony.

Author contributions
LFG conceived the study, carried out field work and analyses; MG provided key materials; both authors contributed to writing and reviewing the manuscript.

Competing financial interests
We have no competing financial interests.
PART III – GENERAL DISCUSSION
In my thesis, I investigated vocal communication in group-living songbirds with individual-level resolution, with a specific focus on calls. To do so, I carried out two studies describing and evaluating different tools (chapters 1 & 3), and two studies applying these tools to current research questions (chapters 2 & 4). Using two different types of recording techniques – transmitters (chapters 1 & 2) and loggers (chapters 3 & 4) – on two different study species – zebra finches (chapters 1 & 2) and jackdaws (chapters 3 & 4) – I have shown that on-board microphones allow recording private communication channels inside groups, from animals behaving freely in biologically meaningful settings, and investigated aspects of call communication previously impossible to address.
SECTION 32

Individual sound recordings from on-animal microphones

32.1. Transmitters versus loggers

Audio transmitters are extremely lightweight and allow long-term, continuous and fully synchronised sound recordings of individuals, making them useful for complex lab settings and even field studies under specific conditions, for example for highly territorial, non-dispersing, or highly predictable animals. However, audio transmitters are rather limited in range (so far). Further, their recordings may depend on the orientation and distance towards a receiving antenna. They may also be influenced by the properties of materials in between the sending and receiving antenna, thus making them susceptible to interferences, degradation or even complete loss of information. Currently, audio transmitter recordings also require a lot of additional technical equipment and logistics (antennas, receivers, cables, A/D converters, computer and appropriate multi-channel recording software etc.).

Because loggers store information on board, they are independent from any receiving base station, making recording quality more stable and comparable between settings. Thus, because they do not require additional equipment for recording, they are very useful for studying vocalisations of animals that move in space unpredictably, in short time or over large distances, or that behave in environments involving large obstacles, such as buildings or mountains (impenetrable for transmitters). However, loggers also have high energy and storage demands, making them heavier than transmitters, and thus limiting recording duration, as well as the range of suitable study species. To my knowledge, audio loggers also do not (yet) allow remote download, due to the large size of audio files, which means that the devices need to be retrieved and physically intact for data access and download. Further, due to clock drift, audio tracks coming from different individuals are never perfectly aligned (see Fig. 33). This can be problematic, especially for studies that focus on fine-scale temporal information of vocal events. To some extent, clock drift may be overcome by post-hoc synchronisation in combination with an external reference, such as infrared pulses, master devices or GPS (Anisimov et al., 2014; Levin et al., 2015; Walker et al., 2015). If drift rates are constant, audio tracks can thus be expanded or compressed to adjust the timing, but not
without affecting spectrotemporal properties of the recorded sounds (e.g. frequency and duration). But sometimes high temporal precision is required, e.g. to investigate tightly coordinated occurrences of calling events between individuals (chapters 1 & 2, Appendix 1) or the relationship with neuronal activity [Ter Maat et al. 2014]. Thus, absolute track alignment makes audio transmitter backpacks highly attractive for investigating temporal patterns of communication, because they allow multi-channel (100% parallel) recordings (see Fig. 3) – which do not require post-hoc synchronisation or external reference techniques that would increase backpack weight or/and may involve reduced environmental complexity [Anisimov et al. 2014].

Figure 33. Preliminary investigation of clock drift: 4 loggers simultaneously recorded the sounds of a stopwatch set to “beep” every 10 minutes. To investigate relative drift, one logger (105) was set as a reference (also because the stopwatch drifted in time as well). Drift rates were more or less constant (apart from timepoints 2 and 3 of logger 267). Over the course of 2.2 hours, loggers drifted between 2.7 and 24.6 seconds relative to the reference logger.

32.2. Drawbacks

Unfortunately, recording quality of on-animal microphone methods is not as high as with state-of-the-art, high-quality external microphone recording techniques. This is due to the miniaturised microphones and to the fact that vocalisations may overlap with animal movement sounds, or might be degraded otherwise (see
above). Therefore, it is not advised to compare recordings from on-animal and external microphones directly, as they show differences in spectro-temporal features (Figs. 6 and 7). Also, full automation of on-board recording analyses is not (yet) advisable, but should rather be supervised, i.e. verified and refined, by a (trained) human analyst. As mentioned in the introduction, individuality is often encoded in animal vocalisations, due to morphological or other individual features [Bradbury and Vehrencamp 2011]. However, in on-board sound recordings, these differences may be confounded with other factors that coincide with acoustic between-individual differences, for example with different backpacks, slight differences in backpack application, or with characteristic background sounds of different animals (e.g. different territories). Thus, investigating individuality calls for external recordings without overlapping movement sounds, with standardised acoustic backgrounds, better quality microphones etc. which may then allow automated analytic tools in some cases. For an investigation of individuality encoded in jackdaw vocalisations, recorded via external microphones under standardised conditions, see Appendix 2.

32.3. Advantage: context

32.3.1. Vocal behaviour in context. But on-board microphones have an immense advantage towards conventional, external recording techniques: they provide individual-level vocalisation recordings in acoustically challenging contexts. Studies in captivity as well as in the field may thus investigate individuals’ vocalisations even if animals vocalise in close proximity to each other (chapters 1 - 4), when using soft vocalisations (chapter 2), or if out of reach of an observer, for example in very private contexts (chapters 2 - 4), in inaccessible terrain (chapters 3 & 4) or when covering large distances (chapter 3). Because on-animal microphones are always at the same distance to the vocalising focal individuals, they also allow a standardised assessment of amplitude. This can be difficult in the field using conventional methods, especially with highly mobile species and directional sounds such as bird song [Brumm and Todt 2002] [Brumm 2004] [Zollinger et al. 2011]. Thus, on-animal microphones may also be useful for investigating vocalisation amplitudes in relation to different environmental factors (social, abiotic), as well as vocal communication in noise. Study animals can now be housed in adequate settings that are meaningful for the species and the research question involved, or may even be recorded in the wild. Thus, after an appropriate habituation period (chapter 1), these methods allow recording focal animals behaving “normally” and exhibiting vocal behaviours, such as soft calls of pair members (chapter 2) or copulation calls (chapter 4), which
are unlikely to occur in overly simplified conditions, but difficult to record using conventional methods. Further, although simple experiments can be useful for answering specific detailed questions (chapter 1, Appendices 1 & 2), investigations of vocal interactions are no longer limited to the pair or group level, or restricted to examinations of isolates' vocal responses towards playbacks. Instead, it becomes possible to investigate individual contributions to naturally occurring vocal interactions, and thus to investigate the role of vocal communication with respect to specific selection pressures. Thus, providing adequate context is key for detailed investigations that reflect a more realistic picture of vocal signals and thus eventually may lead to a more functional understanding of vocal communication.

32.3.2. Acoustic context recognition. During the process of analysing the continuous on-board sound recordings of wild jackdaws, I noticed that these streams of sound contained more information than merely vocalisation data, and may shed some light on the according acoustic contexts in which vocalisations are made. Intrigued by this, I conducted a video-validation study in captivity, and initiated an interdisciplinary collaboration to investigate this further (chapter 3). Observations in captivity and in the field, as well as human-coded, video-validated and machine-learned annotations confirmed that the on-bird sound recordings were useful for disentangling focal versus non-focal vocalisations, and provided information on an animal’s immediate acoustic context: i) its surroundings, such as specific locations (nest-site versus foraging site or roost), or the presence of vocalising conspecifics (or heterospecifics), and ii) also surprisingly detailed aspects of the focal animal's behaviour, from flight or preening up to bill wipes and head turns (chapter 3). Lastly, combining microphone backpacks with other observation techniques even revealed extra-pair copulations (chapter 4).

Of course, the detection and quantification of these contexts rely on the presence of acoustic information, meaning that "silent" events will not be captured. For instance, during observations of the wild jackdaw colony, I realised that encounters with avian predators often resulted in group flight responses without a single loud vocalisation. Thus, as already suggested by Lorenz, jackdaws seem to use further channels of information transfer in addition to vocal communication, for example tail movements as an invitation to joint travel (Lorenz 1931). However, on-animal sound recordings are unique in the amount and type of information they contain: they combine individual vocalisation recordings with information on focal animal behaviour and background context – all by recording a single channel of information (sound).
Other remote sensing techniques (by themselves) do not usually provide extensive information on the social and abiotic environment of an animal. For instance, GPS (Global Positioning System) loggers have become a common tool for studying the spatial location of animals, which is helpful for understanding daily or migratory movement and distribution patterns (review by Wilmers et al., 2015). But unless such tags are applied to multiple individuals or combined with additional techniques, this does not provide information on the animal’s surroundings or current behaviour. More fine-scaled behavioural or physiological data can be obtained via observations or other on-animal devices (e.g. accelerometers, heart-rate loggers etc., review by Wilmers et al., 2015). Therefore, multiple on-animal devices or measuring techniques are required to achieve a similar amount of information content as carried by sound (exceptions to this could be visual data, e.g. Yoda et al., 2011; Rutz and Troscianko, 2013). However, placing multiple devices on animals comes at a cost of animal weight load, and is thus not (yet) appropriate for most birds. Therefore, tapping into the full potential of on-animal sound recordings by extracting their contextual information could be highly rewarding for studying the vocal behaviour of wild birds in more detail. In this way, multiple sources of information could be obtained when observations are impossible, without increasing animal weight load, to investigate vocal behaviours in relation to the immediate bioacoustic contexts in which they occur.

So far, our approach (chapter 3) does not yet allow fully automated recognition of acoustic events without human supervision and solid validation. The development of such tools remains for future study and methodological testing, beyond the focus of this dissertation (and the average biologist’s expertise). However, with technical advancement, on-animal microphone technology is likely to become smaller, lighter and more efficient in terms of battery life and storage capacities. In my opinion, these methods are thus likely to become increasingly incorporated into the biologist’s toolbox for quantifying animal behaviour. Therefore, scientists should start thinking about ways of analysing and optimising the information yield of such large streams of acoustic data, and we provide a first step into doing so.
Calling behaviour of group-living songbirds in naturalistic contexts

33.1. Building a jackdaw call repertoire?

By using on-animal microphones for the first time in wild songbirds in nature, I was able to record previously unknown aspects of calling behaviour in group-living birds (jackdaws), and to investigate aspects of the bioacoustic context in which they were produced. Collecting on-board sound recordings of these cautious, highly cognitive birds [Lorenz, 1931] was not trivial, and required extensive testing and preparations, a lot of tricks and patience. In sum, a pilot phase and two field seasons yielded recordings with sufficient quality from 10 out of 22 audio-tagged individuals, resulting in about 256 hours of individual-level sound recordings. From these, I obtained vocalisations of individuals inside groups, in different contexts (chapter 3), for example at different breeding stages, at the night roost, during flight or foraging, or inside the nest during copulation (chapter 4) or chick provisioning. Preliminary analyses showed that the birds used a wide range of vocalisation types, despite wearing the backpacks. For instance, I recorded different variations of the species’ most characteristic “kjack” calls, for example before feeding chicks (“food calls”, [Lorenz, 1931]), as well as nest defence, alarm and copulation calls. These recordings of individual vocalisations occurring spontaneously in their immediate natural bioacoustic context could thus be a good starting point for a functional description of the jackdaw repertoire, providing insights into the vocal behaviour of this species previously impossible to obtain. However, such a detailed investigation of the complete call repertoire remains for future study, because efficient analytic tools are not available (quite) yet. Therefore, I presented one concrete example of unknown jackdaw vocal behaviour: “copulatory calls” (chapter 4).

33.2. Jackdaw copulatory calls and extra-pair behaviour

Jackdaw copulatory calls are context-specific and rather puzzling vocalisations. They are produced exclusively during copulation, are long, harsh and very loud
Despite being very conspicuous, copulatory calls have not been described in much detail. Because jackdaws usually copulate inside their nest cavities (Dwenger 1989, Cramp and Perrins 1994), they are difficult to observe and to record in this context (Dwenger 1989), and it was previously not possible to identify which sex produced the copulatory vocalisations (Dwenger 1989, Cramp and Perrins 1994). Therefore, this interesting vocal behaviour is unexplored, and its function unknown. Not all animal species vocalise during copulation, and it seems obvious that decreased vigilance in this context, in combination with loud calls, could come at the cost of increased predation risk (Gwynne 1989, Siemers et al. 2012) or conflict with conspecifics (Hauser 1998, Løvlie et al. 2014). However, in some species, copulatory calls have been associated with different functions, ranging from mate attraction (Anoop and Yorzinski 2013, Løvlie et al. 2014) to third-party social information transfer (Hauser 1998, Tobias 2002, Townsend and Zuberbühler 2009). Using the acoustic on-board sound recordings, I was able to identify the respective male as the source of copulation calls (chapter 4), and, in combination with video validation (and field observations of colour-banded birds), revealed extra-pair copulation (EPC) behaviour. Previous studies have reported zero to very low rates of extra-pair fertilisation in this species (Liebers and Peter 1998, Henderson et al. 2000). Throughout the literature, the jackdaw is often cited as a rare example of absolute monogamy in birds (most strongly: Arnold and Griffiths 2003, Ådahl et al. 2004, Emery et al. 2007, von Bayern et al. 2007, Verhulst et al. 2014, Campobello et al. 2015). However, in many cases, this sometimes fundamental assumption is based on the genetic investigation (using DNA-fingerprinting) of a single colony (n = 16 complete families out of 40 nests) in which the authors did not find evidence of extra-pair fertilisations (Henderson et al. 2000). Further, although this study revealed nest-box intrusions through detailed morning observations of tagged individuals, it is not clear what happened inside the intruded nest-boxes. In our study colony (ca. 20 nest sites), I observed multiple instances of EPCs via nest-box cameras and close-up observations of colour-banded birds. Unfortunately, exact numbers and genetic data are not available, as this was not the main focus of my PhD. However, I also obtained EPC recordings from 2 out of 4 males (backpack sound) and from 1 out of the 4 respective females (video). This would be highly unlikely if EPC behaviour were as extremely rare as suggested, especially given my low sample size. Also, not only unsuccessful birds (Henderson et al. 2000) engaged in this behaviour. Interestingly, the one male I managed to acoustically record during both IPCs and EPCs changed its behaviour: while copulation call bout duration during IPCs was consistently

However, the female EPC video suggested severe fighting between the intruding male and resident female.
short, it increased during copulation with an extra-pair mate. An increase in copulation duration could indicate different factors, such as difficulties between non-synchronised copulation partners or even female resistance. However, the underlying factors, and whether this change in behaviour could be a general pattern or was unique to this specific male, remain unclear. Therefore, my results strongly call for state-of-the-art behavioural and genetic analyses to investigate parentage and (social) monogamy in wild jackdaws.

The fact that the loud copulatory vocalisations were not only produced during intra-pair copulations (IPCs), but also during EPCs, strongly suggests that copulatory calls do not exclusively serve as a signal for pair bond strength, opening up the floor for diverse alternative speculation. A study on Galapagos tortoises (Crawford and Awbrey, 1978) suggested that the loud mating bellows previously claimed to “frighten the female” (after Mertens, 1946) could inhibit female movement for successful copulation. This finds support through experiments with devocalised male rats and playbacks, showing that specific male vocalisations during copulation were correlated with the likelihood of females remaining stationary during the act of copulation (White et al., 1998). Especially in dark or crowded places, it would also make sense for males to use vocalisations to unambiguously signal their intention to mate, as opposed to showing aggression, as suggested for bats (Liu et al., 2013). Given that at least some jackdaw vocalisations carry information on individual identity (Lorenz, 1931; Dwenger, 1989; Zandberg et al., 2014, see also Appendix 2), such vocalisations may also provide information on the identity of an approaching mate. However, because copulatory calls are given during the act of copulation, this seems a bit late for avoiding copulation with the “wrong male” (suggested by Liu et al., 2013). In jackdaws, the production of male copulatory calls during IPCs and EPCs makes a function in infidelity avoidance rather unlikely (also, jackdaws copulate inside nest cavities that are strongly defended by the residents, thus making “mistakes” highly unlikely). Instead, individual, context-specific, copulatory calls could provide information for post-copulatory sperm competition (male quality) or social information for third-party individuals. Indeed, in species as diverse as chickens and apes, (female) copulatory vocalisations have been shown to be correlated with dominance of their copulation partner (Pradhan et al., 2006; Townsend and Zuberbühler, 2009; Løvlie et al., 2014), and have been associated with mate-guarding, infanticide prevention, sperm and direct male-male competition, including the attraction of better quality males to drive off those of lower quality (Hauser, 1998; Pradhan et al., 2006). Lastly, copulatory calls could also be related with reproductive stimulation. The chaotic structure of jackdaw copulatory calls could indicate an honest signal of male effort and/or arousal (Wilden et al., 1998; Riede et al., 2007; Zollinger et al.)
In some primates including humans, it has been suggested that female arousal vocalisations during copulation may influence the timing of male ejaculation, thus potentially manipulating the duration of an encounter (Brewer and Hendrie 2011), and possibly even optimising fertilisation (Hamilton and Arrowood 1978). This would be in line with other studies showing that vocalisations (although usually pre-copulatory) can be related with hormonal stimulation and fertility in other animal species (Hauser 1998; Cheng 2003). For instance, early studies showed that the playback of breeding group vocalisations in ring doves (Lehrman and Friedman 1969) and of single male vocalisations in budgerigars (Brockway 1966) resulted in increased ovarian development and egg-laying in isolated females. Also, experiments with devocalised male budgerigars demonstrated that high levels of testicular activity were maintained only if males were able to produce vocalisations themselves (sham operation) (Brockway 1967). Furthermore, various brain areas associated with bird song express high amounts of specific hormone receptors, making the “song control system” highly sensitive to hormones like testosterone, estradiol and melatonin. Hormones are thus involved in shaping vocal behaviour during ontogeny and song learning (Gahr 2014), but are also part of an intricate interplay of hormones, social behaviours and vocal signalling which requires further investigations (Oliveira 2005) – especially since previous studies focused on how hormones were associated with specific acoustic features of vocalisations (mostly song), but not on how hormones may influence interactive vocal communication, i.e. how vocal signals may be differentially integrated in between-individual communication.

### 33.3. Defining calling and responding individuals in vocal communication

An essential part of vocal behaviour is not only the individual-level production, but also how vocalisations are used in communication with other individuals. In animal vocal interactions, the types (Vehrencamp 2001; Miller et al. 2004; Searcy and Beecher 2009; Baker et al. 2012) and timing (Searcy and Nowicki 2006; Brumm and Slater 2007; Sugíjura 2007; Miller et al. 2009; Ter Maat et al. 2014) of vocal events have been shown to be important – not only for the individuals directly involved (Schmidt et al. 2006, 2007), but also by providing information for bystanders (Naguib and Todt 1997; Mennill 2002; Schmidt et al. 2007; Searcy and Yasukawa 2016). Thus, as mentioned above (Fig. 2 of introduction), defining signallers and receivers involved in vocal communication can be challenging, especially with multiple vocalising individuals, as in groups. We used the relative timing of calls to statistically identify initiating and responding individuals in vocal
exchanges, in order to approach an objective definition of senders and receivers in the presence of multiple individuals (chapter 2, Appendix 1). Because this requires full temporal synchronisation of individual audio tracks (chapter 1 and see discussion above), we used perfectly aligned audio recordings of individual zebra finches to investigate the temporal relationships of calling events for the individuals in groups. In chapter 2, we chose all possible pairwise combinations of all group members, and examined cumulative sums and confidence intervals of vocal events in peri-stimulus time histograms (PSTHs), split up by vocalisation type. For each combination, the strength of a vocal response was calculated and subsequently used to build group vocal networks (confusion matrices, Fig. 2.5). Appendix 1 (Stowell et al., 2016) takes a closer look at the statistical tools for investigating the timing of calls in vocal interactions. In this manuscript, a probabilistic model (initially designed for studying neurons) was applied to vocalisation data collected from captive zebra finches under simplified, standardised conditions, as well as to a subset of the more complex dataset described in chapter 2. Reanalysing my data in this way reflected similar overall results, but provided more detailed information while relying on fewer assumptions than the method used in chapter 2 (that pairwise relative timings of vocal events were independent from group). Because the method described in Appendix 1 may be adapted to different conditions and may also be used as a generative model, it seems promising and useful for future studies aiming to investigate temporal aspects of animal vocal communication in further detail.

33.4. Call types, timing and vocal interactions of zebra finch pairs in groups

Both approaches (chapter 2, Appendix 1) revealed that zebra finch calls were exchanged between individuals on a short temporal scale (<500 ms), and that call communication shared between establishing pair members increasingly differed from communication with other members of the group, especially with the onset of breeding.

Although the zebra finch has become a prime model species for studying different aspects of song, much less is known about its calls which are produced in large numbers, almost constantly throughout the day (Zann, 1996; Beckers and Gahr, 2010). To record zebra finch calls, many studies either isolated the birds to obtain individual-level information (e.g. Blaich et al., 1996a; Vignal et al., 2004), or investigated vocalisations at the pair or group level, without individual information (Elie et al., 2010, 2011). In what seems a tremendously laborious process, another zebra finch call repertoire study recorded the vocalisations of small group units
externally, and subsequently assigned them to individuals manually, via observations (Elie and Theunissen, 2016). In this way, a yield of 8000 individual vocal events was recorded over the course of months – a number which can be easily obtained within a few hours of on-bird group recordings using our method. Therefore, it is not surprising that studies investigating the zebra finch call repertoire and vocal interactions from an individual and functional perspective are rare.

A previous study in the wild suggested the quiet calls of zebra finch pairs exchanged at the nest could be a “duet” (Elie et al., 2010). A recent review showed that male-female duets in birds often follow specific rules: tightly coordinated timing, initiation by both individuals, and sex-specific vocalisations (Dahlin and Benedict, 2014). Vocal duetting between pair members in birds is likely to have multiple functions, depending on the species, for example to maintain contact and jointly defend resources, or for mate-guarding (Hall, 2004; Benedict, 2008; Dahlin and Benedict, 2014; Baldassarre et al., 2016). In songbirds, duetting has also been correlated with the absence of migration (Logue and Hall, 2014), and there is debate whether vocal duets were sexually selected, because they coincide phylogenetically with (at least ancestral) female song (Odom et al., 2015). To my knowledge, these songbird studies only investigated vocal exchanges involving learned song (as opposed to “innate” calls) as vocal contributions from at least one partner. Vocal learning in birds is a phylogenetically derived trait (exclusive to songbirds, hummingbirds and parrots) (Jarvis et al., 2014), but duetting is widespread, and may occur also in birds that do not learn their vocalisations (Thorpe et al., 1972; Hall, 2004; Maurer et al., 2008; Hall, 2009; Dahlin and Benedict, 2014; Baldassarre et al., 2016), and even in insects (Bailey, 2003).

Indeed, the soft calling interactions (involving tet, stack and cackle calls) of zebra finch breeding pairs (chapter 2) followed similar rules as the majority of bird duets. As mentioned above, they occurred on a short temporal scale, suggested sex-specific usage of call types in pair vocal interactions, and both males and females could play the role of the “initiating” and “responding” vocal partner. Not all call types were exchanged in this manner, and there were call combinations that never occurred (or were far from reaching the significance criterion), for example loud distance calls (used to maintain contact over distance Blaich et al., 1996a; Zann, 1996) combined with whines (softer calls used in mating context Zann, 1996) (Fig. 2.5). Further, I found that aspects of this specific vocal behaviour predicted whether or not a pair would produce a clutch of eggs during the (three-week) experimental period – which could be crucial for an opportunistically breeding species that relies on fast physiological and behavioural adaptations to unpredictable environmental conditions for successful reproduction (Perfito et al., 2007; Prior et al.)
Indeed, the changes in breeding status were not only associated with altered call usage (call repertoire and calling interactions), but also with changes in hormonal profiles, thus calling for more detailed investigations of opportunistic breeding, hormones and vocal behaviour.

At this point it is not clear whether these pair calling interactions are truly comparable to duetting in other avian species, or follow other rules of vocal communication. In any case, male and female zebra finch pair members did use different calls as interactive social behaviour, and did so differently for pair and group members which supports previous suggestions (Blaich et al., 1996a; Zann, 1996; Elie et al., 2010). One possible scenario could be that, due to the costs of conspicuous signalling (e.g. predation risk Kleindorfer et al., 2016), loud song could have been exchanged functionally by calls in pair vocal duetting, as an already existing but potentially less conspicuous signalling class. More detailed studies, including neural investigations, are required to find out whether this specific call usage is learned or innate in this species, and to unravel its underlying mechanisms. Further research, optimally with wild birds, is necessary to investigate the function in more detail. However, by providing first evidence for a link between pair calling behaviour and successful reproduction in this opportunistically breeding bird, I found strong empirical support for previous theoretical discussions of the potential of bird calls for studies on mechanism and evolution of vocal communication (Marler, 2004; Marler and Evans, 2008).
Overall conclusions

Songbird calls occur in a multitude of contexts, but have been neglected in the scientific literature for a long time, compared to song. Having overcome the challenges of recording calls and assigning their signal source, it is now time to appreciate their full potential, and to study them in further detail. Calls can be discrete vocal classes or gradually modified in acoustic structure, they may differ in how they are exchanged between individuals, and may be combined with other modalities. This multi-layered variation makes them incredibly flexible in their usage, thus giving calls the potential to reveal emotional states up to semantic meanings, and can help us understand production mechanisms and universal paradigms of learned and unlearned communication. The two different approaches I used in my thesis to record individual-level vocal communication both have characteristic features that may carry convincing advantages but also some disadvantages. It is thus reasonable to choose the appropriate method in advance, keeping in mind any requirements related with the scientific question and the study organism’s specific biology. In this way, it is now becoming increasingly feasible to investigate challenging aspects of animal communication previously impossible to address. Further, like other novel techniques, these methods can bring about unexpected, exciting findings. Therefore, using individual-level sound recordings of vocalisations obtained in natural situations, we may not only record quiet calls and disentangle “which animal is saying what” in spontaneously occurring vocal interactions, but even set the vocalisations in their immediate context to catch a glimpse of how and why animals produce specific signals in communication.
Author contributions and Acknowledgments
Author contributions

Chapter 1
PBD, MNA and I jointly conceived the study, carried out experiments, analysed the data, prepared figures and supplementary materials, and wrote the manuscript. HS provided key materials and technical support. MCG and AT provided key materials and were involved in conceiving the study, reviewing and writing the manuscript.

Chapter 2
All authors were involved in conceiving the study and revising the manuscript. I carried out the experiments, analysed the data, prepared figures and supplementary materials, and wrote the paper. WG was involved in data collection, data analysis and study design. AT and MCG initiated the study, and provided essential materials and analysis tools.

Chapter 3
I came up with the initial concept of the project, and DS and I jointly conceived this manuscript. All authors wrote parts of the manuscript, prepared figures, and reviewed the manuscript. DS conceived the machine learning side of it, implemented the classifier-based system, and led on evaluation and manuscript writing. EB implemented the PLCA-based system, and collaborated in performing the evaluation. I conceived and conducted work in the aviary and in the field, provided key materials, performed initial sound analyses, annotated all data, and was involved in the evaluation of the method.

Chapter 4
I conceived the study, performed field work and analyses, prepared figures and supplementary materials, and wrote the manuscript. MCG provided essential materials, and was involved in writing and revising the manuscript.
Acknowledgments

“Lend me your ears and I’ll sing you a song
And I’ll try not to sing out of key
Oh I get by with a little help from my friends”

Lennon & McCartney

I am grateful to my PhD advisor, Wolfgang Goymann, for supporting me at all stages of my PhD, for tips in the field, on statistics and writing, for superfast reviews, listening to my ideas and encouraging me to follow them up. Manfred Gahr, thank you for supporting me all this time, and letting me do my thing - especially when you were not entirely convinced. I would also like to thank my three other PAC(k) members: Andries, for discussions, your patience and support; Henrik, for highly appreciated advice; Auguste, for discussions and pragmatic help, and thank you and your family for letting me work with the wild colony, and build a sound chamber downstairs. Special thanks to Dan, my co-author and new friend, for getting drawn in to my projects, picking up and advancing my questions and ideas, and for being efficient, honest and so much fun.

Thank you to the IMPRS for funding, excellent training and support (Mäggi!), and thanks to everyone who supported me in the field, lab, office and/or a sticky situation: my field assistants Magdalena + Lena; Katrin; Irmgard; Berenika (for not being afraid of ghosts, and for everything else) + Mark for fruitful discussions, and for aggregating to form a larger creature; Christian + the Maschinenring; Monika + Antje for lab work; Ingrid for watching videos; Hannes (der kann es!), Lisa (die Große), Markus (ich will was basteln) for technical support; my other co-authors, Pie + Nico, for being incredibly annoying and fun at the same time; Fränzi! + Kami (the R-tist) for showing me how to wrestle the dRagon; Claudia (for luring me in), Susi (for being you), Moritz (for walks and splinter removals), Maggie, Camila, Rene, Shouwen, Mariana, Ani + the Latin lunch crew for being the icing on the office cake.

Last but not least I thank all my non-scientific friends and my family, above all my parents for always being there for me, especially in rough times; Opa, danke für das Studentenfutter und deinen unschlagbaren Optimismus; Mert, thank you for sound equipment and support, feedback on my work, and for all the warmth, good food and music.
Bibliography


Cohen, J. and Cheng, M.-F. (1979). Role of vocalizations in the reproductive cycle of ring doves (Streptopelia risoria): Effects of hypoglossal nerve section on the


Vandenabeele, S. P., Shepard, E. L., Grogan, A., and Wilson, R. P. (2012). When three per cent may not be three per cent; device-equipped seabirds experience


Appendix
Appendix 1:
Detailed temporal structure of communication networks in groups of songbirds

Dan Stowell *, Lisa F. Gill and David Clayton

Abstract
Animals in groups often exchange calls, in patterns whose temporal structure may be influenced by contextual factors such as physical location and the social network structure of the group. We introduce a model-based analysis for temporal patterns of animal call timing, originally developed for networks of firing neurons. This has advantages over cross-correlation analysis in that it can correctly handle common-cause confounds and provides a generative model of call patterns with explicit parameters for the influences between individuals. It also has advantages over standard Markovian analysis in that it incorporates detailed temporal interactions which affect timing as well as sequencing of calls. Further, a fitted model can be used to generate novel synthetic call sequences. We apply the method to calls recorded from groups of domesticated zebra finch (Taeniopygia guttata) individuals. We find that the communication network in these groups has stable structure that persists from one day to the next, and that ‘kernels’ reflecting the temporal range of influence have a characteristic structure for a calling individual’s effect on itself, its partner and on others in the group. We further find characteristic patterns of influences by call type as well as by individual.

Appendix 2:  
Individual identity in songbirds: signal representations and metric learning for locating the information in complex corvid calls

Dan Stowell *, Veronica Morfi, Lisa F. Gill

Abstract

Bird calls range from simple tones to rich dynamic multi-harmonic structures. The more complex calls are very poorly understood at present, such as those of the scientifically important corvid family (jackdaws, crows, ravens, etc.). Individual birds can recognise familiar individuals from calls, but where in the signal is this identity encoded? We studied the question by applying a combination of feature representations to a dataset of jackdaw calls, including linear predictive coding (LPC) and adaptive discrete Fourier transform (aDFT). We demonstrate through a classification paradigm that we can strongly outperform a standard spectrogram representation for identifying individuals, and we apply metric learning to determine which time-frequency regions contribute most strongly to robust individual identification. Computational methods can help to direct our search for understanding of these complex biological signals.
