Acoustic sequences in non-human animals: A tutorial review and prospectus

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Biological Reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Date Submitted</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Complete List of Authors:

Kershenbaum, Arik; National Institute for Mathematical and Biological Synthesis
Blumstein, Dan; University of California, Ecol Evol Biol
Roch, Marie; San Diego State University, Department of Computer Science
Akcay, Caglar; Cornell University, Lab of Ornithology
Backus, Gregory; North Carolina State University, Department of Biomathematics
Bee, Mark; University of Minnesota, Department of Ecology, Evolution, & Behavior
Bohn, Kirsten; Florida International University, Integrated Science
Cao, Yan; University of Texas at Dallas, Department of Mathematical Sciences
Carter, Gerald; University of Maryland, Department of Ecology, Evolution, & Systematics
Cäsar, Cristiane; University of St. Andrews, Department of Psychology & Neuroscience
Coen, Michael; University of Wisconsin, Department of Biostatistics and Medical Informatics
DeRuiter, Stacy; University of St. Andrews, School of Mathematics and Statistics
Doyle, Laurance; SETI Institute
Edelman, Shimon; Cornell University, Department of Psychology
Ferrer -i-Cancho, Ramon; Universitat Politècnica de Catalunya, Departament de Llenguatges i Sistemes Informàtics
Freeberg, Todd; University of Tennessee, Department of Psychology
Garland, Ellen; AFSC/NOAA, National Marine Mammal Laboratory
Gustison, Morgan; University of Michigan, Department of Psychology
Harley, Heidi; New College of Florida, Department of Psychology
Huetz, Chloe; Université Paris Sud, Department of Bioacoustics
Hughes, Melissa; College of Charleston, Department of Biology
Hyland Bruno, Julia; Hunter College, The City University of New York, Department of Psychology
Ilany, Amiyaal; National Institute for Mathematical and Biological Synthesis,
Jin, Dezhe; Pennsylvania State University, Department of Physics
Johnson, Michael; Marquette University, Department of Electrical and Computer Engineering
Ju, Chenghui; Queen College, The City Univ. of New York, Department of
Biology
Karnowski, Jeremy; University of California San Diego, Department of Cognitive Science
Lohr, Bernard; University of Maryland Baltimore County, Department of Biological Sciences
Manser, Marta; University of Zurich, Department of Evolutionary Biology and Environmental Studies
McCowan, Brenda; University of California Davis, Department of Veterinary Medicine
Mercado, Eduardo; State University of New York, Department of Psychology
Narins, Peter; University of California Los Angeles, Department of Integrative Biology & Physiology
Piel, Alex; University of Cambridge, Department of Biological Anthropology
Rice, Megan; California State University San Marcos, Department of Psychology
Salmi, Roberta; University of Georgia at Athens, Department of Anthropology
Sasahara, Kazutoshi; Nagoya University, Graduate School of Information Science
Sayigh, Laela; Woods Hole Oceanographic Institute, Department of Biology
Shiu, Yu; Cornell University, Lab of Ornithology
Taylor, Charles; University of California, Ecol Evol Biol
Vallejo, Edgar; Monterrey Technical University, Department of Computer Science
Waller, Sara; Montana State University, Department of Philosophy
Zamora-Gutierrez, Veronica; University of Cambridge, Department of Zoology

Keywords: acoustic communication, information, information theory, machine learning, Markov model, meaning, network analysis, sequence analysis, vocalisation
Acoustic sequences in non-human animals: A tutorial review and prospectus

1

2

3

Arik Kershenaum1, Daniel T. Blumstein2, Marie A. Roch3, Çağlar Akçay4, Gregory Backus5, Mark
4
Bee6, Kirsten Bohn7, Yan Cao8, Gerald Carter9, Cristiane Cásar10, Michael Coen11, Stacy L. DeRuiter12,
5
Laurance Doyle13, Shimon Edelman14, Ramon Ferrer-i-Cancho15, Todd M. Freeberg16, Ellen C. Garland17,
6
Morgan Gustison18, Heidi E. Harley19, Chloé Huetz20, Melissa Hughes21, Julia Hyland Brune22, Amiyaal
7
Ilany23, Dezhe Z. Jin24, Michael Johnson25, Chenghui Ju26, Jeremy Karnowski27, Bernard Lohn28, Marta B.
8
Manser29, Brenda McCowan30, Eduardo Mercado III31, Peter M. Narins32, Alex Piel33, Megan Rice34,
9
Roberta Salmi35, Kazutoshi Sasahara36, Laela Sayigh37, Yu Shiu38, Charles Taylor39, Edgar E. Vallejo40,
10
Sara Waller41, Veronica Zamora-Gutierrez42

1 National Institute for Mathematical and Biological Synthesis, Knoxville, USA
2 Department of Ecology and Evolutionary Biology, University of California Los Angeles, USA
3 Department of Computer Science, San Diego State University, USA
4 Lab of Ornithology, Cornell University, USA
5 Department of Biomathematics, North Carolina State University, USA
6 Department of Ecology, Evolution and Behavior, University of Minnesota, USA
7 Integrated Science, Florida International University, USA
8 Department of Mathematical Sciences, University of Texas at Dallas, USA
9 Biological Sciences Graduate Program, University of Maryland, USA
10 Department of Psychology & Neuroscience, University of St. Andrews, UK
11 Department of Biostatistics and Medical Informatics, University of Wisconsin, USA
12 School of Mathematics and Statistics, University of St. Andrews, UK
13 Carl Sagan Center for the Study of Life in the Universe, SETI Institute, USA
14 Department of Psychology, Cornell University, USA
15 Department of Computer Science, Universitat Politècnica de Catalunya, (Catalonia) Spain
16 Department of Psychology, University of Tennessee, USA
17 National Marine Mammal Laboratory, AFSC/NOAA, USA
18 Department of Psychology, University of Michigan, USA
19 Division of Social Sciences, New College of Florida, USA
20 CNPS, CNRS UMR 8195, Université Paris-Sud, Orsay, France
21 Department of Biology, College of Charleston, USA
22 Department of Psychology, Hunter College and the Graduate Center, The City University of New York, USA
23 National Institute for Mathematical and Biological Synthesis, Knoxville, USA
24 Department of Physics, Pennsylvania State University, USA
25 Department of Electrical and Computer Engineering, Marquette University, USA
26 Department of Biology, Queen College, The City Univ. of New York, USA
27 Department of Cognitive Science, University of California San Diego, USA
28 Department of Biological Sciences, University of Maryland Baltimore County, USA
29 Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland
30 Department of Veterinary Medicine, University of California Davis, USA
31 Department of Psychology, Evolution, Ecology, & Behavior, University at Buffalo, The State University of New York, USA
32 Department of Integrative Biology & Physiology, University of California Los Angeles, USA
33 Division of Biological Anthropology, University of Cambridge, UK
34 Department of Psychology, California State University San Marcos, USA
35 Department of Anthropology, University of Georgia at Athens, USA
36 Graduate School of Information Science, Nagoya University, Japan
37 Biology Department, Woods Hole Oceanographic Institution, USA
38 Lab of Ornithology, Cornell University, USA
39 Department of Ecology and Evolutionary Biology, University of California Los Angeles, USA
40 Department of Computer Science, Monterey Institute of Technology, Mexico
41 Department of Philosophy, Montana State University, USA
42 Department of Zoology, University of Cambridge, UK, Centre for Biodiversity and Environmental Research, University College London, UK
ABSTRACT

Animal acoustic communication often takes the form of complex sequences, made up of multiple distinct acoustic units. Apart from the well-known example of birdsong, other animals such as insects, amphibians, and mammals (including bats, rodents, primates, and cetaceans) also generate complex acoustic sequences. Occasionally, such as with birdsong, the adaptive role of these sequences seems clear (e.g., mate attraction and territorial defence). More often however, researchers have only begun to characterise – let alone understand – the significance and meaning of acoustic sequences. Hypotheses abound, but there is little agreement as to how sequences should be defined and analysed. Our work here aims to forge such an agreement on key hypotheses, to outline suitable methods for testing these hypotheses, and to describe the major limitations to our current and near-future knowledge on questions of acoustic sequences.

This review and prospectus is the result of a collaborative effort between 43 scientists from the fields of animal behaviour, ecology and evolution, signal processing, machine learning, quantitative linguistics, and information theory, who gathered for a 2013 workshop entitled, “Analysing vocal sequences in animals”. Our goal is to present not just a review of the state of the art, but to propose a methodological framework that summarises what we suggest are the best practices for research in this field, across taxa and across disciplines. We also provide a tutorial-style introduction to some of the most promising algorithmic approaches for analysing sequences.

We divide our review into three sections: describing the different ways that information can be contained within a sequence, identifying the distinct units of an acoustic sequence, and analysing the structure of that sequence. Each of these sections is further subdivided to address the key questions and approaches in that area.

We propose a uniform, systematic, and comprehensive approach to studying sequences, with the goal of clarifying research terms used in different fields, and facilitating collaboration and comparative
studies. Allowing greater interdisciplinary collaboration will facilitate the investigation of many important questions in the evolution of communication and sociality.

Keywords: acoustic communication, information, information theory, machine learning, Markov model, meaning, network analysis, sequence analysis, vocalisation
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>CONTENTS</td>
</tr>
<tr>
<td>45</td>
<td>I. INTRODUCTION</td>
</tr>
<tr>
<td>46</td>
<td>II. THE CONCEPTS OF INFORMATION AND MEANING</td>
</tr>
<tr>
<td>47</td>
<td>(1) Information embedding paradigms</td>
</tr>
<tr>
<td>48</td>
<td>(2) Six information embedding paradigms</td>
</tr>
<tr>
<td>49</td>
<td>III. ACOUSTIC UNITS</td>
</tr>
<tr>
<td>50</td>
<td>(1) Identifying potential units</td>
</tr>
<tr>
<td>51</td>
<td>(2) Identifying production units</td>
</tr>
<tr>
<td>52</td>
<td>(3) Identifying perceptual units</td>
</tr>
<tr>
<td>53</td>
<td>(4) Identifying analytical units</td>
</tr>
<tr>
<td>54</td>
<td>IV. THE STRUCTURE OF SEQUENCES</td>
</tr>
<tr>
<td>55</td>
<td>(1) Markov chains</td>
</tr>
<tr>
<td>56</td>
<td>(2) Hidden Markov models</td>
</tr>
<tr>
<td>57</td>
<td>(3) Network models</td>
</tr>
<tr>
<td>58</td>
<td>(4) Formal grammars</td>
</tr>
<tr>
<td>59</td>
<td>(5) Temporal structure</td>
</tr>
<tr>
<td>60</td>
<td>V. FUTURE DIRECTIONS</td>
</tr>
<tr>
<td>61</td>
<td>(1) Two case studies</td>
</tr>
<tr>
<td>62</td>
<td>(2) Outstanding questions in animal vocal sequences</td>
</tr>
<tr>
<td>63</td>
<td>VI. CONCLUSIONS</td>
</tr>
<tr>
<td>64</td>
<td>VII. ACKNOWLEDGEMENTS</td>
</tr>
<tr>
<td>65</td>
<td>VIII. REFERENCES</td>
</tr>
</tbody>
</table>
Sequences are everywhere, from the genetic code, to behavioural patterns such as foraging, as well as the sequences that comprise music and language. Often, but not always, sequences convey meaning (and can do so more effectively than other types of signals; Shannon et al., 1949), and individuals can take advantage of the information contained in a sequence to increase their own fitness (Bradbury & Vehrencamp, 2011). Acoustic communication is widespread in the animal world, and very often individuals communicate using a sequence of distinct acoustic elements, the order of which may contain information of potential benefit to the receiver. In some cases, acoustic sequences appear to be ritualised signals where the signaller benefits if the signal is detected and acted upon by a receiver. The most studied examples include birdsong, where males may use sequences to advertise their potential quality to rival males and to receptive females (Catchpole & Slater, 2003). Acoustic sequences can contain information on species identity (e.g., in many frogs and insects; Gerhardt & Huber, 2002), on individual identity and traits (e.g., in starlings *Sturnus vulgaris*, Gentner & Hulse, 1998; wolves *Canis lupus*, Root-Gutteridge et al., 2014; dolphins *Tursiops truncatus*, Sayigh et al., 2007; and hyraxes *Procavia capensis*, Koren & Geffen, 2011), and in some cases, on contextual information such as resource availability (e.g., food calls in chimpanzees *Pan troglodytes*, Slocombe & Zuberbühler, 2006), or predator threats (e.g., in marmots *Marmota spp.*, Blumstein, 2007; primates, Schel, Tranquilli & Zuberbühler, 2009; Cäsar et al., 2012b; and parids, Baker & Becker, 2002). In many cases, however, the ultimate function of communicating in sequences is unclear. Understanding the proximate and ultimate forces driving and constraining the evolution of acoustic sequences, as well as decoding the information contained within them, is a growing field in animal behaviour (Freeberg et al., 2012). New analytical techniques are uncovering characteristics shared between diverse taxa, and offer the potential of describing and interpreting the information within animal communication signals. The field is ripe for a review and a prospectus to guide future empirical research.
Progress in this field has been somewhat hampered in the past partly by inconsistent terminology, conflicting assumptions, and different research goals, both between disciplines (e.g., between biologists and mathematicians), and also between researchers concentrating on different taxa (e.g., ornithologists and primatologists). Therefore, we aim to do more than provide a glossary of terms. Rather, we build a framework that identifies the key conceptual issues common to the study of acoustic sequences of all types, while providing specific definitions useful for clarifying questions and approaches in more narrow fields. Our approach identifies three central questions: What are the units of which the sequence is composed? How do we assess the structure with which these units are combined? How is information contained within the sequence? Figure 1 illustrates a conceptual flow diagram linking these questions, and their sub-components, and should be broadly applicable to any study involving animal acoustic sequences.

Our aims in this review are as follows: (1) to identify the key issues and concepts necessary for the successful analysis of animal acoustic sequences; (2) to describe the commonly used analytical techniques, and importantly, also those underused methods deserving of more attention; (3) to encourage a cross-disciplinary approach to the study of animal acoustic sequences that takes advantage of tools and examples from other fields to create a broader synthesis; and (4) to facilitate the investigation of new questions through the articulation of a solid conceptual framework.

In Section II we ask why sequences are important, and how information may be embedded within them. We present this section first, rather than in the order shown in Figure 1, because it is necessary early on to define and establish the significance of the terminology that will follow in the review. In Section III, we examine the questions of what units make up a sequence and how to identify them. In some applications the choice seems trivial, however in many study species, sequences can be represented at different hierarchical levels of abstraction, and the choice of sequence “unit” may depend on the hypotheses being tested. In Section IV, we examine the structure of the sequence, the mathematical and statistical models that quantify how units are combined, and how these models can be analysed, compared, and assessed. In Section V, we provide some case studies that illustrate our approach, describe
some of the evolutionary and ecological questions that can be addressed by analysing animal acoustic sequences, and look at some promising future directions and new approaches.
II. THE CONCEPTS OF INFORMATION AND MEANING

The complementary terms, “meaning” and “information” in communication, have been variously defined, and have long been the subject of some controversy (Dawkins & Krebs, 1978; Stegmann, 2013). In this section we explore some of the different definitions from different fields, and their significance for research on animal behaviour. The distinction between meaning and information is sometimes portrayed with meaning, on the one hand, as activity, and information, on the other hand, as form, or structure (Bohm, 1989).

Philosophical understanding of meaning is rooted in studies of human language and has a variety of schools of thought. Philosophers consider intension (a meaning or sense, e.g., a chair is something that one sits on) and extension (objects that are instances of an intension, e.g., a particular lounge chair), prototype theory, whether meaning is innate or learned, mental representations, and cognitive content. Philosophers also view meaning as computational/functional, as atomic or holistic, as bound to both speaker and audience, as speech act and performance, as rule bound or as referentially based, as description, as conventional, and as a game dependent on a form of life, among other examples (Christiansen & Chater, 2001; Martinich & Sosa, 2013).

Biologists (particularly behavioural ecologists), and cognitive neuroscientists have different understandings of meaning. For most biologists, meaning relates to the function of signalling. The function of signals is examined in agonistic and affiliative interactions, in courtship and mating decisions, and in communicating about environmental stimuli, such as the detection of predators (Bradbury & Vehrencamp, 2011). Behavioural ecologists study meaning by determining the degree of production specificity, the degree of response specificity, and contextual independence (e.g., Evans, 1997). Cognitive neuroscientists generally understand meaning through mapping behaviour onto structure-function relationships in the brain (Chatterjee, 2005).
Mathematicians understand meaning by developing theories and models to interpret the observed signals. This includes defining and quantifying the variables (observable and unobservable), and the formalism for combining various variables into a coherent framework (e.g., pattern theory; Mumford & Desolneux, 2010). One approach to examining a signal mathematically is to determine the entropy, or amount of structure (or lack thereof) present in a sequence.

The amount of structure measured does not indicate meaning, nor does measuring the amount of structure quantify the complexity of the meaning, if it is present. As an example, the structure may be indicative of meaning, but it could also be related to a strategy to ensure acoustic propagation in an adverse environment. A distinction is often made between a signal’s “content”, or broadcast information, and its “efficacy”, or transmitted information – the characteristics or features of signals that actually reach receivers (Wiley, 1983; Hebets & Papaj, 2005). This is basically the distinction between bearing functional information and getting that information across to receivers in conditions that can be adverse to clear signal propagation. A sequence may also contain elements that do not in themselves contain meaning, but are intended to get the listeners’ attention, in anticipation of future meaningful elements (e.g., Richards, 1981; Call & Tomasello, 2007; Arnold & Zuberbühler, 2013).

Context has a profound influence on signal meaning, and this should apply to the meaning of sequences as well. Context includes internal and external factors that may influence both the production and perception of acoustic sequences; the effects of context can partially be understood by considering how it specifically influences the costs and benefits of producing a particular signal or responding to it. For instance, an individual’s motivational, behavioural, or physiological state may influence response (Lynch et al., 2005; Goldbogen et al., 2013); hungry animals respond differently to signals than satiated ones, and an individual in oestrus or musth may respond differently than ones not in those altered physiological states (Poole, 1999). Gender may influence response as well (Tyack, 1983; Darling, Jones & Nicklin, 2006; Smith et al., 2008; van Schaik, Damerius & Isler, 2013). The social environment may influence the costs and benefits of responding to a particular signal (Bergman et al., 2003; Wheeler, 2010a; Ilany et al., 2011; Wheeler & Hammerschmidt, 2012) as might environmental attributes, such as
temperature or precipitation. Knowledge from other social interactions or environmental experiences can
also play a role in context, e.g., habituation (Krebs, 1976). Context can also alter a behavioural response
when hearing the same signal originate from different spatial locations. For instance in neighbour-stranger
discrimination in songbirds, territorial males typically respond less aggressively toward neighbours
compared with strangers, so long as the two signals are heard coming from the direction of the
neighbour’s territory. If both signals are played back from the centre of the subject’s territory, or from a
neutral location, subjects typically respond equally aggressively to both neighbours and strangers (Falls,
1982; Stoddard, 1996). Identifying and testing for important contextual factors appears to be an essential
step in decoding the meaning of sequences.

Qualitatively, we infer meaning in a sequence if it modifies the receiver’s response in some
predictable way. Quantitatively, information theory measures the amount of information (usually in units
of bits) transmitted and received within a communication system (Shannon et al., 1949). Therefore,
information theory approaches can describe the complexity of the communication system. Information
theory additionally can characterise transmission errors and reception errors, and has been
comprehensively reviewed in the context of animal communication in (Bradbury & Vehrencamp, 2011).

Considerable debate exists over the nature of animal communication and the terminology used in
animal communication research (Owren, Rendall & Ryan, 2010; Seyfarth et al., 2010; Ruxton &
Schaefer, 2011; Stegmann, 2013), and in particular the origin of and relationship between meaning and
information, and their evolutionary significance. For our purposes, we will use the term “meaning” when
discussing behavioural and evolutionary processes, and the term “information” when discussing the
mathematical and statistical properties of sequences. This parallels (but is distinct from) the definitions
given by Ruxton & Schaefer (2011), in particular because we wish to have a single term (“information”)
that describes inherent properties of sequences, without reference to the putative behavioural effects on
receivers, or the ultimate evolutionary processes that caused the sequence to take the form that it does. In
the remainder of this section on information and meaning, we address the question of how information
can be embedded into signal sequences.
(1) Information embedding paradigms

A “sequence” can be defined as an ordered list of units. Animals produce sequences of sounds through a wide range of mechanisms (e.g., vocalisation, stridulation, percussion), and different uses of the sound-producing apparatus can produce different sound “units” with distinct and distinguishable properties. The resulting order of these varied sound units may or may not contain information that can be interpreted by a receiver, irrespective of whether or not the signaller intended to convey meaning. Given that a sequence must consist of more than one “unit” of one or more different types, the delineation and definition of the unit types is clearly of vital importance. We discuss this question at length in Section III. However, assuming that units have been successfully assigned short-hand labels (e.g., A, B, C, etc.), what different methods can be used to arrange these units in a sequence, in such a way that the sequence can contain information?

Although it seems intuitively obvious that a sequence of such labels may contain information, this intuition arises from our own natural human dispensation to language and writing, and may not be particularly useful in identifying information in animal sequences. We appreciate birdsong, for instance, as a complex combination of notes, and may be tempted to compare this animal vocalisation to human music (Baptista & Keister, 2005; Araya-Salas, 2012; Rothenberg et al., 2013). An anthropocentric approach, however, is not likely in all cases to identify structure relevant to animal communication. Furthermore, wide variation can be expected between the structure of sequences generated by different taxa, from the pulse-based stridulation of insects (Gerhardt & Huber, 2002) to song in whales (reviewed in Cholewiak, Sousa-Lima & Cerchio, 2012), and a single analytical paradigm derived from a narrow taxonomic view is also likely to be inadequate. A more rigorous analysis is needed, one that indicates the fundamental structural properties of acoustic sequences, in all their diversity. Looking for information only, say, in the order of units can lead researchers to miss information encoded in unit timing, or pulse rate.
We suggest a classification scheme based on six distinct paradigms for encoding information in sequences (Figure 2): (a) Repetition, where a single unit is repeated more than once; (b) Diversity, where information is represented by the number of distinct units present; (c) Combination, where sets of units have different information from each unit individually; (d) Ordering, where the relative position of units to each other is important; (e) Overlapping, where information is conveyed in the relationship between sequences of two or more individuals; and (f) Timing, where the time gap between units conveys information. This framework can form the basis of much research into sequences, and provides a useful and comprehensive approach for classifying information-bearing sequences. We recommend that in any research into animal acoustic communication with a sequential component, researchers first identify the place(s) of their focal system in this framework, and use this structure to guide the formulation of useful, testable hypotheses. Considering the formal structures of possible information embedding systems may provide supportive insights into the cognitive and evolutionary processes taking place (Chatterjee, 2005; Seyfarth, Cheney & Bergman, 2005). Of course, any particular system might have properties of more than one of the six paradigms in this framework, and the boundaries between them may not always be clearly distinguished. Sperm whale *Physeter macrocephalus* coda exchanges (Watkins & Schevill, 1977) provide an example of this. A coda is a sequence of clicks (Repetition of the acoustic unit) where the Timing between echolocation clicks moderates response. In duet behaviour, Overlap also exists, with one animal producing and another responding with another coda (Schulz et al., 2008). Each of these paradigms is now described in more detail below.

(2) **Six information embedding paradigms**

1. **Repetition:** Sequences are made of repetitions of discrete units, and repetitions of the same unit affect receiver responses. For instance, the information contained in a unit A given in isolation may convey a different meaning to a receiver than an iterated sequence of unit A (e.g., AAAA, etc.). For example, greater numbers of D notes in the *chick-a-dee* calls of chickadee species *Poecile* spp. can be related to the immediacy of threat posed by a detected predator (Krams et al., 2012). Repetition in alarm
calls are related to situation urgency (meerkats *Suricata suricatta*: Manser, 2001, marmots *Marmota spp.*: Blumstein, 2007, colobus monkeys *Colobus spp.*: Schel, Candiotti & Zuberbühler, 2010, Campbell’s monkeys *Cercopithecus campbelli*: Lemasson et al., 2010, lemurs *Lemur catta* and *Varecia variegata*: Macedonia, 1990). As an additional example, many frog species produce pulsatile advertisement calls consisting of the same repeated element. If it is the case that both number of pulses and pulse rate affect receiver responses, as shown in some hylid treefrogs (Gerhardt, 2001), then information is being embedded using both the *Repetition* (1) and the *Timing* (6) paradigms simultaneously. Such use of multiple embedding techniques may be quite common, for instance in intrasexual competitive and intersexual reproductive contexts (Gerhardt & Huber, 2002).

2. Diversity: Sequences of different units (e.g., A, B, C) are produced, but those units are functionally interchangeable, and therefore ordering is unimportant. For instance, many songbirds produce songs with multiple different syllables. In many species, however, the particular syllables are substitutable (e.g., Eens, Pinxten & Verheyen, 1991; Farabaugh & Dooling, 1996, but see Lipkind et al., 2013) and receivers attend to the overall diversity of sounds in the songs or repertoires of signallers (Catchpole & Slater, 2003). Large acoustic repertoires have been proposed to be sexually selected in species such as great reed warblers *Acrocephalus arundinaceus* and common starlings *Sturnus vulgaris* (Eens, Pinxten & Verheyen, 1993; Hasselquist, Bensch & von Schantz, 1996; Eens, 1997), in which case diversity embeds information (that carries meaning) on signaller quality (e.g., Kipper et al., 2006). Acoustic "diversity" has additionally been proposed as a means of preventing habituation on the part of the receiver (Hartshorne, 1956; Hartshorne, 1973; Kroodsma, 1990) as well as a means of avoiding (neuromuscular) "exhaustion" on the part of the sender (Lambruchs & Dhondt, 1987; Lambruchs & Dhondt, 1988). We do note that these explanations remain somewhat controversial, especially if the transitions between acoustic units are, indeed, biologically constrained (Weary & Lemon, 1988; Weary et al., 1988; Weary & Lemon, 1990; Weary, Lambruchs & Krebs, 1991; Riebel & Slater, 2003; Brumm & Slater, 2006).
3. Combination: Sequences may consist of different discrete acoustic units (e.g., A, B, C) each of which is itself meaningful, and the combining of the different units conveys distinct information. Here, order does not matter (in contrast to the Ordering paradigm below) – the sequence of unit A followed by unit B has the same information as the sequence of unit B followed by unit A. For example, titi monkeys *Callicebus nigrifrons* (Cäsar et al., 2013) use semantic alarm combinations, in which interspersing avian predator alarms calls (A-type) with terrestrial predator alarm calls (B-type) indicates the presence of a raptor on the ground. In this case, the number of calls (i.e. Repetition) also appears to influence the information present in each call sequence (Cäsar et al., 2013).

4. Ordering: Sequences of different discrete acoustic units (e.g., A, B, C) each of which is itself meaningful and the specific order of which is meaningful. Here, order matters – and the ordered combination of discrete units may result in emergent responses. For instance, A followed by B may elicit a different response than either A or B alone, or B followed by A. Examples include primate alarm calls which, when combined, elicit different responses related to the context of the predatory threat (Arnold & Zuberbühler, 2006a; Arnold & Zuberbühler, 2008). Human languages are a sophisticated example of ordered information encoding (Hauser, Chomsky & Fitch, 2002).

5. Overlapping: Sequences are combined from two or more individuals into exchanges for which the order of these overlapping sequences has information distinct from each signaller’s signals in isolation. Overlapping can be in the time dimension (i.e., two signals emitted at the same time) or in acoustic space (e.g., song type matching, Krebs, Ashcroft & Orsdol, 1981, and frequency matching, Mennill & Ratcliffe, 2004) For example, in different parid species (Paridae: chickadees, tits, and titmice), females seem to attend to the degree to which their males’ songs are overlapped (in time) by neighbouring males’ songs, and seek extra-pair copulations when their mate is overlapped (Otter et al., 1999; Mennill, Ratcliffe & Boag, 2002). Overlapping is also used for social bonding, spatial perception, and reunion, such as chorus
howls in wolves (Harrington et al., 2003) and sperm whale codas (Schulz et al., 2008). Overlapping as
song type matching (overlapping in acoustic space) is also an aggressive signal in some songbirds (Akçay
et al., 2013), though this may depend on whether it is the sequence or the individual unit that is
overlapped (Searcy & Beecher, 2011). Coordination between the calling of individuals can also give
identity cues (Carter et al., 2008).

6. Timing: The temporal spacing between units in a sequence can contain information. In the
simplest case, pulse rate and interpulse interval can distinguish between different species, for example in
insects and anurans (Gerhardt & Huber, 2002; Nityananda & Bee, 2011), rodents (Randall, 1997), and
primates (Hauser, Agnetta & Perez, 1998). Call timing can indicate fitness and aggressive intent, e.g.,
male howler monkeys *Alouatta pigra* attend to howling delay as an indicator of aggressive escalation
(Kitchen, 2004). Additionally, when sequences are produced by different individuals, a receiver may
interpret the timing differences between the producing individuals to obtain contextual information. For
instance, ground squirrels *Spermophilus richarsonii* use the spatial pattern and temporal sequence of
conspecific alarm calls to provide information on a predator’s movement trajectory (Thompson & Hare,
2010). This information only emerges from the sequence of different callers initiating calls (Blumstein,
Verneyre & Daniel, 2004). Such risk tracking could also emerge from animals responding to sequences of
heterospecific alarm signals produced over time.

We conclude this section with a discussion of two examples of how sequences of acoustic signals
produced by signallers can influence meaning to receivers. These two examples come from primates and
exemplify the *Diversity* and *Ordering* types of sequences illustrated in Figure 2. The example of the
*Diversity* type is the system of serial calls of titi monkeys, *Callicebus moloch*, used in a wide range of
social interactions. Here, the calls comprise several distinct units, many of which are produced in
sequences. Importantly, the units of this call system seem to have meaning primarily in the context of the
sequence – this call system therefore seems to represent the notion of phonological syntax (Marler, 1977).
One sequence has been tested via playback studies – the ‘honks-bellows-pumps’ sequence is used frequently by males that are isolated from and not closely associated with females and may recruit non-paired females (Robinson, 1979). Robinson (1979) played back typical sequences of honks-bellows-pumps sequences and atypical sequences of honks-pumps-bellows and found little evidence that groups of titi monkeys responded differently to the two playbacks (though they gave one call type – a ‘moan’, produced often during disturbances caused by other conspecific or heterospecific monkey groups – more often to the atypical sequences). Unfortunately, the playbacks were not done to groups at a distance to determine whether sequence order mattered to the question of recruitment of females.

The second example relates to the **Ordering** type of sequence (Figure 2), and stems from two common calls of putty-nosed monkeys, *Cercopithecus nictitans martini*. ‘Pyow’ calls can be produced individually or in strings of pyows, and seem to be used by putty-nosed monkeys frequently when leopards are detected in the environment (Arnold & Zuberbühler, 2006a), and more generally as an attention-getting signal related to recruitment of receivers and low level alarm (Arnold & Zuberbühler, 2013). ‘Hack’ calls can also be produced individually or in strings of hacks, and seem to be used frequently when eagles are detected in the environment, and more generally as a higher-level alarm call (Arnold & Zuberbühler, 2013). Importantly, pyow and hack calls are frequently combined into pyow-hack sequences. Both naturalistic observational data as well as experimental call playback results indicate that pyow-hack sequences influence receiver behaviour differently than do pyow or hack sequences alone – pyow-hack sequences seem to mean “let’s go!”, and produce greater movement distances in receivers (Arnold & Zuberbühler, 2006b). The case of the pyow-hack sequence therefore seems to represent something closer to the notion of lexical syntax – individual units and ordered combinations of those units have distinct meanings from one another (Marler, 1977).

These two examples of primate calls illustrate the simple but important point that sequences matter in acoustic signals – combinations or different linear orderings of units (whether those units have meaning individually or not) can have different meanings to receivers. In the case of titi monkeys, the call
sequences seem to serve the function of female attraction for male signallers, whereas in the case of putty-nosed monkeys, the call sequences serve anti-predatory and group cohesion functions.

We have so far been somewhat cavalier in how we have described the structures of call sequences, using terms like notes, units, and, indeed, calls. In the next section of our review, we describe in depth the notion of signalling ‘units’ in the acoustic modality.

III. ACOUSTIC UNITS

Sequences are made of constituent units. Thus the accurate analysis of potential information in animal acoustic sequences depends on appropriately characterising their constituent acoustic units. We recognise, however, that there is no single definition of a unit. Indeed definitions of units, how they are identified, and the semantic labels we assign them vary widely across researchers working with different taxonomic groups (Gerhardt & Huber, 2002) or even within taxonomic groups, as illustrated by the enormous number of names for different units in the songs of songbird species. Our purpose in this section is to discuss issues surrounding the various ways the acoustic units composing a sequence may be characterised.

Units may be identified based on either production mechanisms, which focus on how the sounds are generated by signallers, or by perceptual mechanisms, which focus on how the sounds are interpreted by receivers. How we define a unit will therefore be different if the biological question pertains to production mechanisms or perceptual mechanisms. For example, in birdsong even a fairly simple note may be the result of two physical production pathways, each made on a different side of the syrinx (Catchpole & Slater, 2003). In practice, however, the details of acoustic production and perception are often hidden from the researcher, and so the definition of acoustic units is often carried out on the basis of observed acoustic properties (see Catchpole & Slater, 2003). It is not always clear to what extent these observed acoustic properties accurately represent the production/perceptual constraints on
communication, and the communicative role of the sequence. Identifying units is made all the more
challenging because acoustic units produced by animals often exhibit graded variation in their features
(e.g., absolute frequency, duration, rhythm or tempo, or frequency modulation), but most analytical
methods for unit classification assume that units can be divided into discrete, distinct categories (e.g.,
Clark, Marler & Beeman, 1987). Thus, how we identify units may differ depending on whether the
biological question pertains to production mechanisms, perceptual mechanisms, or acoustical analyses of
information content in the sequences. If the unit classification scheme must reflect animal sound
production or perception, care must be taken to base unit identification on the appropriate features of a
signal, and features that are biologically relevant (e.g., Clemins & Johnson, 2006). In cases where
sequences carry meaning, it is likely that they can be correlated with observational behaviours (possibly
context-dependent) observed over a large number of trials. There is still no guarantee that the sequence
assigned by the researcher is representative of the animal’s perception of the same sequence. To some
degree, this can be tested with playback trials where the signals are manipulated with respect to the
hypothesised unit sequence (Kroodsma, 1989; Fischer, Noser & Hammerschmidt, 2013).

Whatever technique for identifying potential acoustic units is used, we emphasise here that there
are four acoustic properties that are commonly used to delineate potential units (Figure 3). First, the
spectrogram may show a silent gap between two acoustic elements (Figure 3a). When classifying units
“by eye”, separating units by silent gaps is probably the most commonly used criterion. Second,
examination of a spectrogram may show that an acoustic signal changes its properties at a certain time,
without the presence of a silent “gap” (Figure 3b). For example, a pure tone may become harmonic or
noisy, as the result of the animal altering its articulators (e.g., lips), without ceasing sound production in
the source (e.g., larynx). Third, a series of similar sounds may be grouped together as a single unit,
regardless of silent gaps between them, and separated from dissimilar units (Figure 3c). This is
characteristic of pulse trains and “trills”. Finally, there may be a complex hierarchical structure to the
sequence, in which combinations of sounds, which might otherwise be considered fundamental units,
always appear together, giving the impression of a coherent, larger unit of communication (Figure 3d). A
consideration of these four properties together can provide valuable insights into defining units of production, units of perception, and units for sequence analyses.

In Table 1, we give examples of the wide range of studies that have used these different criteria for dividing acoustic sequences into units. Although not intended to be comprehensive, the table shows how all of the four criteria listed above have been used for multiple species and with multiple aims – whether simply characterising the vocalisations, defining units of production/perception, or identifying the functional purpose of the sequences.

(1) Identifying potential units

Before we discuss in more detail how acoustic units may be identified in terms of production, perception, and analysis methods, we point out here that practically all such efforts require scientists to identify potential units at some early stage of their planned investigation or analysis. Two practical considerations are noteworthy.

First, a potential unit can be considered that part of a sequence that can be replaced with a label for analysis purposes (e.g., unit A or unit B), without adversely affecting the results of a planned investigation or analysis. Because animal acoustic sequences are sometimes hierarchical in nature (e.g., humpback whale Megaptera novaengliae song, reviewed in Cholewiak, Sousa-Lima & Cerchio, 2012), distinct sequences of units may themselves be organised into longer, distinctive sequences (i.e., “sequences of sequences”, Berwick et al., 2011). Thus, an important consideration in identifying potential acoustic units for sequence analyses is that they can be hierarchically nested, such that a sequence of units can itself be considered as a unit and replaced with a label.

Second, potential acoustic units are almost always identified based on acoustic features present in a spectrographic representation of the acoustic waveform. Associating combinations of these features with a potential unit can be performed either manually (i.e., examining the spectrograms “by eye”), or automatically by using algorithms for either supervised classification (where sounds are placed in categories according to pre-defined exemplars) or unsupervised clustering (where labelling units is
performed without prior knowledge of the types of units that occur). We return to these analytical methods in a subsequent section, and elaborate here on spectrographic representations.

Spectrograms consisting of discrete Fourier transforms of short, frequently overlapped, segments of the signal are ubiquitous and characterise well those acoustic features related to spectral profile and frequency modulation, many of which are relevant in animal acoustic communication. Examples of such features include minimum and maximum fundamental frequency, slope of the fundamental frequency, number of inflection points, and the presence of harmonics (Oswald et al., 2007) that vary, for example, between individuals (Buck & Tyack, 1993; Blumstein & Munos, 2005; Koren & Geffen, 2011; Ji et al., 2013; Kershenbaum, Sayigh & Janik, 2013; Root-Gutteridge et al., 2014), and in different environmental and behavioural contexts (Matthews et al., 1999; Taylor, Reby & McComb, 2008; Henderson, Hildebrand & Smith, 2011).

Other less used analytical techniques, such as cepstral analysis, may provide additional detail on the nature of acoustic units, and are worth considering for additional analytical depth. Cepstra are the Fourier (or inverse Fourier) transform of the log of the power spectrum (Oppenheim & Schafer, 2004), and can be thought of as producing a spectrum of the power spectrum. Discarding coefficients can yield a compact representation of the spectrum (Figure 4). Further, while Fourier transforms have uniform temporal and frequency resolution, other techniques vary this resolution by using different basis sets, and this provides improved frequency resolution at low frequencies and better temporal resolution at higher frequencies. Examples of these other techniques include multi-taper spectra (Thomson, 1982; Tchemichovski et al., 2000; Baker & Logue, 2003), Wigner-Ville spectra (Martin & Flandrin, 1985; Cohn, 1995), and wavelet analysis (Mallat, 1999). While spectrograms and cepstra are useful for examining frequency-related features of signals, they are less useful when analysing temporal patterns of amplitude modulation. This is an important issue worth bearing in mind, because amplitude modulations are probably critical in signal perception by many animals (Henry et al., 2011), including speech perception by humans (Remez et al., 1994).
(2) Identifying production units

One important approach to identifying acoustic units stems from considering the mechanisms for sound production. In stridulating insects, for example, relatively simple, repeated sounds are typically generated by musculature action that causes hard physical structures to be engaged, such as the file and scraper located on the wings of crickets or the tymbal organs of cicadas (Gerhardt & Huber, 2002). The resulting units, variously termed “chirps,” or, “pulses,” can be organised into longer temporal sequences often termed “trills” or “echemes” (Ragge & Reynolds, 1988). Frogs can produce sounds with temporally structured units in a variety of ways (Martin & Gans, 1972; Martin, 1972; Gerhardt & Huber, 2002). In some species, a single acoustic unit (sometimes called a “pulse,” “note,” or a “call”) is produced by a single contraction of the trunk and laryngeal musculature that induces vibrations in the vocal folds (e.g., Girgenrath & Marsh, 1997). In other instances, frogs can generate short sequences of distinct sound units (also often called “pulses”) produced by the passive expulsion of air forced through the larynx that induces vibrations in structures called arytenoid cartilages, which impose temporal structure on sound (Martin & Gans, 1972; Martin, 1972). Many frogs organise these units into trills (e.g., Gerhardt, 2001), while other species combine acoustically distinct units (e.g., Narins, Lewis & McClelland, 2000; Larson, 2004). In songbirds, coordinated control of the two sides of the syrinx can be used to produce different units of sound, or “notes” (Suthers, 2004). These units can be organised into longer sequences, of “notes,” “trills,” “syllables,” “phrases,” “motifs,” and “songs” (Catchpole & Slater, 2003). In most mammals, sounds are produced as an air source (pressure squeezed from the lungs) causes vibrations in the vocal membranes, which are then filtered by a vocal tract (Peterson & Barney, 1952; Titze, 1994). When resonances occur in the vocal tract, certain frequencies known as formants are reinforced. Formants and formant transitions have been strongly implicated in human perception of vowels and voiced consonants, and may also be used by other species to perceive information (Raemaekers, Raemaekers & Haimoff, 1984; Fitch, 2000).

As the variety in these examples illustrates, there is incredible diversity in the mechanisms animals use to produce the acoustic units that are subsequently organised into sequences. Moreover, there
are additional mechanisms that constrain the production of some of the units. For example, in zebra finches *Taeniopygia guttata*, songs can be interrupted between some of its constitutive units but not others (Cynx, 1990). This suggests that at a neuronal level, certain units share a common, integrated neural production mechanism. Such examples indicate that identifying units based on metrics of audition or visual inspection of spectrograms (e.g., based on silent gaps) may not always be justified, and that there may be essential utility that emerges from a fundamental understanding of unit production. Thus, a key consideration in identifying functional units of production is that doing so may often require knowledge about production mechanisms that can only come about through rigorous experimental studies.

(3) Identifying perceptual units

While there may be fundamental insights gained from identifying units based on a detailed understanding of sound production, there may not always be a one-to-one mapping of the units of production or the units identified in acoustics analyses, onto units of perception (e.g., Blumstein, 1995). Three key considerations should be borne in mind when thinking about units of perception and the analysis of animal acoustic sequences (Figure 5).

First, it is possible that units of production or the units a scientist might identify on a spectrogram are perceptually bound together by receivers into a single unit of perception (Figure 5a). In this sense, a unit of perception is considered a perceptual auditory object in terms familiar to cognitive psychologists and auditory scientists. There are compelling reasons for researchers to consider vocalisations and other sounds as auditory objects (Miller & Cohen, 2010). While the rules governing auditory object formation in humans have been well studied (Griffiths & Warren, 2004; Bizley & Cohen, 2013), the question of precisely how, and to what extent, non-humans group acoustic information into coherent perceptual representations remains a largely open empirical question (Hulse, 2002; Bee & Micheyl, 2008; Miller & Bee, 2012).
Second, studies of categorical perception in humans and other animals (Harnad, 1990) show that continuous variation can nevertheless be perceived as forming discrete categories. In the context of units of perception, this means that the graded variation often seen in spectrograms may nevertheless be perceived categorically by receivers (Figure 5b). Thus, in instances where there are few discrete differences in production mechanisms or in spectrograms, receivers might still perceive distinct units (Nelson & Marler, 1989; Baugh, Akre & Ryan, 2008).

Third, well-known perceptual constraints related to the limits of spectrotemporal resolution may identify units of perception in ways that differ from analytical units and the units of production (Figure 5c). For example, due to temporal integration by the auditory system (Recanzone & Sutter, 2008), some short units of production might be produced so rapidly that they are not perceived as separate units. Instead, they might be integrated into a single percept having a pitch proportional to the repetition rate. For example, in both bottlenose dolphins *Tursiops truncatus* and Atlantic spotted dolphins *Stenella frontalis*, the “squawking” sound that humans perceive as having some tonal qualities is actually a set of rapid echolocation clicks known as a burst pulse (Herzing, 1996). The perceived pitch is related to the repetition rate, the faster the repetition, the higher the pitch. Given the perceptual limits of gap detection (Recanzone & Sutter, 2008), some silent gaps between units of production may be too short to be perceived by the receiver. Clearly, while it may sometimes be desirable or convenient to use “silence” as a way to create analysis boundaries between units, a receiver may not always perceive the silent gaps that we see in our spectrograms. Likewise, some transitions in frequency may reflect units of production that are not perceived because the changes remain unresolved by auditory filters (Moore & Moore, 2003; Recanzone & Sutter, 2008). Indeed, some species may be forced to trade off temporal and spectral resolution to optimise signalling efficiency in different environmental conditions. Frequency modulated signals are more reliable than amplitude modulation in reverberant habitats, such as forests, so woodland birds are adapted to greater frequency resolution and poorer temporal resolution, while the reverse is true of grassland species (Henry & Lucas, 2010; Henry et al., 2011).
The question of what constitutes a unit that is perceptually meaningful to the animal demands rigorous experimental approaches that put this question to the animal itself. There simply is no convenient shortcut to identifying perceptual units. Experimental approaches ranging from operant conditioning (e.g., Dooling et al., 1987; Brown, Dooling & O'Grady, 1988; Dent et al., 1997; Tu, Smith & Dooling, 2011; Ohms et al., 2012; Tu & Dooling, 2012), to field playback experiments, often involving the habituation-discrimination paradigm (e.g., Nelson & Marler, 1989; Wyttenbach, May & Hoy, 1996; Evans, 1997; Searcy, Nowicki & Peters, 1999; Ghazanfar et al., 2001; Weiss & Hauser, 2002) have the potential to identify the boundaries of perceptual units. Playbacks additionally can determine whether units can be discriminated (as in ‘go no-go’ tasks stemming from operant conditioning), or whether they can be recognised and are functionally meaningful to receivers.

Obviously some animals and systems are more tractable than others when it comes to experimentally assessing units of perception, but those not easy to manipulate experimentally (e.g., baleen whales, Balaenopteridae) should not necessarily be excluded from communication sequence research, although the inevitable constraints must be recognised.

(4) Identifying analytical units

In many instances, it is desirable to analyse sequences of identified units in acoustic recordings without having a priori knowledge about how those units may be produced or perceived by the animals themselves. Such analyses are often a fundamental first step toward investigating the potential meaning of acoustic sequences. Before turning to our main discussion of algorithms for analysing sequences of identified units, we briefly discuss methods by which scientists can identify and validate units for sequence analyses from acoustic recordings.

Sounds are typically assigned classifications to units based on the consistency of acoustic characteristics. When feasible, external validation of categories (i.e., comparing animal behavioural
responses to playback experiments) should be performed. Even without directly testing hypotheses of biological significance by playback experiment, there may be other indicators of the validity of a classification scheme based purely on acoustic similarity. For example, naive observers correctly divide dolphin signature whistles into groups corresponding closely to the individuals that produced them (Sayigh et al., 2007), and similar (but poorer) results are achieved using quantitative measures of spectrogram features (Kershenbaum, Sayigh & Janik, 2013).

When classifying units on the basis of their acoustic properties, errors can occur both as the result of perceptual bias, and as the result of poor repeatability. Perceptual bias occurs either when the characteristics of the sound that are used to make the unit assignment are inappropriate for the communication system being studied, or when the classification scheme relies too heavily on those acoustic features that appear important to human observers. For example, analysing spectrograms with a 50 Hz spectral resolution would be appropriate for human speech, but not for Asian elephants (*Elephas maximus*), which produce infrasonic calls that are typically between 14-24 Hz (Payne, Langbauer Jr & Thomas, 1986), as details of the elephant calls would be unobservable. Features that appear important to human observers may include tonal modulation shapes, often posed in terms of geometric descriptors, such as "upsweep", "concave", and "sine" (e.g., Bazúa-Durán & Au, 2002), which are prominent to the human eye, but may or may not be of biological relevance. Poor repeatability, or variance, can occur both in human classification, as inter-observer variability, and in machine learning, where computer classification algorithms can make markedly different decisions after training with data that are very similar (overtraining). Poor repeatability can be a particular problem when the classification scheme ignores, or fails to give sufficient weight to, the features that are of biological significance, or the algorithm (human or machine) places too much emphasis on particular classification cues that are specific to the examples used to learn the categories. Repeatability suffers particularly when analysing signals in the presence of noise, which can mask fine acoustic details (Kershenbaum & Roch, 2013).
Two approaches have been used to classify units by their acoustic properties: visual inspection of spectrograms, and application of automatic algorithms that assign classifications based on mathematical rules.

(a) Visual classification, “by eye”

Traditionally, units are “hand-scored” by humans searching for consistent patterns in spectrograms (or even listening to sound recordings without the aid of a spectrogram). Visual classification has been an effective technique that has led to many important advances in the study both of birdsong (e.g., Kroodsma, 1985; Podos et al., 1992, and reviewed in Catchpole & Slater, 2003), and acoustic sequences in other taxa (e.g., Narins, Lewis & McClelland, 2000; Larson, 2004). Humans are usually considered to be good at visual pattern recognition – and better than most computer algorithms (Ripley, 2007; Duda, Hart & Stork, 2012), which makes visual classification an attractive approach to identifying acoustic units. However, drawbacks to visual classification exist (Clark, Marler & Beeman, 1987). Visual classification is time consuming and prevents taking full advantage of large acoustic data sets generated by automated recorders. Similarly, the difficulty in scoring large data sets means that sample sizes used in research may be too small to draw firm conclusions (Kershenbaum, 2013). Furthermore, visual classification can be prone to subjective errors (Jones, ten Cate & Bijleveld, 2001), and inter-observer reliability should be used (and reported) as a measure of the robustness of the visual assessments (Burghardt et al., 2012).

(b) Automatic classification

As an alternative to visual classification, automated methods may extract specific metrics, or features, from the acoustic data for input to classification algorithms. Although the acoustic features used by automated systems also may not reflect the cues used by the focal species, automated systems have the advantage of being able to handle large data sets. In principle, automatic classification is attractive as it is...
not susceptible to the inter-observer variability of visual classification (Tchernichovski et al., 2000). However, current implementations generally fall short of the performance desired (Janik, 1999), for instance by failing to recognise subtle features that can be detected both by humans, and by the focal animals, and visual classification has been shown to out-perform automated systems in cases where the meaning of acoustic signals is known \textit{a priori} (e.g., Sayigh et al., 2007; Kershenbaum, Sayigh & Janik, 2013). However, once an automatic algorithm is defined, large datasets can be analysed.

A third possibility is to use a hybrid system. Automated techniques can be used to find regions of possible calls that are then verified and corrected by a human analyst (Helble et al., 2012). Machine assistance can allow analysts to process much larger data sets than before, but at the risk of possibly missing calls that they might have been able to detect.

Classification algorithms can accept acoustic data with varying degrees of pre-processing as inputs. For example, in addition to the commonly used spectrograms (Picone, 1993), cepstra (Oppenheim & Schafer, 2004), multi-taper spectra (Thomson, 1982), wavelets (Mallat, 1999), and formants (Fitch, 1997) may be used, as they provide additional information on the acoustic characteristics of units, which may not be well represented by traditional spectrograms (Tchernichovski et al., 2000). Each of these methods provide analysis of the spectral content of a short segment of the acoustic production, and algorithms frequently examine how these parameters are distributed or change over time (e.g., Kogan & Margoliash, 1998).

Units may be classified automatically using supervised algorithms, in which the algorithm is taught to recognise unit types given some \textit{a priori} known exemplars, or clustered using unsupervised algorithms, in which no \textit{a priori} unit type assignment is known (Duda, Hart & Stork, 2012). In both cases, the biological relevance of units must be verified independently because mis-specification of units can obscure sequential patterns. Environmental noise or sounds from other species may be mistakenly classified as an acoustic unit, and genuine units may be incorrectly assigned to unit categories. When
using supervised algorithms, perceptual bias may lead to misinterpreting data when the critical bands, 
temporal resolution, and hearing capabilities of a species are not taken into account, i.e., the exemplars 
themselves may be subject to similar subjective errors that can occur in visual classification. However, 
validation of unsupervised clustering into units is also problematic, where clustering results cannot be 
assessed against known unit categories. The interplay between unit identification and sequence model 
validation is a non-trivial problem (e.g., Jin & Kozhevnikov, 2011). Similarly, estimating uncertainty in 
unit classification and assessing how that uncertainty affects conclusions from a sequence analysis is a 
key part of model assessment (Duda, Hart & Stork, 2012)

When using supervised classification, one appropriate technique for measuring classification 
uncertainty is cross-validation (Arlot & Celisse, 2010). For fully unsupervised clustering algorithms, 
where the desired classification is unknown, techniques exist to quantify the stability of the clustering 
result, as an indicator of clustering quality. Examples include “leave-k-out” (Manning, Raghavan & 
Schütze, 2008), a generalisation of the “leave-one-out” cross-validation, and techniques based on 
normalised mutual information (Zhong & Ghosh, 2005), which measure the similarity between two 
clustering schemes (Fred & Jain, 2005). However, it must be clear that cluster stability (and 
correspondingly, inter-observer reliability) is not evidence that the classification is appropriate (i.e., 
matches the true, unknown, biologically relevant categorisation), or will remain stable upon addition of 
new data (Ben-David, Von Luxburg & Pál, 2006). Other information theoretic tests provide an alternative 
assessment of the validity of unsupervised clustering results, such as checking if units follow Zipf's law of 
abbreviation, which is predicted by a universal principle of compression (Zipf, 1949; Ferrer-i-Cancho et 
al., 2013) or Zipf's law for word frequencies, which is predicted by a compromise between maximizing 
the distinctiveness of units and the cost of producing them (Zipf, 1949; Ferrer-i-Cancho, 2005).
Given that the researcher has successfully determined the units of an acoustic sequence that are appropriate for the hypothesis being tested, one must select and apply appropriate algorithms for analysing the sequence of units. Many algorithms exist for the analysis of sequences: both those produced by animals, and sequences in general (such as DNA, and stock market prices). Selection of an appropriate algorithm can sometimes be guided by the quantity and variability of the data, but there is no clear rule to be followed. In fact, in machine learning, the so-called no free lunch theorem (Wolpert & Macready, 1997) shows that there is no one pattern recognition algorithm that is best for every situation, and any improvement in performance for one class of problems is offset by lower performance in another problem class. In choosing an algorithm for analyses, one should be guided by the variability and quantity of the data for analysis, keeping in mind that models with more parameters require more data to estimate the parameters effectively.

The structure analysis algorithms discussed throughout this section can be used to model the different methods for combining units discussed earlier (Figure 2). Repetition, Diversity, and Ordering are reasonably well captured by models such as Markov chains, hidden Markov models, and grammars. Networks capture structure either with or without order, although much of the application of networks has been done on unordered associations (Combination). Temporal information can be modelled as an attribute of an acoustic unit requiring extensions to the techniques discussed below, or as a separate process.

Here we give a sample of some of the more important and more promising algorithms for animal acoustic sequence analysis, and discuss ways for selecting and evaluating analytical techniques. Selecting appropriate algorithms should involve the following steps. (i) Technique: understand the nature of the models and their mathematical basis. (ii) Suitability: assess the suitability of the models and their constraints with respect to the research questions being asked. (iii) Application: apply the models to the empirical data (training, parameter estimation). (iv) Assessment: extract metrics from the models that
summarise the nature of the sequences analysed. (v) Inference: compare metrics between data sets (or between empirical data and random null-models) to draw ecological, mechanistic, evolutionary, and behavioural inferences. (vi) Validate: determine the goodness of fit of the model to the data and uncertainty of parameter estimates. Bootstrapping techniques can allow validation with sets that were not used in model development.

We consider five models in this section: (1) Markov chains, (2) hidden Markov models, (3) network models, (4) formal grammars, and (5) temporal models.

(1) **Markov chains**

Markov chains, or N-grams models, capture structure in acoustic unit sequences based on the recent history of a finite number of discrete unit types. Thus, the occurrence of a unit (or the probability of occurrence of a unit) is determined by a finite number of previous units. The history length is referred to as the order, and the simplest such model is a 0\(^{th}\) order Markov model, which assumes that each unit is independent of one another, and simply determines the probability of observing any unit with no prior knowledge. A 1\(^{st}\) order Markov model is one in which the probability of each unit occurring is determined only by the preceding unit, together with the “transition probability” from one unit to the next. This transition probability is assumed to be constant (stationary). Higher order Markov models condition the unit probabilities based on more than one preceding units, as determined by the model order. An N-gram model conditions the probability on the N-1 previous units, and is equivalent to an N-1\(^{st}\) order Markov model. A K\(^{th}\) order Markov model of a sequence with C distinct units is defined by a C\(^K\) x C matrix of transition probabilities from each of the C\(^K\) possible preceding sequences, to each of the C possible subsequent units, or equivalently by a state transition diagram (Figure 6).

As the order of the model increases, more and more data are required for the accurate estimation of transition probabilities, i.e., sequences must be longer, and many transitions will have zero counts. This is particularly problematic when looking at new data, which may contain sequences that were not previously encountered, as they will appear to have zero probability. As a result, Markov models with
orders greater than two (trigram, N=3) are rare. In principle, a $K^{th}$ order Markov model requires sufficient data to provide accurate estimates of $C^{K+1}$ transition probabilities.

Closed-form expressions for maximum likelihood estimates of the transition probabilities can be used with conditional counts (Anderson & Goodman, 1957). For example, assuming five acoustic units $(A-E)$, maximum likelihood estimates of the transition probabilities for a first order Markov model (bigram, N=2) can be found directly from the number of occurrences of each transition, e.g.

$$P(B \mid A) = \frac{\text{count}(AB)}{\sum_{i \in \{A,B,C,D,E\}} \text{count}(A,i)}$$

Although not widely used in the animal communication literature, research in human natural language processing has led to the development of methods known as back-off models (Katz, 1987), which account for the underestimated probability of rare sequences using Good-Turing counts (Gale & Sampson, 1995). When a particular state transition is never observed in empirical data, the back-off model offers the minimum probability for this state transition so as not to rule it out automatically during the testing.

Standard freely available tools such as the SRI language modelling toolkit (Stolcke, 2002) implement back-off models and can reduce the effort of adopting these more advanced techniques.

Once Markovian transitions have been calculated and validated, the transition probabilities can be used to calculate a number of summary metrics using information theory (Shannon et al., 1949; Chatfield & Lemon, 1970; Hailman, 2008). For a review on the mathematics underlying information theories, we direct the readers to the overview in McCowan, Hanser & Doyle (1999) or Freeberg & Lucas (2012), which provides the equations as well as a comprehensive reference list to other previous work. Here we will define these quantitative measures with respect to their relevance in analysing of animal acoustic sequences. Zero-order entropy measures repertoire diversity:

$$H_0 = \log_2(C)$$
where $C=|V|$ is the cardinality of the set of acoustic units $V$. First-order entropy begins to measure simple repertoire internal organisational structure by evaluating the relative frequency of use of different signal types in the repertoire:

$$H_1 = \sum_{v_i \in V} -P(v_i) \log_2 P(v_i)$$

where the probability of each acoustic unit is typically estimated based on frequencies of occurrence, as described earlier with N-grams. Higher-order entropies measure internal organisational structure, and thus one form of communication complexity, by examining how signals interact within a repertoire at the two-unit sequence level, the three-unit sequence level, and so forth.

One inferential approach is to calculate the entropic values from first-order and higher-order Markov models to summarise the extent to which sequential structure is present at each order. A random sequence would show no dependence of entropy on Markov order, whereas decreases in entropy as the order is increased would be an indication of sequential organisation, and thus higher communication complexity (Ferrer-i-Cancho & McCowan, 2012). These summary measures can then be further extended to compare the importance of sequential structure across different taxa, social and ecological contexts. These types of comparisons can provide novel insights into the ecological, environmental, social, and contextual properties that shape the structure, organisation, and function of signal repertoires (McCowan, Doyle & Hanser, 2002).

The most common application of the Markov model is to test whether or not units occur independently in a sequence. Model validation techniques include the sequential and $\chi^2$ tests (Anderson & Goodman, 1957). For instance, Narins, Lewis & McClelland (2000) used a permutation test (Adams & Anthony, 1996) to evaluate the hypothesis that a frog with an exceptionally large vocal repertoire, *Bufo madagascariensis*, emitted any call pairs more often than would be expected by chance. Similar techniques were used to show non-random call production by Sayigh et al., (2012) with short-finned pilot whales *Globicephala macrorhynchus*, and by Bohn et al., (2009) with free-tailed bats *Tadarida*.
brasiliensis. However, deviation from statistical independence does not in itself prove a sequence to have been generated by a Markov chain. Other tests, such as N-gram distribution (Jin & Kozhevnikov, 2011) may be more revealing.

(2) **Hidden Markov models**

HMMs are a generalisation of the Markov model. In Markov models, the acoustic unit history (of length \( N \)) can be considered the current “state” of the system. In hidden Markov models (HMMs) (Rabiner, 1989), states are not necessarily associated with acoustic units, but instead represent the state of some possibly unknown and unobservable process. Thus, the system progresses from one state to another, where the nature of each state is unknown to the observer. Each of these states may generate a “signal” (i.e., a unit), but there is not necessarily a one-to-one mapping between state transitions and signals generated. For example, transitioning to state \( X \) might generate unit \( A \), but the same might be true of transitioning to state \( Y \). An observation is generated at each state according to a state-dependent probability density function, and state transitions are governed by a separate probability distribution (Figure 7). HMMs are particularly useful to model very complex systems, while still being computationally tractable.

Extensions to the HMM model also exist, in which the state transition probabilities are non-stationary. For example, the probability of remaining in the same state may decay with time e.g., due to neural depletion, as shown by Jin & Kozhevnikov (2011), or recurrent units may appear more often than expected by a Markov model, particularly where behavioural sequences are non-Markovian (Cane, 1959; Kershenbaum, 2013). Also, HMMs are popular in speech analysis (Rabiner, 1989), where emissions are continuous-valued, rather than discrete.

HMMs have been used fairly extensively in speaker recognition (Lee & Hon, 1989), the identification of acoustic units in birdsong (Trawicki, Johnson & Osiejuk, 2005), and other analyses of bird song sequences. ten Cate, Lachlan & Zuidema (2013) reviewed analytical methods for inferring the structure of birdsong and highlighted the idea that HMM states can be thought of as possibly modelling
an element of an animal’s cognitive state. This makes it possible to build models that have multiple state
distributions for the same acoustic unit sequence. For instance, in the trigram AAC, the probability given
by the 2nd order Markov model, \( P(C|A, A) \) is fixed. There cannot be different distributions for observing
the unit \( C \), if the previous two units are \( A \). Yet cognitive state may have the potential to influence the
probability of observing \( C \), even for identical sequence contexts (AA). Another state variable (\( \theta \)) exists
unobserved, as it reflects cognitive state, rather than sequence history. In this case, \( P(C|A, A, \theta=0) \neq P(C|A, A, \theta=1) \).

Hahnloser, Kozhevnikov & Fee (2002), Katahira et al. (2011) and Jin (2009) have used HMMs to
model the interaction between song and neural substrates in the brain. A more recent example of this can
be seen in the work of Jin & Kozhevnikov (2011), where they used states to model neural units in song
production of the Bengalese finch *Lonchura striata ver. domestica*, restricting each state to the emission
of a single acoustic unit, thus making acoustic units associated with each state deterministic while
retaining the stochastic nature of state transitions.

Because the states of a HMM represent an unobservable process, it is difficult to estimate the
number of states needed to describe the empirical data adequately. Model selection methods and criteria
(for example Akaike and Bayesian information criteria, and others) can be used to estimate model order
(see Hamaker, Ganapathiraju & Picone, 1998, and Zucchini & MacDonald, 2009 for a brief review), so
the number of states is often determined empirically. Increasing the number of states permits the
modelling of more complex underlying sequences (e.g., longer term dependencies), but increases the
amount of data required for proper estimation. The efficiency and accuracy of model fitting depends on
model complexity, so that models with many states, many acoustic units, and perhaps many covariates or
other conditions will take more time and require more data to fit.

During training, HMM parameters are estimated using an optimisation algorithm (Cappé,
Moulines & Rydén, 2005) that finds a combination of hidden states, state transition tables, and state-
dependent distributions that best describe the data. Software libraries for the training of HMMs are
available in many formats (e.g., the Matlab function *hmmtrain*, the R package HMM; R Development
Team, 2012, and the Hidden Markov Model Toolkit; Young & Young, 1994). Similar considerations of
dataset completeness exist to those when generating regular Markov models, most importantly, that long
sequences of data are required.

Although the states of a HMM are sometimes postulated to possess biologically relevant
meaning, the internal states of the HMM represent a hidden process, and do not necessarily refer to
conge concrete behavioural states. Specifically, the training algorithm does not contain an optimisation criterion
that will necessarily associate model states with the functional or ecological states of the animal that a
researcher is interested in observing (e.g., foraging, seeking a mate, etc.). While the functional/ecological
state is likely related to the sequence, each model state may in fact represent a different subsequence of
the data. Therefore, one cannot assume in general that there will be a one-to-one mapping between model
and animal states. Network structures derived from different empirical data are often widely different,
and it can be misleading to make comparisons between HMMs derived from different data sets.
Furthermore, obtaining consistent states requires many examples with respect to the diversity of the
sequence being modelled. An overtrained network will be highly dependent on the data presented to it
and small changes in the training data can result in very different model parameters, making state-based
inference questionable.

(3) Network models

The structure of an acoustic sequence can also be described using a network approach (reviewed in
Newman, 2003 and Baronchelli et al., 2013), as has been done for other behavioural sequences (e.g.,
pollen processing by honeybees; Fewell, 2003). A node in the network represents a type of unit, and a
directional edge connecting two nodes means that one unit comes after the other in the acoustic sequence.
For example, if a bird sings a song in the order: ABCABC; the network representing this song will have
three nodes for A, B, and C, and three edges connecting A to B, B to C, and C to A (Figure 8). The edges
may simply indicate association between units without order (undirected binary network), an ordered
sequence (directed binary network), or a probability of an ordered sequence (directed weighted network),
the latter being equivalent to a Markov chain (Newman, 2009).

The network representation is fundamentally similar to the Markov model, and the basic input for
constructing a binary network is a matrix of unit pairs within the repertoire, which corresponds to the
transition matrix in a Markov model. However, the network representation may be more robust than a
Markov analysis, particularly when a large number of distinct unit types exist, precluding accurate
estimation of transition probabilities (e.g., Sasahara et al., 2012). In this case, binary or simple directed
networks may capture pertinent properties of the sequence, even if transition probabilities are unknown.

One of the attractive features of network analysis is that a large number of quantitative network
measures exist for comparison to other networks (e.g., from different individuals, populations, or species),
or for testing hypotheses. We list a few of the popular algorithms that can be used to infer the structure of
the acoustic sequence using a network approach. We refer the reader to introductory texts to network
analysis for further details (Newman, 2009; Scott & Carrington, 2011).

Degree centrality measures the number of edges directly connected to each node. In a directed
network, each node has an in-degree and an out-degree, corresponding to incoming and outgoing edges.
The weighted version of degree centrality is termed strength centrality, which takes into account the
weights of each edge (Barrat et al., 2004). Degree/strength centrality identifies the central nodes in the
network, corresponding to central elements in the acoustic sequence. For example, in the mockingbird
* Mimus polyglottos*, which imitates sounds of other species, its own song is central in the network,
meaning that it usually separates between other sounds by singing its own song (Gammon & Altizer,
2011).

Betweenness centrality is a measure of the role a central node plays in connecting other nodes.
For example, if an animal usually uses three units before moving to another group of units, a unit that lies
between these groups in the acoustic sequence will have high betweenness centrality. A weighted version
of betweenness centrality was described in Opsahl, Agneessens & Skvoretz (2010).
Clustering coefficient describes how many triads of nodes are closed in the network. For example, if unit A is connected to B, and B is connected to C, a cluster is formed if A is also connected to C.

Directed and weighted versions of the clustering coefficient have been described (Barrat et al., 2004; Fagiolo, 2007).

Mean path length is defined as the average minimum number of connections to be crossed from any arbitrary node to any other. This measures the overall navigability in the network; as this value becomes large, a longer series of steps is required for any node to reach another.

Small-world metric measures the level of connectedness of a network and is the ratio of the clustering coefficient \( C \) to the mean path length \( L \) after normalising each with respect to the clustering coefficient and mean path length of a random network: \( S = \frac{C}{C_{\text{rand}}} \frac{L}{L_{\text{rand}}} \). If \( S > 1 \) the network is regarded as “small-world” (Watts & Strogatz, 1998; Humphries & Gurney, 2008), with the implication that nodes are reasonably well connected and that it does not take a large number of edges to connect most pairs of nodes. Sasahara et al. (2012) demonstrated that the network of California thrasher songs has a small-world structure, in which subsets of phrases are highly grouped and linked with a short mean path length.

Network motifs are recurring structures that serve as building blocks of the network (Milo et al., 2002). For example, a network may feature an overrepresentation of specific types of triads, tetrads, or feed-forward loops. Network motif analysis could be informative in comparing sequence networks from different individuals, populations or species. We refer the reader to three software packages available for motif analysis: FANMOD (Wernicke & Rasche, 2006); MAVisto (Schreiber & Schwöbbermeyer, 2005); and MFinder (Kashtan et al., 2002).

Community detection algorithms offer a method to detect network substructure objectively (Fortunato, 2010). These algorithms identify groups of nodes with dense connections between them but that are sparsely connected to other groups/nodes. Subgroups of nodes in a network can be considered somewhat independent components of it, offering insight into the different subunits of acoustic
sequences. Multi-scale community detection algorithms can be useful for detecting hierarchical sequence structures (Fushing & McAssey, 2010; Chen & Fushing, 2012).

*Exponential family Random Graph Models* (ERGMs) offer a robust analytic approach to evaluate the contribution of multiple factors to the network structure using statistical modelling (Snijders, 2002). These factors may include structural factors (e.g., the tendency to have closed triads in the network), and factors based on node or edge attributes (e.g., a tendency for connections between nodes that are acoustically similar). The goal of ERGMs is to predict the joint probability that a set of edges exist on nodes in a network. The R programming language package *statnet* has tools for model estimation and evaluation, and for model-based network simulation and network visualisation (Handcock et al., 2008).

As with other models, many statistical tests for inference and model assessment require a comparison of the observed network to a set of random networks. For example, the clustering coefficient of an observed network can be compared to those of randomly generated networks, to test if it is significantly smaller or larger than expected. A major concern when constructing random networks is what properties of the observed network should be retained (Croft, James & Krause, 2008). The answer to this question depends on the hypothesis being tested. For example, when testing the significance of the clustering coefficient, it is reasonable to retain the original number of nodes and edges, density and possibly also the degree distribution, such that the observed network is compared to random networks with similar properties.

Several software packages exist that permit the computation of many of the metrics from this section that can be used to make inferences about the network. Examples include UCINet (Borgatti, Everett & Freeman, 2002), Gephi (Bastian, Heymann & Jacomy, 2009), igraph (Csardi & Nepusz, 2006), and Cytoscape (Shannon et al., 2003).

(4) *Formal grammars*

The structure of an acoustic sequence can be described using formal grammars. A grammar consists of a set of rewrite rules (or “productions”), that define the ways in which units can be ordered. Grammar rules
consist of operations performed on “terminals” (in our case, units), which are conventionally denoted with lower case letters, and non-terminals (symbols that must be replaced by terminals before the derivation is complete), conventionally denoted with upper case letters (note that this convention is inconsistent with the upper case convention used for acoustic unit labels). Grammars generate sequences iteratively, by applying rules repeatedly to a growing sequence. For example, the rule “U → a W” means that the nonterminal U can be rewritten with the symbols “a W.” The terminal a is a unit, as we are familiar with, but as W is a non-terminal, and may itself be rewritten by a different rule. For an example, see Figure 9.

Sequences that can be derived by a given grammar are called grammatical with respect to that grammar. The collection of all sequences that could possibly be generated by a grammar is called the language of the grammar. The validation of a grammar consists of verifying that the grammar’s language matches exactly the set of sequences to be modelled. If a species produces sequences that cannot be generated by the grammar, the grammar is deemed “overselective”. A grammar that is “overgeneralising” produces sequences not observed in the empirical data – although it is often unclear whether this represents a true failure of the grammar, or insufficient sampling of observed sequences. In the example given in Figure 9, the grammar is capable of producing the sequence $ABBBBBBBBBBB$, however, since blue whales have not been observed to produce similar sequences in decades of observation, we conclude that this grammar is overgeneralising. It is important to note, however, that formal grammars are deterministic, in contrast to the probabilistic models discussed previously (Markov model, HMM). If one assigned probabilities to each of the rewriting rules, the particular sequence shown above may not have been observed simply because it is very unlikely.

Algorithms known as parsers can be constructed from grammars to estimate the probability that a sequence belongs to the language for which the grammar has been inferred. Inferring a grammar from a collection of sequences is a difficult problem, which, as famously formulated by (Gold, 1967), is intractable for all but a number of restricted cases. Gold’s formulation, however, does not appear to preclude the learning of grammar in real-world examples, and is of questionable direct relevance to the understanding or modelling of the psychology of sequence processing (Johnson, 2004). When restated in
terms that arguably fit better the cognitive tasks faced by humans and other animals, grammar inference becomes possible (Clark, 2010; Clark, Eyraud & Habrard, 2010). Algorithms based on distributional learning, which seek probabilistically motivated phrase structure by recursively aligning and comparing input sequences, are becoming increasingly successful in sequence processing tasks such as modelling language acquisition (Solan et al., 2005; Kolodny, Lotem & Edelman, in press).

A grammar can be classified according to its place in a hierarchy of classes of formal grammars known as the Chomsky hierarchy (Chomsky, 2002) and illustrated in Figure 10. These classes differ in the complexity of languages that can be modelled. The simplest class of grammars are called regular grammars, which are capable of describing the generation of any finite set of sequences or repeating pattern, and are fundamentally similar to Markov models. Figure 9 is an example of a regular grammar. Kakishita et al. (2009) showed that Bengalese finch songs can be modelled by a restricted class of regular grammars, called “k-reversible regular grammars,” which is learnable from only positive samples, i.e., observed and hence permissible sequences, without information on those sequences that are not permissible in the grammar. Context-free grammars are more complex, and are able to retain state information that enable one part of the sequence to affect another; this is usually demonstrated through the ability to create sequences of symbols where each unit is repeated the same number of times $A^nB^n$ where $n$ denotes $n$ repetitions of the terminal unit, e.g., $AAABBB$ ($A^3B^3$). Context sensitive languages allow context dependent rewrite rules that have few restrictions, permitting further reaching dependencies such as in the set of sequences $A^nB^nC^n$. The highest level in the Chomsky hierarchy, recursively enumerable grammars, are more complex still, and rarely have relevance to animal communication studies.

The level of a grammar within the Chomsky hierarchy can give an indication of the complexity of the communication system represented by that grammar. Most animal acoustic sequences are thought to be no more complex than regular grammars (Berwick et al., 2011), whereas complexity greater than the regular grammar is thought to be a unique feature of human language (Hauser, Chomsky & Fitch, 2002). Therefore, indication that any animal communication could not be represented by a regular grammar
would be considered an important discovery, and claims of context-free (but non-regular) sequences in
European starlings *Sturnus vulgaris* (Gentner et al., 2006) have not been widely accepted (Van
Heijningen et al., 2009; Beckers et al., 2012). The deterministic nature of regular grammars – or indeed
any formal grammars – may explain why formal grammars are not sufficiently general to describe the
sequences of many animal species, and formal grammars remain more popular in human linguistic studies
than in animal communication research.

(5)  **Temporal structure**

Information may exist in the relative or absolute timing of acoustic units in a sequence, rather than in the
order of those units. In particular, timing and rhythm information may be of importance, and may be lost
when acoustic sequences are represented as a series of symbols. This section describes two different
approaches to quantifying the temporal structure in acoustic sequences: traditional techniques examining
inter-event interval and pulse statistics (e.g., Randall, 1989; Narins et al., 1992), and recent multi-
timescale rhythm analysis (Saar & Mitra, 2008).

Analyses of temporal structure can be applied to any audio recording, regardless of whether that
recording contains recognisable sequences, individual sounds, or multiple simultaneously vocalising
individuals. Such analyses are most likely to be informative, however, when recurring acoustic patterns
are present, especially if those recurring patterns are rhythmic or produced at a predictable rate.

Variations in interactive sound sequence production during chorusing and cross-individual
synchronisation can be quantified through meter, or prosody analysis, and higher-order sequence structure
can be identified through automated identification of repeating patterns. At the simplest level, it is
possible to analyse the timing of sounds in a sequence, simply by recording when sound energy is above a
fixed threshold. For instance, temporal patterns can be extracted automatically from simpler acoustic
sequences by transforming recordings into sequences of numerical measures of the durations and silent
intervals between sounds (Isaac & Marler, 1963; Catchpole, 1976; Mercado, Herman & Pack, 2003;
Handel, Todd & Zoidis, 2009; Green et al., 2011), song bouts (Eens, Pinxten & Verheyen, 1989; Saar &
Mitra, 2008), or of acoustic energy within successive intervals (Murray, Mercado & Roitblat, 1998; Mercado et al., 2010). Before the invention of the Kay sonograph, which led to the routine analysis of audio spectrograms, temporal dynamics of bird song were often transcribed using musical notation (Saunders, 1951; Nowicki & Marler, 1988).

Inter-pulse interval has been widely used to quantify temporal structure in animal acoustic sequences, for example in kangaroo rats *Dipodomys spectabilis* (Randall, 1989), fruit flies *Drosophila melanogaster* (Bennet-Clark & Ewing, 1969), and rhesus monkeys *Macaca mulatta* (Hauser, Agnetta & Perez, 1998). Variations in pulse intervals can encode individual information such as identity and fitness (Bennet-Clark & Ewing, 1969; Randall, 1989), as well species identity (Randall, 1997; Hauser, Agnetta & Perez, 1998). In these examples, comparing the median inter-pulse interval between two sample populations is often sufficient to uncover significant differences.

More recently developed techniques for analysis of temporal structure require more detailed processing. For example, periodic regularities and repetitions of patterns within recordings of musical performances can be automatically detected and characterised (Paulus, Müller & Klapuri, 2010; Weiss & Bello, 2011). The first step in modern approaches to analysing the temporal structure of sound sequences involves segmenting the recording. The duration and distribution of individual segments can be fixed (e.g., splitting a recording into 100 ms chunks/frames) or variable (e.g., using multiple frame sizes in parallel or adjusting the frame size based on the rate and duration of acoustic events). The acoustic features of individual frames can then be analysed using the same signal processing methods that are applied when measuring the acoustic features of individual sounds, thereby transforming the smaller waveform into a vector of elements that describe features of the segment. Sequences of such frame-describing vectors then would typically be used to form a matrix representing the entire recording. In this matrix, the sequence of columns (or rows) corresponds to the temporal order of individual frames extracted from the recording.

Regularities within the feature matrix generated from frame-describing vectors reflect temporal regularities within the original recording. Thus, the problem of describing and detecting temporal patterns
within a recording is transformed into the more computationally tractable problem of detecting and
identifying structure within a matrix of numbers (as opposed to a sequence of symbols). If each frame is
described by a single number (e.g., mean amplitude), then the resulting sequence of numbers can be
analysed using standard time-frequency analysis techniques to reveal rhythmic patterns (Saar & Mitra,
2008). Alternatively, each frame can be compared with every other frame to detect similarities using
standard measures for quantifying the distance between vectors (Paulus, Müller & Klapuri, 2010). These
distances are then often collected within a second matrix called a self-distance matrix. Temporal
regularities within the original feature matrix are visible as coherent patterns with the self-distance matrix
(typically showing up as patterned blocks or diagonal stripes). Various methods used for describing and
classifying patterns within matrices (or images) can then be used to classify these two-dimensional
patterns.

Different patterns in these matrices can be associated with variations in the novelty or
homogeneity of the temporal regularities over time, as well as the number of repetitions of particular
temporal patterns (Paulus, Müller & Klapuri, 2010). Longitudinal analyses of time-series measures of
temporal structure can also be used to describe the stability or dynamics of rhythmic pattern production
over time (Saar & Mitra, 2008). An alternative approach to identifying temporal structure within the
feature matrix is to decompose it into simpler component matrices that capture the most recurrent features
within the recording (Weiss & Bello, 2011). Similar approaches are common in modern analyses of high-
density EEG recordings (Makeig et al., 2004). Algorithms for analysing the temporal dynamics of brain
waves may thus also be useful for analysing temporal structure within acoustic recordings.

V. FUTURE DIRECTIONS

Many of the central questions in animal communication research focus on the meaning of signals and on
the role of natural, sexual, and social selection on the evolution of communication systems. As shown in
For Review Only

Figure 2, information can exist in a sequence simultaneously via diversity, and order, as well as other less well-studied phenomena. Both natural and sexual selection may act on this information, either through conspecifics or heterospecifics (e.g., predators). This is especially true for animal acoustic sequences because the potential complexity of a sequence may imply greater scope for both meaning and selective pressure. Many new questions – and several old and unanswered ones – can be addressed by the techniques that we have outlined in this review. Some of the most promising avenues for future research are outlined below. First, we illustrate the integration of our framework using two case studies: the songs of rock hyraxes, *Procavia capensis*, and California thrashers, *Toxostoma redivivum*. Then, we discuss four outstanding questions in animal acoustic sequences that can potentially be addressed more effectively using the approaches proposed in this review.

(1) Two case studies

(a) Rock hyraxes

Kershenbaum et al. (2012) examined the syntactic structure of hyrax vocal sequences. They selected units using the separation by silence approach (Figure 3a). A histogram showing the frequency of different lengths of silence in the recordings was strongly bimodal (Kershenbaum et al., 2012: Figure S3), supporting the separation by silence paradigm. From these recordings, five distinct syllable types were identified. Previous studies (Koren & Geffen, 2009; Koren & Geffen, 2011) had identified only three types of hyrax syllables, but Kershenbaum et al. (2012) felt that one of the syllable categories (“wail”) was too general for their data, and that continuous variation in the length and frequency modulation patterns in the “wail” syllable could be perceived as different units by the receiver (Figure 5b). To avoid observation bias, they used an automatic system to measure the length and bandwidth of the purportedly “wail” syllables, and divided them into two further categories, based on fixed thresholds. Inspection “by eye” showed that only 10% of automatically classified syllables needed manual reclassification.

The analysis of these hyrax vocalisations followed the *Ordering* paradigm (Figure 2c). No attempt was made to fit these data to one of the models described in Section IV.
The main finding of this study was to show the presence of geographical dialects, by quantifying the differences in order (syntax) using the edit (or Levenshtein) distance (Garland et al., 2012), via the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970). Using a quantitative tool to compare sequence similarity, and demonstrating that hyrax vocalisations have syntactic structure, the study showed that complex vocalisation syntax is more common in mammals than commonly believed.

(b) California thrashers

Sasahara et al. (2012) examined the song sequences of the California thrasher, a bird with an extensive repertoire (over 180 song types in this study). Acoustic units were separated by silence (Figure 3a), and classified “by eye” against a library catalogue of phrase types. However, to assess the reliability of the classification criteria, the researchers trained a support vector machine (or a support vector network) on a sample of manually classified units, and ensured that the automatic and manual classifications were consistent. Although song phrases were typically monosyllabic, if the time between units was small, units were grouped into a larger “phrase” (Figure 5c). Similarly two or more units were considered a single unit if they were always associated with each other in a series (Figure 2c). Although not analysed in Sasahara et al. (2012), subsequent unpublished work indicated that sequences of phrases can be grouped into higher levels of organisation (Figure 3d), with longer intervals between them.

Although Markov and hidden Markov models work well for some species, these models are unlikely to work for California thrashers because of the very large phrase repertoire; nodes in the model appear to grow without limit, and estimation of the transition probabilities will be inaccurate. Sasahara et al. (2012) used a network model, to bypass these problems.

Figure 11 shows a sample of the analysed data, with the phrase types as nodes, and the colour of directed lines indicating the proportion of observed transitions. The network structure for California thrasher songs varies between individuals, as indicated by network metrics of mean path length, clustering coefficient, and degree centrality (Section IV.3). All individuals showed networks with a “small world”
structure. Individual networks also showed variation in network metrics with time, indicating that song
structure was non-stationary.

Although traditional analytical techniques such as Markov models would have been inappropriate
for these data, by using new and alternative approaches (a network model), the authors extracted
quantitative metrics on song structure, and successfully used these to compare the songs of different
individuals. These techniques are a promising approach for the highly complex sequences of species such
as the thrasher, mockingbird, and humpback whales.

(2) Outstanding questions in animal vocal sequences

(a) How do we define communication complexity?

Perceptual and developmental constraints may drive selection for communication complexity. However,
complexity can exist at any one (or more) of the six levels of information encoding that we have detailed,
often leading to definitions of “communication complexity” that are inconsistent between researchers. As
it is likely that no one definition of communication complexity can be universally valid, any definition
has relevance only after choosing to which of the encoding paradigms described in Figure 2 – or
combination thereof – it applies. Complexity defined, say, for the Repetition paradigm (Figure 2a) and
quantified as pulse rate variation, is not easily compared with Diversity complexity (Figure 2b), typically
quantified as repertoire size.

For example, one hotly debated subject is whether selection from increased social complexity or
sexual selection is associated with increased acoustic complexity (Pollard & Blumstein, 2012), with
“complexity” defined as repertoire size (Figure 2b). Some researchers have proposed the idea that
communicative complexity, again defined as repertoire size, has at least in some species been driven by
the need to encode more information, or redundant information, in a complex social environment
(Freeberg et al., 2012). Alternatively, complexity metrics that measure Ordering (Figure 2d), often based
on non-zero orders of entropy (McCowan, Hanser & Doyle, 1999; Kershenbaum, 2013), may be more
biologically relevant in species that use unit ordering to encode information. Understanding the variety of
sequence types is essential to choosing the relevant acoustic unit definitions, and without this, testing
competitive evolutionary hypotheses becomes problematic.

(b) What is the role of sequence dialects in speciation?

It is tempting to speculate that acoustic sequences may have arisen from earlier selective forces acting on
a communication system based on single units. Alternatively, however, sequences could arise by neutral
processes analogous to drift. A complex interplay between production, perception, and encoding of
information in sequence syntax (along with the large relative differences between different species in
adaptive flexibility, Seyfarth & Cheney, 2010) could lead to adaptive pressures on communication
structure. In a few species, geographic syntactic dialects (Nettle, 1999) have been demonstrated (e.g.,
Rhesus monkeys *Macaca mulatta*, Gouzoules, Gouzoules & Marler, 1984; chimpanzees *Pan troglodytes*,
Arcadi, 1996; Mitani, Hunley & Murdoch, 1999; Crockford & Boesch, 2005; Carolina chickadees *Poecile
carolinensis*, Freeberg, 2012; and rock hyraxes *Procavia capensis*, Kershenbaum et al., 2012) raising the
question of whether sequence syntax has a role in speciation (Wiens, 1982; Nevo et al., 1987; Irwin,
2000; Lachlan et al., 2013). However, the definition of acoustic units is rarely considered. In particular,
perceptual binding (Figure 5a) and the response of the focal species must be considered, as reproductive
isolation cannot occur on the basis of differences that are not perceived by the receiver.

(c) How do individual differences in acoustic sequences arise?

The proximal processes driving individual differences in communicative sequences are rarely
investigated. Likewise, although there is a decades-rich history of song learning studies in songbirds, the
ontogenetic processes giving rise to communicative sequences *per se* have rarely been studied. Neural
models (e.g., Jin, 2009) can provide probabilistic descriptions of sequence generation (e.g., Markov
models, hidden Markov models), but the nature of the underlying stochasticity is unknown. The
California thrasher example given above (Figure 11), shows how an appropriate choice of a model for
sequence structure can allow quantitative comparisons between the parameters of different individuals.
However, model fitting is only valid if unit selection is biologically appropriate (Section III). Other, more abstract, questions can also be addressed. Individual humans use language with varying degrees of efficiency, creativity, and effectiveness. Shakespearean sequences are radically unlike Haiku sequences, political speeches, or the babbling of infants, in part because their communicative purposes differ. While sexual selection and survival provide some purposive contexts through which we can approach meaning, additional operative contexts may suggest other purposes, and give us new frameworks through which to view vocal sequences (Waller, 2012).

(d) How might information exist within units themselves?

Another promising direction lies in studying how animals use concatenated signals with multiple meanings. For example, Jansen, Cant & Manser (2012) provided evidence for temporal segregation of information within a syllable, where one segment of a banded mongoose *Mungos mungo* close call is individually distinct, while the other segment contains meaning about the caller’s activity. Similar results have been demonstrated in the song of the white-crowned sparrow *Zonotrichia leucophrys* (Nelson & Poesel, 2007). Understanding how to divide acoustic units according to criteria other than silent gaps (Figure 3) can change the research approach, as well as the results of a study. The presence of information in sub-divisions of traditional acoustic units is a subject underexplored in the field of animal communication, and an understanding of the production and perceptual constraints on unit definition (Figure 5) is essential.

We conclude by noting that more detailed and rigorous approaches to investigating animal acoustic sequences will allow us to investigate more complex systems that have not been formally studied. A number of directions lack even a basic framework as we have proposed in this review. For example, there is much to be learned from the detailed study of the sequences created by multiple animals vocalising simultaneously, and from the application of sequence analysis to multimodal communication with a combination of acoustic, visual, and perhaps other modalities (e.g., Partan & Marler, 1999; Bradbury &
Vehrencamp, 2011; Munoz & Blumstein, 2012). Eavesdropping, in which non-target receivers (such as predators) gain additional information from listening to the interaction between individuals, has only just begun to be studied in the context of sequence analysis. Finally, the study of non-stationary systems, where the statistical nature of the communicative sequences changes over long or short time scales (such as appears to occur in humpback whale songs) is ripe for exploration. We encourage researchers in these fields to extend treatments such as ours to cover these more complex directions in animal communication research, thereby facilitating quantitative comparisons between fields.

VI. CONCLUSIONS

1. The use of acoustic sequences by animals is widespread across a large number of taxa. As diverse as the sequences themselves is the range of analytical approaches used by researchers. We have proposed a framework for analysing and interpreting such acoustic sequences, based around three central ideas of understanding the information content of sequences, defining the acoustic units that comprise sequences, and proposing analytical algorithms for testing hypotheses on empirical sequence data.

2. We propose use of the term “meaning” to refer to a feature of communication sequences that influences behavioural and evolutionary processes, and the term “information” to refer to the non-random statistical properties of sequences.

3. Information encoding in acoustic sequences can be classified into six non-mutually exclusive paradigms: Repetition, Diversity, Combination, Ordering, Overlapping, and Timing.

4. The constituent units of acoustic sequences can be classified according to production mechanisms, perception mechanisms, or analytical properties.

5. Discrete acoustic units are often delineated by silent intervals. However, changes in the acoustic properties of a continuous sound may also indicate a transition between discrete units, multiple repeated sounds may act as a discrete unit, and more complex hierarchical structure may also be present.
(6) We have reviewed five approaches used for analysing the structure of animal acoustic sequences:
Markov chains, hidden Markov models, network models, formal grammars, and temporal models, discussing their use and relative merits.

(7) Many important questions in the behavioural ecology of acoustic sequences remain to be answered, such as understanding the role of communication complexity, including multimodal sequences, the potential effect of communicative isolation on speciation, and the source of syntactic differences between individuals.

VII. ACKNOWLEDGEMENTS

This review was developed at an investigative workshop, “Analyzing Animal Vocal Communication Sequences” that took place on October 21-23 2013 in Knoxville, Tennessee, sponsored by the National Institute for Mathematical and Biological Synthesis (NIMBioS). NIMBioS is an Institute sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Awards #EF-0832858 and #DBI-1300426, with additional support from The University of Tennessee, Knoxville. In addition to the authors, Vincent Janik participated in the workshop. DTB’s research is currently supported by NSF DEB-1119660. MAB’s research is currently supported by NSF IOS-0842759 and NIH R01DC009582. MAR’s research is supported by ONR N0001411IP20086 and NOPP (ONR/BOEM) N00014-11-1-0697. SLDeR’s research is supported by the U.S. Office of Naval Research. RF-i-C’s research was supported by the grant BASMATI (TIN2011-27479-C04-03) from the Spanish Ministry of Science and Innovation. ECG’s research is currently supported by a National Research Council postdoctoral fellowship. EEV’s research is supported by CONACYT, Mexico, award number I010/214/2012.


FIGURE LEGENDS

Figure 1. Overall flowchart showing a typical analysis of animal acoustic sequences. In this review, we discuss Identifying units, Characterising sequences, and Identifying meaning.

Figure 2. Different ways that units can be combined to encode information in a sequence.

Figure 3. Examples of the different criteria for dividing a spectrogram into units. (a) Separating units by silent gaps is probably the most commonly used criterion. (b) An acoustic signal may change its properties at a certain time, without the presence of a silent “gap”, for instance becoming harmonic or noisy. (c) A series of similar sounds may be grouped together as a single unit, regardless of silent gaps between them; a chirp sequence is labelled as C. (d) A complex hierarchical structure to the sequence, combining sounds that might otherwise be considered fundamental units.

Figure 4. Example of cepstral processing of a grey wolf Canis lupis howl and crickets chirping. Recording was sampled at Fs = 16 kHz, 8 bit quantization. (a) Standard spectrogram analyzed with a 15 ms Blackman-Harris window. (b) Plot of transform to cepstral domain. Lower quefrencies are related to vocal tract information. F0 can be determined from the “cepstral bump” apparent between quefrencies 25-45 and can be derived by Fs/quefrency. (c) Cepstrum (inset) of the frame indicated by an arrow (2.5 s) along with reconstructions of the spectrum created from truncated cepstral sequences. Fidelity improves as the number of cepstra are increased.

Figure 5. Perceptual constraints for the definition of sequence units. (a) Perceptual binding, where two discrete acoustic elements may be perceived by the receiver either as a single element, or as two separate ones. (b) Categorical perception, where continuous variation in acoustic signals may be interpreted by the
receiver as discrete categories. (c) Spectrotemporal constraints, where if the receiver cannot distinguish
small differences in time or frequency, discrete elements may be interpreted as joined.

Figure 6. State transition diagram equivalent to a 2nd order Markov model and trigram model (N=3) for a
sequence containing A’s and B’s.

Figure 7. State transition diagram of a two state (X, Y) hidden Markov model capable of producing
sequences of acoustic units A and B. When in state X, acoustic units emission of signals A and B are
equally likely $Pr(A|X) = Pr(B|X) = 0.5$, and when in state Y, acoustic unit A is much more likely $Pr(A|Y) = 0.9$ than $B|Y) = 0.1$. Transitioning from state X to state Y occurs with probability $Pr(X → Y) = 0.6$, whereas from state Y to state X with probability $Pr(Y → X) = 0.3$.

Figure 8. Simple networks constructed from the sequence of acoustic units ABC. The undirected binary
network (left) simply indicates that A, B, and C are associated with one another without any information
about transition direction. The directed binary network (centre) adds ordering information, for example
that C cannot follow A. The weighted directed network (right) show the probabilities of the transitions
between units based on a bigram model.

Figure 9. Grammar (rewrite rules) for approximating the sequence of acoustic units produced by Eastern
Pacific blue whales *Balaenoptera musculus*. There are three acoustic units, A, B, and D (Oleson, Wiggins
& Hildebrand, 2007), and the sequence begins with a start symbol S. Individual B or D calls may be
produced, or song, which consists of repeated sequences of an A call followed by one or more B calls. The
symbol | indicates a choice, and $\epsilon$, the empty string, indicates that the rule is no longer used. A derivation
is shown for the song $ABBAB$. Grammar produced with contributions from Ana Širović (Scripps
Institution of Oceanography).
Figure 10. The classes of formal grammars known as the Chomsky hierarchy (Chomsky, 2002). Each class is a generalisation of the class it encloses, and is more complex than the enclosed classes. Image publicly available under the Creative Commons Attribution-Share Alike 3.0 Unported license. 

https://commons.wikimedia.org/wiki/File:Wiki_inf_chomskeho_hierarchia.jpg

Figure 11. Network diagram for a sample of California thrasher song data. Nodes indicate distinct phrase types (acoustic units), arrows indicate transitions, with the colour of the arrow indicating the strength of probability of transition between two units (hotter colours represent more likely transitions).
Table 1. Examples of different approaches to unit definition, from different taxa and with different research aims.

<table>
<thead>
<tr>
<th>Unit criterion</th>
<th>Taxon</th>
<th>Goal of division into “units”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Descriptive</strong></td>
</tr>
<tr>
<td>Separated by silence</td>
<td>Birds Swamp sparrow <em>Melospiza georgiana</em> note: (Marler &amp; Pickert, 1984)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Black capped chickadee <em>Poecile atricapillus</em> note: (Nowicki &amp; Nelson, 1990)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zebra finch <em>Taeniopygia guttata</em> syllable: (Cynx, 1990)</td>
</tr>
<tr>
<td></td>
<td>Terrestrial mammals Meerkat <em>Suricata suricatta</em> calls: (Manser, 2001)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gibbon <em>Hyrobates lar</em> phrase: (Raemaekers, Raemaekers &amp; Haimoff, 1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rock hyrax <em>Procavia capensis</em> songs: (Kershenbaum et al., 2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Free-tailed bat <em>Tadarida brasiliensis</em> syllable: (Bohn et al., 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mustached bat <em>Pteronotus parnelli</em> syllable: (Kanwal et al., 1994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Swamp sparrow <em>Melospiza georgiana</em> note: (Nelson &amp; Marler, 1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Black-capped chickadee <em>Poecile atricapillus</em> notes: (Sturdy, Phillmore &amp; Weisman, 2000)</td>
</tr>
<tr>
<td></td>
<td>Terrestrial mammals Meerkat <em>Suricata suricatta</em> calls: (Manser, 2001)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rock hyrax <em>Procavia capensis</em> songs: (Koren &amp; Geffen, 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Free-tailed bat <em>Tadarida brasiliensis</em> syllable: (Bohn et al., 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carolina chickadee <em>Poecile carolinensis</em> and Black-capped chickadee <em>P. atricapillus</em> note composition → <strong>predator, foraging activity, identity</strong>: (Krams et al., 2012)</td>
</tr>
<tr>
<td>Marine mammals</td>
<td>Humpback whale <em>Megaptera novaeangliae</em> unit: (Payne &amp; McVay, 1971)</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Killer whale <em>Orcinus orca</em> calls: (Ford, 1989)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bottlenose dolphin <em>Tursiops truncatus</em> signature whistles: (Caldwell, 1965; McCowan &amp; Reiss, 1995)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bottlenose dolphin <em>Tursiops truncatus</em> signature whistles: (Janik, Sayigh &amp; Wells, 2006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Killer whale <em>Orcinus orca</em> calls: → <em>individual identity</em>: (Sayigh et al., 1999; Harley, 2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Killer whale <em>Orcinus orca</em> calls: → <em>group identity</em>: (Ford, 1989)</td>
<td></td>
</tr>
<tr>
<td>Change in acoustic properties (regardless of silence)</td>
<td>Birds</td>
<td></td>
</tr>
<tr>
<td>Terrestrial mammals</td>
<td>Northern cardinal <em>Cardinalis cardinalis</em>: (Suthers, 1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black-fronted titi monkey <em>Callicebus nigrifrons</em>: (Cäsar et al., 2012b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Western gorilla <em>Gorilla gorilla</em>: (Salmi, Hammerschmidt &amp; Doran-Sheehy, 2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Red titi monkey <em>Callicebus cupreus</em>: (Robinson, 1979)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black-fronted titi monkey <em>Callicebus nigrifrons</em> alarm calls → <em>predator type and behaviour</em>: (Cäsar et al., 2012a)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Western gorilla <em>Gorilla gorilla</em> vocalisations → <em>multiple functions</em>: (Salmi, Hammerschmidt &amp; Doran-Sheehy, 2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tufted capuchin monkeys <em>Sapajus nigritus</em> calls → <em>predator type</em>: (Wheeler, 2010b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Banded mongoose <em>Mungos mungo</em> close calls → <em>individual identity</em>, <em>group cohesion</em>: (Jansen, Cant and Manser, 2012)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrestrial mammals</td>
<td>Black-fronted titi monkey <em>Callicebus nigrifrons</em>: (Cäsar et al., 2012b; Cäsar et al., 2013)</td>
<td>Black-fronted titi monkey <em>Callicebus nigrifrons</em>: (Cäsar et al., 2012a)</td>
</tr>
<tr>
<td></td>
<td>Mustached bat <em>Pteronotus parnellii</em> syllable: (Kanwal et al., 1994)</td>
<td>Colobus <em>Colobus guereza</em> sequences: (Schel, Candiotti &amp; Zuberbühler, 2010)</td>
</tr>
<tr>
<td></td>
<td>Free-tailed bat <em>Tadarida brasiliensis</em> calls: (Bohn et al., 2008)</td>
<td>Tufted capuchin monkey <em>Sapajus nigris</em> bouts:</td>
</tr>
<tr>
<td>Hymenomorphs</td>
<td>Hyrax <em>Procavia capensis</em> social calls: (Ilany et al., 2013)</td>
<td>Chimpanzee <em>Pan troglodytes</em> pant hoots: (Notman &amp; Rendall, 2005)</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>Higher levels of organisation</td>
<td>Swamp sparrow <em>Melospiza georgiana</em> trills: (Podos, 1997)</td>
<td>Song sparrow <em>Melospiza melodia</em> songs: (Searcy et al., 1995)</td>
</tr>
<tr>
<td></td>
<td>Free-tailed bat <em>Tadarida brasiliensis</em> songs: (Bohn et al., 2008)</td>
<td>Black-fronted titi monkey <em>Callicebus nigrifrons</em>: (Cäsar et al., 2008)</td>
</tr>
<tr>
<td></td>
<td>Rhesus-macaque <em>Macaca mulatta</em> vocalisations: (Fitch, 1997)</td>
<td>Black-fronted titi monkey <em>Callicebus nigrifrons</em>: (Cäsar et al., 2008)</td>
</tr>
<tr>
<td>Marine mammals</td>
<td>Humpback whale <em>Megaptera novaeangliae</em> theme and song: (Payne &amp; McVay, 1971)</td>
<td>1979)</td>
</tr>
</tbody>
</table>
Data collection

Identifying units

Characterising sequence

Identifying meaning

Data collection

Raw audio

Preprocessing

Filtering

Time-frequency analysis (e.g. spectrogram)

Time-series analysis

Division into units

Change in acoustic properties

Separated by silence

Series of sounds

Higher level of organisation

Select relevant units

Characterise sequence

Extract features

Identify meaning

Clustering

Repertoire

Testable hypotheses & behavioural experiments

Repetition

Diversity

Composition

Ordering

Overlapping

Temporal
<table>
<thead>
<tr>
<th>Type</th>
<th>Criterion</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Repetition</td>
<td>Single unit repeated more than once</td>
</tr>
<tr>
<td>b</td>
<td>Diversity</td>
<td>A number of distinct units are present. Order is unimportant.</td>
</tr>
<tr>
<td>c</td>
<td>Combination</td>
<td>Set of units has different information from each unit individually. Order is unimportant.</td>
</tr>
<tr>
<td>d</td>
<td>Ordering</td>
<td>Set of units has different information from each unit individually. Order is important</td>
</tr>
<tr>
<td>e</td>
<td>Overlapping</td>
<td>Information conveyed in the relationship between sequences of two or more individuals</td>
</tr>
<tr>
<td>f</td>
<td>Timing</td>
<td>Timing between units (often between different individuals) conveys information</td>
</tr>
</tbody>
</table>
(a) Separated by silence

(b) Change in acoustic properties (regardless of silence)

(c) Series of sounds

(d) Higher levels of organization
Figure 4. Example of cepstral processing of a grey wolf *Canis lupis* howl and crickets chirping. Recording was sampled at $F_s = 16$ kHz, 8 bit quantization. (a) Standard spectrogram analyzed with a 15 ms Blackman-Harris window. (b) Plot of transform to cepstral domain. Lower quefrencies are related to vocal tract information. F0 can be determined from the "cepstral bump" apparent between quefrencies 25-45 and can be derived by $F_s$/quefrency. (c) Cepstrum (inset) of the frame indicated by an arrow (2.5 s) along with reconstructions of the spectrum created from truncated cepstral sequences. Fidelity improves as the number of cepstra are increased.
(a) Perceptual binding. Two discrete acoustic elements may be perceived by the receiver either as a single element, or as two separate ones.

(b) Categorical perception. Continuous variation in acoustic signals may be interpreted by the receiver as discrete categories.

(c) Spectrotemporal constraints. If the receiver cannot distinguish small differences in time or frequency, discrete elements may be interpreted as joined.
\[ \begin{array}{|c|c|c|}
\hline
\text{ } & \text{A} & \text{B} \\
\hline
\text{P(X|AA)} & 0.2 & 0.8 \\
\text{P(X|AB)} & 0.6 & 0.4 \\
\text{P(X|BA)} & 0.3 & 0.7 \\
\text{P(X|BB)} & 0.9 & 0.1 \\
\hline
\end{array} \]

\[ \begin{aligned}
P(B|AB) &= 0.4 \\
P(B|BB) &= 0.1 \\
P(B|BA) &= 0.7 \\
P(B|AA) &= 0.6 \\
P(A|BB) &= 0.9 \\
P(A|BA) &= 0.3 \\
P(A|AA) &= 0.2 \\
\end{aligned} \]
\[ \text{Pr}(X \rightarrow Y) = 0.6 \]
\[ \text{Pr}(Y \rightarrow X) = 0.3 \]
\[ \text{Pr}(A \mid X) = 0.5 \]
\[ \text{Pr}(B \mid X) = 0.5 \]
\[ \text{Pr}(A \mid Y) = 0.9 \]
\[ \text{Pr}(B \mid Y) = 0.1 \]
\[ \text{Pr}(X \rightarrow X) = 0.4 \]
\[ \text{Pr}(Y \rightarrow Y) = 0.7 \]
Figure 9. Grammar (rewrite rules) for approximating the sequence of acoustic units produced by Eastern Pacific blue whales *Balaenoptera musculus*. There are three acoustic units, A, B, and D (Oleson, Wiggins & Hildebrand. 2007), and the sequence begins with a start symbol S. Individual B or D calls may be produced, or song, which consists of repeated sequences of an A call followed by one or more B calls. The symbol | indicates a choice, and ε, the empty string, indicates that the rule is no longer used. A derivation is shown for the song ABBAB. Grammar produced with contributions from Ana Širović (Scripps Institution of Oceanography).
Figure 10. The classes of formal grammars known as the Chomsky hierarchy (Chomsky, 2002). Each class is a generalisation of the class it encloses, and is more complex than the enclosed classes. Image publicly available under the Creative Commons Attribution-Share Alike 3.0 Unported license. https://commons.wikimedia.org/wiki/File:Wiki_inf_chomskeho_hierarchia.jpg 251x181mm (72 x 72 DPI)
Figure 11. Network diagram for a sample of California thrasher song data. Nodes indicate distinct phrase types (acoustic units), arrows indicate transitions, with the colour of the arrow indicating the strength of probability of transition between two units (hotter colours represent more likely transitions).

281x280mm (300 x 300 DPI)