Kershenbaum, A, Blumstein, DT, Roch, MA, Akcay, C, Backus, G, Bee, MA, Bohn, K, Cao, Y, Carter, G, Caesar, C, Coen, M, DeRuiter, SL, Doyle, L, Edelman, S, Ferrer-i-Cancho, R, Freeberg, TM, Garland, EC, Gustison, M, Harley, HE, Huetz, C, Hughes, M, Bruno, JH, Ilany, A, Jin, DZ, Johnson, M, Ju, C, Karnowski, J, Lohr, B, Manser, MB, McCowan, B, Mercado, E, Narins, PM, Piel, AK, Rice, M, Salmi, R, Sasahara, K, Sayigh, L, Shiu, Y, Taylor, C, Vallejo, EE, Waller, S and Zamora-Gutierrez, V

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

http://researchonline.ljmu.ac.uk/id/eprint/2927/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Kershenbaum, A, Blumstein, DT, Roch, MA, Akcay, C, Backus, G, Bee, MA, Bohn, K, Cao, Y, Carter, G, Caesar, C, Coen, M, DeRuiter, SL, Doyle, L, Edelman, S, Ferrer-i-Cancho, R, Freeberg, TM, Garland, EC, Gustison, M, Harley, HE, Huetz, C, Hughes, M, Bruno, JH, Ilany, A, Jin, DZ, Johnson, M,

BIOLOGICAL REVIEWS
Cambridge Philosophical Society

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

Journal:	*Biological Reviews*
Manuscript ID:	BRV-02-2014-0031.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
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

Arik Kershenbaum[1], Daniel T. Blumstein[2], Marie A. Roch[3], Çağlar Akçay[4], Gregory Backus[5], Mark Bee[6], Kirsten Bohn[7], Yan Cao[8], Gerald Carter[9], Cristiane Cäsar[10], Michael Coen[11], Stacy L. DeRuiter[12], Laurance Doyle[13], Shimon Edelman[14], Ramon Ferrer-i-Cancho[15], Todd M. Freeberg[16], Ellen C. Garland[17], Morgan Gustison[18], Heidi E. Harley[19], Chloé Huetz[20], Melissa Hughes[21], Julia Hyland Bruno[22], Amiyaal Ilany[23], Dezhe Z. Jin[24], Michael Johnson[25], Chenghui Ju[26], Jeremy Karnowski[27], Bernard Lohr[28], Marta B. Manser[29], Brenda McCowan[30], Eduardo Mercado III[31], Peter M. Narins[32], Alex Piel[33], Megan Rice[34], Roberta Salmi[35], Kazutoshi Sasahara[36], Laela Sayigh[37], Yu Shiu[38], Charles Taylor[39], Edgar E. Vallejo[40], Sara Waller[41], Veronica Zamora-Gutierrez[42]

---

[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 Politecnica 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, Swizerland*
[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, Monterrey 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 review aims 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: identifying the distinct units of an acoustic sequence, describing the different ways that information can be contained within a 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

CONTENTS

I. INTRODUCTION

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, Dunbar & Ord. 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 could benefit from an approach that can bridge and bring together 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 that compose the sequence? How do we assess the structure governing the composition of these units? 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 what is meant by “information” content and “meaning” in sequences. 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 look at the different ways that units can encode information in sequences. In Section V, 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 VI, we 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 information and meaning is sometimes portrayed with information as the form or structure of some entity on the one hand, and meaning as the resulting activity of a receiver of that information on the other hand (Bohm. 1989).

(1) Philosophy of meaning

The different vocal signals of a species are typically thought to vary in ways associated with factors that are primarily internal (hormonal, motivational, emotional), behavioural (movement, affiliation, agonistic), external (location, resource and threat detection), or combinations of such factors. Much of the variation in vocal signal structure and signal use relates to what W. John Smith called the message of the signal – the “kinds of information that displays enable their users to share” (Smith. 1977, pg. 70). Messages of signals are typically only understandable to us as researchers after considerable observational effort aimed at determining the extent of association between signal structure and use and the factors mentioned above. The receiver of a signal gains information, or meaning, from the structure and use of the signal. Depending on whether the interests of the receiver and the signaller are aligned or opposed, the receiver may benefit, or potentially be fooled or deceived, respectively (Searcy & Nowicki. 2005). The meaning of a signal stems not just from the message or information in the signal itself, but also from the context in which the signal is produced. The context of communication involving a particular signal could relate to a number of features, including signaller characteristics, such as recent signals or cues it has sent, as well as location or physiological state, and receiver characteristics, such as current behavioural activity or recent experience. Context can also relate to joint signaller and receiver characteristics, such as the nature of their relationship (Smith. 1977).

Philosophical understanding of meaning is rooted in studies of human language and offers a variety of schools of thought. The nature of meaning has been theorised in many ways: extensional (based on things in the world, like the set of all animals), intensional (based on thoughts within minds, notions, concepts, ideas), or according to prototype theory (in which objects have meaning through a graded categorisation, e.g. “baldness” is not precisely determined by the number of hairs on the head). The physiological nature of meaning may be innate or learned, in terms of its mental representations and cognitive content. Finally, descriptions of the role of meaning are diverse: meaning may be computational/functional; atomic or holistic; bound to both signaller and receiver, or a speech act of the signaller; rule bound or referentially based; a description, or a convention; or a game dependent on a form of life, among other examples (Christiansen & Chater. 2001; Martinich & Sosa. 2013). This myriad list of philosophical theories is presented to give the reader a sense both of the lack of agreement as to the nature of meaning, and to highlight the lack of connection between theories of human semantics, and theories of animal communication.

(2) Context

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). Sex 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.

In human language, context has been proposed to be either irrelevant to, or crucial to, the meaning of words and sentences. In some cases, a sentence bears the same meaning across cultures, times, and locations, irrespective of context, e.g., "2+2=4" (Quine. 1960). In other cases, meaning is derived at least partially from external factors, e.g. the chemical composition of a substance defines its nature, irrespective of how the substance might be variously conceived by different people (Putnam. 1975). In contrast, indexical terms such as "she" gain meaning only as a function of context, such as physical or implied pointing gestures (Kaplan. 1978). Often, the effect of the signal on the receivers determines its usefulness, and that usefulness is dependent upon situational-contextual forces (Millikan. 2004).

(3) Definitions of meaning

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. An entropy metric places a bound on the maximum amount of information that can be present in a signal, although it does not determine that such information is, in fact, present.

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).

The structure of acoustic signals does not necessarily have meaning *per se*, and so measuring that structure does not necessarily reveal the complexity of meaning. As one example, the structure of an acoustic signal could be related to effective signal transmission through a noisy or reverberant 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).

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.

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).

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 2). First, the spectrogram may show a silent gap between two acoustic elements (Figure 2a). 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 2b). 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 2c). 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 2d). 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 3). 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; Tchernichovski 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 (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 (Peterson & Barney. 1952; 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 4).

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 4a). 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 4b). 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 4c). 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. 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, naïve 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 different sets of 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).

Three approaches have been used to classify units by their acoustic properties: visual classification of spectrograms, quantitative classification using features extracted visually from spectrograms, and fully-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) Classification of manually extracted metrics

As an alternative to visual classification, specific metrics, or features, measured on the acoustic data can be extracted for input to classification algorithms. A variety of time (e.g., duration, pulse repetition rate) and frequency (e.g., minimum, maximum, start, end, and range) components can be measured (extracted) from spectrograms, using varying degrees of automation, or computer assistance for a manual operator. Software tools such as Sound Analysis Pro (Tchernichovski et al. 2000) have been developed to assist with this task. Metrics are then used in classification analyses to identify units, using mathematical techniques such as discriminant function analysis (DFA), principal components analysis (PCA), or classification and regression trees (CART), and these have been applied to many mammalian and avian taxa, e.g., Derégnaucourt et al. (2005); Dunlop et al. (2007); Garland et al. (2012); Grieves, Logue & Quinn (2014). Feature extraction can be conducted using various levels of automation. A human analyst may note specific features for each call, an analyst-guided algorithm can be employed where sounds are identified by the analyst and a box is drawn around the call that automatically extracts a specific number of features, or the process of extraction can be fully automated. 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).

(c) Fully-automatic metric extraction and classification

Fully 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. Visual classification has been shown to out-perform automated systems in cases where the meaning of acoustic signals is known *a priori,* e.g., Sayigh et al. (2007); Kershenbaum, Sayigh & Janik (2013), possibly because the acoustic features used by fully automated systems may not reflect the cues used by the focal species. However, once an automatic algorithm is defined, large datasets can be analysed. 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.

The metrics generated either by manual or automatic extraction must be passed to a classification algorithm, to separate detections into discrete unit types. 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).

(d) Classification algorithms

Units may be classified automatically using supervised algorithms, in which the algorithm is taught to recognise unit types given some *a priori* known exemplars, or clustered using unsupervised algorithms, in which no *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 assigned to incorrect 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, for instance when the exemplars themselves in supervised clustering 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).

(5) Unit choice protocol

The definition of a unit for a particular focal species and a particular research question is necessarily dependent on a large number of factors in each specific project, and cannot be concisely summarised in a review of this length. In particular, availability or otherwise of behavioural information, such as the responses of individuals to playback experiments, is often the determining factor in deciding how to define a sequence unit. However, we provide here a brief protocol that can be used in conjunction with such prior information, or in its absence, to guide the researcher in choosing the definition of a unit. This protocol is also represented graphically in Figure 5. (a) Determine what is known about the production mechanism of the signalling individual. For example, Figure 5a lists eight possible production types that produce notably different sounds, although clearly other categories are also possible. (b) Determine what is known about the perception abilities of the individual. Perceptual limitations may substantially alter the structure of production units. Figure 5b gives examples of typical modifications resulting from reduced temporal or spectral resolution at the receiver. (c) Choose a classification method, such as manual, semi-automatic, or fully-automatic (Figure 5c). Some putative unit types lend themselves more readily to certain classification techniques than others. For example, "separated by silence" is often well-distinguished by manual inspection of spectrograms "by eye" or a band-limited energy detector, whereas "changes in acoustic properties" may benefit from manual extraction of features for passing to a classification algorithm (semi-automatic definition), and "series of sounds" may lend itself to a fully-automatic classification approach.

IV. 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 have discussed 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 that birdsong, for instance, can be described as a complex combination of notes, and we 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.

Although acoustic information can be encoded in many different ways, we consider here only the encoding of information via sequences. We suggest a classification scheme based on six distinct paradigms for encoding information in sequences (Figure 6). (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. Identification of the place for one's study system will stem in part from the nature of the system – a call system comprising a single, highly stereotyped contact note will likely fit neatly into the Repetition and Timing schemes we discuss, but may have little or nothing to do with the other schemes. We hope that our proposed framework will go beyond this, however, to drive researchers to consider additional schemes for their systems of study. For example, birdsong playback studies have long revealed that Diversity and Repetition often influence the behaviour of potential conspecific competitors and mates (Searcy & Nowicki. 2005). Much less is known about the possibility that Ordering, Overlapping, or Timing affect songbird receiver behaviour, largely because researchers simply have yet to assess the possibility in most systems. 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.

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; in 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)).

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 (Lambrechts & Dhondt. 1987; Lambrechts & 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, Lambrechts & 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). When sequences have complex ordering, simple quantitative measures are unlikely to capture the ordering information. Indeed, the Kolmogorov complexity of a sequence indicates how large a descriptor is required to specify the sequence adequately (Denker & Woyczyński. 1998). Instead of quantifying individual sequences, an alternative approach to measuring ordering is to calculate the pairwise similarity or difference between two sequences, using techniques such as the Levenshtein or Edit distance (Garland et al. 2012; Kershenbaum et al. 2012).

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). However, despite the apparent widespread use of overlapping in sequences, few analytical models have been developed to address this mechanism. While this is a promising area for future research, it is currently beyond the purview of this review.

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.

The use of multiple embedding techniques may be quite common, for instance in intrasexual competitive and intersexual reproductive contexts (Gerhardt & Huber. 2002). For 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.

Before hypothesising a specific structuring paradigm, it is frequently useful to perform exploratory data analysis (Figure 7). This might begin by looking at histograms, networks, or low-order Markov models that are based on acoustic units or timing between units. This analysis can be on the raw acoustic units or may involve preprocessing. An example of preprocessing that might be helpful for hypothesising *Repetition* would be to create histograms that count the number of times that acoustic units occur within a contiguous sequence of vocalisations. As an example, if twelve different acoustic units each occurred three times, a histogram bin representing three times would have a value of twelve; for examples, see Jurafsky & Martin (2000). For histograms or networks, visual analysis can be used to determine if there are any patterns that bear further scrutiny. Metrics such as entropy can be used to provide an upper bound on how well a Markov chain model describes a set of vocalisations (smaller numbers are better, as an entropy of zero indicates that we model the data perfectly). If nothing is apparent, it might mean that there is no structure to the acoustic sequences, but it also possible that the quantity of data are insufficient to reveal the structure or that the structure is more complex than what can be revealed through casual exploratory data analysis.

Exploratory data analysis may lead to hypotheses that one or more of the embedding paradigms for acoustic sequences may be appropriate. At this point a greater effort should be put into the modelling and understanding and we provide a suggested flow of techniques (Figure 7). It is important to keep in mind that these are only suggestions. For example, while we suggest that a grammar (section V.4) be modelled if there is evident and easily described structure for *Repetition*, *Diversity*, and *Ordering*, other models could be used effectively and machine learning techniques for generating grammars may be able to do so when the structure is less evident.

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 6. The example of the *Diversity* type is the system of serial calls of titi monkeys, *Callicebus molloch*, 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 (i.e. reordered) 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).

The second example relates to the *Ordering* type of sequence (Figure 6), 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.

V. THE STRUCTURE OF SEQUENCES

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.

We consider five models in this section: (1) Markov chains, (2) hidden Markov models, (3) network models, (4) formal grammars, and (5) temporal models. Each of these models has been growing in popularity among researchers, with the number of publications increasing in recent years. The number of publications in 2013 mentioning both the terms "animal communication" as well as the model name has grown since 2005 by a factor of: "Markov", 4.9; "hidden Markov", 3.3; "network", 2.6; "grammar" 1.7; "timing", 2.3.

The structure analysis algorithms discussed throughout this section can be used to model the different methods for combining units discussed earlier (Figure 6). *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. Table 2 summarises the assumptions and requirements for each of these models.

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.

(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 0th 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 1st 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-1th order Markov model. A K^{th} order Markov model of a sequence with C distinct units is defined by at most 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 8).

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. In many cases, the number of possible transitions is similar to, or larger than, the entire set of empirical data. For example, Briefer et al. (2010) examined very extensive skylark *Alauda arvensis* sequences totalling 16,829 units, but identified over 340 unit types. Since a naïve transition matrix between all unit types would contain 340 x 340 = 115,600 cells, the collected data set would be too small to estimate the entire matrix. A different problem occurs when, as is commonly the case, animal acoustic sequences are short. Kershenbaum et al. (2012) examined hyrax *Procavia capensis* sequences that are composed of just five unit types. However, 81% of the recorded sequences were only five or less units long. For these short sequences, $5^5 = 3125$ different combinations are possible – which is greater than the number of such sequences recorded (2374). In these cases, estimates of model parameters, and conclusions drawn from them, may be quite inaccurate (Cover & Thomas. 1991; Hausser & Strimmer. 2009; Kershenbaum. 2013).

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.

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

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, a method for improving estimated counts for events that occur infrequently (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 χ^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 9). 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 (θ) 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 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. Specific hidden Markov models 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 over-trained 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 10). 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); Weiss et al. (2014); Deslandes et al. (2014). 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: $S=(C/C_{rand})/(L/L_{rand}))$. 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 exists 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 11.

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 “over-selective”. A grammar that is “over-generalising” 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 11, the grammar is capable of producing the sequence *abbbbbbbbbbbb*, 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 12. 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 11 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 than regular grammars 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). Such an ability requires keeping track of a state, e.g. “how many times the unit A has been used”, and this neurological implementation may be lacking in most species (Beckers et al. 2012). Context sensitive languages allow context dependent rewrite rules that have few restrictions, permitting further reaching dependencies such as in the set of sequences $A^nB^nC^n$, and require still more sophisticated neural implementations. 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. For example, Gentner et al. (2006) proposed that European starlings *Sturnus vulgaris* can learn to recognise context-free (but non-regular) sequences, and reject sequences that do not correspond to the learned grammar. However, other authors have pointed out that the observed results could be explained by more simple mechanisms than context-free processing, such as primacy rules (Van Heijningen et al. 2009) in which simple analysis of short substrings is sufficient to distinguish between grammatical and non-grammatical sequences, or acoustic similarity matching (Beckers et al. 2012). Consequently, claims of greater than regular grammar in non-human animals have not been widely accepted. 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.

VI. FUTURE DIRECTIONS

Many of the central research questions in animal communication 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 Figure 6, 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, with some outstanding questions in animal acoustic sequences that can potentially be addressed more effectively using the approaches proposed in this review.

(1) As sequences are composed of units, how might information exist within units themselves?

One 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 2) can change the research approach, as well as the results of a study. The presence of information in subdivisions 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 4) is essential.

(2) How does knowledge and analysis of sequences help us define and understand communication complexity?

There is a long history of mathematical and physical sciences approaches to the question of complexity, which have typically defined complexity in terms of how difficult a system is to describe, how difficult a system is to create, or the extent of the system's disorder or organization (Mitchell. 2009; Page. 2010). This is an area of heavy debate among proponents of different views of complexity, as well as a debate about whether a universal definition of complexity is even possible. In the life and social sciences, the particular arguments are often different from those of the mathematical and physical sciences, but a similar heavy debate about the nature of biological complexity exists (Bonner. 1988; McShea. 1991; Adami. 2002; McShea. 2009).

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. In light of multiple levels of complexity, as well as multiple methods for separating units, we propose that no one definition of communication complexity can be universally suitable, and any definition has relevance only after choosing to which of the encoding paradigms described in Figure 6 – or combination thereof – it applies. Complexity defined, say, for the *Repetition* paradigm (Figure 6a) and quantified as pulse rate variation, is not easily compared with *Diversity* complexity (Figure 6b), typically quantified as repertoire size.

For example, is selection from increased social complexity associated with increased vocal complexity (Pollard & Blumstein. 2012; Freeberg, Dunbar & Ord. 2012), or does some other major selective factors – such as sexual selection or intensity of predation – drive the evolution of vocal complexity? In most of the studies to date on vocal complexity, complexity is defined in terms of repertoire size (Figure 6b). Considerable evidence in diverse taxa indicates that increased social complexity is associated with increased repertoire size, reviewed in Freeberg, Dunbar & Ord (2012). Different views of complexity in this literature are revealed by the fact that social complexity has been measured in terms of group size, group stability, or information-based metrics of group composition, and vocal complexity has been measured in terms of not just repertoire size, but also information-based metrics of acoustic variation in signals. In fact, the work of Pollard & Blumstein (2011) is highly informative to questions of complexity in that different metrics of social complexity can drive different metrics of vocal complexity – these authors have found that group size is associated with greater individual distinctiveness (information) in the calls of species, but the diversity of social roles in groups is more heavily associated with vocal repertoire size. 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, Dunbar & Ord. 2012). Alternatively, complexity metrics that measure *Ordering* (Figure 6d), 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.

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

If we can develop categories for unit types and sequence types that lead to productive vocalisation analysis and a deeper understanding of universal factors of encoded multi-layered messages, then individual differences in sequence production become interesting and puzzling. 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. When an appropriate choice of a model for sequence structure is made, quantitative comparisons can be made between the parameters of different individuals, for example with the California thrasher *Toxostoma redivivum* (Sasahara et al. 2012). 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). In many animals, song syntax may be related to sexual selection. Females of some species such as zebra finches *Taeniopygia guttata* not only prefer individuals with longer songs, but also songs comprising a greater variety of syllables (Searcy & Andersson. 1986; Neubauer. 1999; Holveck et al. 2008); whereas in other species, this preference is not observed (Byers & Kroodsma. 2009). Variation in syntax may also reflect individual differences in intraspecific aggression, for instance in banded wrens *Pheugopedius pleurostictus* (Vehrencamp et al. 2007) and western populations of song sparrows *Melospiza melodia* (Burt, Campbell & Beecher. 2001). Individual syntax may also serve to distinguish neighbours from non-neighbours in song sparrows (Beecher et al. 2000) and skylarks *Alauda arvensis* (Briefer et al. 2008). Male Cassin's vireos *Vireo cassinii* can usually be discriminated by the acoustic features of their song, but are discriminated even better by the sequences of phrases that they sang (Arriaga et al. 2013).

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

In a few species, geographic syntactic dialects (Nettle. 1999) have been demonstrated, including primates, such as Rhesus monkeys *Macaca mulatta* (Gouzoules, Gouzoules & Marler. 1984), chimpanzees *Pan troglodytes* (Arcadi. 1996; Mitani, Hunley & Murdoch. 1999; Crockford & Boesch. 2005); birds, such as Carolina chickadees *Poecile carolinensis* (Freeberg. 2012), swamp sparrows *Melospiza georgiana* (Liu et al. 2008), chaffinches *Fringilla coelebs* (Lachlan et al. 2013); and rock hyraxes *Procavia capensis* (Kershenbaum et al. 2012). This broad taxonomic spread raises the question of whether sequence syntax has a role in speciation (Wiens. 1982; Nevo et al. 1987; Irwin. 2000; Slabbekoorn & Smith. 2002; Lachlan et al. 2013), with some support for such a role in chestnut-tailed antbirds *Myrmeciza hemimelaena* (Seddon & Tobias. 2007), and winter wrens *Troglodytes troglodytes* (Toews & Irwin. 2008). It is tempting to speculate that acoustic sequences may have arisen from earlier selective forces acting on a communication system based on single units, with variation in the sequences of individuals providing differential adaptive benefit. The ability to communicate effectively with some but not others could lead to divergence of groups, and genetic pooling. Conversely, differences in acoustic sequences could be adaptive to ecological variation. It is hard to distinguish retrospectively between sequence dialect shift leading to divergence of sub-groups and eventual speciation, or group separation leading to new communicative strategies that are epiphenomena of species formation. What are the best methods for investigating the relationship between communication and biological change?

A third alternative is that sequence differences 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. However, the definition of acoustic units is rarely considered in this set of questions. In particular, perceptual binding (Figure 4a) 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. As units may be divided at many levels, there may be multiple sequences that convey different information types. Thus, a deeper understanding of units and sequences will contribute productively to questions regarding forces at work in speciation events.

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. For example, acoustic sequences may be constantly evolving sexual displays that are stereotyped within a population at any particular point in time (Payne & McVay. 1971; Payne, Tyack & Payne. 1983). The application of visual classification (Garland et al. 2011) and a variation of the statistical approach (edit distance) employed in the rock hyrax case study highlighted above (Kershenbaum et al. 2012), appears to capture the sequential information present within humpback whale song (Garland et al. 2012; Garland et al. 2013) This work traced the evolution of song lineages, and the movement or horizontal cultural transmission of multiple different versions of the song that were concurrently present across an ocean basin over a decade (Garland et al. 2013). These results are encouraging for the investigation of complex non-stationary systems; however, further refinement of this approach is warranted. 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.

VII. 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.

VIII. 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.

IX. REFERENCES

ADAM, O., CAZAU, D., GANDILHON, N., FABRE, B., LAITMAN, J. T. & REIDENBERG, J. S. (2013). New acoustic model for humpback whale sound production. *Applied Acoustics* **74**, 1182-1190.

ADAMI, C. (2002). What is complexity? *BioEssays* **24**, 1085-1094.

ADAMS, D. C. & ANTHONY, C. D. (1996). Using randomization techniques to analyse behavioural data. *Animal Behaviour* **51**, 733-738.

AKÇAY, Ç, TOM, M. E., CAMPBELL, S. E. & BEECHER, M. D. (2013). Song type matching is an honest early threat signal in a hierarchical animal communication system. *Proceedings of the Royal Society B: Biological Sciences* **280**.

ANDERSON, T. W. & GOODMAN, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics* **28**, 89-110.

ARAYA-SALAS, M. (2012). Is birdsong music? *Significance* **9**, 4-7.

ARCADI, A. C. (1996). Phrase structure of wild chimpanzee pant hoots: patterns of production and interpopulation variability. *American Journal of Primatology* **39**, 159-178.

ARLOT, S. & CELISSE, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys* **4**, 40-79.

ARNOLD, K. & ZUBERBÜHLER, K. (2013). Female putty-nosed monkeys use experimentally altered contextual information to disambiguate the cause of male alarm calls. *PloS One* **8**, e65660.

ARNOLD, K. & ZUBERBÜHLER, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology* **18**, R202-R203.

ARNOLD, K. & ZUBERBÜHLER, K. (2006a). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour* **72**, 643-653.

ARNOLD, K. & ZUBERBÜHLER, K. (2006b). Language evolution: semantic combinations in primate calls. *Nature* **441**, 303-303.

ARRIAGA, J. G., KOSSAN, G., CODY, M. L., VALLEJO, E. E. & TAYLOR, C. E. (2013). Acoustic sensor arrays for understanding bird communication. Identifying Cassin's Vireos using SVMs and HMMs. *ECAL 2013* **12**, 827-828.

ATTARD, M. R., PITCHER, B. J., CHARRIER, I., AHONEN, H. & HARCOURT, R. G. (2010). Vocal discrimination in mate guarding male Australian sea lions: familiarity breeds contempt. *Ethology* **116**, 704-712.

AUBIN, T., MATHEVON, N., STASZEWSKI, V. & BOULINIER, T. (2007). Acoustic communication in the Kittiwake *Rissa tridactyla*: potential cues for sexual and individual signatures in long calls. *Polar Biology* **30**, 1027-1033.

AUBIN, T., JOUVENTIN, P. & HILDEBRAND, C. (2000). Penguins use the two-voice system to recognize each other. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1081-1087.

BAKER, M. C. & BECKER, A. M. (2002). Mobbing calls of black-capped chickadees: effects of urgency on call production. *The Wilson Bulletin* **114**, 510-516.

BAKER, M. C. & LOGUE, D. M. (2003). Population differentiation in a complex bird sound: a comparison of three bioacoustical analysis procedures. *Ethology* **109**, 223-242.

BAPTISTA, L. F. & KEISTER, R. A. (2005). Why birdsong is sometimes like music. *Perspectives in Biology and Medicine* **48**, 426-443.

BARONCHELLI, A., FERRER-I-CANCHO, R., PASTOR-SATORRAS, R., CHATER, N. & CHRISTIANSEN, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences* **17**, 348-360.

BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R. & VESPIGNANI, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences* **101**, 3747-3752.

BASTIAN, M., HEYMANN, S. & JACOMY, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proceedings of the Third International ICWSM Conference* .

BAUGH, A., AKRE, K. & RYAN, M. (2008). Categorical perception of a natural, multivariate signal: mating call recognition in túngara frogs. *Proceedings of the National Academy of Sciences* **105**, 8985-8988.

BAZÚA-DURÁN, C. & AU, W. W. (2002). The whistles of Hawaiian spinner dolphins. *The Journal of the Acoustical Society of America* **112**, 3064-3072.

BECKERS, G. J., BOLHUIS, J. J., OKANOYA, K. & BERWICK, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23**, 139-145.

BEE, M. A. & MICHEYL, C. (2008). The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *Journal of Comparative Psychology* **122**, 235-251.

BEECHER, M. D., CAMPBELL, S. E., BURT, J. M., HILL, C. E. & NORDBY, J. C. (2000). Song-type matching between neighbouring song sparrows. *Animal Behaviour* **59**, 21-27.

BEN-DAVID, S., VON LUXBURG, U. &PÁL, D. (2006). A sober look at clustering stability. In *Learning theory*, pp. 5-19. Springer.

BENNET-CLARK, H. & EWING, A. (1969). Pulse interval as a critical parameter in the courtship song of *Drosophila melanogaster*. *Animal Behaviour* **17**, 755-759.

BERGMAN, T. J., BEEHNER, J. C., CHENEY, D. L. & SEYFARTH, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science* **302**, 1234-1236.

BERWICK, R. C., OKANOYA, K., BECKERS, G. J. L. & BOLHUIS, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences* **15**, 113-121.

BIZLEY, J. K. & COHEN, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience* **14**, 693-707.

BLUMSTEIN, D. T. (2007). The evolution, function, and meaning of marmot alarm communication. *Advances in the Study of Behavior* **37**, 371-401.

BLUMSTEIN, D. T. (1995). Golden-marmot alarm calls. II. Asymmetrical production and perception of situationally specific vocalizations? *Ethology* **101**, 25-32.

BLUMSTEIN, D. T., VERNEYRE, L. & DANIEL, J. C. (2004). Reliability and the adaptive utility of discrimination among alarm callers. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**, 1851-1857.

BLUMSTEIN, D. T. & MUNOS, O. (2005). Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls. *Animal Behaviour* **69**, 353-361.

BOHM, D. (1989). Meaning and information. In *The Search for Meaning, The New Spirit in Science and Philosophy* (ed. P. Pylkkänen), pp. 43-62. Thorsons Publishing Group, Wellingborough.

BOHN, K. M., SCHMIDT-FRENCH, B., MA, S. T. & POLLAK, G. D. (2008). Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *The Journal of the Acoustical Society of America* **124**, 1838-1848.

BOHN, K. M., SCHMIDT-FRENCH, B., SCHWARTZ, C., SMOTHERMAN, M. & POLLAK, G. D. (2009). Versatility and stereotypy of free-tailed bat songs. *PloS One* **4**, e6746.

BONNER, J. T. (1988). *The evolution of complexity by means of natural selection*. Princeton University Press.

BORGATTI, S. P., EVERETT, M. G. & FREEMAN, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Analytic Technologies.

BRADBURY, J. W. & VEHRENCAMP, S. L. (2011). *Principles of Animal Communication.* Sinauer, Sunderland, MA.

BRIEFER, E., OSIEJUK, T. S., RYBAK, F. & AUBIN, T. (2010). Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. *Journal of Theoretical Biology* **262**, 151-164.

BRIEFER, E. F., RYBAK, F. & AUBIN, T. (2013). Does true syntax or simple auditory object support the role of skylark song dialect? *Animal Behaviour* **86**, 1131-1137.

BRIEFER, E., AUBIN, T., LEHONGRE, K. & RYBAK, F. (2008). How to identify dear enemies: the group signature in the complex song of the skylark *Alauda arvensis*. *Journal of Experimental Biology* **211**, 317-326.

BROWN, S. D., DOOLING, R. J. & O'GRADY, K. E. (1988). Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): III. Contact calls. *Journal of Comparative Psychology* **102**, 236-247.

BRUMM, H. & SLATER, P. J. (2006). Ambient noise, motor fatigue, and serial redundancy in chaffinch song. *Behavioral Ecology and Sociobiology* **60**, 475-481.

BUCK, J. R. & TYACK, P. L. (1993). A quantitative measure of similarity for *Tursiops truncatus* signature whistles. *The Journal of the Acoustical Society of America* **94**, 2497-2506.

BURGHARDT, G. M., BARTMESS-LEVASSEUR, J. N., BROWNING, S. A., MORRISON, K. E., STEC, C. L., ZACHAU, C. E. & FREEBERG, T. M. (2012). Perspectives–minimizing observer bias in behavioral studies: A review and recommendations. *Ethology* **118**, 511-517.

BURT, J. M., CAMPBELL, S. E. & BEECHER, M. D. (2001). Song type matching as threat: a test using interactive playback. *Animal Behaviour* **62**, 1163-1170.

BYERS, B. E. & KROODSMA, D. E. (2009). Female mate choice and songbird song repertoires. *Animal Behaviour* **77**, 13-22.

CALDWELL, M. C., CALDWELL, D. K. &TYACK, P. L. (1990). Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin. In *The Bottlenose Dolphin* (eds. S. Leatherwood and R. R. Reeves), pp. 199-234. Academic Press, San Diego.

CALDWELL, M. C. (1965). Individualized whistle contours in bottle-nosed dolphins (*Tursiops truncatus*). *Nature* **207**, 434-435.

CALL, J. E. & TOMASELLO, M. E. (2007). *The gestural communication of apes and monkeys.* Taylor & Francis Group/Lawrence Erlbaum Associates.

CANE, V. R. (1959). Behaviour sequences as semi-Markov chains. *Journal of the Royal Statistical Society. Series B (Methodological)* **21**, 36-58.

CAPPÉ, O., MOULINES, E. & RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer Science Business Media, New York.

CARTER, G. G., SKOWRONSKI, M. D., FAURE, P. A. & FENTON, B. (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Animal Behaviour* **76**, 1343-1355.

CÄSAR, C., BYRNE, R. W., HOPPITT, W., YOUNG, R. J. & ZUBERBÜHLER, K. (2012a). Evidence for semantic communication in titi monkey alarm calls. *Animal Behaviour* **84**, 405-411.

CÄSAR, C., BYRNE, R., YOUNG, R. J. & ZUBERBÜHLER, K. (2012b). The alarm call system of wild black-fronted titi monkeys, *Callicebus nigrifrons*. *Behavioral Ecology and Sociobiology* **66**, 653-667.

CÄSAR, C., ZUBERBÜHLER, K., YOUNG, R. J. & BYRNE, R. W. (2013). Titi monkey call sequences vary with predator location and type. *Biology Letters* **9**, 20130535.

CATCHPOLE, C. K. & SLATER, P. J. B. (2003). *Bird song: biological themes and variations*. Cambridge Univ Press, Cambridge.

CATCHPOLE, C. K. (1976). Temporal and sequential organisation of song in the sedge warbler (*Acrocephalus schoenobaenus*). *Behaviour* **59**, 226-246.

CAZAU, D., ADAM, O., LAITMAN, J. T. & REIDENBERG, J. S. (2013). Understanding the intentional acoustic behavior of humpback whales: A production-based approach. *The Journal of the Acoustical Society of America* **134**, 2268-2273.

CEUGNIET, M. & AUBIN, T. (2001). The rally call recognition in males of two hybridizing partridge species, red-legged (*Alectoris rufa*) and rock (*A. graeca*) partridges. *Behavioural Processes* **55**, 1-12.

CHARRIER, I., AHONEN, H. & HARCOURT, R. G. (2011). What makes an Australian sea lion (*Neophoca cinerea*) male's bark threatening? *Journal of Comparative Psychology* **125**, 385.

CHARRIER, I. & HARCOURT, R. G. (2006). Individual vocal identity in mother and pup Australian sea lions (*Neophoca cinerea*). *Journal of Mammalogy* **87**, 929-938.

CHARRIER, I., LEE, T. T., BLOOMFIELD, L. L. & STURDY, C. B. (2005). Acoustic mechanisms of note-type perception in black-capped chickadee (*Poecile atricapillus*) calls. *Journal of Comparative Psychology* **119**, 371.

CHARRIER, I., MATHEVON, N. & AUBIN, T. (2013). Bearded seal males perceive geographic variation in their trills. *Behavioral Ecology and Sociobiology* **67**, 1679-1689.

CHARRIER, I., MATHEVON, N. & JOUVENTIN, P. (2003). Individuality in the voice of fur seal females: an analysis study of the pup attraction call in *Arctocephalus tropicalis*. *Marine Mammal Science* **19**, 161-172.

CHARRIER, I., PITCHER, B. J. & HARCOURT, R. G. (2009). Vocal recognition of mothers by Australian sea lion pups: individual signature and environmental constraints. *Animal Behaviour* **78**, 1127-1134.

CHATFIELD, C. & LEMON, R. E. (1970). Analysing sequences of behavioural events. *Journal of Theoretical Biology* **29**, 427-445.

CHATTERJEE, A. (2005). A madness to the methods in cognitive neuroscience? *Journal of Cognitive Neuroscience* **17**, 847-849.

CHEN, C. & FUSHING, H. (2012). Multiscale community geometry in a network and its application. *Physical Review E* **86**, 041120.

CHOLEWIAK, D. M., SOUSA-LIMA, R. S. & CERCHIO, S. (2012). Humpback whale song hierarchical structure: Historical context and discussion of current classification issues. *Marine Mammal Science* **29**, E312-E332.

CHOMSKY, N. (2002). *Syntactic structures, 9th edition*. de Gruyter Mouton, The Hague.

CHRISTIANSEN, M. H. & CHATER, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences* **5**, 82-88.

CLARK, A. (2010). Towards general algorithms for grammatical inference. *Lecture Notes in Computer Science* **6331**, 11-30.

CLARK, A., EYRAUD, R. & HABRARD, A. (2010). Using contextual representations to efficiently learn context-free languages. *The Journal of Machine Learning Research* **11**, 2707-2744.

CLARK, C. J. & FEO, T. J. (2010). Why do Calypte hummingbirds "sing" with both their tail and their syrinx? An apparent example of sexual sensory bias. *The American Naturalist* **175**, 27-37.

CLARK, C. W., MARLER, P. & BEEMAN, K. (1987). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology* **76**, 101-115.

CLARK, C. J. & FEO, T. J. (2008). The Anna's hummingbird chirps with its tail: a new mechanism of sonation in birds. *Proceedings of the Royal Society B: Biological Sciences* **275**, 955-962.

CLEMINS, P. J. & JOHNSON, M. T. (2006). Generalized perceptual linear prediction features for animal vocalization analysis. *The Journal of the Acoustical Society of America* **120**, 527-534.

COHN, L. (1995). *Time-Frequency Analysis: Theory and Applications, 1995*. Prentice Hall.

COLLIAS, N. E. (1987). The vocal repertoire of the red junglefowl: a spectrographic classification and the code of communication. *Condor* **89**, 510-524.

COVER, T. M. & THOMAS, J. A. (1991). *Elements of information theory*. John Wiley & Sons, Inc, New York, NY.

CROCKFORD, C. & BOESCH, C. (2005). Call combinations in wild chimpanzees. *Behaviour* **142**, 397-421.

CROFT, D. P., JAMES, R. & KRAUSE, J. (2008). *Exploring animal social networks*. Princeton University Press.

CSARDI, G. & NEPUSZ, T. (2006). The igraph software package for complex network research. *International Journal of Complex Systems* **1695**, 1-9.

CURE, C., AUBIN, T. & MATHEVON, N. (2011). Sex discrimination and mate recognition by voice in the Yelkouan shearwater *Puffinus yelkouan*. *Bioacoustics* **20**, 235-249.

CURÉ, C., MATHEVON, N., MUNDRY, R. & AUBIN, T. (2012). Acoustic cues used for species recognition can differ between sexes and sibling species: evidence in shearwaters. *Animal Behaviour* **84**, 239-250.

CYNX, J. (1990). Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *Journal of Comparative Psychology* **104**, 3-10.

CYNX, J., WILLIAMS, H. & NOTTEBOHM, F. (1990). Timbre discrimination in Zebra finch (*Taeniopygia guttata*) song syllables. *Journal of Comparative Psychology* **104**, 303-308.

DARLING, J. D., JONES, M. E. & NICKLIN, C. P. (2006). Humpback whale songs: Do they organize males during the breeding season? *Behaviour* **143**, 1051-1102.

DAWKINS, R. & KREBS, J. R. (1978). Animal signals: information or manipulation. *Behavioural ecology: An evolutionary approach* **2**, 282-309.

DEECKE, V. B. & JANIK, V. M. (2006). Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. *The Journal of the Acoustical Society of America* **119**, 645-653.

DENKER, M. & WOYCZYŃSKI, W. A. (1998). *Introductory Statistics and Random Phenomena: uncertainty, complexity and chaotic behaviour in engineering and science*. Springer, Boston.

DENT, M. L., BRITTAN-POWELL, E. F., DOOLING, R. J. & PIERCE, A. (1997). Perception of synthetic/ba/–/wa/speech continuum by budgerigars (*Melopsittacus undulatus*). *The Journal of the Acoustical Society of America* **102**, 1891-1897.

DENTRESSANGLE, F., AUBIN, T. & MATHEVON, N. (2012). Males use time whereas females prefer harmony: individual call recognition in the dimorphic blue-footed booby. *Animal Behaviour* **84**, 413-420.

DERÉGNAUCOURT, S., MITRA, P. P., FEHÉR, O., PYTTE, C. & TCHERNICHOVSKI, O. (2005). How sleep affects the developmental learning of bird song. *Nature* **433**, 710-716.

DESLANDES, V., FARIA, L. R., BORGES, M. E. & PIE, M. R. (2014). The structure of an avian syllable syntax network. *Behavioural processes* **106**, 53-59.

DOOLING, R. J., PARK, T. J., BROWN, S. D., OKANOYA, K. & SOLI, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals. *Journal of Comparative Psychology* **101**, 367-381.

DOUPE, A. J. & KONISHI, M. (1991). Song-selective auditory circuits in the vocal control system of the zebra finch. *Proceedings of the National Academy of Sciences* **88**, 11339-11343.

DUDA, R. O., HART, P. E. & STORK, D. G. (2012). *Pattern Classification*. John Wiley & Sons, New York.

DUNLOP, R. A., NOAD, M. J., CATO, D. H. & STOKES, D. (2007). The social vocalization repertoire of east Australian migrating humpback whales (*Megaptera novaeangliae*). *The Journal of the Acoustical Society of America* **122**, 2893-2905.

EENS, M. (1997). Understanding the complex song of the European starling: an integrated ethological approach. *Advances in the Study of Behavior* **26**, 355-434.

EENS, M., PINXTEN, R. & VERHEYEN, R. F. (1993). Function of the song and song repertoire in the European starling (*Sturnus vulgaris*): an aviary experiment. *Behaviour* **125**, 51-66.

EENS, M., PINXTEN, R. & VERHEYEN, R. F. (1991). Male song as a cue for mate choice in the European starling. *Behaviour* **116**, 210-238.

EENS, M., PINXTEN, R. & VERHEYEN, R. F. (1989). Temporal and sequential organization of song bouts in the starling. *Ardea* **77**, 75-86.

ESSER, K., CONDON, C. J., SUGA, N. & KANWAL, J. S. (1997). Syntax processing by auditory cortical neurons in the FM–FM area of the mustached bat*Pteronotus parnellii*. *Proceedings of the National Academy of Sciences* **94**, 14019-14024.

EVANS, C. S. (1997). Referential signals. *Perspectives in Ethology* **12**, 99-143.

EVANS, C. S., EVANS, L. & MARLER, P. (1993). On the meaning of alarm calls: functional reference in an avian vocal system. *Animal Behaviour* **46**, 23-38.

FAGIOLO, G. (2007). Clustering in complex directed networks. *Physical Review E* **76**, 026107.

FALLS, J. B. (1982). Individual recognition by sounds in birds. *Acoustic Communication in Birds* **2**, 237-278.

FARABAUGH, S. &DOOLING, R. (1996). Acoustic communication in parrots: laboratory and field studies of budgerigars, *Melopsittacus undulatus*. In *Ecology and Evolution of Acoustic Communication in Birds*, pp. 97-117. Cornell University Press Ithaca, New York, USA.

FERRER-I-CANCHO, R. (2005). Zipf's law from a communicative phase transition. *The European Physical Journal B-Condensed Matter and Complex Systems* **47**, 449-457.

FERRER-I-CANCHO, R. & MCCOWAN, B. (2012). The span of correlations in dolphin whistle sequences. *Journal of Statistical Mechanics: Theory and Experiment* **2012**, P06002.

FERRER-I-CANCHO, R., HERNÁNDEZ-FERNÁNDEZ, A., LUSSEAU, D., AGORAMOORTHY, G., HSU, M. J. & SEMPLE, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science* **37**, 1565-1578.

FEWELL, J. H. (2003). Social insect networks. *Science* **301**, 1867-1870.

FISCHER, J., NOSER, R. & HAMMERSCHMIDT, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *American Journal of Primatology* **75**, 643-663.

FITCH, W. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences* **4**, 258-267.

FITCH, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America* **102**, 1213-1222.

FORD, J. K. (1989). Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia. *Canadian Journal of Zoology* **67**, 727-745.

FORTUNATO, S. (2010). Community detection in graphs. *Physics Reports* **486**, 75-174.

FRED, A. L. & JAIN, A. K. (2005). Combining multiple clusterings using evidence accumulation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **27**, 835-850.

FREEBERG, T. M., DUNBAR, R. I. M. & ORD, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1785-1801.

FREEBERG, T. M. (2012). Geographic variation in note composition and use of chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *Ethology* **118**, 555-565.

FREEBERG, T. M. & LUCAS, J. R. (2012). Information theoretical approaches to chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *Journal of Comparative Psychology* **126**, 68-81.

FRUMHOFF, P. (1983). Aberrant songs of humpback whales (*Megaptera novaeangliae*): clues to the structure of humpback songs. In *Communication and behavior of whales* (ed. R. Payne), pp. 81-127. Westview Press, Boulder, Colorado.

FUSHING, H. & MCASSEY, M. P. (2010). Time, temperature, and data cloud geometry. *Physical Review E* **82**, 061110.

GALE, W. & SAMPSON, G. (1995). Good-Turing smoothing without tears. *Journal of Quantitative Linguistics* **2**, 217-237.

GAMMON, D. E. & ALTIZER, C. E. (2011). Northern Mockingbirds produce syntactical patterns of vocal mimicry that reflect taxonomy of imitated species. *Journal of Field Ornithology* **82**, 158-164.

GARDNER, T. J., NAEF, F. & NOTTEBOHM, F. (2005). Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science* **308**, 1046-1049.

GARLAND, E. C., GOLDIZEN, A. W., REKDAHL, M. L., CONSTANTINE, R., GARRIGUE, C., HAUSER, N. D., POOLE, M. M., ROBBINS, J. & NOAD, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current Biology* **21**, 687-691.

GARLAND, E. C., LILLEY, M. S., GOLDIZEN, A. W., REKDAHL, M. L., GARRIGUE, C. & NOAD, M. J. (2012). Improved versions of the Levenshtein distance method for comparing sequence information in animals' vocalisations: tests using humpback whale song. *Behaviour* **149**, 1413-1441.

GARLAND, E. C., NOAD, M. J., GOLDIZEN, A. W., LILLEY, M. S., REKDAHL, M. L., GARRIGUE, C., CONSTANTINE, R., HAUSER, N. D., POOLE, M. M. & ROBBINS, J. (2013). Quantifying humpback whale song sequences to understand the dynamics of song exchange at the ocean basin scale. *The Journal of the Acoustical Society of America* **133**, 560-569.

GENTNER, T. Q. & HULSE, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour* **56**, 579-594.

GENTNER, T. Q., FENN, K. M., MARGOLIASH, D. & NUSBAUM, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204-1207.

GERHARDT, H. C. (2001). Acoustic communication in two groups of closely related treefrogs. *Advances in the Study of Behavior* **30**, 99-167.

GERHARDT, H. C. & HUBER, F. (2002). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. University of Chicago Press.

GHAZANFAR, A. A., FLOMBAUM, J. I., MILLER, C. T. & HAUSER, M. D. (2001). The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): playback experiments with long calls. *Journal of Comparative Physiology A* **187**, 27-35.

GIRGENRATH, M. & MARSH, R. (1997). In vivo performance of trunk muscles in tree frogs during calling. *Journal of Experimental Biology* **200**, 3101-3108.

GOLD, E. M. (1967). Language identification in the limit. *Information and control* **10**, 447-474.

GOLDBOGEN, J. A., SOUTHALL, B. L., DERUITER, S. L., CALAMBOKIDIS, J., FRIEDLAENDER, A. S., HAZEN, E. L., FALCONE, E. A., SCHORR, G. S., DOUGLAS, A. & MORETTI, D. J. (2013). Blue whales respond to simulated mid-frequency military sonar. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130657.

GOUZOULES, S., GOUZOULES, H. & MARLER, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour* **32**, 182-193.

GREEN, S. R., MERCADO III, E., PACK, A. A. & HERMAN, L. M. (2011). Recurring patterns in the songs of humpback whales (*Megaptera novaeangliae*). *Behavioural Processes* **86**, 284-294.

GRIEVES, L., LOGUE, D. & QUINN, J. (2014). Joint-nesting smooth-billed anis, *Crotophaga ani*, use a functionally referential alarm call system. *Animal Behaviour* **89**, 215-221.

GRIFFITHS, T. D. & WARREN, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience* **5**, 887-892.

GWILLIAM, J., CHARRIER, I. & HARCOURT, R. G. (2008). Vocal identity and species recognition in male Australian sea lions,*Neophoca cinerea*. *The Journal of Experimental Biology* **211**, 2288-2295.

HAHNLOSER, R. H., KOZHEVNIKOV, A. A. & FEE, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65-70.

HAILMAN, J. P. (2008). *Coding and Redundancy: Man-Made and Animal-Evolved Signals*. Harvard University Press.

HAMAKER, J., GANAPATHIRAJU, A. & PICONE, J. (1998). Information theoretic approaches to model selection. *Proceedings of the International Conference on Spoken Language Processing* **7**, 2931-2934.

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M. & MORRIS, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software* **24**, 1548-1557.

HANDEL, S., TODD, S. K. & ZOIDIS, A. M. (2012). Hierarchical and rhythmic organization in the songs of humpback whales (*Megaptera novaeangliae*). *Bioacoustics* **21**, 141-156.

HANDEL, S., TODD, S. K. & ZOIDIS, A. M. (2009). Rhythmic structure in humpback whale (*Megaptera novaeangliae*) songs: Preliminary implications for song production and perception. *The Journal of the Acoustical Society of America* **125**, EL225-EL230.

HARLEY, H. E. (2008). Whistle discrimination and categorization by the Atlantic bottlenose dolphin (*Tursiops truncatus*): A review of the signature whistle framework and a perceptual test. *Behavioural processes* **77**, 243-268.

HARNAD, S. R. (1990). *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press.

HARRINGTON, F. H., ASA, C. S., MECH, L. &BOITANI, L. (2003). Wolf communication. In *Wolves: Behavior, ecology, and conservation*, pp. 66-103. University of Chicago Press, Chicago.

HARTLEY, R. S. & SUTHERS, R. A. (1989). Airflow and pressure during canary song: direct evidence for mini-breaths. *Journal of Comparative Physiology A* **165**, 15-26.

HARTSHORNE, C. (1973). *Born to sing. An interpretation and world survey of bird songs*. London: Indiana University Press.

HARTSHORNE, C. (1956). The monotony-threshold in singing birds. *The Auk* **73**, 176-192.

HASSELQUIST, D., BENSCH, S. & VON SCHANTZ, T. (1996). Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229-232.

HAUSER, M. D., CHOMSKY, N. & FITCH, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569-1579.

HAUSER, M. D., AGNETTA, B. & PEREZ, C. (1998). Orienting asymmetries in rhesus monkeys: the effect of time-domain changes on acoustic perception. *Animal Behaviour* **56**, 41-47.

HAUSSER, J. & STRIMMER, K. (2009). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *The Journal of Machine Learning Research* **10**, 1469-1484.

HAYES, S. A., KUMAR, A., COSTA, D. P., MELLINGER, D. K., HARVEY, J. T., SOUTHALL, B. L. & LE BOEUF, B. J. (2004). Evaluating the function of the male harbour seal, *Phoca vitulina*, roar through playback experiments. *Animal Behaviour* **67**, 1133-1139.

HEBETS, E. A. & PAPAJ, D. R. (2005). Complex signal function: developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology* **57**, 197-214.

HELBLE, T. A., IERLEY, G. R., GERALD, L., ROCH, M. A. & HILDEBRAND, J. A. (2012). A generalized power-law detection algorithm for humpback whale vocalizations. *The Journal of the Acoustical Society of America* **131**, 2682-2699.

HENDERSON, E. E., HILDEBRAND, J. A. & SMITH, M. H. (2011). Classification of behavior using vocalizations of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*). *The Journal of the Acoustical Society of America* **130**, 557-567.

HENRY, K. S., GALL, M. D., BIDELMAN, G. M. & LUCAS, J. R. (2011). Songbirds tradeoff auditory frequency resolution and temporal resolution. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* **197**, 351-359.

HENRY, K. S. & LUCAS, J. R. (2010). Habitat‐related differences in the frequency selectivity of auditory filters in songbirds. *Functional Ecology* **24**, 614-624.

HERZING, D. L. (1996). Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*. *Aquatic Mammals* **22**, 61-80.

HOLVECK, M., DE CASTRO, ANA CATARINA VIEIRA, LACHLAN, R. F., TEN CATE, C. & RIEBEL, K. (2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behavioral Ecology* **19**, 1267-1281.

HORNING, C. L., BEECHER, M. D., STODDARD, P. K. & CAMPBELL, S. E. (1993). Song perception in the song sparrow: importance of different parts of the song in song type classification. *Ethology* **94**, 46-58.

HULSE, S. H. (2002). Auditory scene analysis in animal communication. *Advances in the Study of Behavior* **31**, 163-200.

HUMPHRIES, M. D. & GURNEY, K. (2008). Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence. *PLoS One* **3**, e0002051.

ILANY, A., BAROCAS, A., KOREN, L., KAM, M. & GEFFEN, E. (2011). Do singing rock hyraxes exploit conspecific calls to gain attention? *PloS One* **6**, e28612.

ILANY, A., BAROCAS, A., KAM, M., ILANY, T. & GEFFEN, E. (2013). The energy cost of singing in wild rock hyrax males: evidence for an index signal. *Animal Behaviour* **85**, 995-1001.

IRWIN, D. E. (2000). Song variation in an avian ring species. *Evolution* **54**, 998-1010.

ISAAC, D. & MARLER, P. (1963). Ordering of sequences of singing behaviour of mistle thrushes in relationship to timing. *Animal Behaviour* **11**, 179-188.

JANIK, V. M. (1999). Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods. *Animal Behaviour* **57**, 133-143.

JANIK, V. M., SAYIGH, L. & WELLS, R. (2006). Signature whistle shape conveys identity information to bottlenose dolphins. *Proceedings of the National Academy of Sciences* **103**, 8293-8297.

JANIK, V. M., KING, S. L., SAYIGH, L. S. & WELLS, R. S. (2013). Identifying signature whistles from recordings of groups of unrestrained bottlenose dolphins (*Tursiops truncatus*). *Marine Mammal Science* **29**, 109-122.

JANSEN, D. A., CANT, M. A. & MANSER, M. B. (2012). Segmental concatenation of individual signatures and con-text cues in banded mongoose (*Mungos mungo*) close calls. *BMC biology* **10**, 97.

JI, A., JOHNSON, M. T., WALSH, E. J., MCGEE, J. & ARMSTRONG, D. L. (2013). Discrimination of individual tigers (*Panthera tigris*) from long distance roars. *The Journal of the Acoustical Society of America* **133**, 1762-1769.

JIN, D. Z. & KOZHEVNIKOV, A. A. (2011). A compact statistical model of the song syntax in Bengalese finch. *PLoS Computational Biology* **7**, e1001108.

JIN, D. Z. (2009). Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. *Physical Review E* **80**, 051902.

JOHNSON, K. (2004). Gold's Theorem and cognitive science. *Philosophy of Science* **71**, 571-592.

JONES, A. E., TEN CATE, C. & BIJLEVELD, C. C. (2001). The interobserver reliability of scoring sonagrams by eye: a study on methods, illustrated on zebra finch songs. *Animal Behaviour* **62**, 791-801.

JOUVENTIN, P., AUBIN, T. & LENGAGNE, T. (1999). Finding a parent in a king penguin colony: the acoustic system of individual recognition. *Animal Behaviour* **57**, 1175-1183.

JURAFSKY, D. & MARTIN, J. H. (2000). *Speech & language processing*. Pearson Education India.

KAKISHITA, Y., SASAHARA, K., NISHINO, T., TAKAHASI, M. & OKANOYA, K. (2009). Ethological data mining: an automata-based approach to extract behavioral units and rules. *Data Mining and Knowledge Discovery* **18**, 446-471.

KANWAL, J. S., MATSUMURA, S., OHLEMILLER, K. & SUGA, N. (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *The Journal of the Acoustical Society of America* **96**, 1229-1254.

KAPLAN, D. (1978). Dthat. *Syntax and semantics* **9**, 221-243.

KASHTAN, N., ITZKOVITZ, S., MILO, R. & ALON, U. (2002). Mfinder tool guide. *Department of Molecular Cell Biology and Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot Israel, Tech.Rep* .

KATAHIRA, K., SUZUKI, K., OKANOYA, K. & OKADA, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PloS One* **6**, e24516.

KATZ, S. (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **35**, 400-401.

KERSHENBAUM, A. (2013). Entropy rate as a measure of animal vocal complexity. *Bioacoustics* .

KERSHENBAUM, A., ILANY, A., BLAUSTEIN, L. & GEFFEN, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2974-2981.

KERSHENBAUM, A. & ROCH, M. A. (2013). An image processing based paradigm for the extraction of tonal sounds in cetacean communications. *Journal of the Acoustical Society of America* **134**, 4435-4445.

KERSHENBAUM, A., SAYIGH, L. S. & JANIK, V. M. (2013). The encoding of individual identity in dolphin signature whistles: how much information is needed? *PLoS One* **8**, e77671.

KIPPER, S., MUNDRY, R., SOMMER, C., HULTSCH, H. & TODT, D. (2006). Song repertoire size is correlated with body measures and arrival date in common nightingales, *Luscinia megarhynchos*. *Animal Behaviour* **71**, 211-217.

KITCHEN, D. M. (2004). Alpha male black howler monkey responses to loud calls: effect of numeric odds, male companion behaviour and reproductive investment. *Animal Behaviour* **67**, 125-139.

KLINCK, H., KINDERMANN, L. & BOEBEL, O. (2008). Detection of leopard seal (*Hydrurga leptonyx*) vocalizations using the Envelope-Spectrogram Technique (TEST) in combination with a Hidden Markov model. *Canadian Acoustics* **36**, 118-124.

KOGAN, J. A. & MARGOLIASH, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *The Journal of the Acoustical Society of America* **103**, 2185-2196.

KOLODNY, O., LOTEM, A. & EDELMAN, S. (in press). Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cognitive Science* .

KOREN, L. & GEFFEN, E. (2011). Individual identity is communicated through multiple pathways in male rock hyrax (*Procavia capensis*) songs. *Behavioral Ecology and Sociobiology* **65**, 675-684.

KOREN, L. & GEFFEN, E. (2009). Complex call in male rock hyrax (*Procavia capensis*): a multi-information distributing channel. *Behavioral Ecology and Sociobiology* **63**, 581-590.

KRAMS, I., KRAMA, T., FREEBERG, T. M., KULLBERG, C. & LUCAS, J. R. (2012). Linking social complexity and vocal complexity: a parid perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1879-1891.

KREBS, J. R., ASHCROFT, R. & ORSDOL, K. V. (1981). Song matching in the great tit*Parus major* L. *Animal Behaviour* **29**, 918-923.

KREBS, J. (1976). Habituation and song repertoires in the great tit. *Behavioral Ecology and Sociobiology* **1**, 215-227.

KROODSMA, D. E. (1989). Suggested experimental designs for song playbacks. *Animal Behaviour* **37**, 600-609.

KROODSMA, D. E. (1985). Development and use of two song forms by the Eastern Phoebe. *The Wilson Bulletin* **97**, 21-29.

KROODSMA, D. E. (1990). Patterns in songbird singing behaviour: Hartshorne vindicated. *Animal Behaviour* **39**, 994-996.

LACHLAN, R. F., VERZIJDEN, M. N., BERNARD, C. S., JONKER, P., KOESE, B., JAARSMA, S., SPOOR, W., SLATER, P. J. & TEN CATE, C. (2013). The progressive loss of syntactical structure in bird song along an island colonization chain. *Current Biology* **19**, 1896-1901.

LAMBRECHTS, M. & DHONDT, A. (1987). Differences in singing performance between male great tits. *Ardea* **75**, 43-52.

LAMBRECHTS, M. & DHONDT, A. A. (1988). The anti-exhaustion hypothesis: a new hypothesis to explain song performance and song switching in the great tit. *Animal Behaviour* **36**, 327-334.

LAMMERS, M. O. & AU, W. W. (2003). Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement? *Marine Mammal Science* **19**, 249-264.

LARSON, K. A. (2004). Advertisement call complexity in northern leopard frogs, *Rana pipiens*. *Journal Information* **2004**.

LEE, K. & HON, H. (1989). Speaker-independent phone recognition using hidden Markov models. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **37**, 1641-1648.

LEHONGRE, K., AUBIN, T., ROBIN, S. & DEL NEGRO, C. (2008). Individual signature in canary songs: contribution of multiple levels of song structure. *Ethology* **114**, 425-435.

LEMASSON, A., OUATTARA, K., BOUCHET, H. & ZUBERBÜHLER, K. (2010). Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften* **97**, 1023-1027.

LENGAGNE, T., AUBIN, T., JOUVENTIN, P. & LAUGA, J. (2000). Perceptual salience of individually distinctive features in the calls of adult king penguins. *The Journal of the Acoustical Society of America* **107**, 508-516.

LENGAGNE, T., LAUGA, J. & AUBIN, T. (2001). Intra-syllabic acoustic signatures used by the king penguin in parent-chick recognition: an experimental approach. *The Journal of Experimental Biology* **204**, 663-672.

LIPKIND, D., MARCUS, G. F., BEMIS, D. K., SASAHARA, K., JACOBY, N., TAKAHASI, M., SUZUKI, K., FEHER, O., RAVBAR, P. & OKANOYA, K. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* **498**, 104-108.

LIU, I. A., LOHR, B., OLSEN, B. & GREENBERG, R. (2008). Macrogeographic vocal variation in subspecies of swamp sparrow. *The Condor* **110**, 102-109.

LOHR, B., ASHBY, S. & WAKAMIYA, S. M. (2013). The function of song types and song components in Grasshopper Sparrows (*Ammodramus savannarum*). *Behaviour* **150**, 1085-1106.

LYNCH, K. S., STANELY RAND, A., RYAN, M. J. & WILCZYNSKI, W. (2005). Plasticity in female mate choice associated with changing reproductive states. *Animal Behaviour* **69**, 689-699.

1848 MACEDONIA, J. M. (1990). What is communicated in the antipredator calls of lemurs: evidence from playback experiments with ringtailed and ruffed lemurs. *Ethology* **86**, 177-190.

1850 MAHURIN, E. J. & FREEBERG, T. M. (2009). Chick-a-dee call variation in Carolina chickadees and recruiting flockmates to food. *Behavioral Ecology* **20**, 111-116.

1852 MAKEIG, S., DEBENER, S., ONTON, J. & DELORME, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences* **8**, 204-210.

1854 MALLAT, S. (1999). *A Wavelet Tour of Signal Processing*. Access Online via Elsevier.

1855 MANNING, C. D., RAGHAVAN, P. & SCHÜTZE, H. (2008). *Introduction to information retrieval*. Cambridge University Press, Cambridge.

1857 MANSER, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**, 2315-2324.

1860 MARLER, P. (1977). The structure of animal communication sounds. In *Recognition of Complex Acoustic Signals* (ed. T. H. Bullock), pp. 17-35. Springer Verlag, Berlin.

1862 MARLER, P. & PICKERT, R. (1984). Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Animal Behaviour* **32**, 673-689.

1864 MARLER, P. & SHERMAN, V. (1985). Innate differences in singing behaviour of sparrows reared in isolation from adult conspecific song. *Animal Behaviour* **33**, 57-71.

1866 MARTIN, W. F. & GANS, C. (1972). Muscular control of the vocal tract during release signaling in the toad *Bufo valliceps*. *Journal of Morphology* **137**, 1-27.

1868 MARTIN, W. P. (1972). Evolution of vocalization in the genus Bufo. In *Evolution in the GenusBufo* (ed. W. F. Blair), pp. 37-70. University of Texas Press, Austin, TX.

1870 MARTIN, W. & FLANDRIN, P. (1985). Wigner-Ville spectral analysis of nonstationary processes. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **33**, 1461-1470.

1872 MARTINICH, A. & SOSA, D. (2013). *The Philosophy of Language*. Oxford University Press, Oxford.

1873 MATHEVON, N., KORALEK, A., WELDELE, M., GLICKMAN, S. E. & THEUNISSEN, F. E. (2010). What the hyena's laugh tells: Sex, age, dominance and individual signature in the giggling call of *Crocuta crocuta*. *BMC Ecology* **10**, 9.

1876 MATHEVON, N. & AUBIN, T. (2001). Sound-based species-specific recognition in the blackcap*Sylvia atricapilla*shows high tolerance to signal modifications. *Behaviour* **138**, 511-524.

1878 MATHEVON, N., AUBIN, T., VIELLIARD, J., DA SILVA, M., SEBE, F. & BOSCOLO, D. (2008). Singing in the rain forest: how a tropical bird song transfers information. *PLoS One* **3**, e1580.

MATTHEWS, J., RENDELL, L., GORDON, J. & MACDONALD, D. (1999). A review of frequency and time parameters of cetacean tonal calls. *Bioacoustics* **10**, 47-71.

MCCOWAN, B., DOYLE, L. R. & HANSER, S. F. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology* **116**, 166-172.

MCCOWAN, B., HANSER, S. F. & DOYLE, L. R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour* **57**, 409-419.

MCCOWAN, B. & REISS, D. (1995). Quantitative Comparison of whistle repertoires from captive adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): a re-evaluation of the signature whistle hypothesis. *Ethology* **100**, 194-209.

MCSHEA, D. W. (2009). The evolution of complexity without natural selection, a possible large-scale trend of the fourth kind.

MCSHEA, D. W. (1991). Complexity and evolution: what everybody knows. *Biology and Philosophy* **6**, 303-324.

MENNILL, D. J. & RATCLIFFE, L. M. (2004). Overlapping and matching in the song contests of black-capped chickadees. *Animal Behaviour* **67**, 441-450.

MENNILL, D. J., RATCLIFFE, L. M. & BOAG, P. T. (2002). Female eavesdropping on male song contests in songbirds. *Science* **296**, 873-873.

MERCADO, E. I., HERMAN, L. M. & PACK, A. A. (2003). Stereotypical sound patterns in humpback whale songs: Usage and function. *Aquatic Mammals* **29**, 37-52.

MERCADO, E. I. & HANDEL, S. (2012). Understanding the structure of humpback whale songs (L). *The Journal of the Acoustical Society of America* **132**, 2947-2950.

MERCADO, E. I., SCHNEIDER, J. N., PACK, A. A. & HERMAN, L. M. (2010). Sound production by singing humpback whales. *The Journal of the Acoustical Society of America* **127**, 2678-2691.

MILLER, C. T. & BEE, M. A. (2012). Receiver psychology turns 20: is it time for a broader approach? *Animal Behaviour* **83**, 331-343.

MILLER, C. T. &COHEN, Y. E. (2010). Vocalizations as auditory objects: behavior and neurophysiology. In *Primate Neuroethology* (eds. M. L. Platt and A. A. Ghazanfar), pp. 237-255. Oxford University Press, Oxford.

MILLER, P. J., SAMARRA, F. I. & PERTHUISON, A. D. (2007). Caller sex and orientation influence spectral characteristics of “two-voice” stereotyped calls produced by free-ranging killer whales. *The Journal of the Acoustical Society of America* **121**, 3932-3937.

MILLIKAN, R. G. (2004). *Varieties of meaning: the 2002 Jean Nicod lectures*. MIT press.

MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. & ALON, U. (2002). Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827.

MITANI, J. C., HUNLEY, K. & MURDOCH, M. (1999). Geographic variation in the calls of wild chimpanzees: a reassessment. *American Journal of Primatology* **47**, 133-151.

MITCHELL, M. (2009). *Complexity: A guided tour*. Oxford University Press.

MOORE, B. C. & MOORE, B. C. (2003). *An Introduction to the Psychology of Hearing*. Academic Press, San Diego.

MULLIGAN, J. A. (1966). *Singing Behavior and its Development in the Song Sparrow Melospiza melodia*. University of California Press.

MUMFORD, D. & DESOLNEUX, A. (2010). *Pattern Theory: The Stochastic Analysis of Real-World Signals*. A. K. Peters Ltd.

MUNOZ, N. E. & BLUMSTEIN, D. T. (2012). Multisensory perception in uncertain environments. *Behavioral Ecology* **23**, 457-462.

MURRAY, S. O., MERCADO, E. I. & ROITBLAT, H. L. (1998). Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations. *The Journal of the Acoustical Society of America* **104**, 1679-1688.

NARINS, P. M., LEWIS, E. R. & MCCLELLAND, B. E. (2000). Hyperextended call note repertoire of the endemic Madagascar treefrog *Boophis madagascariensis* (Rhacophoridae). *Journal of Zoology* **250**, 283-298.

NARINS, P. M., REICHMAN, O., JARVIS, J. U. & LEWIS, E. R. (1992). Seismic signal transmission between burrows of the Cape mole-rat, *Georychus capensis*. *Journal of Comparative Physiology A* **170**, 13-21.

NELSON, D. A. & MARLER, P. (1989). Categorical perception of a natural stimulus continuum: birdsong. *Science* **244**, 976-978.

NELSON, D. A. & POESEL, A. (2007). Segregation of information in a complex acoustic signal: individual and dialect identity in white-crowned sparrow song. *Animal Behaviour* **74**, 1073-1084.

NETTLE, D. (1999). Language variation and the evolution of societies. In *The Evolution of Culture: An Interdisciplinary View* (eds. R. I. M. Dunbar, C. Knight and C. Power), pp. 214-227. Rutgers University Press.

NEUBAUER, R. L. (1999). Super-normal length song preferences of female zebra finches (*Taeniopygia guttata*) and a theory of the evolution of bird song. *Evolutionary Ecology* **13**, 365-380.

NEVO, E., HETH, G., BEILES, A. & FRANKENBERG, E. (1987). Geographic dialects in blind mole rats: role of vocal communication in active speciation. *Proceedings of the National Academy of Sciences* **84**, 3312-3315.

NEWMAN, M. (2009). *Networks: An Introduction*. Oxford University Press, Oxford.

NEWMAN, M. E. (2003). The structure and function of complex networks. *SIAM Review* **45**, 167-256.

NITYANANDA, V. & BEE, M. A. (2011). Finding your mate at a cocktail party: frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. *PloS One* **6**, e21191.

NOTMAN, H. & RENDALL, D. (2005). Contextual variation in chimpanzee pant hoots and its implications for referential communication. *Animal Behaviour* **70**, 177-190.

NOWICKI, S. & MARLER, P. (1988). How do birds sing? *Music Perception* , 391-426.

NOWICKI, S. & NELSON, D. A. (1990). Defining natural categories in acoustic signals: comparison of three methods applied to ‘chick-a-dee’ call notes. *Ethology* **86**, 89-101.

OHMS, V. R., ESCUDERO, P., LAMMERS, K. & TEN CATE, C. (2012). Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception. *Animal Cognition* **15**, 155-161.

OLESON, E. M., WIGGINS, S. M. & HILDEBRAND, J. A. (2007). Temporal separation of blue whale call types on a southern California feeding ground. *Animal Behaviour* **74**, 881-894.

OPPENHEIM, A. V. & SCHAFER, R. W. (2004). From frequency to quefrency: A history of the cepstrum. *Signal Processing Magazine, IEEE* **21**, 95-106.

OPSAHL, T., AGNEESSENS, F. & SKVORETZ, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* **32**, 245-251.

OSWALD, J. N., RANKIN, S., BARLOW, J. & LAMMERS, M. O. (2007). A tool for real-time acoustic species identification of delphinid whistles. *The Journal of the Acoustical Society of America* **122**, 587-595.

OTTER, K., MCGREGOR, P. K., TERRY, A. M., BURFORD, F. R., PEAKE, T. M. & DABELSTEEN, T. (1999). Do female great tits (*Parus major*) assess males by eavesdropping? A field study using interactive song playback. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 1305-1309.

OWREN, M. J., RENDALL, D. & RYAN, M. J. (2010). Redefining animal signaling: influence versus information in communication. *Biology & Philosophy* **25**, 755-780.

PACK, A. A., HERMAN, L. M., HOFFMANN-KUHNT, M. & BRANSTETTER, B. K. (2002). The object behind the echo: dolphins (*Tursiops truncatus*) perceive object shape globally through echolocation. *Behavioural Processes* **58**, 1-26.

PAGE, S. E. (2010). *Diversity and complexity*. Princeton University Press.

PAREJO, D., AVILES, J. M. & RODRIGUEZ, J. (2012). Alarm calls modulate the spatial structure of a breeding owl community. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2135-2141.

PARSONS, S., RISKIN, D. K. & HERMANSON, J. W. (2010). Echolocation call production during aerial and terrestrial locomotion by New Zealand's enigmatic lesser short-tailed bat, *Mystacina tuberculata*. *The Journal of Experimental Biology* **213**, 551-557.

PARTAN, S. & MARLER, P. (1999). Communication goes multimodal. *Science* **283**, 1272-1273.

PAULUS, J., MÜLLER, M. & KLAPURI, A. (2010). State of the Art Report: Audio-Based Music Structure Analysis. *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* , 625-636.

PAYNE, K. B., LANGBAUER JR, W. R. & THOMAS, E. M. (1986). Infrasonic calls of the Asian elephant (*Elephas maximus*). *Behavioral Ecology and Sociobiology* **18**, 297-301.

PAYNE, K., TYACK, P. & PAYNE, R. (1983). Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): a detailed analysis of two seasons in Hawaii. *Communication and behavior of whales* , 9-57.

PAYNE, R. S. & MCVAY, S. (1971). Songs of humpback whales. *Science* **173**, 585-597.

PETERSON, G. E. & BARNEY, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* **24**, 175-184.

PICONE, J. W. (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE* **81**, 1215-1247.

PITCHER, B. J., HARCOURT, R. G. & CHARRIER, I. (2012). Individual identity encoding and environmental constraints in vocal recognition of pups by Australian sea lion mothers. *Animal Behaviour* **83**, 681-690.

PODOS, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution* , 537-551.

PODOS, J., PETERS, S., RUDNICKY, T., MARLER, P. & NOWICKI, S. (1992). The organization of song repertoires in song sparrows: themes and variations. *Ethology* **90**, 89-106.

POLLARD, K. A. & BLUMSTEIN, D. T. (2012). Evolving communicative complexity: insights from rodents and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1869-1878.

POLLARD, K. A. & BLUMSTEIN, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology* **21**, 413-417.

POOLE, J. H. (1999). Signals and assessment in African elephants: evidence from playback experiments. *Animal Behaviour* **58**, 185-193.

PUTNAM, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science* **7**, 131-193.

2014 QUICK, N. J. & JANIK, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2539-2545.

2016 QUINE, W. V. O. (1960). *Word and object*. MIT press.

2017 R DEVELOPMENT TEAM. (2012)*R: A Language and Environment for Statistical Computing.R Foundation for Statistical Computing, Vienna, Austria, 2007* .

2019 RABINER, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257-286.

2021 RAEMAEKERS, J. J., RAEMAEKERS, P. M. & HAIMOFF, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*): repertoire, organisation and context. *Behaviour* **91**, 146-189.

2023 RAGGE, D. & REYNOLDS, W. (1988). The songs and taxonomy of the grasshoppers of the *Chorthippus biguttulus* group in the Iberian Peninsula (Orthoptera: Acrididae). *Journal of Natural History* **22**, 897-929.

2026 RANDALL, J. A. (1997). Species-specific footdrumming in kangaroo rats: *Dipodomys ingens*, *D. deserti*, *D. spectabilis*. *Animal Behaviour* **54**, 1167-1175.

2028 RANDALL, J. A. (1989). Individual footdrumming signatures in banner-tailed kangaroo rats *Dipodomys spectabilis*. *Animal Behaviour* **38**, 620-630.

2030 RECANZONE, G. H. & SUTTER, M. L. (2008). The biological basis of audition. *Annu.Rev.Psychol.* **59**, 119-142.

2032 REMEZ, R. E., RUBIN, P. E., BERNS, S. M., PARDO, J. S. & LANG, J. M. (1994). On the perceptual organization of speech. *Psychological Review* **101**, 129.

2034 RIBEIRO, S., CECCHI, G. A., MAGNASCO, M. O. & MELLO, C. V. (1998). Toward a song code: evidence for a syllabic representation in the canary brain. *Neuron* **21**, 359-371.

2036 RICHARDS, D. G. (1981). Alerting and message components in songs of rufous-sided towhees. *Behaviour* **76**, 223-249.

2038 RIEBEL, K. & SLATER, P. (2003). Temporal variation in male chaffinch song depends on the singer and the song type. *Behaviour* **140**, 269-288.

2040 RIEDE, T., BRONSON, E., HATZIKIROU, H. & ZUBERBÜHLER, K. (2005). Vocal production mechanisms in a non-human primate: morphological data and a model. *Journal of Human Evolution* **48**, 85-96.

2043 RIEDE, T. & FITCH, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *The Journal of Experimental Biology* **202**, 2859-2867.

2045 RIPLEY, B. D. (2007). *Pattern recognition and neural networks*. Cambridge University Press.

ROBINSON, J. G. (1979). An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Zeitschrift für Tierpsychologie* **49**, 381-405.

ROBISSON, P., AUBIN, T. & BREMOND, J. (1993). Individuality in the voice of the emperor penguin*Aptenodytes forsteri*: adaptation to a noisy environment. *Ethology* **94**, 279-290.

ROOT-GUTTERIDGE, H., BENCSIK, M., CHEBLI, M., GENTLE, L. K., TERRELL-NIELD, C., BOURIT, A. & YARNELL, R. W. (2014). Identifying individual wild Eastern grey wolves (*Canis lupus lycaon*) using fundamental frequency and amplitude of howls. *Bioacoustics* **23**, 55-66.

ROTHENBERG, D., ROESKE, T. C., VOSS, H. U., NAGUIB, M. & TCHERNICHOVSKI, O. (2013). Investigation of musicality in birdsong. *Hearing Research* **308**, 71-83.

RUXTON, G. D. & SCHAEFER, H. M. (2011). Resolving current disagreements and ambiguities in the terminology of animal communication. *Journal of Evolutionary Biology* **24**, 2574-2585.

SAAR, S. & MITRA, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *PLoS One* **3**, e1461.

SALMI, R., HAMMERSCHMIDT, K. & DORAN-SHEEHY, D. M. (2013). Western Gorilla Vocal Repertoire and Contextual Use of Vocalizations. *Ethology* **119**, 831-847.

SASAHARA, K., CODY, M. L., COHEN, D. & TAYLOR, C. E. (2012). Structural design principles of complex bird songs: a network-based approach. *PloS One* **7**, e44436.

SAUNDERS, A. A. (1951). The song of the Song Sparrow. *The Wilson Bulletin* , 99-109.

SAYIGH, L., QUICK, N., HASTIE, G. & TYACK, P. (2012). Repeated call types in short-finned pilot whales, *Globicephala macrorhynchus*. *Marine Mammal Science* **29**, 312-324.

SAYIGH, L. S., ESCH, H. C., WELLS, R. S. & JANIK, V. M. (2007). Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*. *Animal Behaviour* **74**, 1631-1642.

SAYIGH, L. S., TYACK, P. L., WELLS, R. S., SOLOW, A. R., SCOTT, M. D. & IRVINE, A. B. (1999). Individual recognition in wild bottlenose dolphins: a field test using playback experiments. *Animal Behaviour* **57**, 41-50.

SCHEL, A. M., CANDIOTTI, A. & ZUBERBÜHLER, K. (2010). Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Animal Behaviour* **80**, 799-808.

SCHEL, A. M., TRANQUILLI, S. & ZUBERBÜHLER, K. (2009). The alarm call system of two species of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *Journal of Comparative Psychology* **123**, 136-150.

SCHREIBER, F. & SCHWÖBBERMEYER, H. (2005). MAVisto: a tool for the exploration of network motifs. *Bioinformatics* **21**, 3572-3574.

SCHULZ, T. M., WHITEHEAD, H., GERO, S. & RENDELL, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Animal Behaviour* **76**, 1977-1988.

SCOTT, J. & CARRINGTON, P. J. (2011). *The SAGE Handbook of Social Network Analysis*. SAGE Publications.

SEARCY, W. A. & ANDERSSON, M. (1986). Sexual selection and the evolution of song. *Annual Review of Ecology and Systematics* , 507-533.

SEARCY, W. A. & BEECHER, M. D. (2011). Continued scepticism that song overlapping is a signal. *Animal Behaviour* **81**, e1-e4.

SEARCY, W. A. & NOWICKI, S. (2005). *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press.

SEARCY, W. A., NOWICKI, S. & PETERS, S. (1999). Song types as fundamental units in vocal repertoires. *Animal Behaviour* **58**, 37-44.

SEARCY, W. A., PODOS, J., PETERS, S. & NOWICKI, S. (1995). Discrimination of song types and variants in song sparrows. *Animal Behaviour* **49**, 1219-1226.

SEDDON, N. & TOBIAS, J. A. (2007). Song divergence at the edge of Amazonia: an empirical test of the peripatric speciation model. *Biological Journal of the Linnean Society* **90**, 173-188.

SEYFARTH, R. M. & CHENEY, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language* **115**, 92-100.

SEYFARTH, R. M., CHENEY, D. L. & BERGMAN, T. J. (2005). Primate social cognition and the origins of language. *Trends in Cognitive Sciences* **9**, 264-266.

SEYFARTH, R. M., CHENEY, D. L., BERGMAN, T., FISCHER, J., ZUBERBÜHLER, K. & HAMMERSCHMIDT, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour* **80**, 3-8.

SHANNON, C. E., WEAVER, W., BLAHUT, R. E. & HAJEK, B. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498-2504.

SHAPIRO, A. D., TYACK, P. L. & SENEFF, S. (2010). Comparing call-based versus subunit-based methods for categorizing Norwegian killer whale, *Orcinus orca*, vocalizations. *Animal Behaviour* **81**, 377-386.

SLABBEKOORN, H. & SMITH, T. B. (2002). Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society B: Biological sciences* **357**, 493-503.

SLOCOMBE, K. E. & ZUBERBÜHLER, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour* **72**, 989-999.

SMITH, J. N., GOLDIZEN, A. W., DUNLOP, R. A. & NOAD, M. J. (2008). Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions. *Animal Behaviour* **76**, 467-477.

SMITH, W. J. (1977). *The behavior of communicating*. Harvard University Press.

SNIJDERS, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure* **3**, 1-40.

SOLAN, Z., HORN, D., RUPPIN, E. & EDELMAN, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences* **102**, 11629-11634.

STEGMANN, U. (2013). *Animal Communication Theory: Information and Influence*. Cambridge University Press.

STODDARD, P. (1996). Vocal recognition of neighbors by territorial passerines. In *Ecology and evolution of acoustic communication in birds* (eds. D. E. Kroodsma and E. L. Miller), pp. 356-374. Cornell University Press, Ithaca, New York.

STOLCKE, A. (2002). SRILM-an extensible language modeling toolkit. *International Conference on Spoken Language Processing* **2**, 901-904.

STURDY, C. B., PHILLMORE, L. S. & WEISMAN, R. G. (2000). Call-note discriminations in black-capped chickadees (*Poecile atricapillus*). *Journal of Comparative Psychology* **114**, 357.

SUTHERS, R. A. (2004). How birds sing and why it matters. In *Nature's Music: The Science of Birdsong* (eds. P. R. Marler and H. Slabbekoorn), pp. 272-295. Elsevier Academic Press, San Diego.

SUTHERS, R. A. (1997). Peripheral control and lateralization of birdsong. *Journal of Neurobiology* **33**, 632-652.

TAYLOR, A. M., REBY, D. & MCCOMB, K. (2008). Human listeners attend to size information in domestic dog growls. *The Journal of the Acoustical Society of America* **123**, 2903-2909.

TCHERNICHOVSKI, O., NOTTEBOHM, F., HO, C. E., PESARAN, B. & MITRA, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour* **59**, 1167-1176.

TEN CATE, C., LACHLAN, R. &ZUIDEMA, W. (2013). Analyzing the Structure of Bird Vocalizations and Language: Finding Common Ground. In *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain* (eds. J. J. Bolhuis and M. Everaert), pp. 243-260. MIT Press.

THOMAS, J. A., ZINNEL, K. C. & FERM, L. M. (1983). Analysis of Weddell seal (*Leptonychotes weddelli*) vocalizations using underwater playbacks. *Canadian Journal of Zoology* **61**, 1448-1456.

THOMPSON, A. B. & HARE, J. F. (2010). Neighbourhood watch: multiple alarm callers communicate directional predator movement in Richardson's ground squirrels, *Spermophilus richardsonii*. *Animal Behaviour* **80**, 269-275.

THOMSON, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* **70**, 1055-1096.

TITZE, I. R. (1994). *Principles of Voice Production*. Prentice Hall Englewood Cliffs.

TODT, D. & HULTSCH, H. (1998). How songbirds deal with large amounts of serial information: retrieval rules suggest a hierarchical song memory. *Biological Cybernetics* **79**, 487-500.

TOEWS, D. P. & IRWIN, D. E. (2008). Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustic analyses. *Molecular Ecology* **17**, 2691-2705.

TRAWICKI, M. B., JOHNSON, M. & OSIEJUK, T. (2005). Automatic song-type classification and speaker identification of Norwegian Ortolan Bunting (*Emberiza hortulana*) vocalizations. *IEEE Workshop on Machine Learning for Signal Processing* , 277-282.

TU, H. & DOOLING, R. J. (2012). Perception of warble song in budgerigars (*Melopsittacus undulatus*): evidence for special processing. *Animal Cognition* **15**, 1151-1159.

TU, H., SMITH, E. W. & DOOLING, R. J. (2011). Acoustic and perceptual categories of vocal elements in the warble song of budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology* **125**, 420-430.

TYACK, P. (1983). Differential response of humpback whales,*Megaptera novaeangliae*, to playback of song or social sounds. *Behavioral Ecology and Sociobiology* **13**, 49-55.

VAN HEIJNINGEN, C. A., DE VISSER, J., ZUIDEMA, W. & TEN CATE, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences* **106**, 20538-20543.

VAN SCHAIK, C. P., DAMERIUS, L. & ISLER, K. (2013). Wild orangutan males plan and communicate their travel direction one day in advance. *PloS One* **8**, e74896.

VEHRENCAMP, S. L., HALL, M. L., BOHMAN, E. R., DEPEINE, C. D. & DALZIELL, A. H. (2007). Song matching, overlapping, and switching in the banded wren: the sender's perspective. *Behavioral Ecology* **18**, 849-859.

WALLER, S. (2012). Science of the Monkey Mind: Primate Penchants and Human Pursuits. In *Experiencing Animal Minds: An Anthology of Animal-human Encounters* (eds. J. A. Smith and R. W. Mitchell).

WATKINS, W. A. & SCHEVILL, W. E. (1977). Sperm whale codas. *The Journal of the Acoustical Society of America* **62**, 1485-1490.

WATTS, D. J. & STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’networks. *Nature* **393**, 440-442.

WEARY, D. M., LAMBRECHTS, M. & KREBS, J. (1991). Does singing exhaust male great tits? *Animal Behaviour* **41**, 540-542.

2180 WEARY, D. M. & LEMON, R. E. (1990). Kroodsma refuted. *Animal Behaviour* **39**, 996-998.

2181 WEARY, D. M. & LEMON, R. E. (1988). Evidence against the continuity-versatility relationship in bird
2182 song. *Animal Behaviour* **36**, 1379-1383.

2183 WEARY, D., KREBS, J., EDDYSHAW, R., MCGREGOR, P. & HORN, A. (1988). Decline in song
2184 output by great tits: Exhaustion or motivation? *Animal Behaviour* **36**, 1242-1244.

2185 WEISS, D. J. & HAUSER, M. D. (2002). Perception of harmonics in the combination long call of
2186 cottontop tamarins, *Saguinus oedipus*. *Animal Behaviour* **64**, 415-426.

2187 WEISS, R. J. & BELLO, J. P. (2011). Unsupervised discovery of temporal structure in music. *Selected*
2188 *Topics in Signal Processing, IEEE Journal of* **5**, 1240-1251.

2189 WEISS, M., HULTSCH, H., ADAM, I., SCHARFF, C. & KIPPER, S. (2014). The use of network
2190 analysis to study complex animal communication systems: a study on nightingale song. *Proceedings of*
2191 *the Royal Society B: Biological Sciences* **281**, 20140460.

2192 WERNICKE, S. & RASCHE, F. (2006). FANMOD: a tool for fast network motif detection.
2193 *Bioinformatics* **22**, 1152-1153.

2194 WHEELER, B. C. & HAMMERSCHMIDT, K. (2012). Proximate factors underpinning receiver
2195 responses to deceptive false alarm calls in wild tufted capuchin monkeys: Is it counterdeception?
2196 *American Journal of Primatology* **75**, 715-725.

2197 WHEELER, B. C. (2010a). Decrease in alarm call response among tufted capuchins in competitive
2198 feeding contexts: possible evidence for counterdeception. *International Journal of Primatology* **31**, 665-
2199 675.

2200 WHEELER, B. C. (2010b). Production and perception of situationally variable alarm calls in wild tufted
2201 capuchin monkeys (*Cebus apella nigritus*). *Behavioral Ecology and Sociobiology* **64**, 989-1000.

2202 WIENS, J. A. (1982). Song pattern variation in the sage sparrow (*Amphispiza belli*): dialects or
2203 epiphenomena? *The Auk* **99**, 208-229.

2204 WILEY, R. H. (1983). The evolution of communication: information and manipulation. *Animal*
2205 *Behaviour* **2**, 156-189.

2206 WOLPERT, D. H. & MACREADY, W. G. (1997). No free lunch theorems for optimization.
2207 *Evolutionary Computation, IEEE Transactions on* **1**, 67-82.

2208 WYTTENBACH, R. A., MAY, M. L. & HOY, R. R. (1996). Categorical perception of sound frequency
2209 by crickets. *Science* **273**, 1542-1544.

2210 YOUNG, S. J. & YOUNG, S. (1994). The htk hidden Markov model toolkit: Design and philosophy.
2211 *Entropic Cambridge Research Laboratory, Ltd* **2**, 2-44.

2212 ZHONG, S. & GHOSH, J. (2005). Generative model-based document clustering: a comparative study.
2213 *Knowledge and Information Systems* **8**, 374-384.

2214 ZIPF, G. K. (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley press, Oxford.

2215 ZUCCHINI, W. & MACDONALD, I. L. (2009). *Hidden Markov models for time series: an introduction*
2216 *using R*. CRC Press.

2217

2218

2219

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. 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 3. 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 analysed 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 4. 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 5. Graphical representation of the process of selecting an appropriate unit definition. (a) Determine what is known about the production mechanism of the signalling individual, from the hierarchy of production mechanisms, and their spectrotemporal differences. (b) Determine what is known about the perception abilities of the receiver (vertical axis), and how this may modify the production characteristics of the sound (horizontal axis). (c) Choose a classification method, suitable for the modified acoustic characteristics.

Figure 6. Different ways that units can be combined to encode information in a sequence.

Figure 7. Flowchart suggesting possible paths for the analysis of sequences of acoustic units. Exploratory data analysis is conducted on the units or timing information using techniques such as histograms, networks, or low-order Markov models. Preliminary embedding paradigm hypotheses are formed based on observations. Depending upon the hypothesised embedding paradigm, various analysis techniques are suggested.

Figure 8. 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 9. 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 $Pr_e(\mathrm{A}|\mathrm{X}) = Pr_e(\mathrm{B}|\mathrm{X}) = 0.5$, and when in state *Y*, acoustic unit A is much more likely $Pr_e(\mathrm{A}|\mathrm{Y}) = 0.9$ than B $Pr_e(\mathrm{B}|\mathrm{Y}) = 0.1$. Transitioning from state *X* to state *Y* occurs with probability $Pr_t(\mathrm{X}\rightarrow\mathrm{Y}) = 0.6$, whereas from state *Y* to state *X* with probability $Pr_t(\mathrm{Y}\rightarrow\mathrm{X}) = 0.3$.

Figure 10. 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 11. 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*. Underlined variables indicate those to be replaced. Grammar produced with contributions from Ana Širović (Scripps Institution of Oceanography).

Figure 12. 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

Table 1. Examples of different approaches to unit definition, from different taxa and with different research aims.

Unit criterion	Taxon	Goal of division into "units"			
		Descriptive	Production	Perception	Function
Separated by silence	Birds	Swamp sparrow *Melospiza georgiana* note: (Marler & Pickert. 1984) Black capped chickadee *Poecile atricapillus* note: (Nowicki & Nelson. 1990) Red legged *Alectoris rufa* and rock *A. graeca* partridges: (Ceugniet & Aubin. 2001)	Zebra finch *Taeniopygia guttata* syllable: (Cynx. 1990) Emperor penguin *Aptenodytes forsteri*: (Robisson, Aubin & Bremond. 1993) Canary *Serinus canaria* breaths: (Hartley & Suthers. 1989)	Swamp sparrow *Melospiza georgiana* note: (Nelson & Marler. 1989) Black-capped chickadee *Poecile atricapillus* notes: (Sturdy, Phillmore & Weisman. 2000; Charrier et al. 2005) King penguin *Aptenodytes patagonicus*: (Lengagne, Lauga & Aubin. 2001)	Carolina chickadee *Poecile carolinensis* and Black-capped chickadee *P. atricapillus* note composition → **predator, foraging activity, identity**:(Freeberg. 2012; Krams et al. 2012) King penguin *Aptenodytes patagonicus*→ **individual identities:** (Jouventin, Aubin & Lengagne. 1999; Lengagne et al. 2000) Emperor penguin *Aptenodytes forsteri*→ **individual identities**: (Aubin, Jouventin & Hildebrand. 2000)
	Terrestrial mammals	Meerkat *Suricata suricatta* calls: (Manser. 2001) Gibbon *Hyrobates lar* phrase: (Raemaekers, Raemaekers & Haimoff. 1984) Rock hyrax *Procavia capensis* songs: (Kershenbaum et al. 2012)	Lesser short-tailed bat *Mystacina tuberculata* pulses: (Parsons, Riskin & Hermanson. 2010)	Meerkat *Suricata suricatta* calls: (Manser. 2001)	Meerkat *Suricata suricatta* calls → **predator type**: (Manser. 2001) Rock hyrax *Procavia capensis* songs → **male quality**: (Koren & Geffen. 2009) Free-tailed bat *Tadarida brasiliensis* syllable → **courtship:** (Bohn et al. 2008; Parsons, Riskin & Hermanson. 2010)

		Free-tailed bat *Tadarida brasiliensis* syllable: (Bohn et al. 2008) Mustached bat *Pteronotus parnellii* syllable: (Kanwal et al. 1994)			
	Marine mammals	Humpback whale *Megaptera novaeangliae* unit: (Payne & McVay. 1971) Killer whale *Orcinus orca* calls: (Ford. 1989) Bottlenose dolphin *Tursiops truncatus* signature whistles: (Caldwell. 1965; McCowan & Reiss. 1995) Australian sea lion *Neophoca cinerea* barking calls: (Gwilliam, Charrier & Harcourt. 2008)	Humpback whale *Megaptera novaeangliae* song: (Adam et al. 2013)	Bottlenose dolphin *Tursiops truncatus* signature whistles: (Janik, Sayigh & Wells. 2006) Subantartic fur seal *Arctocephalus tropicalis* pup attraction call: (Charrier, Mathevon & Jouventin. 2003) Australian sea lion *Neophoca cinerea*: (Charrier & Harcourt. 2006)	Bottlenose dolphin *Tursiops truncatus* signature whistles → **individual identity**: (Sayigh et al. 1999; Harley. 2008) Killer whale *Orcinus orca* calls: → **group identity**: (Ford. 1989) Australian sea lion *Neophoca cinerea* call → **colony identity**: (Attard et al. 2010) Australian sea lion *Neophoca cinerea* call → **threat level**: (Charrier, Ahonen & Harcourt. 2011) Australian sea lion *Neophoca cinerea* call → **individual identity**: (Charrier, Pitcher & Harcourt. 2009; Pitcher, Harcourt & Charrier. 2012)
Change in acoustic properties (regardless of silence)	Birds	Red junglefowl *Gallus gallus* elements: (Collias. 1987)	Northern cardinal *Cardinalis cardinalis:* (Suthers. 1997) Anna hummingbird	Anna hummingbird *Calypte anna* mechanical chirps: (Clark & Feo. 2010)	Blackcap *Sylvia atricapilla* song → **species identity**: (Mathevon & Aubin. 2001) White-browed warbler

			Calypte anna mechanical chirps: (Clark & Feo. 2008)	Male chickens *Gallus gallus* alarm calls: (Evans, Evans & Marler. 1993)	*Basileuterus leucoblepharus* song → **species identity**: (Mathevon et al. 2008) Yelkouan Shearwaters *Puffinus yelkouan* call → **sex and mate identity**: (Cure, Aubin & Mathevon. 2011) Grasshopper sparrow *Ammodramus savannarum* buzz/warble → **territorial/social**: (Lohr, Ashby & Wakamiya. 2013) Rufous-sided towhee *Pipilo erythrophthalmus* song → **species identity**: (Richards. 1981)
	Terrestrial mammals	Black-fronted titi monkey *Callicebus nigrifrons:* (Cäsar et al. 2012b) Western gorilla *Gorilla gorilla*: (Salmi, Hammerschmidt & Doran-Sheehy. 2013) Red titi monkey *Callicebus cupreus:* (Robinson. 1979)	Banded mongoose *Mungos mungo:* (Jansen, Cant & Manser. 2012)	Mustached bat *Pteronotus parnellii:* (Esser et al. 1997)	Black-fronted titi monkey *Callicebus nigrifrons* alarm calls → **predator type and behaviour***:* (Cäsar et al. 2012a) Western gorilla *Gorilla gorilla* vocalisations → **multiple functions**: (Salmi, Hammerschmidt & Doran-Sheehy. 2013) Tufted capuchin monkeys *Sapajus nigritus* calls→ **predator type**: (Wheeler. 2010b) Banded mongoose *Mungos mungo* close calls → **individual**

					identity, group cohesion: (Jansen, Cant and Manser. 2012) Spotted hyena *Crocuta crocuta* call → **sex/age/individual identities**: (Mathevon et al. 2010)
	Marine mammals	Bottlenose dolphin *Tursiops truncatus* whistle loops: (Caldwell, Caldwell & Tyack. 1990) Killer whale *Orcinus orca*, subunit of calls: (Shapiro, Tyack & Seneff. 2010) Humpback whale *Megaptera novaeangliae* subunit: (Payne & McVay. 1971) Leopard seal *Hydrurga leptonyx* calls: (Klinck, Kindermann & Boebel. 2008)	False killer whale *Pseudorca crassidens* vocalisations: (Murray, Mercado & Roitblat. 1998) Bottlenose dolphin *Tursiops truncatus* tonal calls: (Parsons, Riskin & Hermanson. 2010)	Bearded seal *Erignatus barbatus* trills: (Charrier, Mathevon & Aubin. 2013)	Killer whales *Orcinus orca* calls → **sex/orientation:** (Miller, Samarra & Perthuison. 2007) Spinner dolphin *Stenella longirostris* whistles → **movement direction:** (Lammers & Au. 2003)
Series of sounds	Birds	Song sparrow *Melospiza melodia* phrases: (Mulligan. 1966; Marler & Sherman. 1985) Blue-footed booby *Sula*	Emberizid sparrow trills: (Podos. 1997)	Zebra finch *Taeniopygia guttata* syllables: (Cynx, Williams & Nottebohm. 1990) Little owl *Athene*	Carolina chickadee *Poecile carolinensis* D-notes → **food availability**: (Mahurin & Freeberg. 2009) Kittiwake *Rissa tridactyla* call → **sex/individual identities**: (Aubin

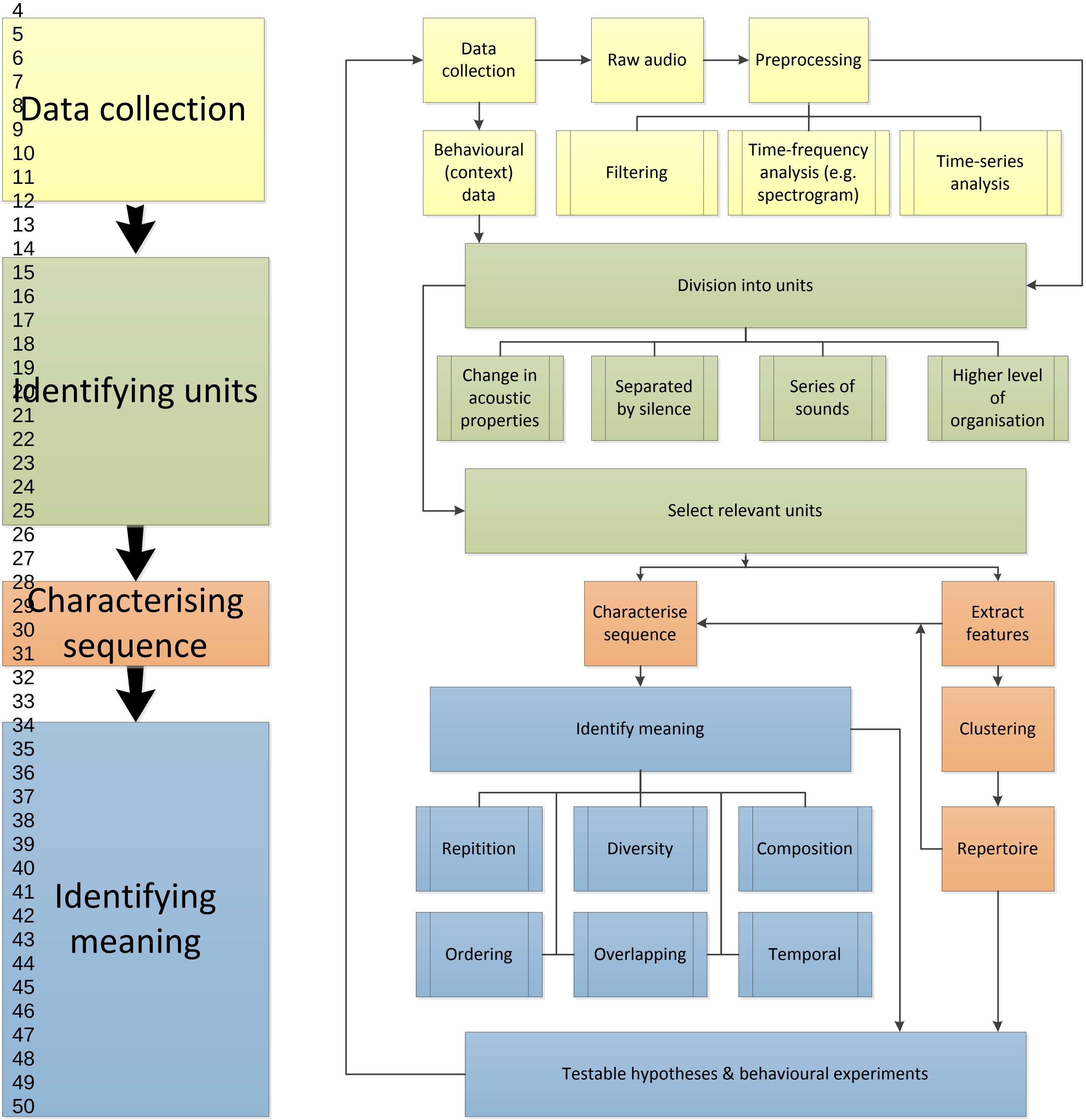
		nebouxii call: (Dentressangle, Aubin & Mathevon. 2012)		*noctua* syllables: (Parejo, Aviles & Rodriguez. 2012) Song sparrow *Melospiza melodia* songs: (Horning et al. 1993)	et al. 2007) Shearwaters *Puffinus yelkouan*, *Puffinus mauretanicus*, *Calonectris d. diomedea* call → **species identity**: (Curé et al. 2012)
	Terrestrial mammals	Black-fronted titi monkey *Callicebus nigrifrons:* (Cäsar et al. 2012b; Cäsar et al. 2013) Mustached bat *Pteronotus parnellii* syllable: (Kanwal et al. 1994) Free-tailed bat *Tadarida brasiliensis* calls: (Bohn et al. 2008) Hyrax *Procavia capensis* social calls: (Ilany et al. 2013) Chimpanzee *Pan troglodytes* pant hoots: (Notman & Rendall. 2005)	Diana monkey *Cercopithecus diana* alarm calls: (Riede et al. 2005) Domestic dog *Canis familiaris* growls: (Riede & Fitch. 1999)	Black-fronted titi monkey *Callicebus nigrifrons:* (Cäsar et al. 2012a) Colobus *Colobus guereza* sequences: (Schel, Candiotti & Zuberbühler. 2010) Tufted capuchin monkey *Sapajus nigritus* bouts: (Wheeler. 2010b)	Chimpanzee *Pan troglodytes* pant hoots → **foraging**: (Notman & Rendall. 2005) Free-tailed bat *Tadarida brasiliensis* calls → **courtship**: (Bohn et al. 2008)

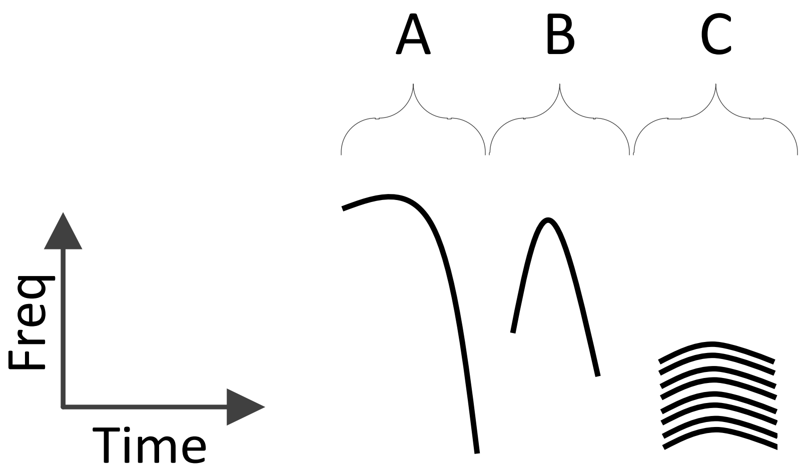
	Marine mammals	Humpback whale *Megaptera novaeangliae* phrases: (Payne & McVay. 1971) Bottlenose dolphin *Tursiops truncatus* whistles: (Deecke & Janik. 2006) Free-tailed bat *Tadarida brasiliensis* syllable: (Bohn et al. 2008)	Humpback whale *Megaptera novaeangliae* songs: (Frumhoff. 1983; Payne, Tyack & Payne. 1983; Mercado et al. 2010; Mercado & Handel. 2012) Bottlenose dolphin *Tursiops truncatus* whistles: (Janik et al. 2013)	Humpback whale *Megaptera novaeangliae* songs: (Handel, Todd & Zoidis. 2009) Bottlenose dolphin *Tursiops truncatus* whistles: (Pack et al. 2002) Weddell seal *Leptonychotes weddelli* vocalisations: (Thomas, Zinnel & Ferm. 1983) Harbour seal *Phoca vitulina* roars: (Hayes et al. 2004)	Bottlenose dolphin *Tursiops truncatus* signature whistles → **individual identity, group cohesion**: (Quick & Janik. 2012) Humpback whale *Megaptera novaeangliae* phrases → **unknown**: (Payne & McVay. 1971)
Higher levels of organisation	Birds	Canary *Serinus canaria* song: (Lehongre et al. 2008)	Swamp sparrow *Melospiza georgiana* trills: (Podos. 1997) Nightingale *Luscinia megarhynchos* song: (Todt & Hultsch. 1998) Canary *Serinus canaria* song: (Gardner, Naef & Nottebohm. 2005)	Song sparrow *Melospiza melodia* songs: (Searcy et al. 1995) Zebra finch *Taeniopygia guttata* song: (Doupe & Konishi. 1991) Canary *Serinus canaria* song: (Ribeiro et al. 1998)	Skylark *Alauda arvensis* songs → **group identity**: (Briefer, Rybak & Aubin. 2013) White-browed warbler *Basileuterus leucoblepharus* song → **individual identity**: (Mathevon et al. 2008)
	Terrestrial mammals	Red titi monkey *Callicebus cupreus* syllable: (Robinson.	Rhesus-macaque *Macaca mulatta* vocalisations: (Fitch. 1997)	Putty-nosed monkey *Cercopithecus nictitans* sequences:	Chimpanzee *Pan troglodytes* phrases → **group identity**: (Arcadi. 1996)

		1979) Free-tailed bat *Tadarida brasiliensis* songs: (Bohn et al. 2008)		(Arnold & Zuberbühler. 2006b) Red titi monkey *Callicebus cupreus* syllable: (Robinson. 1979)	Putty-nosed monkey *Cercopithecus nictitans* sequences → **predators presence, group movement**: (Arnold & Zuberbühler. 2006b) Tufted capuchin monkeys *Sapajus nigritus* calls→ **predator type**: (Wheeler. 2010b) Spotted hyena *Crocuta crocuta* call → **dominance rank identity**: (Mathevon et al. 2010)
	Marine mammals	Humpback whale *Megaptera novaeangliae* theme and song: (Payne & McVay. 1971)	Humpback whale *Megaptera novaeangliae* song: (Cazau et al. 2013)	Humpback whale *Megaptera novaeangliae* song: (Handel, Todd & Zoidis. 2012)	Humpback whale *Megaptera novaeangliae* song → **mating display - female attraction/male-male interactions** (Darling, Jones & Nicklin. 2006; Smith et al. 2008)

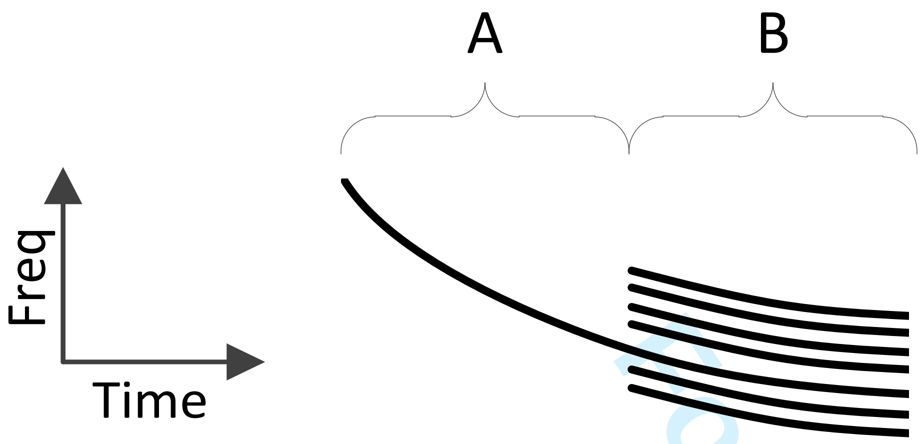
Table 2. A summary of some of the assumptions and requirements for each of the analytical models suggested in the review.

Model type	Embedding type	Data requirements	Typical hypotheses	Assumptions
Markov chain	• Repetition • Diversity • Ordering	• Number of observations required increases greatly as the size of the model grows	• Independence of sequence • Sequential structure	• Stationary transition matrix • Sufficient data for MLE of transition matrix
Hidden Markov model	• Repetition • Diversity • Ordering	• Number of observations required increases greatly as the size of the model grows	• Non-stationary transitions of observable states • Long-range correlations • Existence of cognitive states	• Sufficient data to estimate hidden states
Network	• Combination • Ordering	• Many unit types	• Network metrics have biological meaning • Comparison of motifs	• Network paths are meaningful
Formal grammar	• Repetition • Diversity • Ordering	• Few requirements	• Linguistic hypotheses • Deterministic sequences • Place in Chomsky hierarchy	• Deterministic transition rules
Temporal structure	• Overlapping • Timing	• Timing information exists • No need to define units	• Production/perception mechanisms • Changes with time/affect	• Temporal variations are perceived by receiver

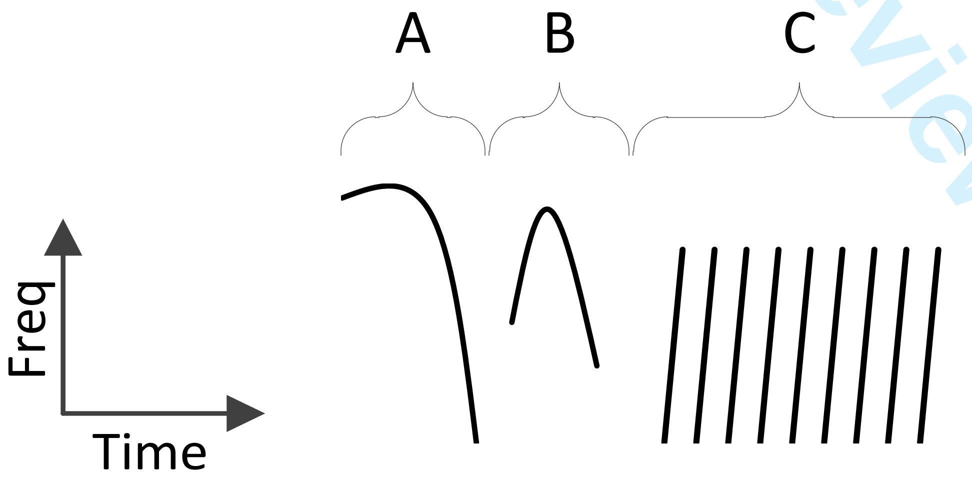

Data collection
Identifying units
Characterising sequence
Identifying meaning
Data collection
Raw audio
Preprocessing
Behavioural (context) data
Filtering
Time-frequency analysis (e.g. spectrogram)
Time-series analysis
Division into units
Change in acoustic properties
Separated by silence
Series of sounds
Higher level of organisation
Select relevant units
Characterise sequence
Extract features
Identify meaning
Clustering
Repitition
Diversity
Composition
Repertoire
Ordering
Overlapping
Temporal
Testable hypotheses & behavioural experiments



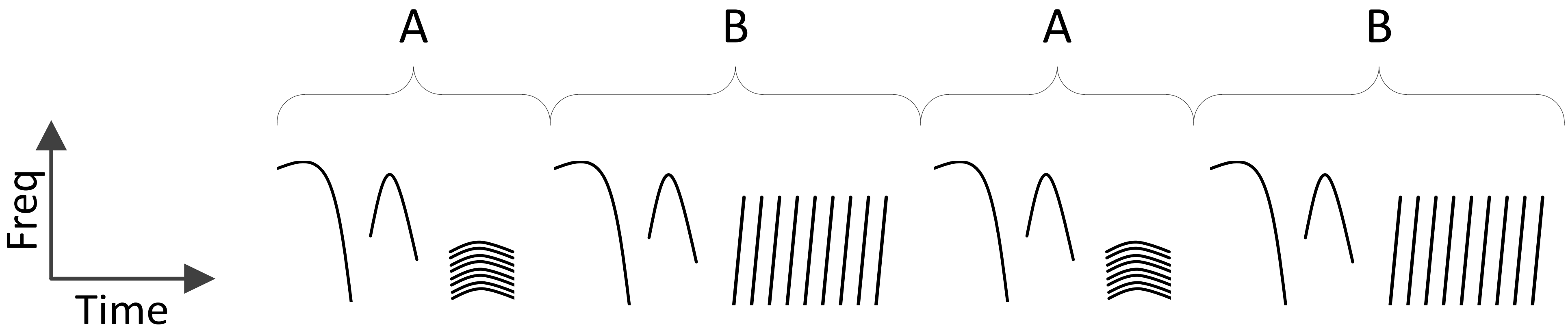
(a) Separated by silence



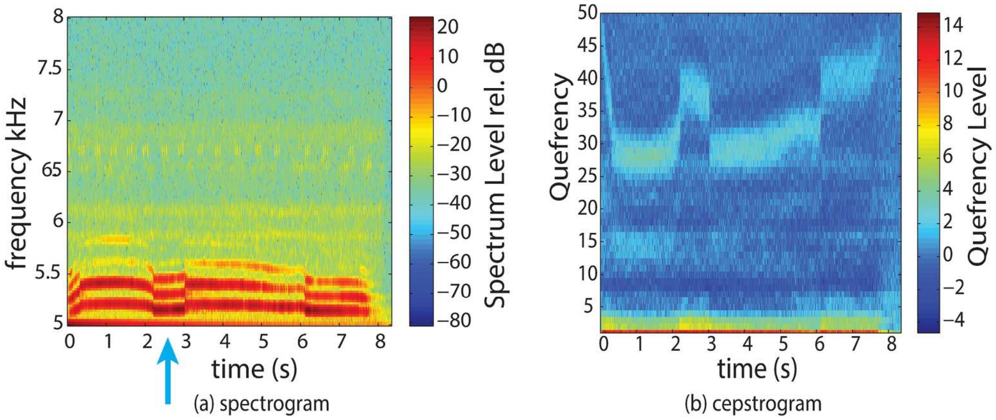
(b) Change in acoustic properties (regardless of silence)



(c) Series of sounds

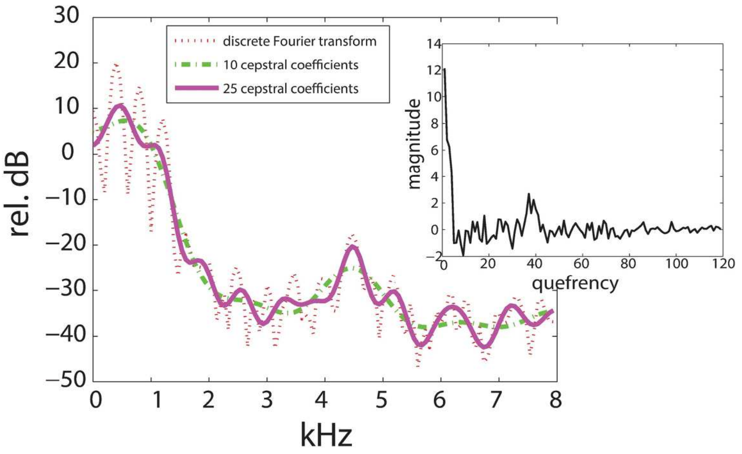


(d) Higher levels of organization



(a) spectrogram

(b) cepstrogram

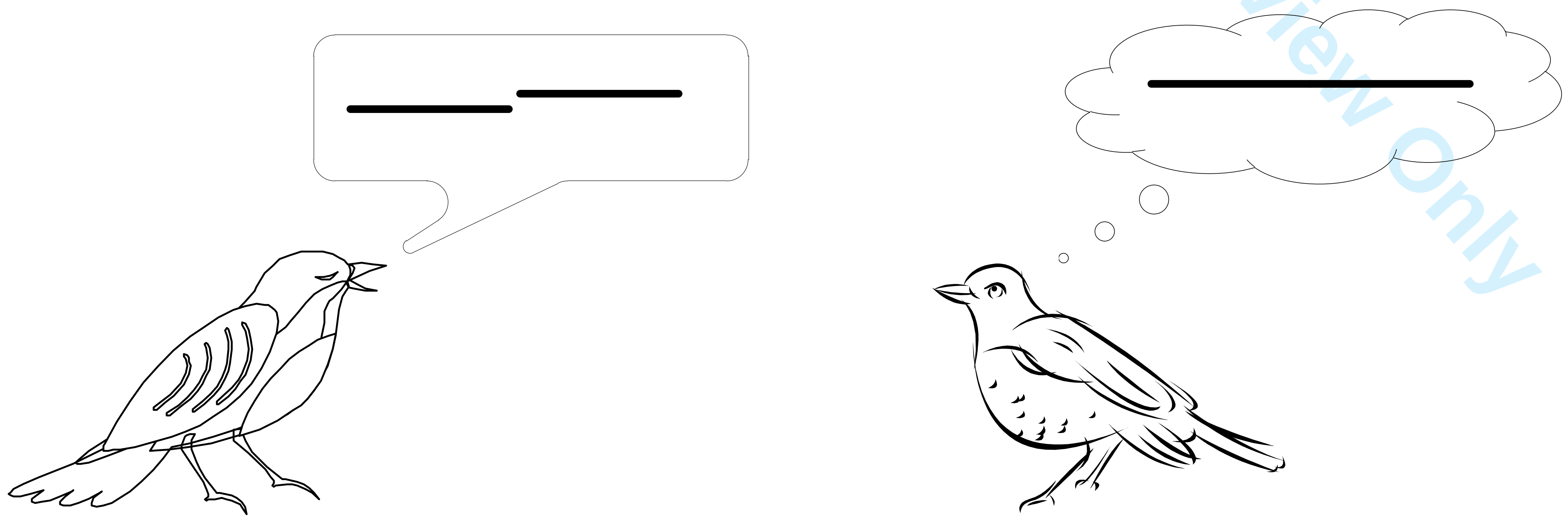


(c) cepstrum at 2.5 s (inset) and spectral reconstructions from truncated cepstral coefficients

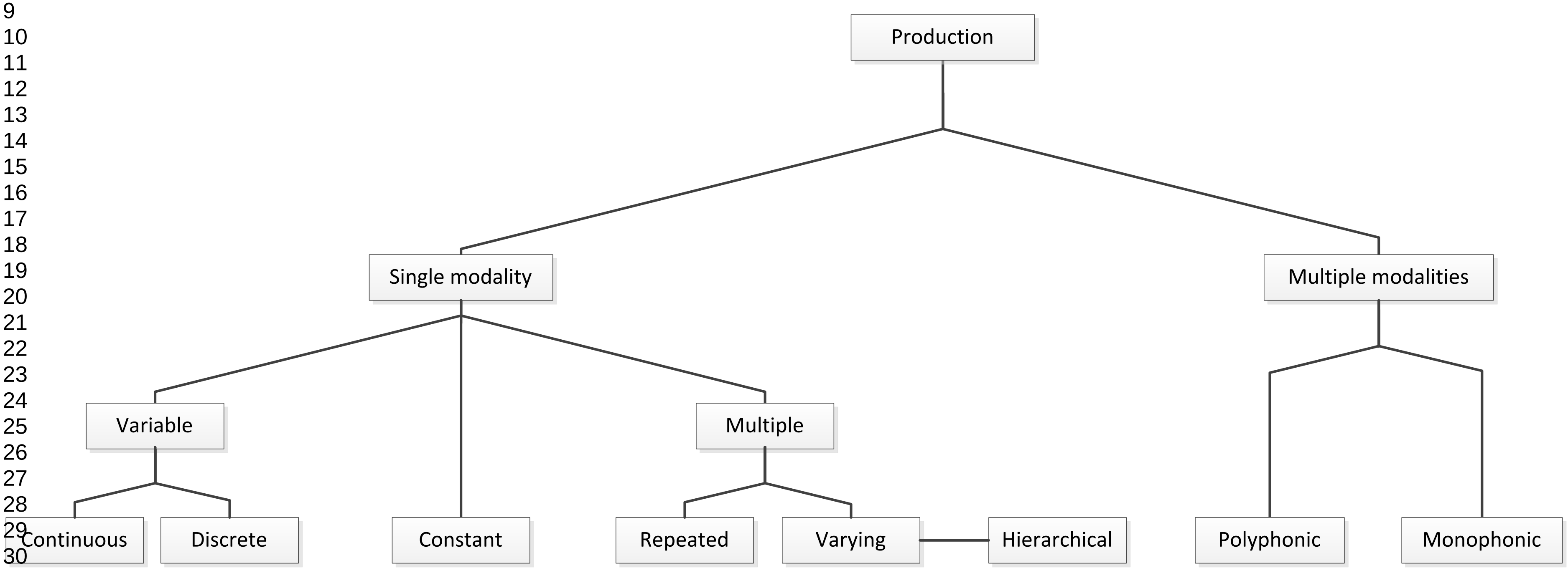
Figure 3. Example of cepstral processing of a grey wolf *Canis lupis* howl and crickets chirping. Recording was sampled at Fs = 16 kHz, 8 bit quantization. (a) Standard spectrogram analysed with a 15 ms Blackman-Harris window. (b) Plot of transform to cepstral domain. Lower quefrencies are related to 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.
190x178mm (300 x 300 DPI)

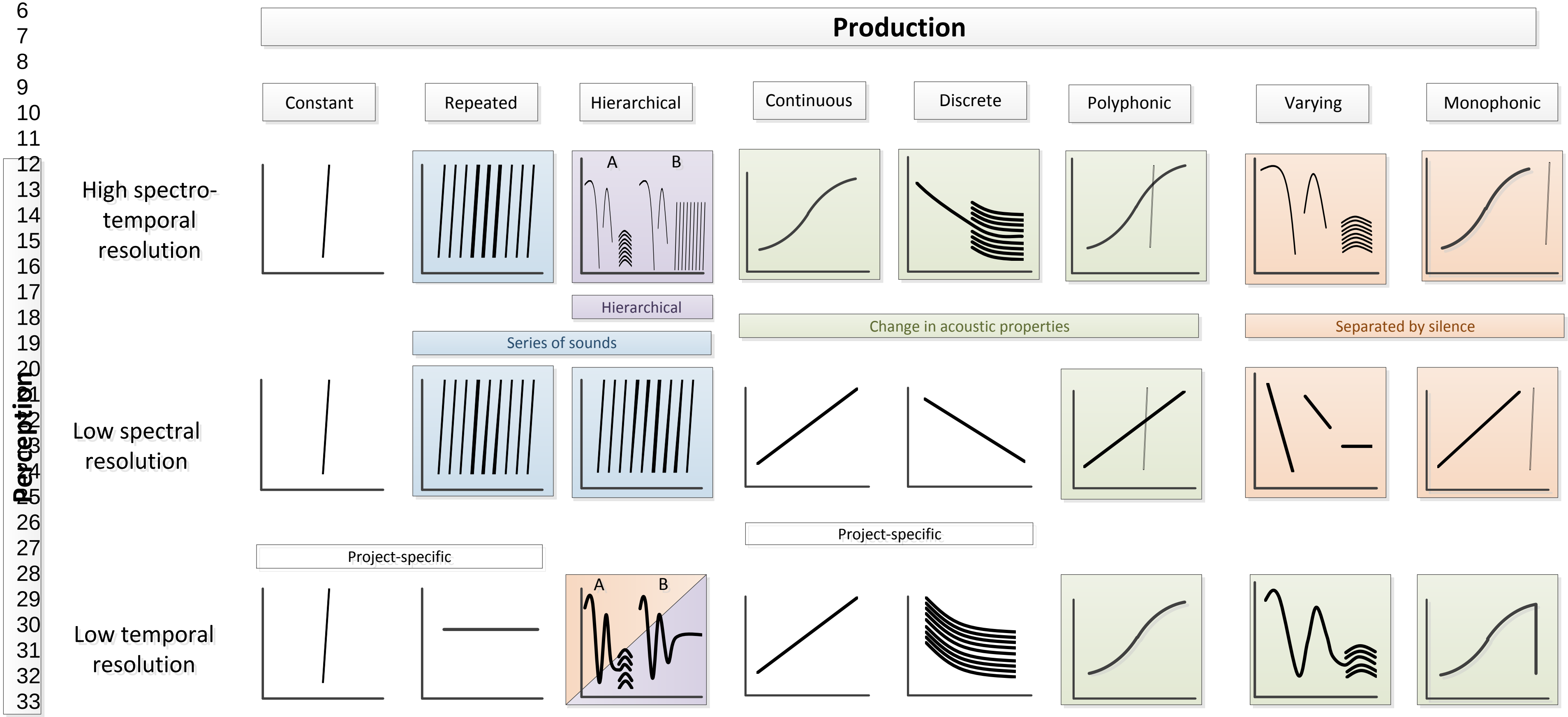
(a) Perceptual binding. Two discrete acoustic elements may be perceived by the receiver either as a single element, or as two separate ones

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



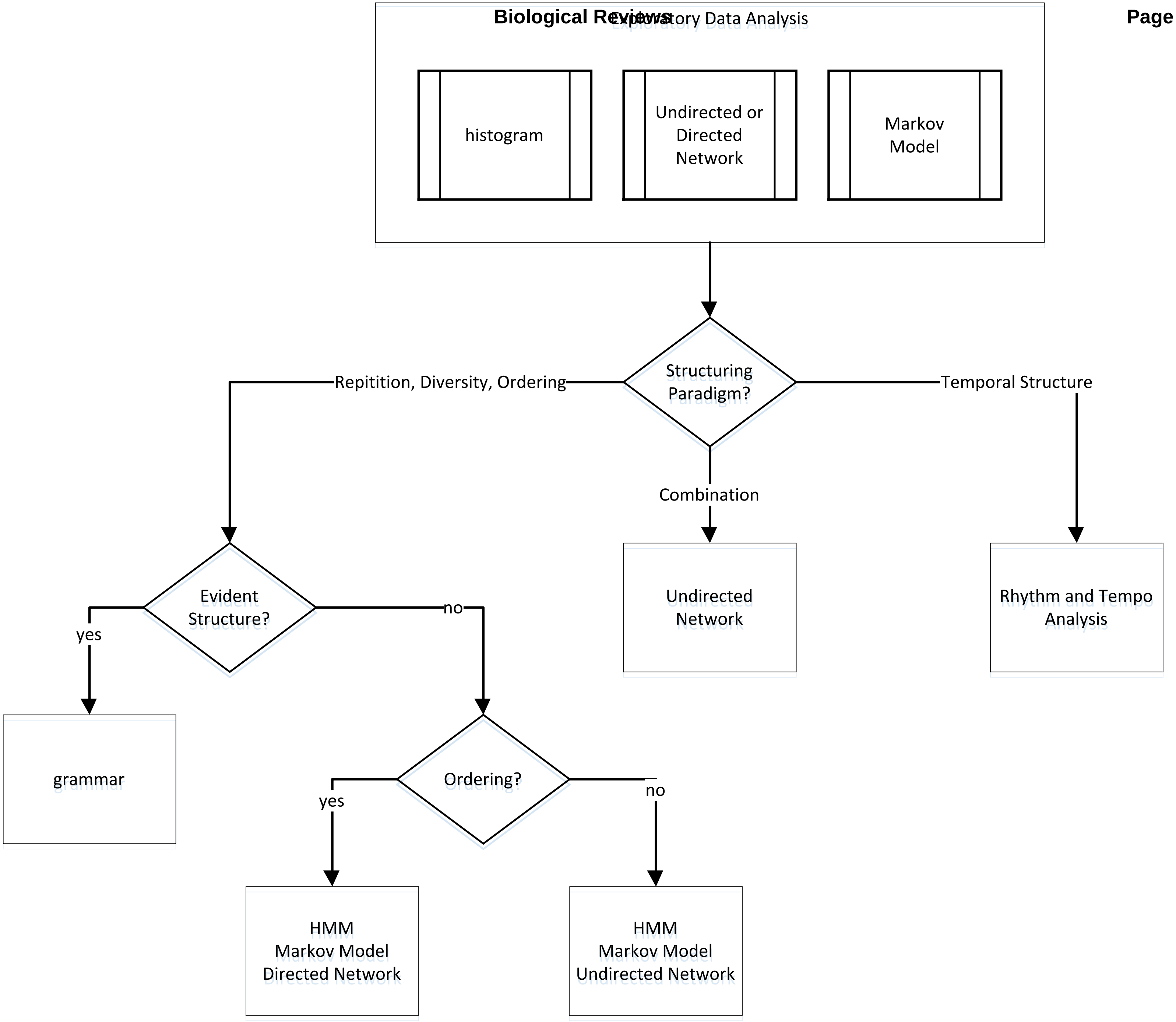
(c) Spectrotemporal constraints. If the receiver cannot distinguish small differences in time or frequency, discrete elements may be interpreted as joined

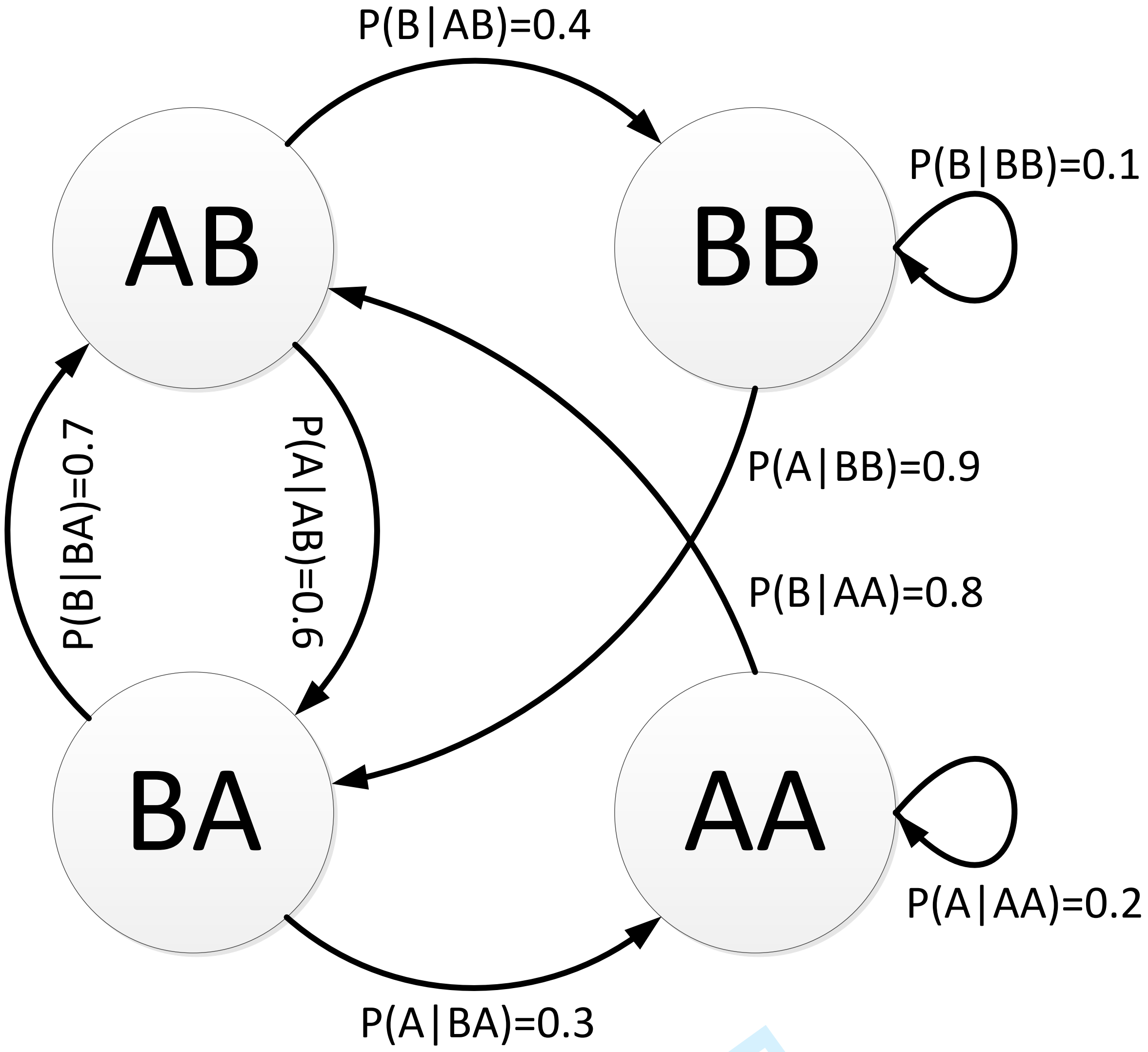

Production
Single modality
Multiple modalities
Variable
Multiple
Continuous
Discrete
Constant
Repeated
Varying
Hierarchical
Polyphonic
Monophonic


Production
Constant
Repeated
Hierarchical
Continuous
Discrete
Polyphonic
Varying
Monophonic
Perception
High spectro-temporal resolution
A
B
Hierarchical
Series of sounds
Change in acoustic properties
Separated by silence
Low spectral resolution
Project-specific
Project-specific
Low temporal resolution
A
B

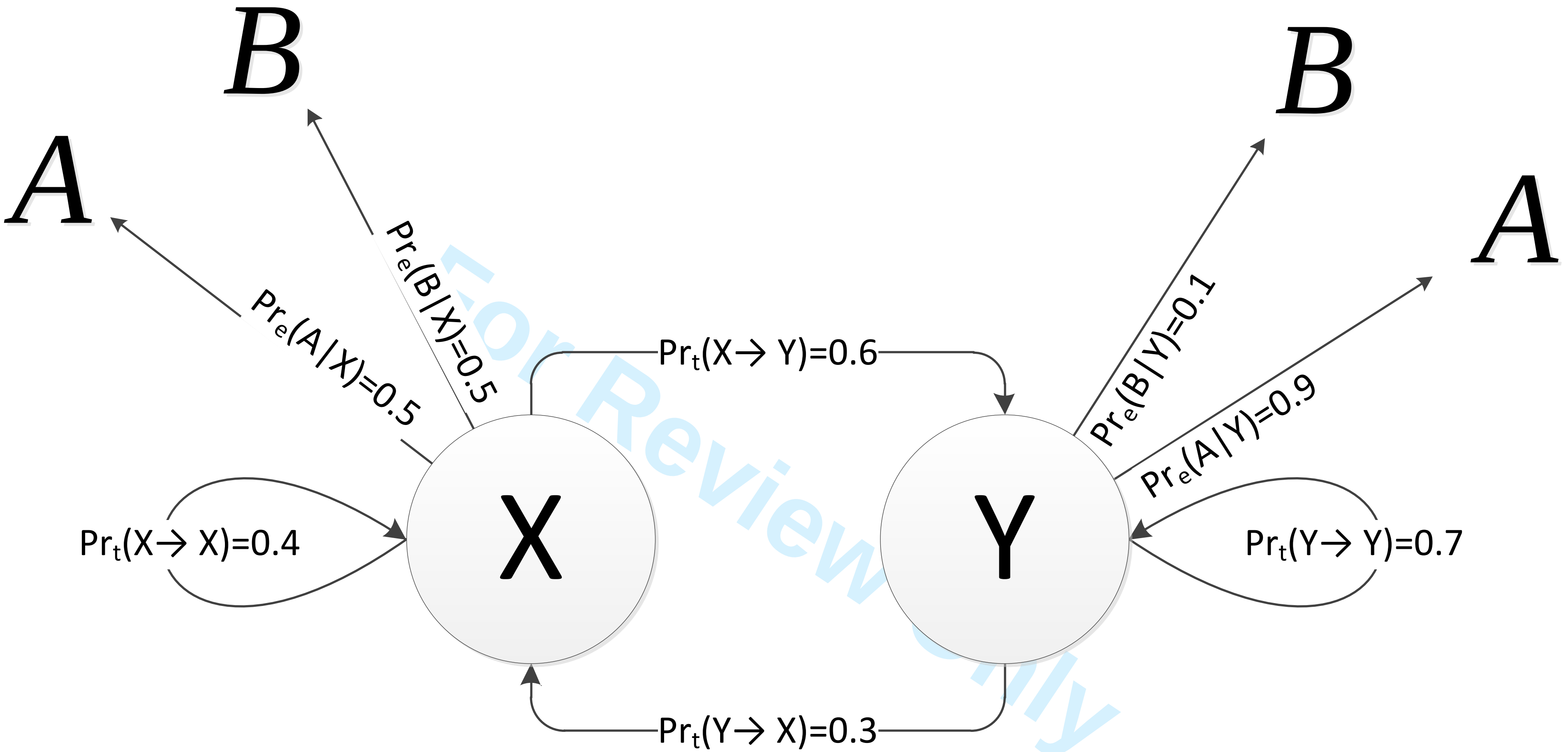
	Separated by silence	Change in acoustic properties	Series of sounds	Hierarchical
Manual classification “by eye”	√	~	~	√
Manual feature extraction + classification (semi-automatic)	~	√	X	X
Fully-automatic classification	~	~	√	~

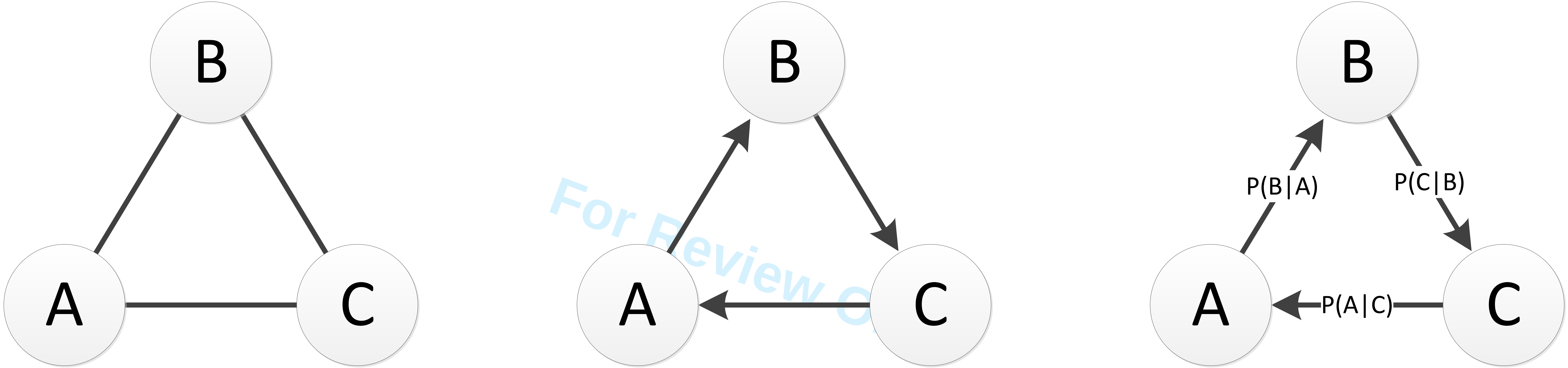
	Type	Criterion	Example
a	Repetition	Single unit repeated more than once	Chickadee D-note mobbing call (Baker & Becker. 2002)
b	Diversity	A number of distinct units are present. Order is unimportant.	Birdsong repertoire size (Searcy. 1992)
c	Combination	Set of units has different information from each unit individually. Order is unimportant.	Banded mongoose close calls (Jansen, Cant & Manser. 2012)
d	Ordering	Set of units has different information from each unit individually. Order is important	Human language, Humpback song (Payne & McVay. 1971)
e	Overlapping	Information conveyed in the relationship between sequences of two or more individuals	Sperm whale codas (Schulz et al. 2008)
f	Timing	Timing between units (often between different individuals) conveys information	Group alarm calling (Thompson & Hare. 2010)


Exploratory Data Analysis
histogram
Undirected or Directed Network
Markov Model
Structuring Paradigm?
Repitition, Diversity, Ordering
Temporal Structure
Combination
Undirected Network
Rhythm and Tempo Analysis
Evident Structure?
yes
no
grammar
Ordering?
yes
no
HMM
Markov Model
Directed Network
HMM
Markov Model
Undirected Network

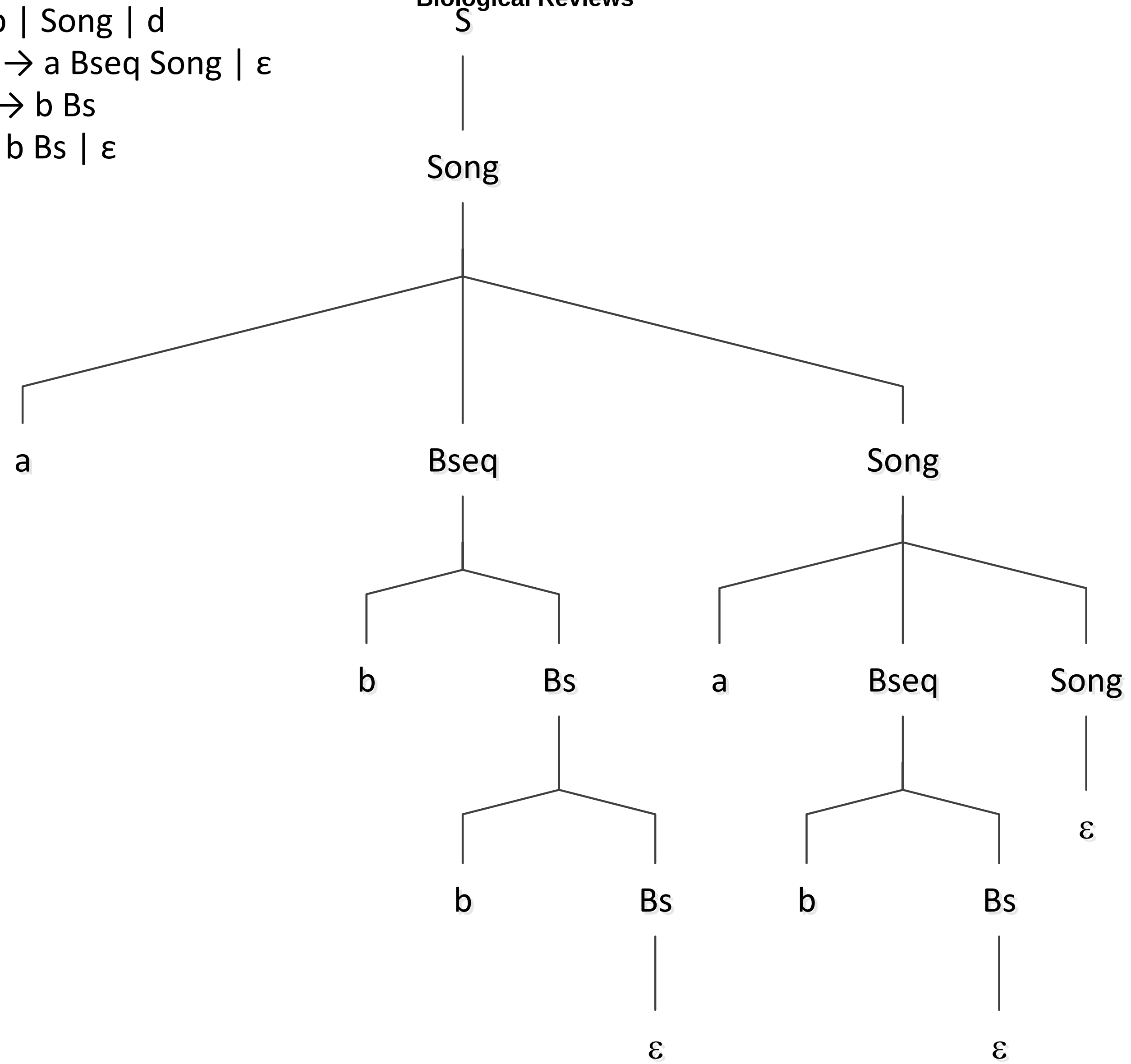


	A	B
P(X|AA)	**0.2**	**0.8**
P(X|AB)	**0.6**	**0.4**
P(X|BA)	**0.3**	**0.7**
P(X|BB)	**0.9**	**0.1**


B
A
$Pr_e(B|X)=0.5$
$Pr_e(A|X)=0.5$
$Pr_t(X\to X)=0.4$
X
$Pr_t(X\to Y)=0.6$
$Pr_t(Y\to X)=0.3$
Y
B
A
$Pr_e(B|Y)=0.1$
$Pr_e(A|Y)=0.9$
$Pr_t(Y\to Y)=0.7$


B
A
C
B
A
C
B
P(B|A)
P(C|B)
A
P(A|C)
C

S → b | Song | d
Song → a Bseq Song | ε
Bseq→ b Bs
Bs → b Bs | ε



$$S \Rightarrow \underline{\text{Song}} \Rightarrow \text{a } \underline{\text{Bseq}} \text{ Song} \Rightarrow \text{a b } \underline{\text{Bs}} \text{ Song} \Rightarrow \text{a b b } \underline{\text{Bs}} \text{ Song} \Rightarrow \text{a b b } \underline{\text{Song}}$$
$$\Rightarrow \text{a b b a } \underline{\text{Bseq}} \text{ Song} \Rightarrow \text{a b b a b } \underline{\text{Bs}} \text{ Song} \Rightarrow \text{a b b a b } \underline{\text{Song}} \Rightarrow \text{a b b a b}$$

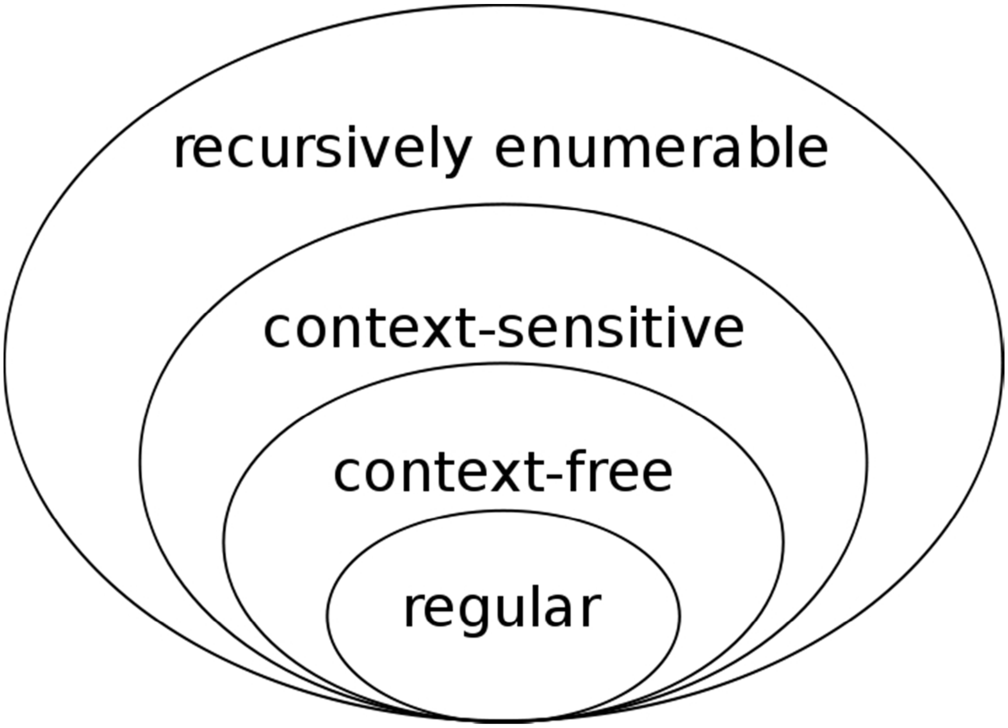


Figure 12. The classes of formal grammars known as the Chomsky hierarchy (Chomsky. 2002). Each class is a generalisation of the class it encloses, and is more complex than the enclosed classes. Image publicly available under the Creative Commons Attribution-Share Alike 3.0 Unported license.
https://commons.wikimedia.org/wiki/File:Wiki_inf_chomskeho_hierarchia.jpg
251x181mm (72 x 72 DPI)