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

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

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

2 **prospectus**

3 Arik Kershenbaum[1,2,*], Daniel T. Blumstein[3], Marie A. Roch[4], Çağlar Akçay[5],

4 Gregory Backus[6], Mark A. Bee[7], Kirsten Bohn[8], Yan Cao[9], Gerald Carter[10],

5 Cristiane Cäsar[11], Michael Coen[12], Stacy L. DeRuiter[13], Laurance Doyle[14], Shimon

6 Edelman[15], Ramon Ferrer-i-Cancho[16], Todd M. Freeberg[17], Ellen C. Garland[18],

7 Morgan Gustison[19], Heidi E. Harley[20], Chloé Huetz[21], Melissa Hughes[22], Julia

8 Hyland Bruno[23], Amiyaal Ilany[1], Dezhe Z. Jin[24], Michael Johnson[25], Chenghui

9 Ju[26], Jeremy Karnowski[27], Bernard Lohr[28], Marta B. Manser[29], Brenda McCowan[30],

10 Eduardo Mercado III[31], Peter M. Narins[32], Alex Piel[33], Megan Rice[34], Roberta

11 Salmi[35], Kazutoshi Sasahara[36], Laela Sayigh[37], Yu Shiu[5], Charles Taylor[3], Edgar E.

12 Vallejo[38], Sara Waller[39], Veronica Zamora-Gutierrez[2,40]

13

14 [1]National Institute for Mathematical and Biological Synthesis, 1122 Volunteer Blvd., Suite 106,

15 University of Tennessee, Knoxville, TN 37996-3410, USA

16 [2] Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

17 [3] Department of Ecology and Evolutionary Biology, University of California Los Angeles, 621

18 Charles E. Young Drive South, Los Angeles, CA 90095-1606 , USA

19 [4] Department of Computer Science, San Diego State University, 5500 Campanile Dr, San Diego,

20 CA 92182, USA

21 [5] Lab of Ornithology, Cornell University, 159 Sapsucker Woods Rd, Ithaca, NY 14850, USA

22   [6] Department of Biomathematics, North Carolina State University, Raleigh, NC 27607, USA

23   [7] Department of Ecology, Evolution and Behavior, University of Minnesota, 100 Ecology

24   Building, 1987 Upper Buford Cir, Falcon Heights, MN 55108, USA

25   [8] Integrated Science, Florida International University, Modesto Maidique Campus, 11200 SW

26   8th Street, AHC-4, 351, Miami, FL 33199, USA

27   [9] Department of Mathematical Sciences, University of Texas at Dallas, 800 W Campbell Rd,

28   Richardson, TX 75080, USA

29   [10] Biological Sciences Graduate Program, University of Maryland, College Park, MD 20742,

30   USA

31   [11] Department of Psychology & Neuroscience, University of St. Andrews, St Mary's Quad

32   South Street, St Andrews, Fife, KY16 9JP, UK

33   [12] Department of Biostatistics and Medical Informatics, University of Wisconsin, K6/446

34   Clinical Sciences Center, 600 Highland Avenue, Madison, WI 53792-4675, USA

35   [13] School of Mathematics and Statistics, University of St. Andrews, St Andrews, KY16 9SS, UK

36   [14] Carl Sagan Center for the Study of Life in the Universe, SETI Institute, 189 Bernardo Ave,

37   Suite 100, Mountain View, CA 94043, USA

38   [15] Department of Psychology, Cornell University, 211 Uris Hall, Ithaca, NY 14853-7601, USA

39   [16] Department of Computer Science, Universitat Politecnica de Catalunya, (Catalonia), Calle

40   Jordi Girona, 31, 08034 Barcelona, Spain

41   [17] Department of Psychology, University of Tennessee, Austin Peay Building, Knoxville,

42   Tennessee 37996, USA

43   [18] National Marine Mammal Laboratory, AFSC/NOAA, 7600 Sand Point Way N.E., Seattle,

44   Washington 98115, USA

45 [19] Department of Psychology, University of Michigan, 530 Church St, Ann Arbor, MI 48109,

46 USA

47 [20] Division of Social Sciences, New College of Florida, 5800 Bay Shore Rd, Sarasota, FL 34243,

48 USA

49 [21] CNPS, CNRS UMR 8195, Université Paris-Sud, UMR 8195, Batiments 440-447, Rue Claude

50 Bernard, 91405 Orsay, France

51 [22] Department of Biology, College of Charleston, 66 George St, Charleston, SC 29424, USA

52 [23] Department of Psychology, Hunter College and the Graduate Center, The City University of

53 New York, 365 Fifth Avenue, New York, NY 10016, USA

54 [24] Department of Physics, Pennsylvania State University, 104 Davey Lab, University Park, PA

55 16802-6300, USA

56 [25] Department of Electrical and Computer Engineering, Marquette University, 1515 W.

57 Wisconsin Ave., Milwaukee, WI 53233, USA

58 [26] Department of Biology, Queen College, The City Univ. of New York, 65-30 Kissena Blvd.,

59 Flushing, New York 11367, USA

60 [27] Department of Cognitive Science, University of California San Diego, 9500 Gilman Drive, La

61 Jolla, CA 92093-0515, USA

62 [28] Department of Biological Sciences, University of Maryland Baltimore County, 1000 Hilltop

63 Circle, Baltimore, MD 21250, USA

64 [29] Institute of Evolutionary Biology and Environmental Studies, University of Zurich,

65 Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

66 [30] Department of Veterinary Medicine, University of California Davis, 1 Peter J Shields Ave,

67 Davis, CA 95616, USA

68   [31] Department of Psychology; Evolution, Ecology, & Behavior, University at Buffalo, The State

69   University of New York, Park Hall Room 204, Buffalo, NY 14260-4110, USA

70   [32] Department of Integrative Biology & Physiology, University of California Los Angeles, 612

71   Charles E. Young Drive East, Los Angeles, CA 90095-7246, USA

72   [33] Division of Biological Anthropology, University of Cambridge, Pembroke Street

73   Cambridge, CB2 3QG, UK

74   [34] Department of Psychology, California State University San Marcos, 333 S. Twin Oaks Valley

75   Rd., San Marcos, CA 92096-0001, USA

76   [35] Department of Anthropology, University of Georgia at Athens, 355 S Jackson St, Athens, GA

77   30602, USA

78   [36] Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya,

79   464-8601, Japan

80   [37] Biology Department, Woods Hole Oceanographic Institution, 86 Water St, Woods Hole, MA

81   02543, USA

82   [38] Department of Computer Science, Monterrey Institute of Technology, Ave. Eugenio Garza

83   Sada 2501 Sur Col. Tecnológico C.P. 64849, Monterrey, Nuevo León,Mexico

84   [39] Department of Philosophy, Montana State University, 2-155 Wilson Hall, Bozeman , Montana

85   59717, USA

86   [40]Centre for Biodiversity and Environmental Research, University College London, London

87   WC1H 0AG, UK

88

89   *Author for correspondence (E-mail: arik.kershenbaum@gmail.com ; Tel. +44-1223-3336682).

90   ABSTRACT

91   Animal acoustic communication often takes the form of complex sequences, made up of multiple

92   distinct acoustic units. Apart from the well-known example of birdsong, other animals such as

93   insects, amphibians, and mammals (including bats, rodents, primates, and cetaceans) also

94   generate complex acoustic sequences. Occasionally, such as with birdsong, the adaptive role of

95   these sequences seems clear (e.g. mate attraction and territorial defence). More often however,

96   researchers have only begun to characterise – let alone understand – the significance and

97   meaning of acoustic sequences. Hypotheses abound, but there is little agreement as to how

98   sequences should be defined and analysed. Our review aims to outline suitable methods for

99   testing these hypotheses, and to describe the major limitations to our current and near-future

100   knowledge on questions of acoustic sequences.

101   This review and prospectus is the result of a collaborative effort between 43 scientists

102   from the fields of animal behaviour, ecology and evolution, signal processing, machine learning,

103   quantitative linguistics, and information theory, who gathered for a 2013 workshop entitled,

104   "Analysing vocal sequences in animals". Our goal is to present not just a review of the state of

105   the art, but to propose a methodological framework that summarises what we suggest are the best

106   practices for research in this field, across taxa and across disciplines. We also provide a tutorial-

107   style introduction to some of the most promising algorithmic approaches for analysing

108   sequences.

109   We divide our review into three sections: identifying the distinct units of an acoustic

110   sequence, describing the different ways that information can be contained within a sequence, and

111   analysing the structure of that sequence. Each of these sections is further subdivided to address

112   the key questions and approaches in that area.

113   We propose a uniform, systematic, and comprehensive approach to studying sequences,

114   with the goal of clarifying research terms used in different fields, and facilitating collaboration

115   and comparative studies. Allowing greater interdisciplinary collaboration will facilitate the

116   investigation of many important questions in the evolution of communication and sociality.

117

118   *Key words*: acoustic communication, information, information theory, machine learning, Markov

119   model, meaning, network analysis, sequence analysis, vocalisation.

120

121   CONTENTS

162

I. INTRODUCTION

164   Sequences are everywhere, from the genetic code, to behavioural patterns such as foraging, as

165   well as the sequences that comprise music and language. Often, but not always, sequences

166   convey meaning, and can do so more effectively than other types of signals (Shannon *et al*.,

167   1949), and individuals can take advantage of the information contained in a sequence to increase

168   their own fitness (Bradbury & Vehrencamp, 2011). Acoustic communication is widespread in the

169   animal world, and very often individuals communicate using a sequence of distinct acoustic

170   elements, the order of which may contain information of potential benefit to the receiver. In

171   some cases, acoustic sequences appear to be ritualised signals where the signaller benefits if the

172   signal is detected and acted upon by a receiver. The most studied examples include birdsong,

173   where males may use sequences to advertise their potential quality to rival males and to receptive

174   females (Catchpole & Slater, 2003). Acoustic sequences can contain information on species

175   identity, e.g. in many frogs and insects (Gerhardt & Huber, 2002), on individual identity and

176   traits, e.g. in starlings *Sturnus vulgaris* (Gentner & Hulse, 1998), wolves *Canis lupus* (Root-

177   Gutteridge *et al*., 2014), dolphins *Tursiops truncatus* (Sayigh *et al*., 2007), and hyraxes *Procavia*

178   *capensis* (Koren & Geffen. 2011), and in some cases, on contextual information such as resource

179   availability, e.g. food calls in chimpanzees *Pan troglodytes* (Slocombe & Zuberbühler, 2006), or

180   predator threats, e.g. in marmots *Marmota* spp. (Blumstein, 2007), primates (Schel, Tranquilli &

181   Zuberbühler, 2009; Cäsar *et al*., 2012*b*), and parids (Baker & Becker, 2002). In many cases,

182 however, the ultimate function of communicating in sequences is unclear. Understanding the

183 proximate and ultimate forces driving and constraining the evolution of acoustic sequences, as

184 well as decoding the information contained within them, is a growing field in animal behaviour

185 (Freeberg, Dunbar & Ord, 2012). New analytical techniques are uncovering characteristics

186 shared among diverse taxa, and offer the potential of describing and interpreting the information

187 within animal communication signals. The field is ripe for a review and a prospectus to guide

188 future empirical research.

189   Progress in this field could benefit from an approach that can bridge and bring together

190 inconsistent terminology, conflicting assumptions, and different research goals, both between

191 disciplines (e.g. between biologists and mathematicians), and also between researchers

192 concentrating on different taxa (e.g. ornithologists and primatologists). Therefore, we aim to do

193 more than provide a glossary of terms. Rather, we build a framework that identifies the key

194 conceptual issues common to the study of acoustic sequences of all types, while providing

195 specific definitions useful for clarifying questions and approaches in more narrow fields. Our

196 approach identifies three central questions: what are the units that compose the sequence? How is

197 information contained within the sequence? How do we assess the structure governing the

198 composition of these units? Fig. 1 illustrates a conceptual flow diagram linking these questions,

199 and their sub-components, and should be broadly applicable to any study involving animal

200 acoustic sequences.

201   Our aims in this review are as follows: (1) to identify the key issues and concepts

202 necessary for the successful analysis of animal acoustic sequences; (2) to describe the commonly

203 used analytical techniques, and importantly, also those underused methods deserving of more

204 attention; (3) to encourage a cross-disciplinary approach to the study of animal acoustic

205   sequences that takes advantage of tools and examples from other fields to create a broader

206   synthesis; and (4) to facilitate the investigation of new questions through the articulation of a

207   solid conceptual framework.

208   In Section II we ask why sequences are important, and what is meant by "information"

209   content and "meaning" in sequences. In Section III, we examine the questions of what units

210   make up a sequence and how to identify them. In some applications the choice seems trivial,

211   however in many study species, sequences can be represented at different hierarchical levels of

212   abstraction, and the choice of sequence "unit" may depend on the hypotheses being tested. In

213   Section IV, we look at the different ways that units can encode information in sequences. In

214   Section V, we examine the structure of the sequence, the mathematical and statistical models that

215   quantify how units are combined, and how these models can be analysed, compared, and

216   assessed. In Section VI, we describe some of the evolutionary and ecological questions that can

217   be addressed by analysing animal acoustic sequences, and look at some promising future

218   directions and new approaches.

219

220   **II. THE CONCEPTS OF INFORMATION AND MEANING**

221   The complementary terms, "meaning" and "information" in communication, have been variously

222   defined, and have long been the subject of some controversy (Dawkins & Krebs, 1978;

223   Stegmann, 2013). In this section we explore some of the different definitions from different

224   fields, and their significance for research on animal behaviour. The distinction between

225   information and meaning is sometimes portrayed with information as the form or structure of

226   some entity on the one hand, and meaning as the resulting activity of a receiver of that

227   information on the other hand (Bohm, 1989).

228

(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, p. 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. As an example, we present a list of some of these philosophical theories to give the reader a sense both of the lack of agreement as to the nature of

11

251   meaning, and to highlight the lack of connection between theories of human semantics, and

252   theories of animal communication. The nature of meaning has been theorised in many ways:

253   extensional (based on things in the world, like "animals"), intensional (based on thoughts within

254   minds, notions, concepts, ideas), or according to prototype theory (in which objects have

255   meaning through a graded categorisation, e.g. "baldness" is not precisely determined by the

256   number of hairs on the head). The physiological nature of meaning may be innate or learned, in

257   terms of its mental representations and cognitive content. Finally, descriptions of the role of

258   meaning are diverse: meaning may be computational/functional; atomic or holistic; bound to

259   both signaller and receiver, or a speech act of the signaller; rule bound or referentially based; a

260   description, or a convention; or a game dependent on a form of life, among other examples

261   (Christiansen & Chater, 2001; Martinich & Sosa, 2013).

262

263   **(2) Context**

264   Context has a profound influence on signal meaning, and this should apply to the

265   meaning of sequences as well. Context includes internal and external factors that may influence

266   both the production and perception of acoustic sequences; the effects of context can partially be

267   understood by considering how it specifically influences the costs and benefits of producing a

268   particular signal or responding to it. For instance, an individual's motivational, behavioural, or

269   physiological state may influence response (Lynch *et al*., 2005; Goldbogen *et al*., 2013); hungry

270   animals respond differently to signals than satiated ones, and an individual in oestrus or musth

271   may respond differently than ones not in those altered physiological states (Poole, 1999). Sex

272   may influence response as well (Tyack, 1983; Darling, Jones & Nicklin, 2006; Smith *et al*.,

273   2008; van Schaik, Damerius & Isler, 2013). The social environment may influence the costs and

274　benefits of responding to a particular signal (Bergman *et al*., 2003; Wheeler, 2010*a*; Ilany *et al*.,

275　2011; Wheeler & Hammerschmidt, 2012) as might environmental attributes, such as temperature

276　or precipitation. Knowledge from other social interactions or environmental experiences can also

277　play a role in context, e.g. habituation (Krebs, 1976). Context can also alter a behavioural

278　response when hearing the same signal originate from different spatial locations. For instance in

279　neighbour–stranger discrimination in songbirds, territorial males typically respond less

280　aggressively toward neighbours compared with strangers, so long as the two signals are heard

281　coming from the direction of the neighbour's territory. If both signals are played back from the

282　centre of the subject's territory, or from a neutral location, subjects typically respond equally

283　aggressively to both neighbours and strangers (Falls, 1982; Stoddard, 1996). Identifying and

284　testing for important contextual factors appears to be an essential step in decoding the meaning

285　of sequences.

286　　　　In human language, context has been proposed to be either irrelevant to, or crucial to, the

287　meaning of words and sentences. In some cases, a sentence bears the same meaning across

288　cultures, times, and locations, irrespective of context, e.g. "2+2=4" (Quine, 1960). In other cases,

289　meaning is derived at least partially from external factors, e.g. the chemical composition of a

290　substance defines its nature, irrespective of how the substance might be variously conceived by

291　different people (Putnam, 1975). By contrast, indexical terms such as "she" gain meaning only as

292　a function of context, such as physical or implied pointing gestures (Kaplan, 1978). Often, the

293　effect of the signal on the receivers determines its usefulness, and that usefulness is dependent

294　upon situational-contextual forces (Millikan, 2004).

295

296　**(3) Contrasting definitions of meaning**

297   Biologists (particularly behavioural ecologists), and cognitive neuroscientists have

298   different understandings of meaning. For most biologists, meaning relates to the function of

299   signalling. The function of signals is examined in agonistic and affiliative interactions, in

300   courtship and mating decisions, and in communicating about environmental stimuli, such as the

301   detection of predators (Bradbury & Vehrencamp, 2011). Behavioural ecologists study meaning

302   by determining the degree of production specificity, the degree of response specificity, and

303   contextual independence, e.g. Evans (1997). Cognitive neuroscientists generally understand

304   meaning through mapping behaviour onto structure–function relationships in the brain

305   (Chatterjee, 2005).

306   Mathematicians understand meaning by developing theories and models to interpret the

307   observed signals. This includes defining and quantifying the variables (observable and

308   unobservable), and the formalism for combining various variables into a coherent framework,

309   e.g. pattern theory (Mumford & Desolneux, 2010). One approach to examining a signal

310   mathematically is to determine the entropy, or amount of structure (or lack thereof) present in a

311   sequence. An entropy metric places a bound on the maximum amount of information that can be

312   present in a signal, although it does not determine that such information is, in fact, present.

313   Qualitatively, we infer meaning in a sequence if it modifies the receiver's response in

314   some predictable way. Quantitatively, information theory measures the amount of information

315   (usually in units of bits) transmitted and received within a communication system (Shannon *et*

316   *al.*, 1949). Therefore, information theory approaches can describe the complexity of the

317   communication system. Information theory additionally can characterise transmission errors and

318   reception errors, and has been comprehensively reviewed in the context of animal

319   communication in Bradbury & Vehrencamp (2011).

320   The structure of acoustic signals does not necessarily have meaning *per se*, and so

321   measuring that structure does not necessarily reveal the complexity of meaning. As one example,

322   the structure of an acoustic signal could be related to effective signal transmission through a

323   noisy or reverberant environment. A distinction is often made between a signal's "content", or

324   broadcast information, and its "efficacy", or transmitted information – the characteristics or

325   features of signals that actually reach receivers (Wiley, 1983; Hebets & Papaj, 2005). This is

326   basically the distinction between bearing functional information and getting that information

327   across to receivers in conditions that can be adverse to clear signal propagation. A sequence may

328   also contain elements that do not in themselves contain meaning, but are intended to get the

329   listeners' attention, in anticipation of future meaningful elements (e.g. Richards, 1981; Call &

330   Tomasello, 2007; Arnold & Zuberbühler, 2013).

331   Considerable debate exists over the nature of animal communication and the terminology

332   used in animal communication research (Owren, Rendall & Ryan, 2010; Seyfarth *et al*., 2010;

333   Ruxton & Schaefer, 2011; Stegmann, 2013), and in particular the origin of and relationship

334   between meaning and information, and their evolutionary significance. For our purposes, we will

335   use the term "meaning" when discussing behavioural and evolutionary processes, and the term

336   "information" when discussing the mathematical and statistical properties of sequences. This

337   parallels (but is distinct from) the definitions given by Ruxton & Schaefer (2011), in particular

338   because we wish to have a single term ("information") that describes inherent properties of

339   sequences, without reference to the putative behavioural effects on receivers, or the ultimate

340   evolutionary processes that caused the sequence to take the form that it does.

341   We have so far been somewhat cavalier in how we have described the structures of call

342 sequences, using terms like notes, units, and, indeed, calls. In the next section of our review, we

343 describe in depth the notion of signalling 'units' in the acoustic modality.

344

III. ACOUSTIC UNITS

346 Sequences are made of constituent units. Thus, the accurate analysis of potential information in

347 animal acoustic sequences depends on appropriately characterising their constituent acoustic

348 units. We recognise, however, that there is no single definition of a unit. Indeed, definitions of

349 units, how they are identified, and the semantic labels we assign them vary widely across

350 researchers working with different taxonomic groups (Gerhardt & Huber, 2002) or even within

351 taxonomic groups, as illustrated by the enormous number of names for different units in the

352 songs of songbird species. Our purpose in this section is to discuss issues surrounding the various

353 ways the acoustic units composing a sequence may be characterised.

354   Units may be identified based on either production mechanisms, which focus on how the

355 sounds are generated by signallers, or by perceptual mechanisms, which focus on how the sounds

356 are interpreted by receivers. How we define a unit will therefore be different if the biological

357 question pertains to production mechanisms or perceptual mechanisms. For example, in birdsong

358 even a fairly simple note may be the result of two physical production pathways, each made on a

359 different side of the syrinx (Catchpole & Slater, 2003). In practice, however, the details of

360 acoustic production and perception are often hidden from the researcher, and so the definition of

361 acoustic units is often carried out on the basis of observed acoustic properties: see Catchpole &

362 Slater (2003). It is not always clear to what extent these observed acoustic properties accurately

363 represent the production/perceptual constraints on communication, and the communicative role

364   of the sequence. Identifying units is made all the more challenging because acoustic units

365   produced by animals often exhibit graded variation in their features (e.g. absolute frequency,

366   duration, rhythm or tempo, or frequency modulation), but most analytical methods for unit

367   classification assume that units can be divided into discrete, distinct categories (e.g. Clark,

368   Marler & Beeman, 1987).

369   How we identify units may differ depending on whether the biological question pertains

370   to production mechanisms, perceptual mechanisms, or acoustical analyses of information content

371   in the sequences. If the unit classification scheme must reflect animal sound production or

372   perception, care must be taken to base unit identification on the appropriate features of a signal,

373   and features that are biologically relevant, e.g. Clemins & Johnson (2006). In cases where

374   sequences carry meaning, it is likely that they can be correlated with observational behaviours

375   (possibly context-dependent) observed over a large number of trials. There is still no guarantee

376   that the sequence assigned by the researcher is representative of the animal's perception of the

377   same sequence. To some degree, this can be tested with playback trials where the signals are

378   manipulated with respect to the hypothesised unit sequence (Kroodsma, 1989; Fischer, Noser &

379   Hammerschmidt, 2013).

380   Whatever technique for identifying potential acoustic units is used, we emphasise here

381   that there are four acoustic properties that are commonly used to delineate potential units (Fig.

382   2). First, the spectrogram may show a silent gap between two acoustic elements (Fig. 2A). When

383   classifying units "by eye", separating units by silent gaps is probably the most commonly used

384   criterion. Second, examination of a spectrogram may show that an acoustic signal changes its

385   properties at a certain time, without the presence of a silent "gap" (Fig. 2B). For example, a pure

386   tone may become harmonic or noisy, as the result of the animal altering its articulators (e.g. lips),

387   without ceasing sound production in the source (e.g. larynx). Third, a series of similar sounds

388   may be grouped together as a single unit, regardless of silent gaps between them, and separated

389   from dissimilar units (Fig. 2C). This is characteristic of pulse trains and "trills". Finally, there

390   may be a complex hierarchical structure to the sequence, in which combinations of sounds,

391   which might otherwise be considered fundamental units, always appear together, giving the

392   impression of a coherent, larger unit of communication (Fig. 2D). A consideration of these four

393   properties together can provide valuable insights into defining units of production, units of

394   perception, and units for sequence analyses.

395   In Table 1, we give examples of the wide range of studies that have used these different

396   criteria for dividing acoustic sequences into units. Although not intended to be comprehensive,

397   the table shows how all of the four criteria listed above have been used for multiple species and

398   with multiple aims – whether simply characterising the vocalisations, defining units of

399   production/perception, or identifying the functional purpose of the sequences.

400

401   **(1) Identifying potential units**

402   Before we discuss in more detail how acoustic units may be identified in terms of production,

403   perception, and analysis methods, we point out here that practically all such efforts require

404   scientists to identify potential units at some early stage of their planned investigation or analysis.

405   Two practical considerations are noteworthy.

406   First, a potential unit can be considered that part of a sequence that can be replaced with a

407   label for analysis purposes (e.g. unit *A* or unit *B*), without adversely affecting the results of a

408   planned investigation or analysis. Because animal acoustic sequences are sometimes hierarchical

409   in nature, e.g. humpback whale *Megaptera novaengliae* song, reviewed in Cholewiak, Sousa-

410   Lima & Cerchio (2012), distinct sequences of units may themselves be organised into longer,

411   distinctive sequences, i.e. "sequences of sequences" (Berwick *et al.*, 2011). Thus, an important

412   consideration in identifying potential acoustic units for sequence analyses is that they can be

413   hierarchically nested, such that a sequence of units can itself be considered as a unit and replaced

414   with a label.

415   Second, potential acoustic units are almost always identified based on acoustic features

416   present in a spectrographic representation of the acoustic waveform. Associating combinations of

417   these features with a potential unit can be performed either manually (i.e. examining the

418   spectrograms "by eye"), or automatically by using algorithms for either supervised classification

419   (where sounds are placed in categories according to pre-defined exemplars) or unsupervised

420   clustering (where labelling units is performed without prior knowledge of the types of units that

421   occur). We return to these analytical methods in Section III-4, and elaborate here on

422   spectrographic representations.

423   Spectrograms (consisting of discrete Fourier transforms of short, frequently overlapped,

424   segments of the signal) are ubiquitous and characterise well those acoustic features related to

425   spectral profile and frequency modulation, many of which are relevant in animal acoustic

426   communication. Examples of such features include minimum and maximum fundamental

427   frequency, slope of the fundamental frequency, number of inflection points, and the presence of

428   harmonics (Oswald *et al.*, 2007) that vary, for example, between individuals (Buck & Tyack,

429   1993; Blumstein & Munos, 2005; Koren & Geffen, 2011; Ji *et al.*, 2013; Kershenbaum, Sayigh

430   & Janik, 2013; Root-Gutteridge *et al.*, 2014), and in different environmental and behavioural

431   contexts (Matthews *et al.*, 1999; Taylor, Reby & McComb, 2008; Henderson, Hildebrand &

432   Smith, 2011).

433   Other less-used analytical techniques, such as cepstral analysis, may provide additional

434 detail on the nature of acoustic units, and are worth considering for additional analytical depth.

435 Cepstra are the Fourier (or inverse Fourier) transform of the log of the power spectrum

436 (Oppenheim & Schafer, 2004), and can be thought of as producing a spectrum of the power

437 spectrum. Discarding coefficients can yield a compact representation of the spectrum (Fig. 3).

438 Further, while Fourier transforms have uniform temporal and frequency resolution, other

439 techniques vary this resolution by using different basis sets, and this provides improved

440 frequency resolution at low frequencies and better temporal resolution at higher frequencies.

441 Examples of these other techniques include multi-taper spectra (Thomson, 1982; Tchernichovski

442 *et al*., 2000; Baker & Logue, 2003), Wigner–Ville spectra (Martin & Flandrin, 1985; Cohn,

443 1995), and wavelet analysis (Mallat, 1999). While spectrograms and cepstra are useful for

444 examining frequency-related features of signals, they are less useful when analysing temporal

445 patterns of amplitude modulation. This is an important issue worth bearing in mind, because

446 amplitude modulations are probably critical in signal perception by many animals (Henry *et al*.,

447 2011), including speech perception by humans (Remez *et al*., 1994).

448

(2) Identifying production units

450 One important approach to identifying acoustic units stems from considering the mechanisms for

451 sound production. In stridulating insects, for example, relatively simple, repeated sounds are

452 typically generated by musculature action that causes hard physical structures to be engaged,

453 such as the file and scraper located on the wings of crickets or the tymbal organs of cicadas

454 (Gerhardt & Huber, 2002). The resulting units, variously termed "chirps," or, "pulses," can be

455 organised into longer temporal sequences often termed "trills" or "echemes" (Ragge &

456 Reynolds, 1988). Frogs can produce sounds with temporally structured units in a variety of ways

457 (Martin & Gans, 1972; Martin, 1972; Gerhardt & Huber, 2002). In some species, a single

458 acoustic unit (sometimes called a "pulse," "note," or a "call") is produced by a single contraction

459 of the trunk and laryngeal musculature that induces vibrations in the vocal folds (e.g. Girgenrath

460 & Marsh, 1997). In other instances, frogs can generate short sequences of distinct sound units

461 (also often called "pulses") produced by the passive expulsion of air forced through the larynx

462 that induces vibrations in structures called arytenoid cartilages, which impose temporal structure

463 on sound (Martin & Gans, 1972; Martin, 1972). Many frogs organise these units into trills (e.g.

464 Gerhardt, 2001), while other species combine acoustically distinct units (e.g. Narins, Lewis &

465 McClelland, 2000; Larson, 2004). In songbirds, coordinated control of the two sides of the syrinx

466 can be used to produce different units of sound, or "notes" (Suthers, 2004). These units can be

467 organised into longer sequences, of "notes," "trills," "syllables," "phrases," "motifs," and

468 "songs" (Catchpole & Slater, 2003). In most mammals, sounds are produced as an air source

469 (pressure squeezed from the lungs) causes vibrations in the vocal membranes, which are then

470 filtered by a vocal tract (Titze, 1994). When resonances occur in the vocal tract, certain

471 frequencies known as formants are reinforced. Formants and formant transitions have been

472 strongly implicated in human perception of vowels and voiced consonants, and may also be used

473 by other species to perceive information (Peterson & Barney, 1952; Raemaekers, Raemaekers &

474 Haimoff, 1984; Fitch, 2000).

475   As the variety in these examples illustrates, there is incredible diversity in the

476 mechanisms animals use to produce the acoustic units that are subsequently organised into

477 sequences. Moreover, there are additional mechanisms that constrain the production of some of

478 the units. For example, in zebra finches *Taeniopygia guttata*, songs can be interrupted between

479   some of its constitutive units but not others (Cynx, 1990). This suggests that at a neuronal level,

480   certain units share a common, integrated neural production mechanism. Such examples indicate

481   that identifying units based on metrics of audition or visual inspection of spectrograms (e.g.

482   based on silent gaps) may not always be justified, and that there may be essential utility that

483   emerges from a fundamental understanding of unit production. Thus, a key consideration in

484   identifying functional units of production is that doing so may often require knowledge about

485   production mechanisms that can only come about through rigorous experimental studies.

486

487   **(3) Identifying perceptual units**

488   While there may be fundamental insights gained from identifying units based on a detailed

489   understanding of sound production, there may not always be a one-to-one mapping of the units

490   of production or the units identified in acoustics analyses, onto units of perception (e.g.

491   Blumstein, 1995). Three key considerations should be borne in mind when thinking about units

492   of perception and the analysis of animal acoustic sequences (Fig. 4).

493   First, it is possible that units of production or the units a scientist might identify on a

494   spectrogram are *perceptually bound* together by receivers into a single unit of perception (Fig.

495   4A). In this sense, a unit of perception is considered a perceptual auditory object in terms

496   familiar to cognitive psychologists and auditory scientists. There are compelling reasons for

497   researchers to consider vocalisations and other sounds as auditory objects (Miller & Cohen,

498   2010). While the rules governing auditory object formation in humans have been well studied

499   (Griffiths & Warren, 2004; Bizley & Cohen, 2013), the question of precisely how, and to what

500   extent, non-humans group acoustic information into coherent perceptual representations remains

501   a largely open empirical question (Hulse, 2002; Bee & Micheyl, 2008; Miller & Bee, 2012).

502   Second, studies of *categorical perception* in humans and other animals (Harnad, 1990)

503   show that continuous variation can nevertheless be perceived as forming discrete categories. In

504   the context of units of perception, this means that the graded variation often seen in spectrograms

505   may nevertheless be perceived categorically by receivers (Fig. 4B). Thus, in instances where

506   there are few discrete differences in production mechanisms or in spectrograms, receivers might

507   still perceive distinct units (Nelson & Marler, 1989; Baugh, Akre & Ryan, 2008).

508   Third, well-known *perceptual constraints* related to the limits of spectrotemporal

509   resolution may identify units of perception in ways that differ from analytical units and the units

510   of production (Fig. 4C). For example, due to temporal integration by the auditory system

511   (Recanzone & Sutter, 2008), some short units of production might be produced so rapidly that

512   they are not perceived as separate units. Instead, they might be integrated into a single percept

513   having a pitch proportional to the repetition rate. For example, in both bottlenose dolphins

514   *Tursiops truncatus* and Atlantic spotted dolphins *Stenella frontalis*, the "squawking" sound that

515   humans perceive as having some tonal qualities is actually a set of rapid echolocation clicks

516   known as a burst pulse (Herzing, 1996). The perceived pitch is related to the repetition rate, the

517   faster the repetition, the higher the pitch. Given the perceptual limits of gap detection

518   (Recanzone & Sutter, 2008), some silent gaps between units of production may be too short to be

519   perceived by the receiver. Clearly, while it may sometimes be desirable or convenient to use

520   "silence" as a way to create analysis boundaries between units, a receiver may not always

521   perceive the silent gaps that we see in our spectrograms. Likewise, some transitions in frequency

522   may reflect units of production that are not perceived because the changes remain unresolved by

523   auditory filters (Moore & Moore, 2003; Recanzone & Sutter, 2008). Indeed, some species may

524   be forced to trade off temporal and spectral resolution to optimise signalling efficiency in

525 different environmental conditions. Frequency modulated signals are more reliable than

526 amplitude modulation in reverberant habitats, such as forests, so woodland birds are adapted to

527 greater frequency resolution and poorer temporal resolution, while the reverse is true of

528 grassland species (Henry & Lucas, 2010; Henry *et al.*, 2011).

529 The question of what constitutes a unit that is perceptually meaningful to the animal

530 demands rigorous experimental approaches that put this question to the animal itself. There

531 simply is no convenient shortcut to identifying perceptual units. Experimental approaches

532 ranging from operant conditioning (e.g. Dooling *et al.*, 1987; Brown, Dooling & O'Grady, 1988;

533 Dent *et al.*, 1997; Tu, Smith & Dooling, 2011; Ohms *et al.*, 2012; Tu & Dooling, 2012), to field

534 playback experiments, often involving the habituation-discrimination paradigm (e.g. Nelson &

535 Marler, 1989; Wyttenbach, May & Hoy, 1996; Evans, 1997; Searcy, Nowicki & Peters, 1999;

536 Ghazanfar *et al.*, 2001; Weiss & Hauser, 2002). Such approaches have the potential to identify

537 the boundaries of perceptual units. Playbacks additionally can determine whether units can be

538 discriminated (as in 'go no-go' tasks stemming from operant conditioning), or whether they can

539 be recognised and are functionally meaningful to receivers.

540 Obviously some animals and systems are more tractable than others when it comes to

541 assessing units of perception experimentally, but those not easy to manipulate experimentally

542 (e.g. baleen whales, Balaenopteridae) should not necessarily be excluded from communication

543 sequence research, although the inevitable constraints must be recognised.

544

545 **(4) Identifying analytical units**

546 In many instances, it is desirable to analyse sequences of identified units in acoustic recordings

547 without having *a priori* knowledge about how those units may be produced or perceived by the

548   animals themselves. Such analyses are often a fundamental first step toward investigating the

549   potential meaning of acoustic sequences. We briefly discuss methods by which scientists can

550   identify and validate units for sequence analyses from acoustic recordings.

551   Sounds are typically assigned classifications to units based on the consistency of acoustic

552   characteristics. When feasible, external validation of categories (i.e. comparing animal

553   behavioural responses to playback experiments) should be performed. Even without directly

554   testing hypotheses of biological significance by playback experiment, there may be other

555   indicators of the validity of a classification scheme based purely on acoustic similarity. For

556   example, naïve human observers correctly divide dolphin signature whistles into groups

557   corresponding closely to the individuals that produced them (Sayigh *et al*., 2007), and similar

558   (but poorer) results are achieved using quantitative measures of spectrogram features

559   (Kershenbaum *et al*., 2013).

560   When classifying units on the basis of their acoustic properties, errors can occur both as

561   the result of perceptual bias, and as the result of poor repeatability. Perceptual bias occurs either

562   when the characteristics of the sound that are used to make the unit assignment are inappropriate

563   for the communication system being studied, or when the classification scheme relies too heavily

564   on those acoustic features that appear important to human observers. For example, analysing

565   spectrograms with a 50 Hz spectral resolution would be appropriate for human speech, but not

566   for Asian elephants *Elephas maximus*, which produce infrasonic calls that are typically between

567   14 and 24 Hz (Payne, Langbauer & Thomas, 1986), as details of the elephant calls would be

568   unobservable. Features that appear important to human observers may include tonal modulation

569   shapes, often posed in terms of geometric descriptors, such as "upsweep", "concave", and "sine"

570   (e.g. Bazúa-Durán & Au, 2002), which are prominent to the human eye, but may or may not be

571   of biological relevance. Poor repeatability, or variance, can occur both in human classification,

572   as inter-observer variability, and in machine learning, where computer classification algorithms

573   can make markedly different decisions after training with different sets of data that are very

574   similar (overtraining). Poor repeatability can be a particular problem when the classification

575   scheme ignores, or fails to give sufficient weight to, the features that are of biological

576   significance, or the algorithm (human or machine) places too much emphasis on particular

577   classification cues that are specific to the examples used to learn the categories. Repeatability

578   suffers particularly when analysing signals in the presence of noise, which can mask fine

579   acoustic details (Kershenbaum & Roch, 2013).

580   Three approaches have been used to classify units by their acoustic properties: visual

581   classification of spectrograms, quantitative classification using features extracted visually from

582   spectrograms, and fully automatic algorithms that assign classifications based on mathematical

583   rules.

584

585   (*a*) *Visual classification, "by eye"*

586   Traditionally, units are "hand-scored" by humans searching for consistent patterns in

587   spectrograms (or even listening to sound recordings without the aid of a spectrogram). Visual

588   classification has been an effective technique that has led to many important advances in the

589   study both of birdsong (e.g. Kroodsma, 1985; Podos *et al*., 1992; reviewed in Catchpole &

590   Slater, 2003), and acoustic sequences in other taxa (e.g. Narins *et al*., 2000; Larson, 2004).

591   Humans are usually considered to be good at visual pattern recognition – and better than most

592   computer algorithms (Ripley, 2007; Duda, Hart & Stork, 2012), which makes visual

593   classification an attractive approach to identifying acoustic units. However, drawbacks to visual

594 classification exist (Clark *et al.*, 1987). Visual classification is time consuming and prevents

595 taking full advantage of large acoustic data sets generated by automated recorders. Similarly, the

596 difficulty in scoring large data sets means that sample sizes used in research may be too small to

597 draw firm conclusions (Kershenbaum, 2013). Furthermore, visual classification can be prone to

598 subjective errors (Jones, ten Cate & Bijleveld, 2001), and inter-observer reliability should be

599 used (and reported) as a measure of the robustness of the visual assessments (Burghardt *et al.*,

600 2012).

601

602 (*b*) *Classification of manually extracted metrics*

603 As an alternative to visual classification, specific metrics, or features, measured on the acoustic

604 data can be extracted for input to classification algorithms. A variety of time (e.g. duration, pulse

605 repetition rate) and frequency (e.g. minimum, maximum, start, end, and range) components can

606 be measured (extracted) from spectrograms, using varying degrees of automation, or computer

607 assistance for a manual operator. Software tools such as Sound Analysis Pro (Tchernichovski *et*

608 *al.*, 2000), Raven (Charif, Ponirakis & Krein, 2006), and Avisoft (Specht, 2004) have been

609 developed to assist with this task. Metrics are then used in classification analyses to identify

610 units, using mathematical techniques such as discriminant function analysis (DFA), principal

611 components analysis (PCA), or classification and regression trees (CART), and these have been

612 applied to many mammalian and avian taxa (e.g. Derégnaucourt *et al.*, 2005; Dunlop *et al.*, 2007;

613 Garland *et al.*, 2012; Grieves, Logue & Quinn, 2014). Feature extraction can be conducted using

614 various levels of automation. A human analyst may note specific features for each call, an

615 analyst-guided algorithm can be employed (where sounds are identified by the analyst placing a

616 bounding box around the call, followed by automatic extraction of a specific number of features),

617   or the process of extraction can be fully automated. Automated techniques can be used to find

618   regions of possible calls that are then verified and corrected by a human analyst (Helble *et al.*,

619   2012).

620

621   (*c*) *Fully automatic metric extraction and classification*

622   Fully automated systems have the advantage of being able to handle large data sets. In principle,

623   automatic classification is attractive as it is not susceptible to the inter-observer variability of

624   visual classification (Tchernichovski *et al.*, 2000). However, current implementations generally

625   fall short of the performance desired (Janik, 1999), for instance by failing to recognise subtle

626   features that can be detected both by humans, and by the focal animals. Visual classification has

627   been shown to out-perform automated systems in cases where the meaning of acoustic signals is

628   known *a priori* (e.g. Sayigh *et al.*, 2007; Kershenbaum *et al.*, 2013), possibly because the

629   acoustic features used by fully automated systems may not reflect the cues used by the focal

630   species. However, once an automatic algorithm is defined, large data sets can be analysed.

631   Machine assistance can allow analysts to process much larger data sets than before, but at the

632   risk of possibly missing calls that they might have been able to detect.

633   The metrics generated either by manual or automatic extraction must be passed to a

634   classification algorithm, to separate detections into discrete unit types. Classification algorithms

635   can accept acoustic data with varying degrees of pre-processing as inputs. For example, in

636   addition to the commonly used spectrograms (Picone, 1993), cepstra (Oppenheim & Schafer,

637   2004), multi-taper spectra (Thomson, 1982), wavelets (Mallat, 1999), and formants (Fitch, 1997)

638   may be used, as they provide additional information on the acoustic characteristics of units,

639   which may not be well represented by traditional spectrograms (Tchernichovski *et al.*, 2000).

640   Each of these methods provide analysis of the spectral content of a short segment of the acoustic

641   production, and algorithms frequently examine how these parameters are distributed or change

642   over time (e.g. Kogan & Margoliash, 1998).

643

644   (*d*) *Classification algorithms*

645   Units may be classified automatically using supervised algorithms, in which the algorithm is

646   taught to recognise unit types given some *a priori* known exemplars, or clustered using

647   unsupervised algorithms, in which no *a priori* unit type assignment is known (Duda *et al*., 2012).

648   In both cases, the biological relevance of units must be verified independently because mis-

649   specification of units can obscure sequential patterns. Environmental noise or sounds from other

650   species may be mistakenly classified as an acoustic unit, and genuine units may be assigned to

651   incorrect unit categories. When using supervised algorithms, perceptual bias may lead to

652   misinterpreting data when the critical bands, temporal resolution, and hearing capabilities of a

653   species are not taken into account. For instance, the exemplars themselves used in supervised

654   clustering may be subject to similar subjective errors that can occur in visual classification.

655   However, validation of unsupervised clustering into units is also problematic, where clustering

656   results cannot be assessed against known unit categories. The interplay between unit

657   identification and sequence model validation is a non-trivial problem (e.g. Jin & Kozhevnikov,

658   2011). Similarly, estimating uncertainty in unit classification and assessing how that uncertainty

659   affects conclusions from a sequence analysis is a key part of model assessment (Duda *et al*.,

660   2012)

661   When using supervised classification, one appropriate technique for measuring

662   classification uncertainty is cross-validation (Arlot & Celisse, 2010). For fully unsupervised

29

663   clustering algorithms, where the desired classification is unknown, techniques exist to quantify

664   the stability of the clustering result, as an indicator of clustering quality. Examples include

665   "leave-k-out" (Manning, Raghavan & Schütze, 2008), a generalisation of the "leave-one-out"

666   cross-validation, and techniques based on normalised mutual information (Zhong & Ghosh,

667   2005), which measure the similarity between two clustering schemes (Fred & Jain, 2005).

668   However, it must be clear that cluster stability (and correspondingly, inter-observer reliability) is

669   not evidence that the classification is appropriate (i.e. matches the true, unknown, biologically

670   relevant categorisation), or will remain stable upon addition of new data (Ben-David, Von

671   Luxburg & Pál, 2006). Other information theoretic tests provide an alternative assessment of the

672   validity of unsupervised clustering results, such as checking if units follow Zipf's law of

673   abbreviation, which is predicted by a universal principle of compression (Zipf, 1949; Ferrer-i-

674   Cancho *et al*., 2013) or Zipf's law for word frequencies, which is predicted by a compromise

675   between maximizing the distinctiveness of units and the cost of producing them (Zipf, 1949;

676   Ferrer-i-Cancho, 2005).

677

678   **(5) Unit choice protocol**

679   The definition of a unit for a particular focal species and a particular research question is

680   necessarily dependent on a large number of factors in each specific project, and cannot be

681   concisely summarised in a review of this length. In particular, availability or otherwise of

682   behavioural information, such as the responses of individuals to playback experiments, is often

683   the determining factor in deciding how to define a sequence unit. However, we provide here a

684   brief protocol that can be used in conjunction with such prior information, or in its absence, to

685   guide the researcher in choosing the definition of a unit. This protocol is also represented

686 graphically in Fig. 5. (*a*) Determine what is known about the production mechanism of the

687 signalling individual. For example, Fig. 5A lists eight possible production types that produce

688 notably different sounds, although clearly other categories are also possible. (*b*) Determine what

689 is known about the perception abilities of the receiving individual. Perceptual limitations may

690 substantially alter the structure of production units. Fig. 5B gives examples of typical

691 modifications resulting from reduced temporal or spectral resolution at the receiver. (*c*) Choose a

692 classification method, such as manual, semi-automatic, or fully automatic (Fig. 5C). Some

693 putative unit types lend themselves more readily to certain classification techniques than others.

694 For example, "separated by silence" is often well distinguished by manual inspection of

695 spectrograms "by eye" or a band-limited energy detector, whereas "changes in acoustic

696 properties" may benefit from manual extraction of features for passing to a classification

697 algorithm (semi-automatic definition), and "series of sounds" may lend itself to a fully automatic

698 classification approach.

699

700 **IV. INFORMATION-EMBEDDING PARADIGMS**

701 A "sequence" can be defined as an ordered list of units. Animals produce sequences of sounds

702 through a wide range of mechanisms (e.g. vocalisation, stridulation, percussion), and different

703 uses of the sound-producing apparatus can produce different sound "units" with distinct and

704 distinguishable properties. The resulting order of these varied sound units may or may not

705 contain information that can be interpreted by a receiver, irrespective of whether or not the

706 signaller intended to convey meaning. Given that a sequence must consist of more than one

707 "unit" of one or more different types, the delineation and definition of the unit types is clearly of

708 vital importance. We have discussed this question at length in Section III. However, assuming

709   that units have been successfully assigned short-hand labels (e.g. *A, B, C*, etc.), what different

710   methods can be used to arrange these units in a sequence, in such a way that the sequence can

711   contain information?

712   Although it seems intuitively obvious that a sequence of such labels may contain

713   information, this intuition arises from our own natural human dispensation to language and

714   writing, and may not be particularly useful in identifying information in animal sequences. We

715   appreciate that birdsong, for instance, can be described as a complex combination of notes, and

716   we may be tempted to compare this animal vocalisation to human music (Baptista & Keister,

717   2005; Araya-Salas, 2012; Rothenberg *et al*., 2013). An anthropocentric approach, however, is not

718   likely in all cases to identify structure relevant to animal communication. Furthermore, wide

719   variation can be expected between the structure of sequences generated by different taxa, from

720   the pulse-based stridulation of insects (Gerhardt & Huber, 2002) to song in whales (reviewed in

721   Cholewiak *et al*., 2012), and a single analytical paradigm derived from a narrow taxonomic view

722   is also likely to be inadequate. A more rigorous analysis is needed, one that indicates the

723   fundamental structural properties of acoustic sequences, in all their diversity. Looking for

724   information only, say, in the order of units can lead researchers to miss information encoded in

725   unit timing, or pulse rate.

726   Although acoustic information can be encoded in many different ways, we consider here

727   only the encoding of information *via* sequences. We suggest a classification scheme based on six

728   distinct paradigms for encoding information in sequences (Fig. 6): (*a*) *Repetition*, where a single

729   unit is repeated more than once; (*b*) *Diversity*, where information is represented by the number of

730   distinct units present; (*c*) *Combination*, where sets of units have different information from each

731   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 believe 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 that

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

755   producing and another responding with another coda (Schulz *et al*., 2008). Each of these

756   paradigms is now described in more detail below.

757

(1) Repetition

759   Sequences are made of repetitions of discrete units, and repetitions of the same unit affect

760   receiver responses. For instance, the information contained in a unit *A* given in isolation may

761   convey a different meaning to a receiver than an iterated sequence of unit *A* (e.g. *AAAA*, etc.).

762   For example, greater numbers of D notes in the *chick-a-dee* calls of chickadee species *Poecile*

763   spp. can be related to the immediacy of threat posed by a detected predator (Krams *et al*., 2012).

764   Repetition in alarm calls is related to situation urgency in meerkats *Suricata suricatta* (Manser,

765   2001), marmots *Marmota spp.* (Blumstein, 2007), colobus monkeys *Colobus* spp. (Schel,

766   Candiotti & Zuberbühler, 2010), Campbell's monkeys *Cercopithecus campbelli* (Lemasson *et*

767   *al*., 2010) and lemurs *Lemur catta* and *Varecia variegata* (Macedonia, 1990).

768

(2) Diversity

770   Sequences of different units (e.g. *A, B, C*) are produced, but those units are functionally

771   interchangeable, and therefore ordering is unimportant. For instance, many songbirds produce

772   songs with multiple different syllables. In many species, however, the particular syllables are

773   substitutable (e.g. Eens, Pinxten & Verheyen, 1991; Farabaugh & Dooling, 1996; but see

774   Lipkind *et al*., 2013), and receivers attend to the overall diversity of sounds in the songs or

775   repertoires of signallers (Catchpole & Slater, 2003). Large acoustic repertoires have been

776   proposed to be sexually selected in species such as great reed warblers *Acrocephalus*

777   *arundinaceus* and common starlings *Sturnus vulgaris* (Eens, Pinxten & Verheyen, 1993;

778   Hasselquist, Bensch & von Schantz, 1996; Eens, 1997), in which case diversity embeds

779   information (that carries meaning) on signaller quality (e.g. Kipper *et al*., 2006). Acoustic

780   "diversity" has additionally been proposed as a means of preventing habituation on the part of

781   the receiver (Hartshorne, 1956, 1973; Kroodsma. 1990) as well as a means of avoiding

782   (neuromuscular) "exhaustion" on the part of the sender (Lambrechts & Dhondt, 1987, 1988). We

783   do note that these explanations remain somewhat controversial, especially if the transitions

784   between acoustic units are, indeed, biologically constrained (Weary & Lemon, 1988, 1990;

785   Weary *et al*., 1988; Weary, Lambrechts & Krebs, 1991; Riebel & Slater, 2003; Brumm & Slater,

786   2006).

787

788   **(3) Combination**

789   Sequences may consist of different discrete acoustic units (e.g. *A, B, C*) each of which is itself

790   meaningful, and the combining of the different units conveys distinct information. Here, order

791   does not matter (in contrast to the *Ordering* paradigm below) – the sequence of unit *A* followed

792   by unit *B* has the same information as the sequence of unit *B* followed by unit *A*. For example,

793   titi monkeys *Callicebus nigrifrons* (Cäsar *et al*., 2013) use semantic alarm combinations, in

794   which interspersing avian predator alarms calls (A-type) with terrestrial predator alarm calls (B-

795   type) indicates the presence of a raptor on the ground. In this case, the number of calls (i.e.

796   *Repetition*) also appears to influence the information present in each call sequence (Cäsar *et al*.,

797   2013).

798

799   **(4) Ordering**

35

800   Sequences of different discrete acoustic units (e.g. *A, B, C*) each of which is itself meaningful

801   and the specific order of which is meaningful. Here, order matters – and the ordered combination

802   of discrete units may result in emergent responses. For instance, *A* followed by *B* may elicit a

803   different response than either *A* or *B* alone, or *B* followed by *A*. Examples include primate alarm

804   calls which, when combined, elicit different responses related to the context of the predatory

805   threat (Arnold & Zuberbühler, 2006*b*, 2008). Human languages are a sophisticated example of

806   ordered information encoding (Hauser, Chomsky & Fitch, 2002). When sequences have complex

807   ordering, simple quantitative measures are unlikely to capture the ordering information. Indeed,

808   the Kolmogorov complexity of a sequence indicates how large a descriptor is required to specify

809   the sequence adequately (Denker & Woyczyński, 1998). Instead of quantifying individual

810   sequences, an alternative approach to measuring ordering is to calculate the pairwise similarity or

811   difference between two sequences, using techniques such as the Levenshtein or Edit distance

812   (Garland *et al*., 2012; Kershenbaum *et al*., 2012).

813

814   **(5) Overlapping**

815   Sequences are combined from two or more individuals into exchanges for which the order of

816   these overlapping sequences has information distinct from each signaller's signals in isolation.

817   Overlapping can be in the time dimension (i.e. two signals emitted at the same time) or in

818   acoustic space, e.g. song-type matching (Krebs, Ashcroft & Orsdol, 1981), and frequency

819   matching (Mennill & Ratcliffe, 2004). For example, in different parid species (Paridae:

820   chickadees, tits, and titmice), females seem to attend to the degree to which their males' songs

821   are overlapped (in time) by neighbouring males' songs, and seek extra-pair copulations when

822   their mate is overlapped (Otter *et al*., 1999; Mennill, Ratcliffe & Boag, 2002). Overlapping is

823 also used for social bonding, spatial perception, and reunion, such as chorus howls in wolves

824 (Harrington *et al.*, 2003) and sperm whale codas (Schulz *et al.*, 2008). Overlapping as song-type

825 matching (overlapping in acoustic space) is also an aggressive signal in some songbirds (Akçay

826 *et al.*, 2013), although this may depend on whether it is the sequence or the individual unit that is

827 overlapped (Searcy & Beecher, 2011). Coordination between the calling of individuals can also

828 give identity cues (Carter *et al.*, 2008). However, despite the apparent widespread use of

829 overlapping in sequences, few analytical models have been developed to address this

830 mechanism. While this is a promising area for future research, it is currently beyond the purview

831 of this review.

832

833 **(6) Timing**

834 The temporal spacing between units in a sequence can contain information. In the simplest case,

835 pulse rate and interpulse interval can distinguish between different species, for example in

836 insects and anurans (Gerhardt & Huber, 2002; Nityananda & Bee, 2011), rodents (Randall,

837 1997), and primates (Hauser, Agnetta & Perez, 1998). Call timing can indicate fitness and

838 aggressive intent, e.g. male howler monkeys *Alouatta pigra* attend to howling delay as an

839 indicator of aggressive escalation (Kitchen, 2004). Additionally, when sequences are produced

840 by different individuals, a receiver may interpret the timing differences between the producing

841 individuals to obtain contextual information. For instance, ground squirrels *Spermophilus*

842 *richarsonii* use the spatial pattern and temporal sequence of conspecific alarm calls to provide

843 information on a predator's movement trajectory (Thompson & Hare, 2010). This information

844 only emerges from the sequence of different callers initiating calls (Blumstein, Verneyre &

845 Daniel, 2004). Such risk tracking could also emerge from animals responding to sequences of

846 heterospecific alarm signals produced over time.

847

(7) Information-embedding paradigms: conclusions

849 The use of multiple embedding techniques may be quite common, for instance in intrasexual

850 competitive and intersexual reproductive contexts (Gerhardt & Huber, 2002). For example, many

851 frog species produce pulsatile advertisement calls consisting of the same repeated element. If it is

852 the case that both number of pulses and pulse rate affect receiver responses, as shown in some

853 hylid treefrogs (Gerhardt, 2001), then information is being embedded using both the *Repetition*

854 and the *Timing* paradigms simultaneously.

855 Before hypothesising a specific structuring paradigm, it is frequently useful to perform

856 exploratory data analysis (Fig. 7). This might begin by looking at histograms, networks, or low-

857 order Markov models that are based on acoustic units or timing between units. This analysis can

858 be on the raw acoustic units or may involve preprocessing. An example of preprocessing that

859 might be helpful for hypothesising *Repetition* would be to create histograms that count the

860 number of times that acoustic units occur within a contiguous sequence of vocalisations. As an

861 example, if 12 different acoustic units each occurred three times, a histogram bin representing

862 three times would have a value of 12; for examples, see Jurafsky & Martin (2000). For

863 histograms or networks, visual analysis can be used to determine if there are any patterns that

864 bear further scrutiny. Metrics such as entropy can be used to provide an upper bound on how

865 well a Markov chain model describes a set of vocalisations (smaller numbers are better, as an

866 entropy of zero indicates that we model the data perfectly). If nothing is apparent, it might mean

867 that there is no structure to the acoustic sequences, but it also possible that the quantity of data

868   are insufficient to reveal the structure or that the structure is more complex than what can be

869   revealed through casual exploratory data analysis.

870   Exploratory data analysis may lead to hypotheses that one or more of the embedding

871   paradigms for acoustic sequences may be appropriate. At this point a greater effort should be put

872   into the modelling and understanding and we provide a suggested flow of techniques (Fig. 7). It

873   is important to keep in mind that these are only suggestions. For example, while we suggest that

874   a grammar (Section V.4) be modelled if there is evident and easily described structure for

875   *Repetition*, *Diversity*, and *Ordering*, other models could be used effectively and machine

876   learning techniques for generating grammars may be able to do so when the structure is less

877   evident.

878   We conclude this section with a discussion of two examples of how sequences of acoustic

879   signals produced by signallers can influence meaning to receivers. These two examples come

880   from primates and exemplify the *Diversity* and *Ordering* types of sequences illustrated in Fig. 6.

881   The example of the *Diversity* type is the system of serial calls of titi monkeys, *Callicebus*

882   *molloch*, used in a wide range of social interactions. Here, the calls comprise several distinct

883   units, many of which are produced in sequences. Importantly, the units of this call system seem

884   to have meaning primarily in the context of the sequence – this call system therefore seems to

885   represent the notion of phonological syntax (Marler, 1977). One sequence has been tested *via*

886   playback studies – the 'honks–bellows–pumps' sequence is used frequently by males that are

887   isolated from and not closely associated with females and may recruit non-paired females

888   (Robinson, 1979). Robinson (1979) played back typical sequences of honks–bellows–pumps

889   sequences and atypical (i.e. reordered) sequences of honks–pumps–bellows and found little

890   evidence that groups of titi monkeys responded differently to the two playbacks (although they

891   gave one call type – a 'moan', produced often during disturbances caused by other conspecific or

892   heterospecific monkey groups – more often to the atypical sequences).

893   The second example relates to the *Ordering* type of sequence (Fig. 6), and stems from

894   two common calls of putty-nosed monkeys, *Cercopithecus nictitans martini*. 'Pyow' calls can be

895   produced individually or in strings of pyows, and seem to be used by putty-nosed monkeys

896   frequently when leopards are detected in the environment (Arnold & Zuberbühler, 2006*b*), and

897   more generally as an attention-getting signal related to recruitment of receivers and low-level

898   alarm (Arnold & Zuberbühler, 2013). 'Hack' calls can also be produced individually or in strings

899   of hacks, and seem to be used frequently when eagles are detected in the environment, and more

900   generally as a higher-level alarm call (Arnold & Zuberbühler, 2013). Importantly, pyow and

901   hack calls are frequently combined into pyow–hack sequences. Both naturalistic observational

902   data as well as experimental call playback results indicate that pyow–hack sequences influence

903   receiver behaviour differently than do pyow or hack sequences alone – pyow–hack sequences

904   seem to mean "let's go!" and produce greater movement distances in receivers (Arnold &

905   Zuberbühler, 2006*a*). The case of the pyow–hack sequence therefore seems to represent

906   something closer to the notion of lexical syntax – individual units and ordered combinations of

907   those units have distinct meanings from one another (Marler, 1977).

908   These two examples of primate calls illustrate the simple but important point that

909   sequences matter in acoustic signals – combinations or different linear orderings of units

910   (whether those units have meaning individually or not) can have different meanings to receivers.

911   In the case of titi monkeys, the call sequences seem to serve the function of female attraction for

912   male signallers, whereas in the case of putty-nosed monkeys, the call sequences serve anti-

913   predatory and group-cohesion functions.

914

V. ANALYSIS OF SEQUENCES

916 Given that the researcher has successfully determined the units of an acoustic sequence that are

917 appropriate for the hypothesis being tested, one must select and apply appropriate algorithms for

918 analysing the sequence of units. Many algorithms exist for the analysis of sequences: both those

919 produced by animals, and sequences in general (such as DNA, and stock market prices).

920 Selection of an appropriate algorithm can sometimes be guided by the quantity and variability of

921 the data, but there is no clear rule to be followed. In fact, in machine learning, the so-called 'no

922 free lunch' theorem (Wolpert & Macready, 1997) shows that there is no one pattern-recognition

923 algorithm that is best for every situation, and any improvement in performance for one class of

924 problems is offset by lower performance in another problem class. In choosing an algorithm for

925 analyses, one should be guided by the variability and quantity of the data for analysis, keeping in

926 mind that models with more parameters require more data to estimate the parameters effectively.

927   We consider five models in this section: (1) Markov chains, (2) hidden Markov models,

928 (3) network models, (4) formal grammars, and (5) temporal models. Each of these models has

929 been growing in popularity among researchers, with the number of publications increasing in

930 recent years. The number of publications in 2013 mentioning both the terms "animal

931 communication" as well as the model name has grown since 2005 by a factor of: "Markov", 4.9;

932 "hidden Markov", 3.3; "network", 2.6; "grammar" 1.7; "timing", 2.3.

933   The structure-analysis algorithms discussed throughout this section can be used to model

934 the different methods for combining units discussed earlier (Fig. 6). *Repetition*, *Diversity*, and

935 *Ordering* are reasonably well captured by models such as Markov chains, hidden Markov

936 models, and grammars. Networks capture structure either with or without order, although much

41

937   of the application of networks has been done on unordered associations (*Combination*).

938   Temporal information can be modelled as an attribute of an acoustic unit requiring extensions to

939   the techniques discussed below, or as a separate process. Table 2 summarises the assumptions

940   and requirements for each of these models.

941   Here we give a sample of some of the more important and more promising algorithms for

942   animal acoustic sequence analysis, and discuss ways for selecting and evaluating analytical

943   techniques. Selecting appropriate algorithms should involve the following steps. (*i*) Technique:

944   understand the nature of the models and their mathematical basis. (*ii*) Suitability: assess the

945   suitability of the models and their constraints with respect to the research questions being asked.

946   (*iii*) Application: apply the models to the empirical data (training, parameter estimation). (*iv*)

947   Assessment: extract metrics from the models that summarise the nature of the sequences

948   analysed. (*v*) Inference: compare metrics between data sets (or between empirical data and

949   random null-models) to draw ecological, mechanistic, evolutionary, and behavioural inferences.

950   (*vi*) Validate: determine the goodness of fit of the model to the data and uncertainty of parameter

951   estimates. Bootstrapping techniques can allow validation with sets that were not used in model

952   development.

953

(1) Markov chains

955   Markov chains, or N-grams models, capture structure in acoustic unit sequences based on the

956   recent history of a finite number of discrete unit types. Thus, the occurrence of a unit (or the

957   probability of occurrence of a unit) is determined by a finite number of previous units. The

958   history length is referred to as the order, and the simplest such model is a 0th order Markov

959   model, which assumes that each unit is independent of another, and simply determines the

42

960 probability of observing any unit with no prior knowledge. A 1st order Markov model is one in

961 which the probability of each unit occurring is determined only by the preceding unit, together

962 with the "transition probability" from one unit to the next. This transition probability is assumed

963 to be constant (stationary). Higher order Markov models condition the unit probabilities based on

964 more than one preceding units, as determined by the model order. An N-gram model conditions

965 the probability on the $N–1$ previous units, and is equivalent to an $N–1^{th}$ order Markov model. A

966 K^{th} order Markov model of a sequence with C distinct units is defined by at most a $C^K \times C$ matrix

967 of transition probabilities from each of the C^K possible preceding sequences, to each of the C

968 possible subsequent units, or equivalently by a state transition diagram (Fig. 8).

969   As the order of the model increases, more and more data are required for the accurate

970 estimation of transition probabilities, i.e. sequences must be longer, and many transitions will

971 have zero counts. This is particularly problematic when looking at new data, which may contain

972 sequences that were not previously encountered, as they will appear to have zero probability. As

973 a result, Markov models with orders greater than two (trigram, $N=3$) are rare. In principle, a Kth

974 order Markov model requires sufficient data to provide accurate estimates of C^{K+1} transition

975 probabilities. In many cases, the number of possible transitions is similar to, or larger than, the

976 entire set of empirical data. For example, Briefer *et al*. (2010) examined very extensive skylark

977 *Alauda arvensis* sequences totalling 16,829 units, but identified over 340 unit types. As a naïve

978 transition matrix between all unit types would contain $340 \times 340 = 115,600$ cells, the collected

979 data set would be too small to estimate the entire matrix. A different problem occurs when, as is

980 commonly the case, animal acoustic sequences are short. Kershenbaum *et al*. (2012) examined

981 hyrax *Procavia capensis* sequences that are composed of just five unit types. However, 81% of

982 the recorded sequences were only five or less units long. For these short sequences, $5^5 = 3125$

983 different combinations are possible – which is greater than the number of such sequences

984 recorded (2374). In these cases, estimates of model parameters, and conclusions drawn from

985 them, may be quite inaccurate (Cover & Thomas, 1991; Hausser & Strimmer, 2009;

986 Kershenbaum, 2013).

987   Closed-form expressions for maximum-likelihood estimates of the transition probabilities

988 can be used with conditional counts (Anderson & Goodman, 1957). For example, assuming five

989 acoustic units (*A–E*), maximum-likelihood estimates of the transition probabilities for a first-

990 order Markov model (bigram, *N*=2) can be found directly from the number of occurrences of

991 each transition, e.g.

992
$$P(B \mid A) = \frac{count(AB)}{\sum_{i \in \{A,B,C,D,E\}} count(A,i)}$$
(1)

993 Although not widely used in the animal communication literature, research in human natural

994 language processing has led to the development of methods known as back-off models (Katz,

995 1987), which account for the underestimated probability of rare sequences using Good–Turing

996 counts, a method for improving estimated counts for events that occur infrequently (Gale &

997 Sampson, 1995). When a particular state transition is never observed in empirical data, the back-

998 off model offers the minimum probability for this state transition so as not to rule it out

999 automatically during the testing. Standard freely available tools, such as the SRI language

1000 modelling toolkit (Stolcke, 2002), implement back-off models and can reduce the effort of

1001 adopting these more advanced techniques.

1002   Once Markovian transitions have been calculated and validated, the transition

1003 probabilities can be used to calculate a number of summary metrics using information theory

1004 (Shannon *et al.*, 1949; Chatfield & Lemon, 1970; Hailman, 2008). For a review on the

1005 mathematics underlying information theories, we direct the readers to the overview in McCowan,

1006 Hanser & Doyle (1999) or Freeberg & Lucas (2012), which provides the equations as well as a

1007 comprehensive reference list to other previous work. Here we will define these quantitative

1008 measures with respect to their relevance in analysing animal acoustic sequences. Zero-order

1009 entropy measures repertoire diversity:

$$H_0 = \log_2(C) \tag{2}$$

1011 where, $C=|V|$ is the cardinality of the set of acoustic units V. First-order entropy H_1 begins to

1012 measure simple repertoire internal organisational structure by evaluating the relative frequency

1013 of use of different signal types in the repertoire:

$$H_1 = \sum_{v_i \in V} -P(v_i) \log_2 P(v_i) \tag{3}$$

1015 where the probability $P(v_i)$ of each acoustic unit i is typically estimated based on frequencies of

1016 occurrence, as described earlier with N-grams. Higher-order entropies measure internal

1017 organisational structure, and thus one form of communication complexity, by examining how

1018 signals interact within a repertoire at the two-unit sequence level, the three-unit sequence level,

1019 and so forth.

1020 One inferential approach is to calculate the entropic values from first-order and higher-

1021 order Markov models to summarise the extent to which sequential structure is present at each

1022 order. A random sequence would show no dependence of entropy on Markov order, whereas

1023 decreases in entropy as the order is increased would be an indication of sequential organisation,

1024 and thus higher communication complexity (Ferrer-i-Cancho & McCowan, 2012). These

1025 summary measures can then be further extended to compare the importance of sequential

1026 structure across different taxa, social and ecological contexts. These types of comparisons can

1027 provide novel insights into the ecological, environmental, social, and contextual properties that

1028   shape the structure, organisation, and function of signal repertoires (McCowan, Doyle & Hanser,

1029   2002).

1030   The most common application of the Markov model is to test whether or not units occur

1031   independently in a sequence. Model validation techniques include the sequential and χ^2 tests

1032   (Anderson & Goodman, 1957). For instance, Narins *et al*. (2000) used a permutation test (Adams

1033   & Anthony, 1996) to evaluate the hypothesis that a frog with an exceptionally large vocal

1034   repertoire, *Bufo madagascariensis*, emitted any call pairs more often than would be expected by

1035   chance. Similar techniques were used to show non-random call production by Sayigh *et al*.

1036   (2012) with short-finned pilot whales *Globicephala macrorhynchus*, and by Bohn *et al*. (2009)

1037   with free-tailed bats *Tadarida brasiliensis*. However, deviation from statistical independence

1038   does not in itself prove a sequence to have been generated by a Markov chain. Other tests, such

1039   as N-gram distribution (Jin & Kozhevnikov, 2011) may be more revealing.

1040

1041   **(2) Hidden Markov models**

1042   Hidden Markov models (HMMs) are a generalisation of the Markov model. In Markov models,

1043   the acoustic unit history (of length *N*) can be considered the current "state" of the system. In

1044   HMMs (Rabiner, 1989), states are not necessarily associated with acoustic units, but instead

1045   represent the state of some possibly unknown and unobservable process. Thus, the system

1046   progresses from one state to another, where the nature of each state is unknown to the observer.

1047   Each of these states may generate a "signal" (i.e. a unit), but there is not necessarily a one-to-one

1048   mapping between state transitions and signals generated. For example, transitioning to state *X*

1049   might generate unit *A*, but the same might be true of transitioning to state *Y*. An observation is

1050   generated at each state according to a state-dependent probability density function, and state

1051   transitions are governed by a separate probability distribution (Fig. 9). HMMs are particularly

1052   useful to model very complex systems, while still being computationally tractable.

1053   Extensions to the HMM model also exist, in which the state transition probabilities are

1054   non-stationary. For example, the probability of remaining in the same state may decay with time

1055   e.g. due to neural depletion, as shown by Jin & Kozhevnikov (2011), or recurrent units may

1056   appear more often than expected by a Markov model, particularly where behavioural sequences

1057   are non-Markovian (Cane, 1959; Kershenbaum, 2013; Kershenbaum *et al*., 2014). Also, HMMs

1058   are popular in speech analysis (Rabiner, 1989), where emissions are continuous-valued, rather

1059   than discrete.

1060   HMMs have been used fairly extensively in speaker recognition (Lee & Hon, 1989), the

1061   identification of acoustic units in birdsong (Trawicki, Johnson & Osiejuk, 2005), and other

1062   analyses of birdsong sequences. ten Cate, Lachlan & Zuidema (2013) reviewed analytical

1063   methods for inferring the structure of birdsong and highlighted the idea that HMM states can be

1064   thought of as possibly modelling an element of an animal's cognitive state. This makes it

1065   possible to build models that have multiple state distributions for the same acoustic unit

1066   sequence. For instance, in the trigram *AAC*, the probability given by the 2nd order Markov model,

1067 $P(C|A, A)$ is fixed. There cannot be different distributions for observing the unit *C*, if the

1068   previous two units are *A*. Yet cognitive state may have the potential to influence the probability

1069   of observing *C*, even for identical sequence contexts (*AA*). Another state variable (θ) exists

1070   unobserved, as it reflects cognitive state, rather than sequence history. In this case, $P(C|A,$

1071 $A,\theta=0)\neq P(C|A, A,\theta=1)$. Hahnloser, Kozhevnikov & Fee (2002), Katahira *et al*. (2011), and Jin

1072   (2009) have used HMMs to model the interaction between song and neural substrates in the

1073   brain. A more recent example of this can be seen in the work of Jin & Kozhevnikov (2011),

1074   where they used states to model neural units in song production of the Bengalese finch *Lonchura*

1075   *striata* ver. *domestica*, restricting each state to the emission of a single acoustic unit, thus making

1076   acoustic units associated with each state deterministic while retaining the stochastic nature of

1077   state transitions.

1078   Because the states of a HMM represent an unobservable process, it is difficult to estimate

1079   the number of states needed to describe the empirical data adequately. Model selection methods

1080   and criteria (for example Akaike and Bayesian information criteria, and others) can be used to

1081   estimate model order – see Hamaker, Ganapathiraju & Picone (1998) and Zucchini &

1082   MacDonald (2009) for a brief review – so the number of states is often determined empirically.

1083   Increasing the number of states permits the modelling of more complex underlying sequences

1084   (e.g. longer term dependencies), but increases the amount of data required for proper estimation.

1085   The efficiency and accuracy of model fitting depends on model complexity, so that models with

1086   many states, many acoustic units, and perhaps many covariates or other conditions will take

1087   more time and require more data to fit.

1088   During training, HMM parameters are estimated using an optimisation algorithm (Cappé,

1089   Moulines & Rydén, 2005) that finds a combination of hidden states, state transition tables, and

1090   state-dependent distributions that best describe the data. Software libraries for the training of

1091   HMMs are available in many formats, e.g. the Matlab function *hmmtrain*, the R package HMM

1092   (R Development Team, 2012), and the Hidden Markov Model Toolkit (Young & Young, 1994).

1093   Similar considerations of data set completeness exist to those when generating regular Markov

1094   models, most importantly, that long sequences of data are required.

1095   Although the states of a HMM are sometimes postulated to possess biologically relevant

1096   meaning, the internal states of the HMM represent a hidden process, and do not necessarily refer

1097   to concrete behavioural states. Specifically, the training algorithm does not contain an

1098   optimisation criterion that will necessarily associate model states with the functional or

1099   ecological states of the animal that a researcher is interested in observing (e.g. foraging, seeking

1100   a mate, etc.). While the functional/ecological state is likely related to the sequence, each model

1101   state may in fact represent a different subsequence of the data. Therefore, one cannot assume in

1102   general that there will be a one-to-one mapping between model and animal states. Specific

1103   hidden Markov models derived from different empirical data are often widely different, and it

1104   can be misleading to make comparisons between HMMs derived from different data sets.

1105   Furthermore, obtaining consistent states requires many examples with respect to the diversity of

1106   the sequence being modelled. An over-trained network will be highly dependent on the data

1107   presented to it and small changes in the training data can result in very different model

1108   parameters, making state-based inference questionable.

1109

(3) Network models

1111   The structure of an acoustic sequence can also be described using a network approach –

1112   reviewed in Newman (2003) and Baronchelli *et al*. (2013) – as has been done for other

1113   behavioural sequences, e.g. pollen processing by honeybees (Fewell, 2003). A node in the

1114   network represents a type of unit, and a directional edge connecting two nodes means that one

1115   unit comes after the other in the acoustic sequence. For example, if a bird sings a song in the

1116   order: *ABCABC*; the network representing this song will have three nodes for *A, B*, and *C*, and

1117   three edges connecting *A* to *B, B* to *C*, and *C* to *A* (Fig. 10). The edges may simply indicate

1118   association between units without order (undirected binary network), an ordered sequence

1119   (directed binary network), or a probability of an ordered sequence (directed weighted network),

1120   the latter being equivalent to a Markov chain (Newman, 2009).

1121   The network representation is fundamentally similar to the Markov model, and the basic

1122   input for constructing a binary network is a matrix of unit pairs within the repertoire, which

1123   corresponds to the transition matrix in a Markov model. However, the network representation

1124   may be more widely applicable than a Markov analysis, particularly when a large number of

1125   distinct unit types exist, precluding accurate estimation of transition probabilities (e.g. Sasahara

1126   *et al*., 2012; Weiss *et al*., 2014; Deslandes *et al*., 2014). In this case, binary or simple directed

1127   networks may capture pertinent properties of the sequence, even if transition probabilities are

1128   unknown.

1129   One of the attractive features of network analysis is that a large number of quantitative

1130   network measures exist for comparison to other networks (e.g. from different individuals,

1131   populations, or species), or for testing hypotheses. We list a few of the popular algorithms that

1132   can be used to infer the structure of the acoustic sequence using a network approach. We refer

1133   the reader to introductory texts to network analysis for further details (Newman, 2009; Scott &

1134   Carrington, 2011).

1135   *Degree centrality* measures the number of edges directly connected to each node. In a

1136   directed network, each node has an in-degree and an out-degree, corresponding to incoming and

1137   outgoing edges. The weighted version of degree centrality is termed strength centrality, which

1138   takes into account the weights of each edge (Barrat *et al*., 2004). Degree/strength centrality

1139   identifies the central nodes in the network, corresponding to central elements in the acoustic

1140   sequence. For example, in the mockingbird *Mimus polyglottos*, which imitates sounds of other

1141   species, its own song is central in the network, meaning that it usually separates between other

1142   sounds by singing its own song (Gammon & Altizer, 2011).

1143   *Betweenness centrality* is a measure of the role a central node plays in connecting other

1144   nodes. For example, if an animal usually uses three units before moving to another group of

1145   units, a unit that lies between these groups in the acoustic sequence will have high betweenness

1146   centrality. A weighted version of betweenness centrality was described in Opsahl, Agneessens &

1147   Skvoretz (2010).

1148   *Clustering coefficient* describes how many triads of nodes are closed in the network. For

1149   example, if unit *A* is connected to *B*, and *B* is connected to *C*, a cluster is formed if *A* is also

1150   connected to *C*. Directed and weighted versions of the clustering coefficient have also been

1151   described (Barrat *et al*., 2004; Fagiolo, 2007).

1152   *Mean path length* is defined as the average minimum number of connections to be

1153   crossed from any arbitrary node to any other. This measures the overall navigability in the

1154   network; as this value becomes large, a longer series of steps is required for any node to reach

1155   another.

1156   *Small-world metric* measures the level of connectedness of a network and is the ratio of

1157   the clustering coefficient *C* to the mean path length *L* after normalising each with respect to the

1158   clustering coefficient and mean path length of a random network: $S=(C/C_{rand})/(L/L_{rand})$. If $S > 1$

1159   the network is regarded as "small-world" (Watts & Strogatz, 1998; Humphries & Gurney, 2008),

1160   with the implication that nodes are reasonably well connected and that it does not take a large

1161   number of edges to connect most pairs of nodes. Sasahara *et al*. (2012) demonstrated that the

1162   network of California thrasher *Toxostoma redivivum* songs has a small-world structure, in which

1163   subsets of phrases are highly grouped and linked with a short mean path length.

1164   *Network motifs* are recurring structures that serve as building blocks of the network (Milo

1165   *et al*., 2002). For example, a network may feature an overrepresentation of specific types of

1166   triads, tetrads, or feed-forward loops. Network motif analysis could be informative in comparing

1167   sequence networks from different individuals, populations or species. We refer the reader to

1168   three software packages available for motif analysis: FANMOD (Wernicke & Rasche, 2006);

1169   MAVisto (Schreiber & Schwöbbermeyer, 2005); and MFinder (Kashtan *et al*., 2002).

1170   *Community detection* algorithms offer a method to detect network substructure

1171   objectively (Fortunato, 2010). These algorithms identify groups of nodes with dense connections

1172   between them but that are sparsely connected to other groups/nodes. Subgroups of nodes in a

1173   network can be considered somewhat independent components of it, offering insight into the

1174   different subunits of acoustic sequences. Multi-scale community detection algorithms can be

1175   useful for detecting hierarchical sequence structures (Fushing & McAssey, 2010; Chen &

1176   Fushing, 2012).

1177   *Exponential family Random Graph Models* (ERGMs) offer a robust analytic approach to

1178   evaluate the contribution of multiple factors to the network structure using statistical modelling

1179   (Snijders, 2002). These factors may include structural factors (e.g. the tendency to have closed

1180   triads in the network), and factors based on node or edge attributes (e.g. a tendency for

1181   connections between nodes that are acoustically similar). The goal of ERGMs is to predict the

1182   joint probability that a set of edges exists on nodes in a network. The R programming language

1183   package *statnet* has tools for model estimation and evaluation, and for model-based network

1184   simulation and network visualisation (Handcock *et al*., 2008).

1185   As with other models, many statistical tests for inference and model assessment require a

1186   comparison of the observed network to a set of random networks. For example, the clustering

1187 coefficient of an observed network can be compared to those of randomly generated networks, to

1188 test if it is significantly smaller or larger than expected. A major concern when constructing

1189 random networks is what properties of the observed network should be retained (Croft, James &

1190 Krause, 2008). The answer to this question depends on the hypothesis being tested. For example,

1191 when testing the significance of the clustering coefficient, it is reasonable to retain the original

1192 number of nodes and edges, density and possibly also the degree distribution, such that the

1193 observed network is compared to random networks with similar properties.

1194   Several software packages exist that permit the computation of many of the metrics from

1195 this section that can be used to make inferences about the network. Examples include UCINet

1196 (Borgatti, Everett & Freeman, 2002), Gephi (Bastian, Heymann & Jacomy, 2009), igraph (Csardi

1197 & Nepusz, 2006) and Cytoscape (Shannon *et al.*, 2003).

1198

(4) Formal grammars

1200 The structure of an acoustic sequence can be described using formal grammars. A grammar

1201 consists of a set of rewrite rules (or "productions") that define the ways in which units can be

1202 ordered. Grammar rules consist of operations performed on "terminals" (in our case, units),

1203 which are conventionally denoted with lower case letters, and non-terminals (symbols that must

1204 be replaced by terminals before the derivation is complete), conventionally denoted with upper

1205 case letters (note that this convention is inconsistent with the upper case convention used for

1206 acoustic unit labels). Grammars generate sequences iteratively, by applying rules repeatedly to a

1207 growing sequence. For example, the rule "$\mathbf{U} \rightarrow a\,\mathbf{W}$" means that the nonterminal $\mathbf{U}$ can be

1208 rewritten with the symbols "$a\,\mathbf{W}$." The terminal a is a unit, as we are familiar with, but as $\mathbf{W}$ is a

1209 non-terminal, and may itself be rewritten by a different rule. For an example, see Fig. 11.

1210   Sequences that can be derived by a given grammar are called grammatical with respect to

1211   that grammar. The collection of all sequences that could possibly be generated by a grammar is

1212   called the language of the grammar. The validation of a grammar consists of verifying that the

1213   grammar's language matches exactly the set of sequences to be modelled. If a species produces

1214   sequences that cannot be generated by the grammar, the grammar is deemed "over-selective". A

1215   grammar that is "over-generalising" produces sequences not observed in the empirical data –

1216   although it is often unclear whether this represents a true failure of the grammar, or insufficient

1217   sampling of observed sequences. In the example given in Fig. 11, the grammar is capable of

1218   producing the sequence *abbbbbbbbbbbb*, however, since blue whales have not been observed to

1219   produce similar sequences in decades of observation, we conclude that this grammar is

1220   overgeneralising. It is important to note, however, that formal grammars are deterministic, in

1221   contrast to the probabilistic models discussed previously (Markov model, HMM). If one assigned

1222   probabilities to each of the rewriting rules, the particular sequence shown above may not have

1223   been observed simply because it is very unlikely.

1224   Algorithms known as parsers can be constructed from grammars to determine whether a

1225   sequence belongs to the language for which the grammar has been inferred. Inferring a grammar

1226   from a collection of sequences is a difficult problem, which, as famously formulated by Gold

1227   (1967), is intractable for all but a number of restricted cases. Gold's formulation, however, does

1228   not appear to preclude the learning of grammar in real-world examples, and is of questionable

1229   direct relevance to the understanding or modelling of the psychology of sequence processing

1230   (Johnson, 2004). When restated in terms that arguably fit better the cognitive tasks faced by

1231   humans and other animals, grammar inference becomes possible (Clark, 2010; Clark, Eyraud &

1232   Habrard, 2010). Algorithms based on distributional learning, which seek probabilistically

1233 motivated phrase structure by recursively aligning and comparing input sequences, are becoming

1234 increasingly successful in sequence-processing tasks such as modelling language acquisition

1235 (Solan *et al.*, 2005; Kolodny, Lotem & Edelman, in press).

1236   A grammar can be classified according to its place in a hierarchy of classes of formal

1237 grammars known as the Chomsky hierarchy (Chomsky, 1957) and illustrated in Fig. 12. These

1238 classes differ in the complexity of languages that can be modelled. The simplest class of

1239 grammars are called regular grammars, which are capable of describing the generation of any

1240 finite set of sequences or repeating pattern, and are fundamentally similar to Markov models.

1241 Fig. 11 is an example of a regular grammar. Kakishita *et al.* (2009) showed that Bengalese finch

1242 *Lonchura striata* ver. *domestica* songs can be modelled by a restricted class of regular grammars,

1243 called "k-reversible regular grammars," which is learnable from only positive samples, i.e.

1244 observed and hence permissible sequences, without information on those sequences that are not

1245 permissible in the grammar. Context-free grammars are more complex than regular grammars

1246 and are able to retain state information that enable one part of the sequence to affect another; this

1247 is usually demonstrated through the ability to create sequences of symbols where each unit is

1248 repeated the same number of times A^nB^n where n denotes n repetitions of the terminal unit, e.g.

1249 *AAABBB* (A^3B^3). Such an ability requires keeping track of a state, e.g. "how many times the unit

1250 A has been used", and a neurological implementation may be lacking in most species (Beckers *et*

1251 *al.*, 2012). Context-sensitive languages allow context-dependent rewrite rules that have few

1252 restrictions, permitting further reaching dependencies such as in the set of sequences $A^nB^nC^n$, and

1253 require still more sophisticated neural implementations. The highest level in the Chomsky

1254 hierarchy, recursively enumerable grammars, are more complex still, and rarely have relevance

1255 to animal communication studies.

1256   The level of a grammar within the Chomsky hierarchy can give an indication of the

1257   complexity of the communication system represented by that grammar. Most animal acoustic

1258   sequences are thought to be no more complex than regular grammars (Berwick *et al.*, 2011),

1259   whereas complexity greater than the regular grammar is thought to be a unique feature of human

1260   language (Hauser *et al.*, 2002). Therefore, indication that any animal communication could not

1261   be represented by a regular grammar would be considered an important discovery. For example,

1262   Gentner *et al.* (2006) proposed that European starlings *Sturnus vulgaris* can learn to recognise

1263   context-free (but non-regular) sequences, and reject sequences that do not correspond to the

1264   learned grammar. However, other authors have pointed out that the observed results could be

1265   explained by more simple mechanisms than context-free processing, such as primacy rules (Van

1266   Heijningen *et al.*, 2009) in which simple analysis of short substrings is sufficient to distinguish

1267   between grammatical and non-grammatical sequences, or acoustic similarity matching (Beckers

1268   *et al.*, 2012). Consequently, claims of greater than regular grammar in non-human animals have

1269   not been widely accepted. The deterministic nature of regular grammars – or indeed any formal

1270   grammars – may explain why formal grammars are not sufficiently general to describe the

1271   sequences of many animal species, and formal grammars remain more popular in human

1272   linguistic studies than in animal communication research.

1273

1274   **(5) Temporal structure**

1275   Information may exist in the relative or absolute timing of acoustic units in a sequence, rather

1276   than in the order of those units. In particular, timing and rhythm information may be of

1277   importance, and may be lost when acoustic sequences are represented as a series of symbols.

1278   This section describes two different approaches to quantifying the temporal structure in acoustic

1279   sequences: traditional techniques examining inter-event interval and pulse statistics (e.g. Randall,

1280   1989; Narins *et al.*, 1992), and recent multi-timescale rhythm analysis (Saar & Mitra, 2008).

1281   Analyses of temporal structure can be applied to any audio recording, regardless of

1282   whether that recording contains recognisable sequences, individual sounds, or multiple

1283   simultaneously vocalising individuals. Such analyses are most likely to be informative, however,

1284   when recurring acoustic patterns are present, especially if those recurring patterns are rhythmic

1285   or produced at a predictable rate. Variations in interactive sound-sequence production during

1286   chorusing and cross-individual synchronisation can be quantified through meter, or prosody

1287   analysis, and higher-order sequence structure can be identified through automated identification

1288   of repeating patterns. At the simplest level, it is possible to analyse the timing of sounds in a

1289   sequence, simply by recording when sound energy is above a fixed threshold. For instance,

1290   temporal patterns can be extracted automatically from simpler acoustic sequences by

1291   transforming recordings into sequences of numerical measures of the durations and silent

1292   intervals between sounds (Isaac & Marler, 1963; Catchpole, 1976; Mercado, Herman & Pack,

1293   2003; Handel, Todd & Zoidis, 2009; Green *et al.*, 2011), song bouts (Eens, Pinxten & Verheyen,

1294   1989; Saar & Mitra, 2008), or of acoustic energy within successive intervals (Murray, Mercado

1295   & Roitblat, 1998; Mercado *et al.*, 2010). Before the invention of the Kay sonograph, which led to

1296   the routine analysis of audio spectrograms, temporal dynamics of birdsong were often

1297   transcribed using musical notation (Saunders, 1951; Nowicki & Marler, 1988).

1298   Inter-pulse interval has been widely used to quantify temporal structure in animal

1299   acoustic sequences, for example in kangaroo rats *Dipodomys spectabilis* (Randall, 1989), fruit

1300   flies *Drosophila melanogaster* (Bennet-Clark & Ewing, 1969), and rhesus monkeys *Macaca*

1301   *mulatta* (Hauser *et al.*, 1998). Variations in pulse intervals can encode individual information

1302   such as identity and fitness (Bennet-Clark & Ewing, 1969; Randall, 1989), as well species

1303   identity (Randall, 1997; Hauser *et al*., 1998). In these examples, comparing the median inter-

1304   pulse interval between two sample populations is often sufficient to uncover significant

1305   differences.

1306   More recently developed techniques for analysis of temporal structure require more

1307   detailed processing. For example, periodic regularities and repetitions of patterns within

1308   recordings of musical performances can be automatically detected and characterised (Paulus,

1309   Müller & Klapuri, 2010; Weiss & Bello, 2011). The first step in modern approaches to analysing

1310   the temporal structure of sound sequences involves segmenting the recording. The duration and

1311   distribution of individual segments can be fixed (e.g. splitting a recording into 100 ms

1312   chunks/frames) or variable (e.g. using multiple frame sizes in parallel or adjusting the frame size

1313   based on the rate and duration of acoustic events). The acoustic features of individual frames can

1314   then be analysed using the same signal-processing methods that are applied when measuring the

1315   acoustic features of individual sounds, thereby transforming the smaller waveform into a vector

1316   of elements that describe features of the segment. Sequences of such frame-describing vectors

1317   then would typically be used to form a matrix representing the entire recording. In this matrix,

1318   the sequence of columns (or rows) corresponds to the temporal order of individual frames

1319   extracted from the recording.

1320   Regularities within the feature matrix generated from frame-describing vectors reflect

1321   temporal regularities within the original recording. Thus, the problem of describing and detecting

1322   temporal patterns within a recording is transformed into the more computationally tractable

1323   problem of detecting and identifying structure within a matrix of numbers (as opposed to a

1324   sequence of symbols). If each frame is described by a single number (e.g. mean amplitude), then

1325   the resulting sequence of numbers can be analysed using standard time–frequency analysis

1326   techniques to reveal rhythmic patterns (Saar & Mitra, 2008). Alternatively, each frame can be

1327   compared with every other frame to detect similarities using standard measures for quantifying

1328   the distance between vectors (Paulus *et al.*, 2010). These distances are then often collected

1329   within a second matrix called a self-distance matrix. Temporal regularities within the original

1330   feature matrix are visible as coherent patterns with the self-distance matrix (typically showing up

1331   as patterned blocks or diagonal stripes). Various methods used for describing and classifying

1332   patterns within matrices (or images) can then be used to classify these two-dimensional patterns.

1333   Different patterns in these matrices can be associated with variations in the novelty or

1334   homogeneity of the temporal regularities over time, as well as the number of repetitions of

1335   particular temporal patterns (Paulus *et al.*, 2010). Longitudinal analyses of time-series measures

1336   of temporal structure can also be used to describe the stability or dynamics of rhythmic pattern

1337   production over time (Saar & Mitra, 2008). An alternative approach to identifying temporal

1338   structure within the feature matrix is to decompose it into simpler component matrices that

1339   capture the most recurrent features within the recording (Weiss & Bello, 2011). Similar

1340   approaches are common in modern analyses of high-density electroencephalograph (EEG)

1341   recordings (Makeig *et al.*, 2004). Algorithms for analysing the temporal dynamics of brain waves

1342   may thus also be useful for analysing temporal structure within acoustic recordings.

1343

VI.   FUTURE DIRECTIONS

1345   Many of the central research questions in animal communication focus on the meaning of signals

1346   and on the role of natural, sexual, and social selection in the evolution of communication

1347   systems. As shown in Fig. 6, information can exist in a sequence simultaneously *via* diversity,

1348 and order, as well as other less well-studied phenomena. Both natural and sexual selection may

1349 act on this information, either through conspecifics or heterospecifics (e.g. predators). This is

1350 especially true for animal acoustic sequences because the potential complexity of a sequence

1351 may imply greater scope for both meaning and selective pressure. Many new questions – and

1352 several old and unanswered ones – can be addressed by the techniques that we have outlined

1353 herein. Some of the most promising avenues for future research are outlined below, with some

1354 outstanding questions in animal acoustic sequences that can potentially be addressed more

1355 effectively using the approaches proposed in this review.

1356

1357 (1) As sequences are composed of units, how might information exist within units

1358 themselves?

1359 One promising direction lies in studying how animals use concatenated signals with multiple

1360 meanings. For example, Jansen, Cant & Manser (2012) provided evidence for temporal

1361 segregation of information within a syllable, where one segment of a banded mongoose *Mungos*

1362 *mungo* close call is individually distinct, while the other segment contains meaning about the

1363 caller's activity. Similar results have been demonstrated in the song of the white-crowned

1364 sparrow *Zonotrichia leucophrys* (Nelson & Poesel, 2007). Understanding how to divide acoustic

1365 units according to criteria other than silent gaps (Fig. 2) can change the research approach, as

1366 well as the results of a study. The presence of information in sub-divisions of traditional acoustic

1367 units is a subject underexplored in the field of animal communication, and an understanding of

1368 the production and perceptual constraints on unit definition (Fig. 4) is essential.

1369

(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 organisation (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, 2009; Adami, 2002).

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 among 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 Fig. 6 – or combination thereof – it applies. Complexity defined, say, for the *Repetition* paradigm (Fig. 6A) and quantified as pulse rate variation, is not easily compared with *Diversity* complexity (Fig. 6B), typically quantified as repertoire size.

For example, is selection from increased social complexity associated with increased vocal complexity (Pollard & Blumstein, 2012; Freeberg *et al*., 2012), or do 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

1393   repertoire size (Fig. 6B). Considerable evidence in diverse taxa indicates that increased social

1394   complexity is associated with increased repertoire size (reviewed in Freeberg *et al*., 2012).

1395   Different views of complexity in this literature are revealed by the fact that social complexity has

1396   been measured in terms of group size, group stability, or information-based metrics of group

1397   composition, and vocal complexity has been measured in terms of not just repertoire size, but

1398   also information-based metrics of acoustic variation in signals. In fact, the work of Pollard &

1399   Blumstein (2011) is highly informative to questions of complexity, in that different metrics of

1400   social complexity can drive different metrics of vocal complexity – these authors have found that

1401   group size is associated with greater individual distinctiveness (information) in the calls of

1402   species, but the diversity of social roles in groups is more heavily associated with vocal

1403   repertoire size. Some researchers have proposed the idea that communicative complexity, again

1404   defined as repertoire size, has at least in some species been driven by the need to encode more

1405   information, or redundant information, in a complex social environment (Freeberg *et al*., 2012).

1406   Alternatively, complexity metrics that measure *Ordering* (Fig. 6D), often based on non-zero

1407   orders of entropy (McCowan *et al*., 1999; Kershenbaum, 2013)*,* may be more biologically

1408   relevant in species that use unit ordering to encode information. Understanding the variety of

1409   sequence types is essential to choosing the relevant acoustic unit definitions, and without this,

1410   testing competitive evolutionary hypotheses becomes problematic.

1411

1412   **(3) How do individual differences in acoustic sequences arise?**

1413   If we can develop categories for unit types and sequence types that lead to productive

1414   vocalisation analysis and a deeper understanding of universal factors of encoded multi-layered

1415   messages, then individual differences in sequence production become interesting and puzzling.

1416 The proximal processes driving individual differences in communicative sequences are rarely

1417 investigated. Likewise, although there is a decades-rich history of song-learning studies in

1418 songbirds, the ontogenetic processes giving rise to communicative sequences *per se* have rarely

1419 been studied. Neural models, e.g. Jin (2009) can provide probabilistic descriptions of sequence

1420 generation (e.g. Markov models, hidden Markov models), but the nature of the underlying

1421 stochasticity is unknown. When an appropriate choice of a model for sequence structure is made,

1422 quantitative comparisons can be carried out between the parameters of different individuals, for

1423 example with the California thrasher *Toxostoma redivivum* (Sasahara *et al*., 2012). However,

1424 model fitting is only valid if unit selection is biologically appropriate (Section III). Other, more

1425 abstract, questions can also be addressed. Individual humans use language with varying degrees

1426 of efficiency, creativity, and effectiveness. Shakespearean sequences are radically unlike Haiku

1427 sequences, political speeches, or the babbling of infants, in part because their communicative

1428 purposes differ. While sexual selection and survival provide some purposive contexts through

1429 which we can approach meaning, additional operative contexts may suggest other purposes, and

1430 give us new frameworks through which to view vocal sequences (Waller, 2012). In many

1431 animals, song syntax may be related to sexual selection. Females of some species such as zebra

1432 finches *Taeniopygia guttata* not only prefer individuals with longer songs, but also songs

1433 comprising a greater variety of syllables (Searcy & Andersson, 1986; Neubauer, 1999; Holveck

1434 *et al*., 2008); whereas in other species, this preference is not observed (Byers & Kroodsma,

1435 2009). Variation in syntax may also reflect individual differences in intraspecific aggression, for

1436 instance in banded wrens *Pheugopedius pleurostictus* (Vehrencamp *et al*., 2007) and western

1437 populations of song sparrows *Melospiza melodia* (Burt, Campbell & Beecher, 2001). Individual

1438 syntax may also serve to distinguish neighbours from non-neighbours in song sparrows (Beecher

1439 *et al*., 2000) and skylarks *Alauda arvensis* (Briefer *et al*., 2008). Male Cassin's vireos *Vireo*

1440 *cassinii* can usually be discriminated by the acoustic features of their song, but are discriminated

1441 even better by the sequences of phrases that they sang (Arriaga *et al*., 2013).

1442

1443 **(4) What is the role of sequence dialects in speciation?**

1444 In a few species, geographic syntactic dialects (Nettle, 1999) have been demonstrated, including

1445 primates, such as Rhesus monkeys *Macaca mulatta* (Gouzoules, Gouzoules & Marler, 1984) and

1446 chimpanzees *Pan troglodytes* (Arcadi, 1996; Mitani, Hunley & Murdoch, 1999; Crockford &

1447 Boesch, 2005), birds, such as Carolina chickadees *Poecile carolinensis* (Freeberg, 2012), swamp

1448 sparrows *Melospiza georgiana* (Liu *et al*., 2008) and chaffinches *Fringilla coelebs* (Lachlan *et*

1449 *al*., 2013) and in rock hyraxes *Procavia capensis* (Kershenbaum *et al*., 2012). This broad

1450 taxonomic spread raises the question of whether sequence syntax has a role in speciation (Wiens,

1451 1982; Nevo *et al*., 1987; Irwin, 2000; Slabbekoorn & Smith, 2002; Lachlan *et al*., 2013), with

1452 some support for such a role in chestnut-tailed antbirds *Myrmeciza hemimelaena* (Seddon &

1453 Tobias, 2007) and winter wrens *Troglodytes troglodytes* (Toews & Irwin, 2008). It is tempting to

1454 speculate that acoustic sequences may have arisen from earlier selective forces acting on a

1455 communication system based on single units, with variation in the sequences of individuals

1456 providing differential adaptive benefit. The ability to communicate effectively with some but not

1457 others could lead to divergence of groups, and genetic pooling. Conversely, differences in

1458 acoustic sequences could be adaptive to ecological variation. It is hard to distinguish

1459 retrospectively between sequence dialect shift leading to divergence of sub-groups and eventual

1460 speciation, or group separation leading to new communicative strategies that are epiphenomena

64

1461   of species formation. What are the best methods for investigating the relationship between

1462   communication and biological change?

1463   A third alternative is that sequence differences could arise by neutral processes analogous

1464   to drift. A complex interplay between production, perception, and encoding of information in

1465   sequence syntax, along with the large relative differences between different species in adaptive

1466   flexibility (Seyfarth & Cheney, 2010), could lead to adaptive pressures on communication

1467   structure. However, the definition of acoustic units is rarely considered in this set of questions. In

1468   particular, perceptual binding (Fig. 4A) and the response of the focal species must be considered,

1469   as reproductive isolation cannot occur on the basis of differences that are not perceived by the

1470   receiver. As units may be divided at many levels, there may be multiple sequences that convey

1471   different information types. Thus, a deeper understanding of units and sequences will contribute

1472   productively to questions regarding forces at work in speciation events.

1473

1474   **(5) Future directions: conclusions**

1475   We conclude by noting that more detailed and rigorous approaches to investigating animal

1476   acoustic sequences will allow us to investigate more complex systems that have not been

1477   formally studied. A number of directions lack even a basic framework as we have proposed in

1478   this review. For example, there is much to be learned from the detailed study of the sequences

1479   created by multiple animals vocalising simultaneously, and from the application of sequence

1480   analysis to multimodal communication with a combination of acoustic, visual, and perhaps other

1481   modalities (e.g. Partan & Marler, 1999; Bradbury & Vehrencamp, 2011; Munoz & Blumstein,

1482   2012). Eavesdropping, in which non-target receivers (such as predators) gain additional

1483   information from listening to the interaction between individuals, has only just begun to be

1484   studied in the context of sequence analysis. Finally, the study of non-stationary systems, where

1485   the statistical nature of the communicative sequences changes over long or short time scales

1486   (such as appears to occur in humpback whale songs) is ripe for exploration. For example,

1487   acoustic sequences may be constantly evolving sexual displays that are stereotyped within a

1488   population at any particular point in time (Payne & McVay, 1971; Payne, Tyack & Payne, 1983).

1489   The application of visual classification (Garland *et al*., 2011) and a statistical approach based on

1490   edit distance (e.g. Kershenbaum *et al*., 2012) appears to capture the sequential information

1491   present within humpback whale song (Garland *et al*., 2012, 2013). This work traced the

1492   evolution of song lineages, and the movement or horizontal cultural transmission of multiple

1493   different versions of the song that were concurrently present across an ocean basin over a decade

1494   (Garland *et al*., 2013). These results are encouraging for the investigation of complex non-

1495   stationary systems; however, further refinement of this approach is warranted. We encourage

1496   researchers in these fields to extend treatments such as ours to cover these more complex

1497   directions in animal communication research, thereby facilitating quantitative comparisons

1498   between fields.

1499

1500   **VII. CONCLUSIONS**

1501   (1) The use of acoustic sequences by animals is widespread across a large number of taxa. As

1502   diverse as the sequences themselves is the range of analytical approaches used by researchers.

1503   We have proposed a framework for analysing and interpreting such acoustic sequences, based

1504   around three central ideas of understanding the information content of sequences, defining the

1505   acoustic units that comprise sequences, and proposing analytical algorithms for testing

1506   hypotheses on empirical sequence data.

1507   (2) We propose use of the term "meaning" to refer to a feature of communication sequences that

1508   influences behavioural and evolutionary processes, and the term "information" to refer to the

1509   non-random statistical properties of sequences.

1510   (3) Information encoding in acoustic sequences can be classified into six non-mutually exclusive

1511   paradigms: *Repetition, Diversity, Combination, Ordering, Overlapping,* and *Timing.*

1512   (4) The constituent units of acoustic sequences can be classified according to production

1513   mechanisms, perception mechanisms, or analytical properties.

1514   (5) Discrete acoustic units are often delineated by silent intervals. However, changes in the

1515   acoustic properties of a continuous sound may also indicate a transition between discrete units,

1516   multiple repeated sounds may act as a discrete unit, and more complex hierarchical structure may

1517   also be present.

1518   (6) We have reviewed five approaches used for analysing the structure of animal acoustic

1519   sequences: Markov chains, hidden Markov models, network models, formal grammars, and

1520   temporal models, discussing their use and relative merits.

1521   (7) Many important questions in the behavioural ecology of acoustic sequences remain to be

1522   answered, such as understanding the role of communication complexity, including multimodal

1523   sequences, the potential effect of communicative isolation on speciation, and the source of

1524   syntactic differences among individuals.

1525

1526   **VIII.   ACKNOWLEDGEMENTS**

1543

IX. REFERENCES

1545   ADAM, O., CAZAU, D., GANDILHON, N., FABRE, B., LAITMAN, J. T. & REIDENBERG, J. S. (2013).
1546   New acoustic model for humpback whale sound production. *Applied Acoustics* **74**, 1182-1190.

1547   ADAMI, C. (2002). What is complexity? *BioEssays* **24**, 1085-1094.

1548   ADAMS, D. C. & ANTHONY, C. D. (1996). Using randomization techniques to analyse behavioural
1549   data. *Animal Behaviour* **51**, 733-738.

1550   AKÇAY, Ç, TOM, M. E., CAMPBELL, S. E. & BEECHER, M. D. (2013). Song type matching is an
1551   honest early threat signal in a hierarchical animal communication system. *Proceedings of the
1552   Royal Society B: Biological Sciences* **280**.

1553   ANDERSON, T. W. & GOODMAN, L. A. (1957). Statistical inference about Markov chains. *The
1554   Annals of Mathematical Statistics* **28**, 89-110.

1555   ARAYA-SALAS, M. (2012). Is birdsong music? *Significance* **9**, 4-7.

1556 ARCADI, A. C. (1996). Phrase structure of wild chimpanzee pant hoots: patterns of production
1557 and interpopulation variability. *American Journal of Primatology* **39**, 159-178.

1558 ARLOT, S. & CELISSE, A. (2010). A survey of cross-validation procedures for model selection.
1559 *Statistics Surveys* **4**, 40-79.

1560 ARNOLD, K. & ZUBERBÜHLER, K. (2006*a*). Language evolution: semantic combinations in
1561 primate calls. *Nature* **441**, 303-303.

1562 ARNOLD, K. & ZUBERBÜHLER, K. (2006*b*). The alarm-calling system of adult male putty-nosed
1563 monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour* **72**, 643-653.

1564 ARNOLD, K. & ZUBERBÜHLER, K. (2008). Meaningful call combinations in a non-human primate.
1565 *Current Biology* **18**, R202-R203.

1566 ARNOLD, K. & ZUBERBÜHLER, K. (2013). Female putty-nosed monkeys use experimentally
1567 altered contextual information to disambiguate the cause of male alarm calls. *PloS One* **8**,
1568 e65660.

1569 ARRIAGA, J. G., KOSSAN, G., CODY, M. L., VALLEJO, E. E. & TAYLOR, C. E. (2013). Acoustic
1570 sensor arrays for understanding bird communication. Identifying Cassin's Vireos using SVMs
1571 and HMMs. *ECAL 2013* **12**, 827-828.

1572 ATTARD, M. R., PITCHER, B. J., CHARRIER, I., AHONEN, H. & HARCOURT, R. G. (2010). Vocal
1573 discrimination in mate guarding male Australian sea lions: familiarity breeds contempt. *Ethology*
1574 **116**, 704-712.

1575 AUBIN, T., JOUVENTIN, P. & HILDEBRAND, C. (2000). Penguins use the two-voice system to
1576 recognize each other. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1081-1087.

1577 AUBIN, T., MATHEVON, N., STASZEWSKI, V. & BOULINIER, T. (2007). Acoustic communication in
1578 the Kittiwake *Rissa tridactyla*: potential cues for sexual and individual signatures in long calls.
1579 *Polar Biology* **30**, 1027-1033.

1580 BAKER, M. C. & BECKER, A. M. (2002). Mobbing calls of black-capped chickadees: effects of
1581 urgency on call production. *The Wilson Bulletin* **114**, 510-516.

1582 BAKER, M. C. & LOGUE, D. M. (2003). Population differentiation in a complex bird sound: a
1583 comparison of three bioacoustical analysis procedures. *Ethology* **109**, 223-242.

1584 BAPTISTA, L. F. & KEISTER, R. A. (2005). Why birdsong is sometimes like music. *Perspectives in
1585 Biology and Medicine* **48**, 426-443.

1586 BARONCHELLI, A., FERRER-I-CANCHO, R., PASTOR-SATORRAS, R., CHATER, N. & CHRISTIANSEN,
1587 M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences* **17**, 348-360.

1588   BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R. & VESPIGNANI, A. (2004). The
1589   architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*
1590   **101**, 3747-3752.

1591   BASTIAN, M., HEYMANN, S. & JACOMY, M. (2009). Gephi: an open source software for exploring
1592   and manipulating networks. *Proceedings of the Third International ICWSM Conference*.

1593   BAUGH, A., AKRE, K. & RYAN, M. (2008). Categorical perception of a natural, multivariate
1594   signal: mating call recognition in túngara frogs. *Proceedings of the National Academy of*
1595   *Sciences* **105**, 8985-8988.

1596   BAZÚA-DURÁN, C. & AU, W. W. (2002). The whistles of Hawaiian spinner dolphins. *The Journal*
1597   *of the Acoustical Society of America* **112**, 3064-3072.

1598   BECKERS, G. J., BOLHUIS, J. J., OKANOYA, K. & BERWICK, R. C. (2012). Birdsong
1599   neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23**, 139-145.

1600   BEE, M. A. & MICHEYL, C. (2008). The cocktail party problem: what is it? How can it be solved?
1601   And why should animal behaviorists study it? *Journal of Comparative Psychology* **122**, 235-251.

1602   BEECHER, M. D., CAMPBELL, S. E., BURT, J. M., HILL, C. E. & NORDBY, J. C. (2000). Song-type
1603   matching between neighbouring song sparrows. *Animal Behaviour* **59**, 21-27.

1604   BEN-DAVID, S., VON LUXBURG, U. & PÁL, D. (2006). A sober look at clustering stability. In
1605   *Learning theory*, pp. 5-19. Springer.

1606   BENNET-CLARK, H. & EWING, A. (1969). Pulse interval as a critical parameter in the courtship
1607   song of *Drosophila melanogaster*. *Animal Behaviour* **17**, 755-759.

1608   BERGMAN, T. J., BEEHNER, J. C., CHENEY, D. L. & SEYFARTH, R. M. (2003). Hierarchical
1609   classification by rank and kinship in baboons. *Science* **302**, 1234-1236.

1610   BERWICK, R. C., OKANOYA, K., BECKERS, G. J. L. & BOLHUIS, J. J. (2011). Songs to syntax: the
1611   linguistics of birdsong. *Trends in Cognitive Sciences* **15**, 113-121.

1612   BIZLEY, J. K. & COHEN, Y. E. (2013). The what, where and how of auditory-object perception.
1613   *Nature Reviews Neuroscience* **14**, 693-707.

1614   BLUMSTEIN, D. T. & MUNOS, O. (2005). Individual, age and sex-specific information is contained
1615   in yellow-bellied marmot alarm calls. *Animal Behaviour* **69**, 353-361.

1616   BLUMSTEIN, D. T. (1995). Golden-marmot alarm calls. II. Asymmetrical production and
1617   perception of situationally specific vocalizations? *Ethology* **101**, 25-32.

1618   BLUMSTEIN, D. T. (2007). The evolution, function, and meaning of marmot alarm
1619   communication. *Advances in the Study of Behavior* **37**, 371-401.

1620 BLUMSTEIN, D. T., VERNEYRE, L. & DANIEL, J. C. (2004). Reliability and the adaptive utility of
1621 discrimination among alarm callers. *Proceedings of the Royal Society of London. Series B:*
1622 *Biological Sciences* **271**, 1851-1857.

1623 BOHM, D. (1989). Meaning and information. In *The Search for Meaning, The New Spirit in*
1624 *Science and Philosophy* (ed. P. Pylkkänen), pp. 43-62. Thorsons Publishing Group,
1625 Wellingborough.

1626 BOHN, K. M., SCHMIDT-FRENCH, B., MA, S. T. & POLLAK, G. D. (2008). Syllable acoustics,
1627 temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats.
1628 *The Journal of the Acoustical Society of America* **124**, 1838-1848.

1629 BOHN, K. M., SCHMIDT-FRENCH, B., SCHWARTZ, C., SMOTHERMAN, M. & POLLAK, G. D. (2009).
1630 Versatility and stereotypy of free-tailed bat songs. *PloS One* **4**, e6746.

1631 BONNER, J. T. (1988). *The evolution of complexity by means of natural selection*. Princeton
1632 University Press.

1633 BORGATTI, S. P., EVERETT, M. G. & FREEMAN, L. C. (2002). *Ucinet for Windows: Software for*
1634 *social network analysis*. Analytic Technologies.

1635 BRADBURY, J. W. & VEHRENCAMP, S. L. (2011). *Principles of Animal Communication*. Sinauer,
1636 Sunderland, MA.

1637 BRIEFER, E. F., RYBAK, F. & AUBIN, T. (2013). Does true syntax or simple auditory object
1638 support the role of skylark song dialect? *Animal Behaviour* **86**, 1131-1137.

1639 BRIEFER, E., AUBIN, T., LEHONGRE, K. & RYBAK, F. (2008). How to identify dear enemies: the
1640 group signature in the complex song of the skylark *Alauda arvensis*. *Journal of Experimental*
1641 *Biology* **211**, 317-326.

1642 BRIEFER, E., OSIEJUK, T. S., RYBAK, F. & AUBIN, T. (2010). Are bird song complexity and song
1643 sharing shaped by habitat structure? An information theory and statistical approach. *Journal of*
1644 *Theoretical Biology* **262**, 151-164.

1645 BROWN, S. D., DOOLING, R. J. & O'GRADY, K. E. (1988). Perceptual organization of acoustic
1646 stimuli by budgerigars (*Melopsittacus undulatus*): III. Contact calls. *Journal of Comparative*
1647 *Psychology* **102**, 236-247.

1648 BRUMM, H. & SLATER, P. J. (2006). Ambient noise, motor fatigue, and serial redundancy in
1649 chaffinch song. *Behavioral Ecology and Sociobiology* **60**, 475-481.

1650 BUCK, J. R. & TYACK, P. L. (1993). A quantitative measure of similarity for *Tursiops truncatus*
1651 signature whistles. *The Journal of the Acoustical Society of America* **94**, 2497-2506.

1652　BURGHARDT, G. M., BARTMESS-LEVASSEUR, J. N., BROWNING, S. A., MORRISON, K. E., STEC, C.
1653　L., ZACHAU, C. E. & FREEBERG, T. M. (2012). Perspectives–minimizing observer bias in
1654　behavioral studies: A review and recommendations. *Ethology* **118**, 511-517.

1655　BURT, J. M., CAMPBELL, S. E. & BEECHER, M. D. (2001). Song type matching as threat: a test
1656　using interactive playback. *Animal Behaviour* **62**, 1163-1170.

1657　BYERS, B. E. & KROODSMA, D. E. (2009). Female mate choice and songbird song repertoires.
1658　*Animal Behaviour* **77**, 13-22.

1659　CALDWELL, M. C. (1965). Individualized whistle contours in bottle-nosed dolphins (*Tursiops*
1660　*truncatus*). *Nature* **207**, 434-435.

1661　CALDWELL, M. C., CALDWELL, D. K. & TYACK, P. L. (1990). Review of the signature-whistle
1662　hypothesis for the Atlantic bottlenose dolphin. In *The Bottlenose Dolphin* (eds. S. Leatherwood
1663　and R. R. Reeves), pp. 199-234. Academic Press, San Diego.

1664　CALL, J. E. & TOMASELLO, M. E. (2007). *The gestural communication of apes and monkeys.*
1665　Taylor & Francis Group/Lawrence Erlbaum Associates.

1666　CANE, V. R. (1959). Behaviour sequences as semi-Markov chains. *Journal of the Royal*
1667　*Statistical Society. Series B (Methodological)* **21**, 36-58.

1668　CAPPÉ, O., MOULINES, E. & RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer
1669　Science Business Media, New York.

1670　CARTER, G. G., SKOWRONSKI, M. D., FAURE, P. A. & FENTON, B. (2008). Antiphonal calling
1671　allows individual discrimination in white-winged vampire bats. *Animal Behaviour* **76**, 1343-
1672　1355.

1673　CÄSAR, C., BYRNE, R. W., HOPPITT, W., YOUNG, R. J. & ZUBERBÜHLER, K. (2012*a*). Evidence for
1674　semantic communication in titi monkey alarm calls. *Animal Behaviour* **84**, 405-411.

1675　CÄSAR, C., BYRNE, R., YOUNG, R. J. & ZUBERBÜHLER, K. (2012*b*). The alarm call system of wild
1676　black-fronted titi monkeys, *Callicebus nigrifrons*. *Behavioral Ecology and Sociobiology* **66**, 653-
1677　667.

1678　CÄSAR, C., ZUBERBÜHLER, K., YOUNG, R. J. & BYRNE, R. W. (2013). Titi monkey call sequences
1679　vary with predator location and type. *Biology Letters* **9**, 20130535.

1680　CATCHPOLE, C. K. & SLATER, P. J. B. (2003). *Bird song: biological themes and variations*.
1681　Cambridge Univ Press, Cambridge.

1682　CATCHPOLE, C. K. (1976). Temporal and sequential organisation of song in the sedge warbler
1683　(*Acrocephalus schoenobaenus*). *Behaviour* **59**, 226-246.

1684 CAZAU, D., ADAM, O., LAITMAN, J. T. & REIDENBERG, J. S. (2013). Understanding the intentional
1685 acoustic behavior of humpback whales: A production-based approach. *The Journal of the*
1686 *Acoustical Society of America* **134**, 2268-2273.

1687 CEUGNIET, M. & AUBIN, T. (2001). The rally call recognition in males of two hybridizing
1688 partridge species, red-legged (*Alectoris rufa*) and rock (*A. graeca*) partridges. *Behavioural*
1689 *Processes* **55**, 1-12.

1690 CHARIF, R., PONIRAKIS, D. & KREIN, T. (2006). Raven Lite 1.0 user's guide. Cornell Laboratory
1691 of Ornithology, Ithaca, NY .

1692 CHARRIER, I. & HARCOURT, R. G. (2006). Individual vocal identity in mother and pup Australian
1693 sea lions (*Neophoca cinerea*). *Journal of Mammalogy* **87**, 929-938.

1694 CHARRIER, I., AHONEN, H. & HARCOURT, R. G. (2011). What makes an Australian sea lion
1695 (*Neophoca cinerea*) male's bark threatening? *Journal of Comparative Psychology* **125**, 385.

1696 CHARRIER, I., LEE, T. T., BLOOMFIELD, L. L. & STURDY, C. B. (2005). Acoustic mechanisms of
1697 note-type perception in black-capped chickadee (*Poecile atricapillus*) calls. *Journal of*
1698 *Comparative Psychology* **119**, 371.

1699 CHARRIER, I., MATHEVON, N. & AUBIN, T. (2013). Bearded seal males perceive geographic
1700 variation in their trills. *Behavioral Ecology and Sociobiology* **67**, 1679-1689.

1701 CHARRIER, I., MATHEVON, N. & JOUVENTIN, P. (2003). Individuality in the voice of fur seal
1702 females: an analysis study of the pup attraction call in *Arctocephalus tropicalis*. *Marine Mammal*
1703 *Science* **19**, 161-172.

1704 CHARRIER, I., PITCHER, B. J. & HARCOURT, R. G. (2009). Vocal recognition of mothers by
1705 Australian sea lion pups: individual signature and environmental constraints. *Animal Behaviour*
1706 **78**, 1127-1134.

1707 CHATFIELD, C. & LEMON, R. E. (1970). Analysing sequences of behavioural events. *Journal of*
1708 *Theoretical Biology* **29**, 427-445.

1709 CHATTERJEE, A. (2005). A madness to the methods in cognitive neuroscience? *Journal of*
1710 *Cognitive Neuroscience* **17**, 847-849.

1711 CHEN, C. & FUSHING, H. (2012). Multiscale community geometry in a network and its
1712 application. *Physical Review E* **86**, 041120.

1713 CHOLEWIAK, D. M., SOUSA-LIMA, R. S. & CERCHIO, S. (2012). Humpback whale song
1714 hierarchical structure: Historical context and discussion of current classification issues. *Marine*
1715 *Mammal Science* **29**, E312-E332.

1716 CHOMSKY, N. (1957). *Syntactic structures*. de Gruyter Mouton, The Hague.

1717   CHRISTIANSEN, M. H. & CHATER, N. (2001). Connectionist psycholinguistics: Capturing the
1718   empirical data. *Trends in Cognitive Sciences* **5**, 82-88.

1719   CLARK, A. (2010). Towards general algorithms for grammatical inference. *Lecture Notes in
1720   Computer Science* **6331**, 11-30.

1721   CLARK, A., EYRAUD, R. & HABRARD, A. (2010). Using contextual representations to efficiently
1722   learn context-free languages. *The Journal of Machine Learning Research* **11**, 2707-2744.

1723   CLARK, C. J. & FEO, T. J. (2008). The Anna's hummingbird chirps with its tail: a new mechanism
1724   of sonation in birds. *Proceedings of the Royal Society B: Biological Sciences* **275**, 955-962.

1725   CLARK, C. J. & FEO, T. J. (2010). Why do Calypte hummingbirds "sing" with both their tail and
1726   their syrinx? An apparent example of sexual sensory bias. *The American Naturalist* **175**, 27-37.

1727   CLARK, C. W., MARLER, P. & BEEMAN, K. (1987). Quantitative analysis of animal vocal
1728   phonology: an application to swamp sparrow song. *Ethology* **76**, 101-115.

1729   CLEMINS, P. J. & JOHNSON, M. T. (2006). Generalized perceptual linear prediction features for
1730   animal vocalization analysis. *The Journal of the Acoustical Society of America* **120**, 527-534.

1731   COHN, L. (1995). *Time-Frequency Analysis: Theory and Applications, 1995*. Prentice Hall.

1732   COLLIAS, N. E. (1987). The vocal repertoire of the red junglefowl: a spectrographic classification
1733   and the code of communication. *Condor* **89**, 510-524.

1734   COVER, T. M. & THOMAS, J. A. (1991). *Elements of information theory*. John Wiley & Sons, Inc,
1735   New York, NY.

1736   CROCKFORD, C. & BOESCH, C. (2005). Call combinations in wild chimpanzees. *Behaviour* **142**,
1737   397-421.

1738   CROFT, D. P., JAMES, R. & KRAUSE, J. (2008). *Exploring animal social networks*. Princeton
1739   University Press.

1740   CSARDI, G. & NEPUSZ, T. (2006). The igraph software package for complex network research.
1741   *International Journal of Complex Systems* **1695**, 1-9.

1742   CURÉ, C., AUBIN, T. & MATHEVON, N. (2011). Sex discrimination and mate recognition by voice
1743   in the Yelkouan shearwater *Puffinus yelkouan*. *Bioacoustics* **20**, 235-249.

1744   CURÉ, C., MATHEVON, N., MUNDRY, R. & AUBIN, T. (2012). Acoustic cues used for species
1745   recognition can differ between sexes and sibling species: evidence in shearwaters. *Animal
1746   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.

1779 EENS, M. (1997). Understanding the complex song of the European starling: an integrated
1780 ethological approach. *Advances in the Study of Behavior* **26**, 355-434.

1781 EENS, M., PINXTEN, R. & VERHEYEN, R. F. (1989). Temporal and sequential organization of song
1782 bouts in the starling. *Ardea* **77**, 75-86.

1783 EENS, M., PINXTEN, R. & VERHEYEN, R. F. (1991). Male song as a cue for mate choice in the
1784 European starling. *Behaviour* **116**, 210-238.

1785 EENS, M., PINXTEN, R. & VERHEYEN, R. F. (1993). Function of the song and song repertoire in
1786 the European starling (*Sturnus vulgaris*): an aviary experiment. *Behaviour* **125**, 51-66.

1787 ESSER, K., CONDON, C. J., SUGA, N. & KANWAL, J. S. (1997). Syntax processing by auditory
1788 cortical neurons in the FM–FM area of the mustached bat*Pteronotus parnellii*. *Proceedings of
1789 the National Academy of Sciences* **94**, 14019-14024.

1790 EVANS, C. S. (1997). Referential signals. *Perspectives in Ethology* **12**, 99-143.

1791 EVANS, C. S., EVANS, L. & MARLER, P. (1993). On the meaning of alarm calls: functional
1792 reference in an avian vocal system. *Animal Behaviour* **46**, 23-38.

1793 FAGIOLO, G. (2007). Clustering in complex directed networks. *Physical Review E* **76**, 026107.

1794 FALLS, J. B. (1982). Individual recognition by sounds in birds. *Acoustic Communication in Birds*
1795 **2**, 237-278.

1796 FARABAUGH, S. & DOOLING, R. (1996). Acoustic communication in parrots: laboratory and field
1797 studies of budgerigars, *Melopsittacus undulatus*. In *Ecology and Evolution of Acoustic
1798 Communication in Birds*, pp. 97-117. Cornell University Press Ithaca, New York, USA.

1799 FERRER-I-CANCHO, R. & MCCOWAN, B. (2012). The span of correlations in dolphin whistle
1800 sequences. *Journal of Statistical Mechanics: Theory and Experiment* **2012**, P06002.

1801 FERRER-I-CANCHO, R. (2005). Zipf's law from a communicative phase transition. *The European
1802 Physical Journal B-Condensed Matter and Complex Systems* **47**, 449-457.

1803 FERRER-I-CANCHO, R., HERNÁNDEZ-FERNÁNDEZ, A., LUSSEAU, D., AGORAMOORTHY, G., HSU,
1804 M. J. & SEMPLE, S. (2013). Compression as a universal principle of animal behavior. *Cognitive
1805 Science* **37**, 1565-1578.

1806 FEWELL, J. H. (2003). Social insect networks. *Science* **301**, 1867-1870.

1807 FISCHER, J., NOSER, R. & HAMMERSCHMIDT, K. (2013). Bioacoustic field research: A primer to
1808 acoustic analyses and playback experiments with primates. *American Journal of Primatology* **75**,
1809 643-663.

1810   FITCH, W. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*
1811   **4**, 258-267.

1812   FITCH, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body
1813   size in rhesus macaques. *The Journal of the Acoustical Society of America* **102**, 1213-1222.

1814   FORD, J. K. (1989). Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver
1815   Island, British Columbia. *Canadian Journal of Zoology* **67**, 727-745.

1816   FORTUNATO, S. (2010). Community detection in graphs. *Physics Reports* **486**, 75-174.

1817   FRED, A. L. & JAIN, A. K. (2005). Combining multiple clusterings using evidence accumulation.
1818   *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **27**, 835-850.

1819   FREEBERG, T. M. & LUCAS, J. R. (2012). Information theoretical approaches to chick-a-dee calls
1820   of Carolina chickadees (*Poecile carolinensis*). *Journal of Comparative Psychology* **126**, 68-81.

1821   FREEBERG, T. M. (2012). Geographic variation in note composition and use of chick-a-dee calls
1822   of Carolina chickadees (*Poecile carolinensis*). *Ethology* **118**, 555-565.

1823   FREEBERG, T. M., DUNBAR, R. I. M. & ORD, T. J. (2012). Social complexity as a proximate and
1824   ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B:*
1825   *Biological Sciences* **367**, 1785-1801.

1826   FRUMHOFF, P. (1983). Aberrant songs of humpback whales (*Megaptera novaeangliae*): clues to
1827   the structure of humpback songs. In *Communication and behavior of whales* (ed. R. Payne), pp.
1828   81-127. Westview Press, Boulder, Colorado.

1829   FUSHING, H. & MCASSEY, M. P. (2010). Time, temperature, and data cloud geometry. *Physical*
1830   *Review E* **82**, 061110.

1831   GALE, W. & SAMPSON, G. (1995). Good-Turing smoothing without tears. *Journal of Quantitative*
1832   *Linguistics* **2**, 217-237.

1833   GAMMON, D. E. & ALTIZER, C. E. (2011). Northern Mockingbirds produce syntactical patterns of
1834   vocal mimicry that reflect taxonomy of imitated species. *Journal of Field Ornithology* **82**, 158-
1835   164.

1836   GARDNER, T. J., NAEF, F. & NOTTEBOHM, F. (2005). Freedom and rules: the acquisition and
1837   reprogramming of a bird's learned song. *Science* **308**, 1046-1049.

1838   GARLAND, E. C., GOLDIZEN, A. W., REKDAHL, M. L., CONSTANTINE, R., GARRIGUE, C., HAUSER,
1839   N. D., POOLE, M. M., ROBBINS, J. & NOAD, M. J. (2011). Dynamic horizontal cultural
1840   transmission of humpback whale song at the ocean basin scale. *Current Biology* **21**, 687-691.

1841 GARLAND, E. C., LILLEY, M. S., GOLDIZEN, A. W., REKDAHL, M. L., GARRIGUE, C. & NOAD, M. J.
1842 (2012). Improved versions of the Levenshtein distance method for comparing sequence
1843 information in animals' vocalisations: tests using humpback whale song. *Behaviour* **149**, 1413-
1844 1441.

1845 GARLAND, E. C., NOAD, M. J., GOLDIZEN, A. W., LILLEY, M. S., REKDAHL, M. L., GARRIGUE, C.,
1846 CONSTANTINE, R., HAUSER, N. D., POOLE, M. M. & ROBBINS, J. (2013). Quantifying humpback
1847 whale song sequences to understand the dynamics of song exchange at the ocean basin scale. *The*
1848 *Journal of the Acoustical Society of America* **133**, 560-569.

1849 GENTNER, T. Q. & HULSE, S. H. (1998). Perceptual mechanisms for individual vocal recognition
1850 in European starlings, *Sturnus vulgaris*. *Animal Behaviour* **56**, 579-594.

1851 GENTNER, T. Q., FENN, K. M., MARGOLIASH, D. & NUSBAUM, H. C. (2006). Recursive syntactic
1852 pattern learning by songbirds. *Nature* **440**, 1204-1207.

1853 GERHARDT, H. C. & HUBER, F. (2002). *Acoustic Communication in Insects and Anurans:*
1854 *Common Problems and Diverse Solutions*. University of Chicago Press.

1855 GERHARDT, H. C. (2001). Acoustic communication in two groups of closely related treefrogs.
1856 *Advances in the Study of Behavior* **30**, 99-167.

1857 GHAZANFAR, A. A., FLOMBAUM, J. I., MILLER, C. T. & HAUSER, M. D. (2001). The units of
1858 perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*):
1859 playback experiments with long calls. *Journal of Comparative Physiology A* **187**, 27-35.

1860 GIRGENRATH, M. & MARSH, R. (1997). In vivo performance of trunk muscles in tree frogs during
1861 calling. *Journal of Experimental Biology* **200**, 3101-3108.

1862 GOLD, E. M. (1967). Language identification in the limit. *Information and control* **10**, 447-474.

1863 GOLDBOGEN, J. A., SOUTHALL, B. L., DERUITER, S. L., CALAMBOKIDIS, J., FRIEDLAENDER, A. S.,
1864 HAZEN, E. L., FALCONE, E. A., SCHORR, G. S., DOUGLAS, A. & MORETTI, D. J. (2013). Blue
1865 whales respond to simulated mid-frequency military sonar. *Proceedings of the Royal Society B:*
1866 *Biological Sciences* **280**, 20130657.

1867 GOUZOULES, S., GOUZOULES, H. & MARLER, P. (1984). Rhesus monkey (*Macaca mulatta*)
1868 screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour* **32**,
1869 182-193.

1870 GREEN, S. R., MERCADO III, E., PACK, A. A. & HERMAN, L. M. (2011). Recurring patterns in the
1871 songs of humpback whales (*Megaptera novaeangliae*). *Behavioural Processes* **86**, 284-294.

1872 GRIEVES, L., LOGUE, D. & QUINN, J. (2014). Joint-nesting smooth-billed anis, *Crotophaga ani*,
1873 use a functionally referential alarm call system. *Animal Behaviour* **89**, 215-221.

1874 GRIFFITHS, T. D. & WARREN, J. D. (2004). What is an auditory object? *Nature Reviews*
1875 *Neuroscience* **5**, 887-892.

1876 GWILLIAM, J., CHARRIER, I. & HARCOURT, R. G. (2008). Vocal identity and species recognition in
1877 male Australian sea lions,*Neophoca cinerea. The Journal of Experimental Biology* **211**, 2288-
1878 2295.

1879 HAHNLOSER, R. H., KOZHEVNIKOV, A. A. & FEE, M. S. (2002). An ultra-sparse code underlies the
1880 generation of neural sequences in a songbird. *Nature* **419**, 65-70.

1881 HAILMAN, J. P. (2008). *Coding and Redundancy: Man-Made and Animal-Evolved Signals*.
1882 Harvard University Press.

1883 HAMAKER, J., GANAPATHIRAJU, A. & PICONE, J. (1998). Information theoretic approaches to
1884 model selection. *Proceedings of the International Conference on Spoken Language Processing*
1885 **7**, 2931-2934.

1886 HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M. & MORRIS, M. (2008).
1887 statnet: Software tools for the representation, visualization, analysis and simulation of network
1888 data. *Journal of Statistical Software* **24**, 1548-1557.

1889 HANDEL, S., TODD, S. K. & ZOIDIS, A. M. (2009). Rhythmic structure in humpback whale
1890 (*Megaptera novaeangliae*) songs: Preliminary implications for song production and perception.
1891 *The Journal of the Acoustical Society of America* **125**, EL225-EL230.

1892 HANDEL, S., TODD, S. K. & ZOIDIS, A. M. (2012). Hierarchical and rhythmic organization in the
1893 songs of humpback whales (*Megaptera novaeangliae*). *Bioacoustics* **21**, 141-156.

1894 HARLEY, H. E. (2008). Whistle discrimination and categorization by the Atlantic bottlenose
1895 dolphin (*Tursiops truncatus*): A review of the signature whistle framework and a perceptual test.
1896 *Behavioural processes* **77**, 243-268.

1897 HARNAD, S. R. (1990). *Categorical Perception: The Groundwork of Cognition*. Cambridge
1898 University Press.

1899 HARRINGTON, F. H., ASA, C. S., MECH, L. & BOITANI, L. (2003). Wolf communication. In
1900 *Wolves: Behavior, ecology, and conservation*, pp. 66-103. University of Chicago Press, Chicago.

1901 HARTLEY, R. S. & SUTHERS, R. A. (1989). Airflow and pressure during canary song: direct
1902 evidence for mini-breaths. *Journal of Comparative Physiology A* **165**, 15-26.

1903 HARTSHORNE, C. (1956). The monotony-threshold in singing birds. *The Auk* **73**, 176-192.

1904 HARTSHORNE, C. (1973). *Born to sing. An interpretation and world survey of bird songs*.
1905 London: Indiana University Press.

1906　HASSELQUIST, D., BENSCH, S. & VON SCHANTZ, T. (1996). Correlation between male song
1907　repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229-
1908　232.

1909　HAUSER, M. D., AGNETTA, B. & PEREZ, C. (1998). Orienting asymmetries in rhesus monkeys: the
1910　effect of time-domain changes on acoustic perception. *Animal Behaviour* **56**, 41-47.

1911　HAUSER, M. D., CHOMSKY, N. & FITCH, W. (2002). The faculty of language: What is it, who has
1912　it, and how did it evolve? *Science* **298**, 1569-1579.

1913　HAUSSER, J. & STRIMMER, K. (2009). Entropy inference and the James-Stein estimator, with
1914　application to nonlinear gene association networks. *The Journal of Machine Learning Research*
1915　**10**, 1469-1484.

1916　HAYES, S. A., KUMAR, A., COSTA, D. P., MELLINGER, D. K., HARVEY, J. T., SOUTHALL, B. L. &
1917　LE BOEUF, B. J. (2004). Evaluating the function of the male harbour seal, *Phoca vitulina*, roar
1918　through playback experiments. *Animal Behaviour* **67**, 1133-1139.

1919　HEBETS, E. A. & PAPAJ, D. R. (2005). Complex signal function: developing a framework of
1920　testable hypotheses. *Behavioral Ecology and Sociobiology* **57**, 197-214.

1921　HELBLE, T. A., IERLEY, G. R., GERALD, L., ROCH, M. A. & HILDEBRAND, J. A. (2012). A
1922　generalized power-law detection algorithm for humpback whale vocalizations. *The Journal of*
1923　*the Acoustical Society of America* **131**, 2682-2699.

1924　HENDERSON, E. E., HILDEBRAND, J. A. & SMITH, M. H. (2011). Classification of behavior using
1925　vocalizations of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*). *The Journal of the*
1926　*Acoustical Society of America* **130**, 557-567.

1927　HENRY, K. S. & LUCAS, J. R. (2010). Habitat-related differences in the frequency selectivity of
1928　auditory filters in songbirds. *Functional Ecology* **24**, 614-624.

1929　HENRY, K. S., GALL, M. D., BIDELMAN, G. M. & LUCAS, J. R. (2011). Songbirds tradeoff auditory
1930　frequency resolution and temporal resolution. *Journal of Comparative Physiology A:*
1931　*Neuroethology, Sensory, Neural, and Behavioral Physiology* **197**, 351-359.

1932　HERZING, D. L. (1996). Vocalizations and associated underwater behavior of free-ranging
1933　Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*. *Aquatic*
1934　*Mammals* **22**, 61-80.

1935　HOLVECK, M., DE CASTRO, ANA CATARINA VIEIRA, LACHLAN, R. F., TEN CATE, C. & RIEBEL, K.
1936　(2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra
1937　finches. *Behavioral Ecology* **19**, 1267-1281.

1938 HORNING, C. L., BEECHER, M. D., STODDARD, P. K. & CAMPBELL, S. E. (1993). Song perception
1939 in the song sparrow: importance of different parts of the song in song type classification.
1940 *Ethology* **94**, 46-58.

1941 HULSE, S. H. (2002). Auditory scene analysis in animal communication. *Advances in the Study of*
1942 *Behavior* **31**, 163-200.

1943 HUMPHRIES, M. D. & GURNEY, K. (2008). Network 'small-world-ness': a quantitative method for
1944 determining canonical network equivalence. *PLoS One* **3**, e0002051.

1945 ILANY, A., BAROCAS, A., KAM, M., ILANY, T. & GEFFEN, E. (2013). The energy cost of singing in
1946 wild rock hyrax males: evidence for an index signal. *Animal Behaviour* **85**, 995-1001.

1947 ILANY, A., BAROCAS, A., KOREN, L., KAM, M. & GEFFEN, E. (2011). Do singing rock hyraxes
1948 exploit conspecific calls to gain attention? *PloS One* **6**, e28612.

1949 IRWIN, D. E. (2000). Song variation in an avian ring species. *Evolution* **54**, 998-1010.

1950 ISAAC, D. & MARLER, P. (1963). Ordering of sequences of singing behaviour of mistle thrushes
1951 in relationship to timing. *Animal Behaviour* **11**, 179-188.

1952 JANIK, V. M. (1999). Pitfalls in the categorization of behaviour: a comparison of dolphin whistle
1953 classification methods. *Animal Behaviour* **57**, 133-143.

1954 JANIK, V. M., KING, S. L., SAYIGH, L. S. & WELLS, R. S. (2013). Identifying signature whistles
1955 from recordings of groups of unrestrained bottlenose dolphins (*Tursiops truncatus*). *Marine*
1956 *Mammal Science* **29**, 109-122.

1957 JANIK, V. M., SAYIGH, L. & WELLS, R. (2006). Signature whistle shape conveys identity
1958 information to bottlenose dolphins. *Proceedings of the National Academy of Sciences* **103**, 8293-
1959 8297.

1960 JANSEN, D. A., CANT, M. A. & MANSER, M. B. (2012). Segmental concatenation of individual
1961 signatures and con-text cues in banded mongoose (*Mungos mungo*) close calls. *BMC biology* **10**,
1962 97.

1963 JI, A., JOHNSON, M. T., WALSH, E. J., MCGEE, J. & ARMSTRONG, D. L. (2013). Discrimination of
1964 individual tigers (*Panthera tigris*) from long distance roars. *The Journal of the Acoustical Society*
1965 *of America* **133**, 1762-1769.

1966 JIN, D. Z. & KOZHEVNIKOV, A. A. (2011). A compact statistical model of the song syntax in
1967 Bengalese finch. *PLoS Computational Biology* **7**, e1001108.

1968 JIN, D. Z. (2009). Generating variable birdsong syllable sequences with branching chain
1969 networks in avian premotor nucleus HVC. *Physical Review E* **80**, 051902.

1970   JOHNSON, K. (2004). Gold's Theorem and cognitive science. *Philosophy of Science* **71**, 571-592.

1971   JONES, A. E., TEN CATE, C. & BIJLEVELD, C. C. (2001). The interobserver reliability of scoring
1972   sonagrams by eye: a study on methods, illustrated on zebra finch songs. *Animal Behaviour* **62**,
1973   791-801.

1974   JOUVENTIN, P., AUBIN, T. & LENGAGNE, T. (1999). Finding a parent in a king penguin colony: the
1975   acoustic system of individual recognition. *Animal Behaviour* **57**, 1175-1183.

1976   JURAFSKY, D. & MARTIN, J. H. (2000). *Speech & language processing*. Pearson Education India.

1977   KAKISHITA, Y., SASAHARA, K., NISHINO, T., TAKAHASI, M. & OKANOYA, K. (2009). Ethological
1978   data mining: an automata-based approach to extract behavioral units and rules. *Data Mining and*
1979   *Knowledge Discovery* **18**, 446-471.

1980   KANWAL, J. S., MATSUMURA, S., OHLEMILLER, K. & SUGA, N. (1994). Analysis of acoustic
1981   elements and syntax in communication sounds emitted by mustached bats. *The Journal of the*
1982   *Acoustical Society of America* **96**, 1229-1254.

1983   KAPLAN, D. (1978). Dthat. *Syntax and semantics* **9**, 221-243.

1984   KASHTAN, N., ITZKOVITZ, S., MILO, R. & ALON, U. (2002). Mfinder tool guide. *Department of*
1985   *Molecular Cell Biology and Computer Science and Applied Mathematics, Weizmann Institute of*
1986   *Science, Rehovot Israel, Tech.Rep* .

1987   KATAHIRA, K., SUZUKI, K., OKANOYA, K. & OKADA, M. (2011). Complex sequencing rules of
1988   birdsong can be explained by simple hidden Markov processes. *PloS One* **6**, e24516.

1989   KATZ, S. (1987). Estimation of probabilities from sparse data for the language model component
1990   of a speech recognizer. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **35**, 400-
1991   401.

1992   KERSHENBAUM, A. & ROCH, M. A. (2013). An image processing based paradigm for the
1993   extraction of tonal sounds in cetacean communications. *Journal of the Acoustical Society of*
1994   *America* **134**, 4435-4445.

1995   KERSHENBAUM, A. (2013). Entropy rate as a measure of animal vocal complexity. *Bioacoustics* .

1996   KERSHENBAUM, A., BOWLES, A. E., FREEBERG, T. M., JIN, D. Z., LAMEIRA, A. R. & BOHN, K.
1997   (2014). Animal vocal sequences: not the Markov chains we thought they were. *Proceedings of*
1998   *the Royal Society B: Biological Sciences***281**, 20141370

1999   KERSHENBAUM, A., ILANY, A., BLAUSTEIN, L. & GEFFEN, E. (2012). Syntactic structure and
2000   geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B:*
2001   *Biological Sciences* **279**, 2974-2981.

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

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.

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

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

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

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

2034 KROODSMA, D. E. (1990). Patterns in songbird singing behaviour: Hartshorne vindicated. *Animal*
2035 *Behaviour* **39**, 994-996.

2036 LACHLAN, R. F., VERZIJDEN, M. N., BERNARD, C. S., JONKER, P., KOESE, B., JAARSMA, S., SPOOR,
2037 W., SLATER, P. J. & TEN CATE, C. (2013). The progressive loss of syntactical structure in bird
2038 song along an island colonization chain. *Current Biology* **19**, 1896-1901.

2039 LAMBRECHTS, M. & DHONDT, A. (1987). Differences in singing performance between male great
2040 tits. *Ardea* **75**, 43-52.

2041 LAMBRECHTS, M. & DHONDT, A. A. (1988). The anti-exhaustion hypothesis: a new hypothesis to
2042 explain song performance and song switching in the great tit. *Animal Behaviour* **36**, 327-334.

2043 LAMMERS, M. O. & AU, W. W. (2003). Directionality in the whistles of Hawaiian spinner
2044 dolphins (*Stenella longirostris*): A signal feature to cue direction of movement? *Marine Mammal*
2045 *Science* **19**, 249-264.

2046 LARSON, K. A. (2004). Advertisement call complexity in northern leopard frogs, *Rana pipiens*.
2047 *Journal Information* **2004**.

2048 LEE, K. & HON, H. (1989). Speaker-independent phone recognition using hidden Markov
2049 models. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **37**, 1641-1648.

2050 LEHONGRE, K., AUBIN, T., ROBIN, S. & DEL NEGRO, C. (2008). Individual signature in canary
2051 songs: contribution of multiple levels of song structure. *Ethology* **114**, 425-435.

2052 LEMASSON, A., OUATTARA, K., BOUCHET, H. & ZUBERBÜHLER, K. (2010). Speed of call delivery
2053 is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften* **97**,
2054 1023-1027.

2055 LENGAGNE, T., AUBIN, T., JOUVENTIN, P. & LAUGA, J. (2000). Perceptual salience of individually
2056 distinctive features in the calls of adult king penguins. *The Journal of the Acoustical Society of*
2057 *America* **107**, 508-516.

2058 LENGAGNE, T., LAUGA, J. & AUBIN, T. (2001). Intra-syllabic acoustic signatures used by the king
2059 penguin in parent-chick recognition: an experimental approach. *The Journal of Experimental*
2060 *Biology* **204**, 663-672.

2061 LIPKIND, D., MARCUS, G. F., BEMIS, D. K., SASAHARA, K., JACOBY, N., TAKAHASI, M., SUZUKI,
2062 K., FEHER, O., RAVBAR, P. & OKANOYA, K. (2013). Stepwise acquisition of vocal combinatorial
2063 capacity in songbirds and human infants. *Nature* **498**, 104-108.

2064 LIU, I. A., LOHR, B., OLSEN, B. & GREENBERG, R. (2008). Macrogeographic vocal variation in
2065 subspecies of swamp sparrow. *The Condor* **110**, 102-109.

2066 LOHR, B., ASHBY, S. & WAKAMIYA, S. M. (2013). The function of song types and song
2067 components in Grasshopper Sparrows (*Ammodramus savannarum*). *Behaviour* **150**, 1085-1106.

2068 LYNCH, K. S., STANELY RAND, A., RYAN, M. J. & WILCZYNSKI, W. (2005). Plasticity in female
2069 mate choice associated with changing reproductive states. *Animal Behaviour* **69**, 689-699.

2070 MACEDONIA, J. M. (1990). What is communicated in the antipredator calls of lemurs: evidence
2071 from playback experiments with ringtailed and ruffed lemurs. *Ethology* **86**, 177-190.

2072 MAHURIN, E. J. & FREEBERG, T. M. (2009). Chick-a-dee call variation in Carolina chickadees and
2073 recruiting flockmates to food. *Behavioral Ecology* **20**, 111-116.

2074 MAKEIG, S., DEBENER, S., ONTON, J. & DELORME, A. (2004). Mining event-related brain
2075 dynamics. *Trends in Cognitive Sciences* **8**, 204-210.

2076 MALLAT, S. (1999). *A Wavelet Tour of Signal Processing*. Access Online via Elsevier.

2077 MANNING, C. D., RAGHAVAN, P. & SCHÜTZE, H. (2008). *Introduction to information retrieval*.
2078 Cambridge University Press, Cambridge.

2079 MANSER, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type
2080 and the level of response urgency. *Proceedings of the Royal Society of London. Series B:*
2081 *Biological Sciences* **268**, 2315-2324.

2082 MARLER, P. & PICKERT, R. (1984). Species-universal microstructure in the learned song of the
2083 swamp sparrow (*Melospiza georgiana*). *Animal Behaviour* **32**, 673-689.

2084 MARLER, P. & SHERMAN, V. (1985). Innate differences in singing behaviour of sparrows reared
2085 in isolation from adult conspecific song. *Animal Behaviour* **33**, 57-71.

2086 MARLER, P. (1977). The structure of animal communication sounds. In *Recognition of Complex*
2087 *Acoustic Signals* (ed. T. H. Bullock), pp. 17-35. Springer Verlag, Berlin.

2088 MARTIN, W. & FLANDRIN, P. (1985). Wigner-Ville spectral analysis of nonstationary processes.
2089 *Acoustics, Speech and Signal Processing, IEEE Transactions on* **33**, 1461-1470.

2090 MARTIN, W. F. & GANS, C. (1972). Muscular control of the vocal tract during release signaling in
2091 the toad *Bufo valliceps*. *Journal of Morphology* **137**, 1-27.

2092 MARTIN, W. P. (1972). Evolution of vocalization in the genus Bufo. In *Evolution in the*
2093 *GenusBufo* (ed. W. F. Blair), pp. 37-70. University of Texas Press, Austin, TX.

2094 MARTINICH, A. & SOSA, D. (2013). *The Philosophy of Language*. Oxford University Press,
2095 Oxford.

2096 MATHEVON, N. & AUBIN, T. (2001). Sound-based species-specific recognition in the
2097 blackcap*Sylvia atricapilla*shows high tolerance to signal modifications. *Behaviour* **138**, 511-524.

2098 MATHEVON, N., AUBIN, T., VIELLIARD, J., DA SILVA, M., SEBE, F. & BOSCOLO, D. (2008).
2099 Singing in the rain forest: how a tropical bird song transfers information. *PLoS One* **3**, e1580.

2100 MATHEVON, N., KORALEK, A., WELDELE, M., GLICKMAN, S. E. & THEUNISSEN, F. E. (2010).
2101 What the hyena's laugh tells: Sex, age, dominance and individual signature in the giggling call of
2102 *Crocuta crocuta*. *BMC Ecology* **10**, 9.

2103 MATTHEWS, J., RENDELL, L., GORDON, J. & MACDONALD, D. (1999). A review of frequency and
2104 time parameters of cetacean tonal calls. *Bioacoustics* **10**, 47-71.

2105 MCCOWAN, B. & REISS, D. (1995). Quantitative Comparison of whistle repertoires from captive
2106 adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): a re-evaluation of the signature
2107 whistle hypothesis. *Ethology* **100**, 194-209.

2108 MCCOWAN, B., DOYLE, L. R. & HANSER, S. F. (2002). Using information theory to assess the
2109 diversity, complexity, and development of communicative repertoires. *Journal of Comparative
2110 Psychology* **116**, 166-172.

2111 MCCOWAN, B., HANSER, S. F. & DOYLE, L. R. (1999). Quantitative tools for comparing animal
2112 communication systems: information theory applied to bottlenose dolphin whistle repertoires.
2113 *Animal Behaviour* **57**, 409-419.

2114 MCSHEA, D. W. (1991). Complexity and evolution: what everybody knows. *Biology and
2115 Philosophy* **6**, 303-324.

2116 MCSHEA, D. W. (2009). The evolution of complexity without natural selection, a possible large-
2117 scale trend of the fourth kind.

2118 MENNILL, D. J. & RATCLIFFE, L. M. (2004). Overlapping and matching in the song contests of
2119 black-capped chickadees. *Animal Behaviour* **67**, 441-450.

2120 MENNILL, D. J., RATCLIFFE, L. M. & BOAG, P. T. (2002). Female eavesdropping on male song
2121 contests in songbirds. *Science* **296**, 873-873.

2122 MERCADO, E. I. & HANDEL, S. (2012). Understanding the structure of humpback whale songs
2123 (L). *The Journal of the Acoustical Society of America* **132**, 2947-2950.

2124 MERCADO, E. I., HERMAN, L. M. & PACK, A. A. (2003). Stereotypical sound patterns in
2125 humpback whale songs: Usage and function. *Aquatic Mammals* **29**, 37-52.

2126 MERCADO, E. I., SCHNEIDER, J. N., PACK, A. A. & HERMAN, L. M. (2010). Sound production by
2127 singing humpback whales. *The Journal of the Acoustical Society of America* **127**, 2678-2691.

2128 MILLER, C. T. & BEE, M. A. (2012). Receiver psychology turns 20: is it time for a broader
2129 approach? *Animal Behaviour* **83**, 331-343.

2130 MILLER, C. T. & COHEN, Y. E. (2010). Vocalizations as auditory objects: behavior and
2131 neurophysiology. In *Primate Neuroethology* (eds. M. L. Platt and A. A. Ghazanfar), pp. 237-255.
2132 Oxford University Press, Oxford.

2133 MILLER, P. J., SAMARRA, F. I. & PERTHUISON, A. D. (2007). Caller sex and orientation influence
2134 spectral characteristics of "two-voice" stereotyped calls produced by free-ranging killer whales.
2135 *The Journal of the Acoustical Society of America* **121**, 3932-3937.

2136 MILLIKAN, R. G. (2004). *Varieties of meaning: the 2002 Jean Nicod lectures*. MIT press.

2137 MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. & ALON, U. (2002).
2138 Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827.

2139 MITANI, J. C., HUNLEY, K. & MURDOCH, M. (1999). Geographic variation in the calls of wild
2140 chimpanzees: a reassessment. *American Journal of Primatology* **47**, 133-151.

2141 MITCHELL, M. (2009). *Complexity: A guided tour*. Oxford University Press.

2142 MOORE, B. C. & MOORE, B. C. (2003). *An Introduction to the Psychology of Hearing*. Academic
2143 Press, San Diego.

2144 MULLIGAN, J. A. (1966). *Singing Behavior and its Development in the Song Sparrow Melospiza
2145 melodia*. University of California Press.

2146 MUMFORD, D. & DESOLNEUX, A. (2010). *Pattern Theory: The Stochastic Analysis of Real-World
2147 Signals*. A. K. Peters Ltd.

2148 MUNOZ, N. E. & BLUMSTEIN, D. T. (2012). Multisensory perception in uncertain environments.
2149 *Behavioral Ecology* **23**, 457-462.

2150 MURRAY, S. O., MERCADO, E. I. & ROITBLAT, H. L. (1998). Characterizing the graded structure
2151 of false killer whale (*Pseudorca crassidens*) vocalizations. *The Journal of the Acoustical Society
2152 of America* **104**, 1679-1688.

2153 NARINS, P. M., LEWIS, E. R. & MCCLELLAND, B. E. (2000). Hyperextended call note repertoire of
2154 the endemic Madagascar treefrog *Boophis madagascariensis* (Rhacophoridae). *Journal of
2155 Zoology* **250**, 283-298.

2156 NARINS, P. M., REICHMAN, O., JARVIS, J. U. & LEWIS, E. R. (1992). Seismic signal transmission
2157 between burrows of the Cape mole-rat, *Georychus capensis*. *Journal of Comparative Physiology
2158 A* **170**, 13-21.

2159   NELSON, D. A. & MARLER, P. (1989). Categorical perception of a natural stimulus continuum:
2160   birdsong. *Science* **244**, 976-978.

2161   NELSON, D. A. & POESEL, A. (2007). Segregation of information in a complex acoustic signal:
2162   individual and dialect identity in white-crowned sparrow song. *Animal Behaviour* **74**, 1073-1084.

2163   NETTLE, D. (1999). Language variation and the evolution of societies. In *The Evolution of*
2164   *Culture: An Interdisciplinary View* (eds. R. I. M. Dunbar, C. Knight and C. Power), pp. 214-227.
2165   Rutgers University Press.

2166   NEUBAUER, R. L. (1999). Super-normal length song preferences of female zebra finches
2167   (*Taeniopygia guttata*) and a theory of the evolution of bird song. *Evolutionary Ecology* **13**, 365-
2168   380.

2169   NEVO, E., HETH, G., BEILES, A. & FRANKENBERG, E. (1987). Geographic dialects in blind mole
2170   rats: role of vocal communication in active speciation. *Proceedings of the National Academy of*
2171   *Sciences* **84**, 3312-3315.

2172   NEWMAN, M. (2009). *Networks: An Introduction*. Oxford University Press, Oxford.

2173   NEWMAN, M. E. (2003). The structure and function of complex networks. *SIAM Review* **45**, 167-
2174   256.

2175   NITYANANDA, V. & BEE, M. A. (2011). Finding your mate at a cocktail party: frequency
2176   separation promotes auditory stream segregation of concurrent voices in multi-species frog
2177   choruses. *PloS One* **6**, e21191.

2178   NOTMAN, H. & RENDALL, D. (2005). Contextual variation in chimpanzee pant hoots and its
2179   implications for referential communication. *Animal Behaviour* **70**, 177-190.

2180   NOWICKI, S. & MARLER, P. (1988). How do birds sing? *Music Perception* , 391-426.

2181   NOWICKI, S. & NELSON, D. A. (1990). Defining natural categories in acoustic signals:
2182   comparison of three methods applied to 'chick-a-dee' call notes. *Ethology* **86**, 89-101.

2183   OHMS, V. R., ESCUDERO, P., LAMMERS, K. & TEN CATE, C. (2012). Zebra finches and Dutch
2184   adults exhibit the same cue weighting bias in vowel perception. *Animal Cognition* **15**, 155-161.

2185   OLESON, E. M., WIGGINS, S. M. & HILDEBRAND, J. A. (2007). Temporal separation of blue whale
2186   call types on a southern California feeding ground. *Animal Behaviour* **74**, 881-894.

2187   OPPENHEIM, A. V. & SCHAFER, R. W. (2004). From frequency to quefrency: A history of the
2188   cepstrum. *Signal Processing Magazine, IEEE* **21**, 95-106.

2189   OPSAHL, T., AGNEESSENS, F. & SKVORETZ, J. (2010). Node centrality in weighted networks:
2190   Generalizing degree and shortest paths. *Social Networks* **32**, 245-251.

2191 OSWALD, J. N., RANKIN, S., BARLOW, J. & LAMMERS, M. O. (2007). A tool for real-time acoustic
2192 species identification of delphinid whistles. *The Journal of the Acoustical Society of America*
2193 **122**, 587-595.

2194 OTTER, K., MCGREGOR, P. K., TERRY, A. M., BURFORD, F. R., PEAKE, T. M. & DABELSTEEN, T.
2195 (1999). Do female great tits (*Parus major*) assess males by eavesdropping? A field study using
2196 interactive song playback. *Proceedings of the Royal Society of London. Series B: Biological*
2197 *Sciences* **266**, 1305-1309.

2198 OWREN, M. J., RENDALL, D. & RYAN, M. J. (2010). Redefining animal signaling: influence versus
2199 information in communication. *Biology & Philosophy* **25**, 755-780.

2200 PACK, A. A., HERMAN, L. M., HOFFMANN-KUHNT, M. & BRANSTETTER, B. K. (2002). The object
2201 behind the echo: dolphins (*Tursiops truncatus*) perceive object shape globally through
2202 echolocation. *Behavioural Processes* **58**, 1-26.

2203 PAGE, S. E. (2010). *Diversity and complexity*. Princeton University Press.

2204 PAREJO, D., AVILES, J. M. & RODRIGUEZ, J. (2012). Alarm calls modulate the spatial structure of
2205 a breeding owl community. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2135-
2206 2141.

2207 PARSONS, S., RISKIN, D. K. & HERMANSON, J. W. (2010). Echolocation call production during
2208 aerial and terrestrial locomotion by New Zealand's enigmatic lesser short-tailed bat, *Mystacina*
2209 *tuberculata*. *The Journal of Experimental Biology* **213**, 551-557.

2210 PARTAN, S. & MARLER, P. (1999). Communication goes multimodal. *Science* **283**, 1272-1273.

2211 PAULUS, J., MÜLLER, M. & KLAPURI, A. (2010). State of the Art Report: Audio-Based Music
2212 Structure Analysis. *Proceedings of the International Conference on Music Information Retrieval*
2213 *(ISMIR)*, 625-636.

2214 PAYNE, K. B., LANGBAUER JR, W. R. & THOMAS, E. M. (1986). Infrasonic calls of the Asian
2215 elephant (*Elephas maximus*). *Behavioral Ecology and Sociobiology* **18**, 297-301.

2216 PAYNE, K., TYACK, P. & PAYNE, R. (1983). Progressive changes in the songs of humpback
2217 whales (*Megaptera novaeangliae*): a detailed analysis of two seasons in Hawaii. *Communication*
2218 *and behavior of whales*, 9-57.

2219 PAYNE, R. S. & MCVAY, S. (1971). Songs of humpback whales. *Science* **173**, 585-597.

2220 PETERSON, G. E. & BARNEY, H. L. (1952). Control methods used in a study of the vowels. *The*
2221 *Journal of the Acoustical Society of America* **24**, 175-184.

2222 PICONE, J. W. (1993). Signal modeling techniques in speech recognition. *Proceedings of the*
2223 *IEEE* **81**, 1215-1247.

2224 PITCHER, B. J., HARCOURT, R. G. & CHARRIER, I. (2012). Individual identity encoding and
2225 environmental constraints in vocal recognition of pups by Australian sea lion mothers. *Animal*
2226 *Behaviour* **83**, 681-690.

2227 PODOS, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird
2228 family (Passeriformes: Emberizidae). *Evolution* , 537-551.

2229 PODOS, J., PETERS, S., RUDNICKY, T., MARLER, P. & NOWICKI, S. (1992). The organization of
2230 song repertoires in song sparrows: themes and variations. *Ethology* **90**, 89-106.

2231 POLLARD, K. A. & BLUMSTEIN, D. T. (2011). Social group size predicts the evolution of
2232 individuality. *Current Biology* **21**, 413-417.

2233 POLLARD, K. A. & BLUMSTEIN, D. T. (2012). Evolving communicative complexity: insights from
2234 rodents and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**,
2235 1869-1878.

2236 POOLE, J. H. (1999). Signals and assessment in African elephants: evidence from playback
2237 experiments. *Animal Behaviour* **58**, 185-193.

2238 PUTNAM, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*
2239 **7**, 131-193.

2240 QUICK, N. J. & JANIK, V. M. (2012). Bottlenose dolphins exchange signature whistles when
2241 meeting at sea. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2539-2545.

2242 QUINE, W. V. O. (1960). *Word and object*. MIT press.

2243 R DEVELOPMENT TEAM (2012). *R: A Language and Environment for Statistical Computing.R*
2244 *Foundation for Statistical Computing, Vienna, Austria, 2007* .

2245 RABINER, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech
2246 recognition. *Proceedings of the IEEE* **77**, 257-286.

2247 RAEMAEKERS, J. J., RAEMAEKERS, P. M. & HAIMOFF, E. H. (1984). Loud calls of the gibbon
2248 (*Hylobates lar*): repertoire, organisation and context. *Behaviour* **91**, 146-189.

2249 RAGGE, D. & REYNOLDS, W. (1988). The songs and taxonomy of the grasshoppers of the
2250 *Chorthippus biguttulus* group in the Iberian Peninsula (Orthoptera: Acrididae). *Journal of*
2251 *Natural History* **22**, 897-929.

2252 RANDALL, J. A. (1989). Individual footdrumming signatures in banner-tailed kangaroo rats
2253 *Dipodomys spectabilis*. *Animal Behaviour* **38**, 620-630.

2254 RANDALL, J. A. (1997). Species-specific footdrumming in kangaroo rats: *Dipodomys ingens*, *D.*
2255 *deserti*, *D. spectabilis*. *Animal Behaviour* **54**, 1167-1175.

2256 RECANZONE, G. H. & SUTTER, M. L. (2008). The biological basis of audition. *Annu.Rev.Psychol.*
2257 **59**, 119-142.

2258 REMEZ, R. E., RUBIN, P. E., BERNS, S. M., PARDO, J. S. & LANG, J. M. (1994). On the perceptual
2259 organization of speech. *Psychological Review* **101**, 129.

2260 RIBEIRO, S., CECCHI, G. A., MAGNASCO, M. O. & MELLO, C. V. (1998). Toward a song code:
2261 evidence for a syllabic representation in the canary brain. *Neuron* **21**, 359-371.

2262 RICHARDS, D. G. (1981). Alerting and message components in songs of rufous-sided towhees.
2263 *Behaviour* **76**, 223-249.

2264 RIEBEL, K. & SLATER, P. (2003). Temporal variation in male chaffinch song depends on the
2265 singer and the song type. *Behaviour* **140**, 269-288.

2266 RIEDE, T. & FITCH, T. (1999). Vocal tract length and acoustics of vocalization in the domestic
2267 dog (*Canis familiaris*). *The Journal of Experimental Biology* **202**, 2859-2867.

2268 RIEDE, T., BRONSON, E., HATZIKIROU, H. & ZUBERBÜHLER, K. (2005). Vocal production
2269 mechanisms in a non-human primate: morphological data and a model. *Journal of Human*
2270 *Evolution* **48**, 85-96.

2271 RIPLEY, B. D. (2007). *Pattern recognition and neural networks*. Cambridge University Press.

2272 ROBINSON, J. G. (1979). An analysis of the organization of vocal communication in the titi
2273 monkey *Callicebus moloch*. *Zeitschrift für Tierpsychologie* **49**, 381-405.

2274 ROBISSON, P., AUBIN, T. & BREMOND, J. (1993). Individuality in the voice of the emperor
2275 penguin*Aptenodytes forsteri*: adaptation to a noisy environment. *Ethology* **94**, 279-290.

2276 ROOT-GUTTERIDGE, H., BENCSIK, M., CHEBLI, M., GENTLE, L. K., TERRELL-NIELD, C., BOURIT,
2277 A. & YARNELL, R. W. (2014). Identifying individual wild Eastern grey wolves (*Canis lupus*
2278 *lycaon*) using fundamental frequency and amplitude of howls. *Bioacoustics* **23**, 55-66.

2279 ROTHENBERG, D., ROESKE, T. C., VOSS, H. U., NAGUIB, M. & TCHERNICHOVSKI, O. (2013).
2280 Investigation of musicality in birdsong. *Hearing Research* **308**, 71-83.

2281 RUXTON, G. D. & SCHAEFER, H. M. (2011). Resolving current disagreements and ambiguities in
2282 the terminology of animal communication. *Journal of Evolutionary Biology* **24**, 2574-2585.

2283 SAAR, S. & MITRA, P. P. (2008). A technique for characterizing the development of rhythms in
2284 bird song. *PLoS One* **3**, e1461.

2285 SALMI, R., HAMMERSCHMIDT, K. & DORAN-SHEEHY, D. M. (2013). Western Gorilla Vocal
2286 Repertoire and Contextual Use of Vocalizations. *Ethology* **119**, 831-847.

2287 SASAHARA, K., CODY, M. L., COHEN, D. & TAYLOR, C. E. (2012). Structural design principles of
2288 complex bird songs: a network-based approach. *PloS One* **7**, e44436.

2289 SAUNDERS, A. A. (1951). The song of the Song Sparrow. *The Wilson Bulletin* , 99-109.

2290 SAYIGH, L. S., ESCH, H. C., WELLS, R. S. & JANIK, V. M. (2007). Facts about signature whistles
2291 of bottlenose dolphins, *Tursiops truncatus*. *Animal Behaviour* **74**, 1631-1642.

2292 SAYIGH, L. S., TYACK, P. L., WELLS, R. S., SOLOW, A. R., SCOTT, M. D. & IRVINE, A. B. (1999).
2293 Individual recognition in wild bottlenose dolphins: a field test using playback experiments.
2294 *Animal Behaviour* **57**, 41-50.

2295 SAYIGH, L., QUICK, N., HASTIE, G. & TYACK, P. (2012). Repeated call types in short-finned pilot
2296 whales, *Globicephala macrorhynchus*. *Marine Mammal Science* **29**, 312-324.

2297 SCHEL, A. M., CANDIOTTI, A. & ZUBERBÜHLER, K. (2010). Predator-deterring alarm call
2298 sequences in Guereza colobus monkeys are meaningful to conspecifics. *Animal Behaviour* **80**,
2299 799-808.

2300 SCHEL, A. M., TRANQUILLI, S. & ZUBERBÜHLER, K. (2009). The alarm call system of two species
2301 of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *Journal of
2302 Comparative Psychology* **123**, 136-150.

2303 SCHREIBER, F. & SCHWÖBBERMEYER, H. (2005). MAVisto: a tool for the exploration of network
2304 motifs. *Bioinformatics* **21**, 3572-3574.

2305 SCHULZ, T. M., WHITEHEAD, H., GERO, S. & RENDELL, L. (2008). Overlapping and matching of
2306 codas in vocal interactions between sperm whales: insights into communication function. *Animal
2307 Behaviour* **76**, 1977-1988.

2308 SCOTT, J. & CARRINGTON, P. J. (2011). *The SAGE Handbook of Social Network Analysis*. SAGE
2309 Publications.

2310 SEARCY, W. A. (1992). Song repertoire and mate choice in birds. *American Zoologist* **32**, 71-80.

2311 SEARCY, W. A. & ANDERSSON, M. (1986). Sexual selection and the evolution of song. *Annual
2312 Review of Ecology and Systematics* **17**, 507-533.

2313 SEARCY, W. A. & BEECHER, M. D. (2011). Continued scepticism that song overlapping is a
2314 signal. *Animal Behaviour* **81**, e1-e4.

2315 SEARCY, W. A. & NOWICKI, S. (2005). *The evolution of animal communication: reliability and
2316 deception in signaling systems*. Princeton University Press.

2317 SEARCY, W. A., NOWICKI, S. & PETERS, S. (1999). Song types as fundamental units in vocal
2318 repertoires. *Animal Behaviour* **58**, 37-44.

2319 SEARCY, W. A., PODOS, J., PETERS, S. & NOWICKI, S. (1995). Discrimination of song types and
2320 variants in song sparrows. *Animal Behaviour* **49**, 1219-1226.

2321 SEDDON, N. & TOBIAS, J. A. (2007). Song divergence at the edge of Amazonia: an empirical test
2322 of the peripatric speciation model. *Biological Journal of the Linnean Society* **90**, 173-188.

2323 SEYFARTH, R. M. & CHENEY, D. L. (2010). Production, usage, and comprehension in animal
2324 vocalizations. *Brain and Language* **115**, 92-100.

2325 SEYFARTH, R. M., CHENEY, D. L. & BERGMAN, T. J. (2005). Primate social cognition and the
2326 origins of language. *Trends in Cognitive Sciences* **9**, 264-266.

2327 SEYFARTH, R. M., CHENEY, D. L., BERGMAN, T., FISCHER, J., ZUBERBÜHLER, K. &
2328 HAMMERSCHMIDT, K. (2010). The central importance of information in studies of animal
2329 communication. *Animal Behaviour* **80**, 3-8.

2330 SHANNON, C. E., WEAVER, W., BLAHUT, R. E. & HAJEK, B. (1949). *The Mathematical Theory of
2331 Communication*. University of Illinois Press, Urbana.

2332 SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N.,
2333 SCHWIKOWSKI, B. & IDEKER, T. (2003). Cytoscape: a software environment for integrated
2334 models of biomolecular interaction networks. *Genome Research* **13**, 2498-2504.

2335 SHAPIRO, A. D., TYACK, P. L. & SENEFF, S. (2010). Comparing call-based versus subunit-based
2336 methods for categorizing Norwegian killer whale, *Orcinus orca*, vocalizations. *Animal
2337 Behaviour* **81**, 377-386.

2338 SLABBEKOORN, H. & SMITH, T. B. (2002). Bird song, ecology and speciation. *Philosophical
2339 Transactions of the Royal Society B: Biological sciences* **357**, 493-503.

2340 SLOCOMBE, K. E. & ZUBERBÜHLER, K. (2006). Food-associated calls in chimpanzees: responses
2341 to food types or food preferences? *Animal Behaviour* **72**, 989-999.

2342 SMITH, J. N., GOLDIZEN, A. W., DUNLOP, R. A. & NOAD, M. J. (2008). Songs of male humpback
2343 whales, *Megaptera novaeangliae*, are involved in intersexual interactions. *Animal Behaviour* **76**,
2344 467-477.

2345 SMITH, W. J. (1977). *The behavior of communicating*. Harvard University Press.

2346 SNIJDERS, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph
2347 models. *Journal of Social Structure* **3**, 1-40.

2348 SOLAN, Z., HORN, D., RUPPIN, E. & EDELMAN, S. (2005). Unsupervised learning of natural
2349 languages. *Proceedings of the National Academy of Sciences* **102**, 11629-11634.

2350 SPECHT, R. (2004). Avisoft-SASLab Pro. Avisoft, Berlin .

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

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.

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.

2384 TOEWS, D. P. & IRWIN, D. E. (2008). Cryptic speciation in a Holarctic passerine revealed by
2385 genetic and bioacoustic analyses. *Molecular Ecology* **17**, 2691-2705.

2386 TRAWICKI, M. B., JOHNSON, M. & OSIEJUK, T. (2005). Automatic song-type classification and
2387 speaker identification of Norwegian Ortolan Bunting (*Emberiza hortulana*) vocalizations. *IEEE*
2388 *Workshop on Machine Learning for Signal Processing* , 277-282.

2389 TU, H. & DOOLING, R. J. (2012). Perception of warble song in budgerigars (*Melopsittacus*
2390 *undulatus*): evidence for special processing. *Animal Cognition* **15**, 1151-1159.

2391 TU, H., SMITH, E. W. & DOOLING, R. J. (2011). Acoustic and perceptual categories of vocal
2392 elements in the warble song of budgerigars (*Melopsittacus undulatus*). *Journal of Comparative*
2393 *Psychology* **125**, 420-430.

2394 TYACK, P. (1983). Differential response of humpback whales,*Megaptera novaeangliae*, to
2395 playback of song or social sounds. *Behavioral Ecology and Sociobiology* **13**, 49-55.

2396 VAN HEIJNINGEN, C. A., DE VISSER, J., ZUIDEMA, W. & TEN CATE, C. (2009). Simple rules can
2397 explain discrimination of putative recursive syntactic structures by a songbird species.
2398 *Proceedings of the National Academy of Sciences* **106**, 20538-20543.

2399 VAN SCHAIK, C. P., DAMERIUS, L. & ISLER, K. (2013). Wild orangutan males plan and
2400 communicate their travel direction one day in advance. *PloS One* **8**, e74896.

2401 VEHRENCAMP, S. L., HALL, M. L., BOHMAN, E. R., DEPEINE, C. D. & DALZIELL, A. H. (2007).
2402 Song matching, overlapping, and switching in the banded wren: the sender's perspective.
2403 *Behavioral Ecology* **18**, 849-859.

2404 WALLER, S. (2012). Science of the Monkey Mind: Primate Penchants and Human Pursuits. In
2405 *Experiencing Animal Minds: An Anthology of Animal-human Encounters* (eds. J. A. Smith and
2406 R. W. Mitchell).

2407 WATKINS, W. A. & SCHEVILL, W. E. (1977). Sperm whale codas. *The Journal of the Acoustical*
2408 *Society of America* **62**, 1485-1490.

2409 WATTS, D. J. & STROGATZ, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*
2410 **393**, 440-442.

2411 WEARY, D. M. & LEMON, R. E. (1988). Evidence against the continuity-versatility relationship in
2412 bird song. *Animal Behaviour* **36**, 1379-1383.

2413 WEARY, D. M. & LEMON, R. E. (1990). Kroodsma refuted. *Animal Behaviour* **39**, 996-998.

2414 WEARY, D. M., LAMBRECHTS, M. & KREBS, J. (1991). Does singing exhaust male great tits?
2415 *Animal Behaviour* **41**, 540-542.

2416   WEARY, D., KREBS, J., EDDYSHAW, R., MCGREGOR, P. & HORN, A. (1988). Decline in song
2417   output by great tits: Exhaustion or motivation? *Animal Behaviour* **36**, 1242-1244.

2418   WEISS, D. J. & HAUSER, M. D. (2002). Perception of harmonics in the combination long call of
2419   cottontop tamarins, *Saguinus oedipus*. *Animal Behaviour* **64**, 415-426.

2420   WEISS, M., HULTSCH, H., ADAM, I., SCHARFF, C. & KIPPER, S. (2014). The use of network
2421   analysis to study complex animal communication systems: a study on nightingale song.
2422   *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140460.

2423   WEISS, R. J. & BELLO, J. P. (2011). Unsupervised discovery of temporal structure in music.
2424   *Selected Topics in Signal Processing, IEEE Journal of* **5**, 1240-1251.

2425   WERNICKE, S. & RASCHE, F. (2006). FANMOD: a tool for fast network motif detection.
2426   *Bioinformatics* **22**, 1152-1153.

2427   WHEELER, B. C. & HAMMERSCHMIDT, K. (2012). Proximate factors underpinning receiver
2428   responses to deceptive false alarm calls in wild tufted capuchin monkeys: Is it counterdeception?
2429   *American Journal of Primatology* **75**, 715-725.

2430   WHEELER, B. C. (2010*a*). Decrease in alarm call response among tufted capuchins in competitive
2431   feeding contexts: possible evidence for counterdeception. *International Journal of Primatology*
2432   **31**, 665-675.

2433   WHEELER, B. C. (2010*b*). Production and perception of situationally variable alarm calls in wild
2434   tufted capuchin monkeys (*Cebus apella nigritus*). *Behavioral Ecology and Sociobiology* **64**, 989-
2435   1000.

2436   WIENS, J. A. (1982). Song pattern variation in the sage sparrow (*Amphispiza belli*): dialects or
2437   epiphenomena? *The Auk* **99**, 208-229.

2438   WILEY, R. H. (1983). The evolution of communication: information and manipulation. *Animal*
2439   *Behaviour* **2**, 156-189.

2440   WOLPERT, D. H. & MACREADY, W. G. (1997). No free lunch theorems for optimization.
2441   *Evolutionary Computation, IEEE Transactions on* **1**, 67-82.

2442   WYTTENBACH, R. A., MAY, M. L. & HOY, R. R. (1996). Categorical perception of sound
2443   frequency by crickets. *Science* **273**, 1542-1544.

2444   YOUNG, S. J. & YOUNG, S. (1994). The htk hidden Markov model toolkit: Design and philosophy.
2445   *Entropic Cambridge Research Laboratory, Ltd* **2**, 2-44.

2446   ZHONG, S. & GHOSH, J. (2005). Generative model-based document clustering: a comparative
2447   study. *Knowledge and Information Systems* **8**, 374-384.

2448 ZIPF, G. K. (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley press,
2449 Oxford.

2450 ZUCCHINI, W. & MACDONALD, I. L. (2009). *Hidden Markov models for time series: an*
2451 *introduction using R.* CRC Press.
2452
2453
2454

2455   FIGURE LEGENDS

2456

2457   **Fig. 1.** Flowchart showing a typical analysis of animal acoustic sequences. In this review, we

2458   discuss identifying units, characterising sequences, and identifying meaning.

2459

2460   **Fig. 2.** Examples of the different criteria for dividing a spectrogram into units. (A) Separating

2461   units by silent gaps is probably the most commonly used criterion. (B) An acoustic signal may

2462   change its properties at a certain time, without the presence of a silent "gap", for instance

2463   becoming harmonic or noisy. (C) A series of similar sounds may be grouped together as a single

2464   unit, regardless of silent gaps between them; a chirp sequence is labelled as C. (D) A complex

2465   hierarchical structure to the sequence, combining sounds that might otherwise be considered

2466   fundamental units.

2467

2468   **Fig. 3.** Example of cepstral processing of a grey wolf *Canis lupis* howl (below 6 kHz) and

2469   crickets chirping (above 6.5 kHz). Recording was sampled at $F_s = 16$ kHz, 8 bit quantization. (A)

2470   Standard spectrogram analysed with a 15 ms Blackman-Harris window. (B) Plot of transform to

2471   cepstral domain. Lower quefrencies are related to vocal tract information. F_0 can be determined

2472   from the "cepstral bump" apparent between quefrencies 25–45 and can be derived by

2473 F_s/quefrency. (C) Cepstrum (inset) of the frame indicated by an arrow in A(2.5 s) along with

2474   reconstructions of the spectrum created from truncated cepstral sequences. Fidelity improves as

2475   the number of cepstra are increased.

2476

2477   **Fig. 4.** Perceptual constraints for the definition of sequence units. (A) Perceptual binding, where

2478   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.

Fig. 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 ($\sqrt{}$ indicates suitable, $\times$ indicates unsuitable, $\sim$

indicates neutral).

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

Fig. 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. HMM, hidden Markov model.

Fig. 8. State transition diagram equivalent to a 2nd order Markov model and trigram model (N=3)

for a sequence containing As and Bs.

99

2502

Fig. 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 $P_e(A|X)= P_e(B|X)=0.5$, and when in state *Y*, acoustic unit *A* is much more likely

$P_e(A|Y)=0.9$ than B $P_e(B|Y)=0.1$. Transitioning from state *X* to state *Y* occurs with probability

$P_t(X{\rightarrow}Y)=0.6$, whereas from state *Y* to state *X* with probability $P_t(Y{\rightarrow}X)=0.3$.

2508

Fig. 10. Simple networks constructed from the sequence of acoustic units *A*, *B* and *C*. 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.

2514

Fig. 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 *et al*., 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).

2523

2524 **Fig. 12.** The classes of formal grammars known as the Chomsky hierarchy (Chomsky, 1957).

2525 Each class is a generalisation of the class it encloses, and is more complex than the enclosed

2526 classes. Image publicly available under the Creative Commons Attribution-Share Alike 3.0

2527 Unported license

2528 (https://commons.wikimedia.org/wiki/File:Wiki_inf_chomskeho_hierarchia.jpg).

2529

2530
2531

1

2 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 (in bold)
Separated by silence	Birds	Swamp sparrow *Melospiza georgiana* note (Marler & Pickert, 1984) Black-capped chickadee *Poecile atricapillus* note (Nowicki & Nelson, 1990) Red-legged partridge *Alectoris rufa* and rock partridge *A. graeca* (Ceugniet & Aubin, 2001)	Zebra finch *Taeniopygia guttata* syllable (Cynx, 1990) Emperor penguin *Aptenodytes forsteri* (Robisson *et al.*, 1993) Canary *Serinus canaria* breaths (Hartley & Suthers, 1989)	Swamp sparrow *Melospiza georgiana* note (Nelson & Marler, 1989) Black-capped chickadee *Poecile atricapillus* notes (Sturdy *et al.*, 2000; Charrier *et al.*, 2005) King penguin *Aptenodytes patagonicus* (Lengagne *et al.*, 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 *et al.*, 1999; Lengagne *et al.*, 2000) Emperor penguin *Aptenodytes forsteri*→ **individual identities** (Aubin *et al.*, 2000)
	Terrestrial mammals	Meerkat *Suricata suricatta* calls (Manser, 2001) Gibbon *Hylobates lar* phrase (Raemaekers *et al.*, 1984) Rock hyrax *Procavia capensis* songs (Kershenbaum *et al.*, 2012) Free-tailed bat *Tadarida brasiliensis* syllable (Bohn *et al.*,	Lesser short-tailed bat *Mystacina tuberculata* pulses (Parsons *et al.*, 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 *et al.*, 2010)

1

		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 *et al.*, 2008)	Humpback whale *Megaptera novaeangliae* song (Adam *et al.*, 2013)	Bottlenose dolphin *Tursiops truncatus* signature whistles (Janik *et al.*, 2006) Subantartic fur seal *Arctocephalus tropicalis* pup attraction call (Charrier *et al.*, 2003) Australian sea lion *Neophoca cinerea* calls (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 *et al.*, 2011) Australian sea lion *Neophoca cinerea* call → **individual identity** (Charrier *et al.*, 2009; Pitcher *et al.*, 2012)
Change in acoustic properties (regardless of silence)	Birds	Red junglefowl *Gallus gallus* elements (Collias, 1987)	Northern cardinal *Cardinalis cardinalis* song (Suthers, 1997) Anna hummingbird *Calypte anna* mechanical chirps (Clark & Feo, 2008)	Anna hummingbird *Calypte anna* mechanical chirps (Clark & Feo, 2010) Male chickens *Gallus gallus* alarm calls (Evans *et al.*, 1993)	Blackcap *Sylvia atricapilla* song → **species identity** (Mathevon & Aubin, 2001) White-browed warbler *Basileuterus leucoblepharus* song → **species identity** (Mathevon *et al.*, 2008) Yelkouan Shearwaters *Puffinus*

2

				yelkouan call → **sex and mate identity** (Cure *et al.*, 2011) Grasshopper sparrow *Ammodramus savannarum* buzz/warble → **territorial/social** (Lohr *et al.*, 2013) Rufous-sided towhee *Pipilo erythrophthalmus* song → **species identity** (Richards, 1981)
	Terrestrial mammals	Black-fronted titi monkey *Callicebus nigrifrons* alarm calls (Cäsar *et al*., 2012b) Western gorilla *Gorilla gorilla* calls (Salmi *et al.*, 2013) Red titi monkey *Callicebus cupreus* calls (Robinson, 1979)	Banded mongoose *Mungos mungo* (Jansen, Cant & Manser, 2012)	Mustached bat *Pteronotus parnellii* composites (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 *et al.*, 2013) Tufted capuchin monkeys *Sapajus nigritus* calls→ **predator type** (Wheeler, 2010b) Banded mongoose *Mungos mungo* close calls → **individual identity, group cohesion** (Jansen *et al.*, 2012) Spotted hyena *Crocuta crocuta* call → **sex/age/individual identities** (Mathevon *et al*., 2010)

3

	Marine mammals	Bottlenose dolphin *Tursiops truncatus* whistle loops (Caldwell *et al.*, 1990) Killer whale *Orcinus orca*, subunit of calls (Shapiro *et al.*, 2010) Humpback whale *Megaptera novaeangliae* subunit (Payne & McVay, 1971) Leopard seal *Hydrurga leptonyx* calls (Klinck *et al.*, 2008)	False killer whale *Pseudorca crassidens* vocalisations (Murray *et al.*, 1998) Bottlenose dolphin *Tursiops truncatus* tonal calls (Parsons *et al.*, 2010)	Bearded seal *Erignatus barbatus* trills (Charrier *et al.*, 2013)	Killer whales *Orcinus orca* calls → **sex/orientation** (Miller *et al.*, 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 nebouxii* call (Dentressangle *et al.*, 2012)	Emberizid sparrow trills (Podos, 1997)	Zebra finch *Taeniopygia guttata* syllables (Cynx *et al.*, 1990) Little owl *Athene noctua* syllables (Parejo *et al.*, 2012) Song sparrow *Melospiza melodia* songs (Horning *et al.*, 1993)	Carolina chickadee *Poecile carolinensis* D-notes → **food availability** (Mahurin & Freeberg, 2009) Kittiwake *Rissa tridactyla* call → **sex/individual identities** (Aubin *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* alarm calls (Cäsar *et al.*, 2012b,	Diana monkey *Cercopithecus diana* alarm calls (Riede *et al.*, 2005)	Black-fronted titi monkey *Callicebus nigrifrons* (Cäsar *et al.*, 2012a)	Chimpanzee *Pan troglodytes* pant hoots → **foraging** (Notman & Rendall, 2005)

4

		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)	Domestic dog *Canis familiaris* growls (Riede & Fitch, 1999)	Colobus *Colobus guereza* sequences (Schel *et al.*, 2010) Tufted capuchin monkey *Sapajus nigritus* bouts (Wheeler, 2010*b*)	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 *et al.*, 1983; Mercado *et al.*, 2010; Mercado & Handel, 2012) Bottlenose dolphin *Tursiops truncatus* whistles (Janik *et al.*, 2013)	Humpback whale *Megaptera novaeangliae* songs (Handel *et al.*, 2009) Bottlenose dolphin *Tursiops truncatus* whistles (Pack *et al.*, 2002) Weddell seal *Leptonychotes weddelli* vocalisations (Thomas *et al.*, 1983) Harbour seal *Phoca*	Bottlenose dolphin *Tursiops truncatus* signature whistles → **individual identity, group cohesion** (Quick & Janik, 2012) Humpback whale *Megaptera novaeangliae* phrases → **unknown** (Payne & McVay, 1971)

5

				vitulina roars (Hayes *et al.*, 2004)	
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 *et al.*, 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 *et al.*, 2013) White-browed warbler *Basileuterus leucoblepharus* song → **individual identity** (Mathevon *et al.*, 2008)
	Terrestrial mammals	Red titi monkey *Callicebus cupreus* syllable (Robinson, 1979) Free-tailed bat *Tadarida brasiliensis* songs (Bohn *et al.*,	Rhesus-macaque *Macaca mulatta* vocalisations (Fitch, 1997)	Putty-nosed monkey *Cercopithecus nictitans* sequences (Arnold & Zuberbühler, 2006*a*) Red titi monkey *Callicebus cupreus*	Chimpanzee *Pan troglodytes* phrases → **group identity** (Arcadi, 1996) Putty-nosed monkey *Cercopithecus nictitans* sequences → **predators presence, group movement** (Arnold &

		2008)		syllable (Robinson, 1979)	Zuberbühler, 2006*a*) Tufted capuchin monkeys *Sapajus nigritus* calls→ **predator type** (Wheeler, 2010*b*) 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 *et al.*, 2012)	Humpback whale *Megaptera novaeangliae* song → **mating display - female attraction/male-male interactions** (Darling *et al.*, 2006; Smith *et al.*, 2008)

7

Table 2. A summary of the assumptions and requirements for each of the five different structure analysis models suggested in the review.

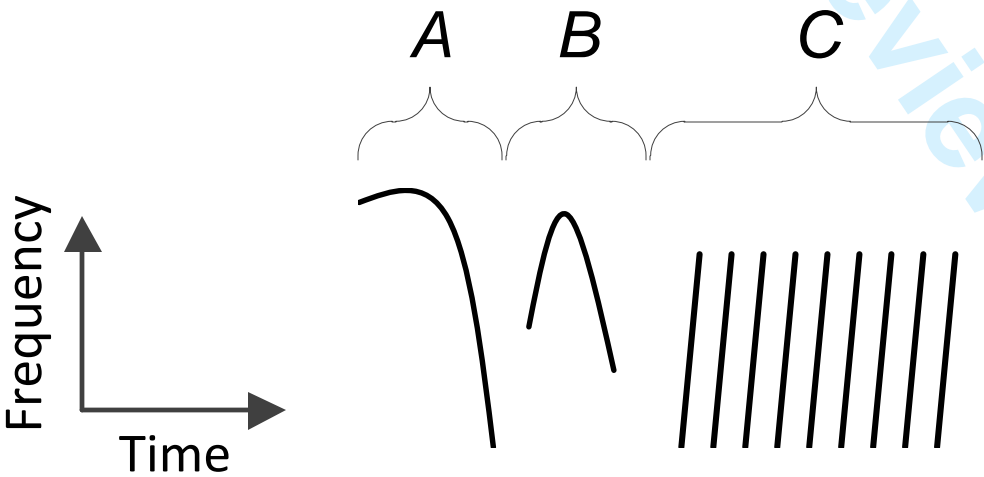| Model type	Embedding type	Data requirements	Typical hypotheses	Assumptions
Markov chain	• Repetition • Diversity • Ordering	• Number of observations required increases greatly as the size of the model grows	• Independence of sequence • Sequential structure	• Stationary transition matrix • Sufficient data for maximum likelihood estimator 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	• The properties of relations between units 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/effect	• 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

Repetition

Diversity

Composition

Repertoire

Ordering

Overlapping

Temporal

Testable hypotheses and behavioural experiments

(A) Separated by silence

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

(C) Series of sounds

(D) Higher levels of organisation

(A) Spectrogram

(B) Cepstrogram
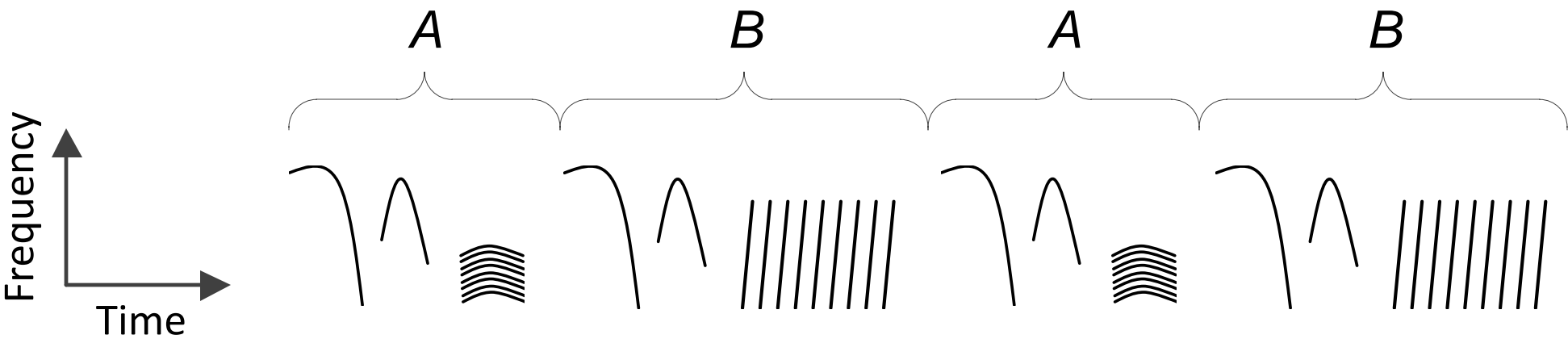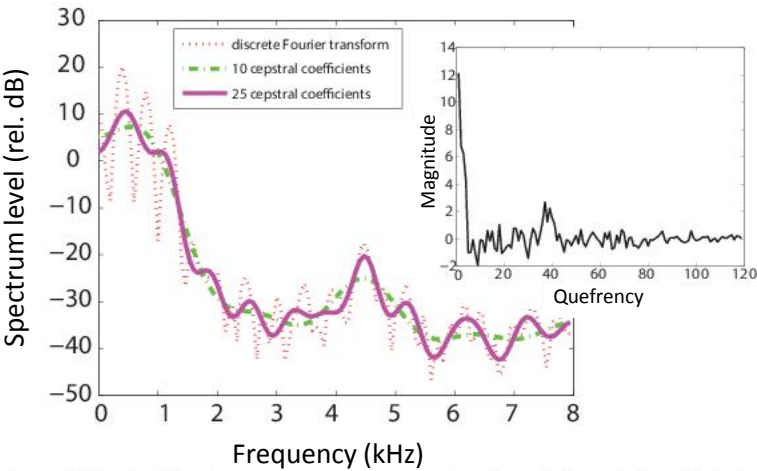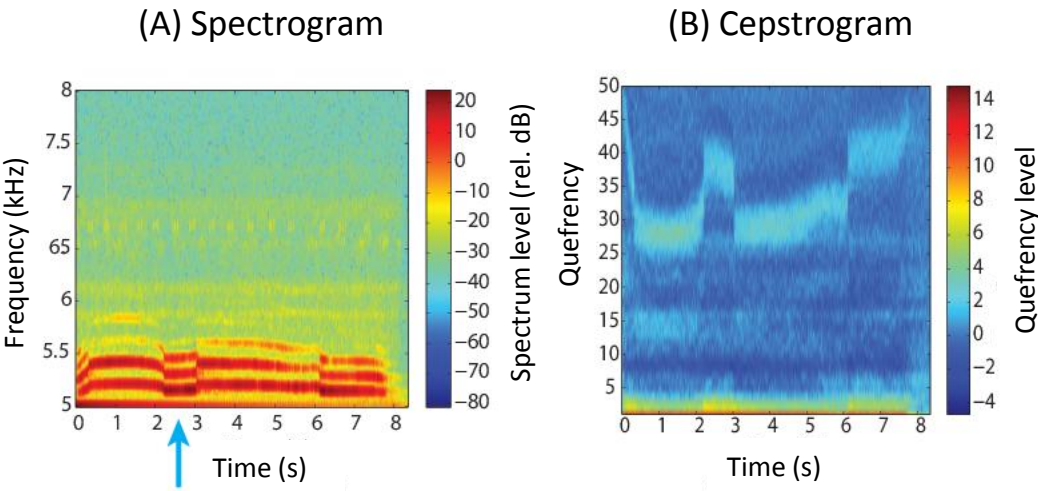
(C) Cepstrum at 2.5 s (inset) and spectral reconstructions from truncated cepstral coefficients

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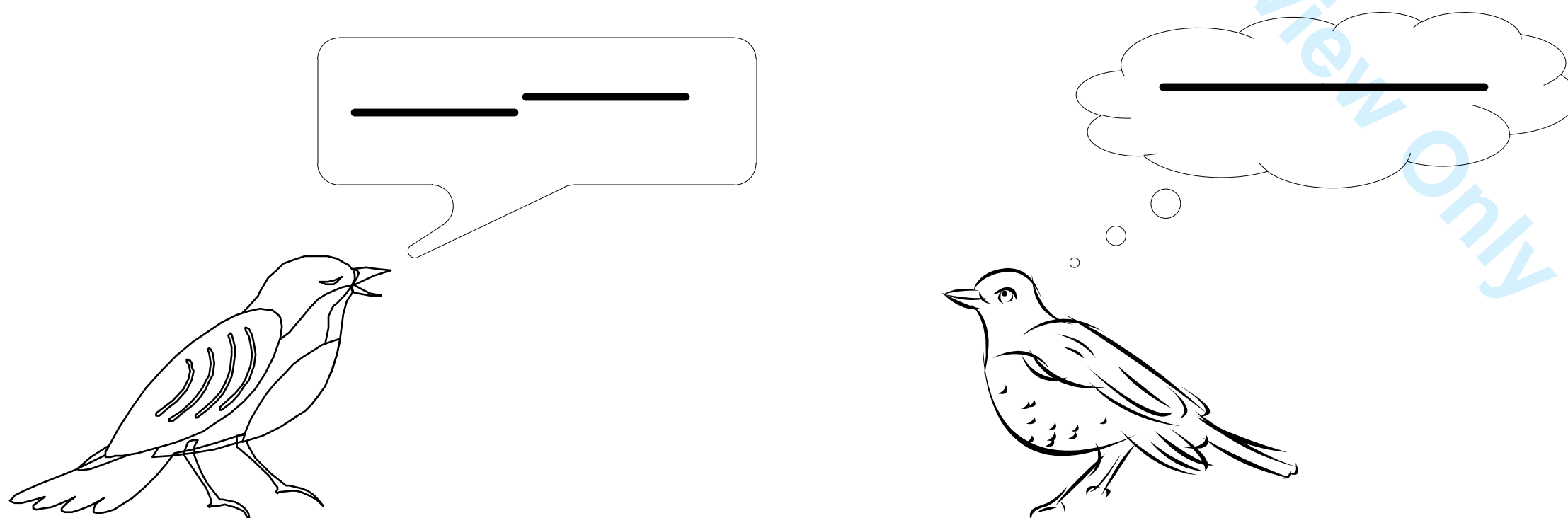
(C) Spectrotemporal constraints. If the receiver cannot distinguish small differences in time or frequency, discrete elements may be interpreted as joined

(A) Repetition

(B) Diversity

(C) Combination
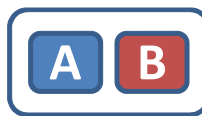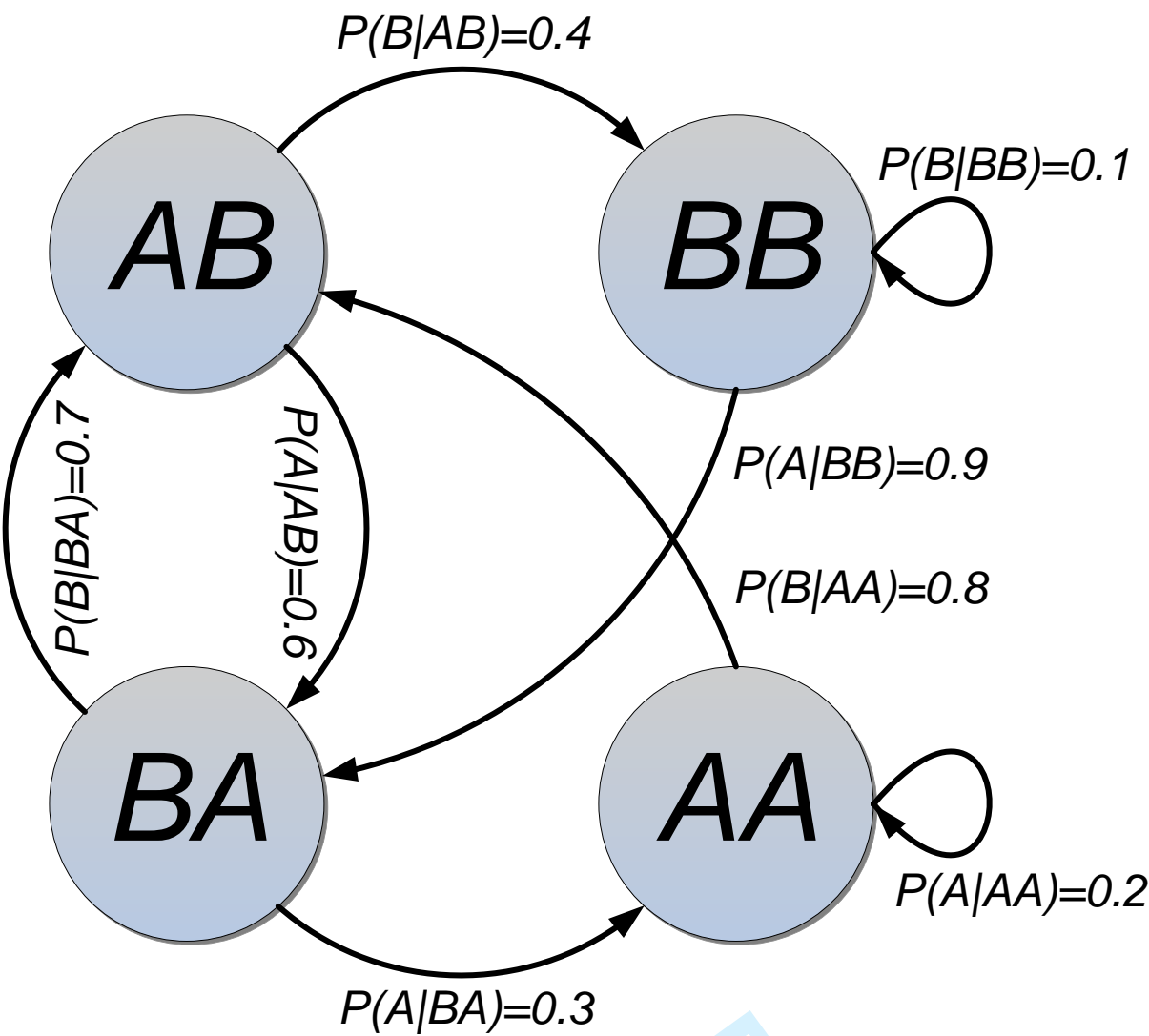
(D) Ordering

(E) Overlapping
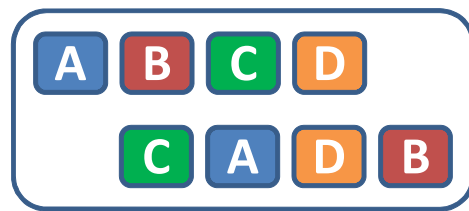
(F) Timing

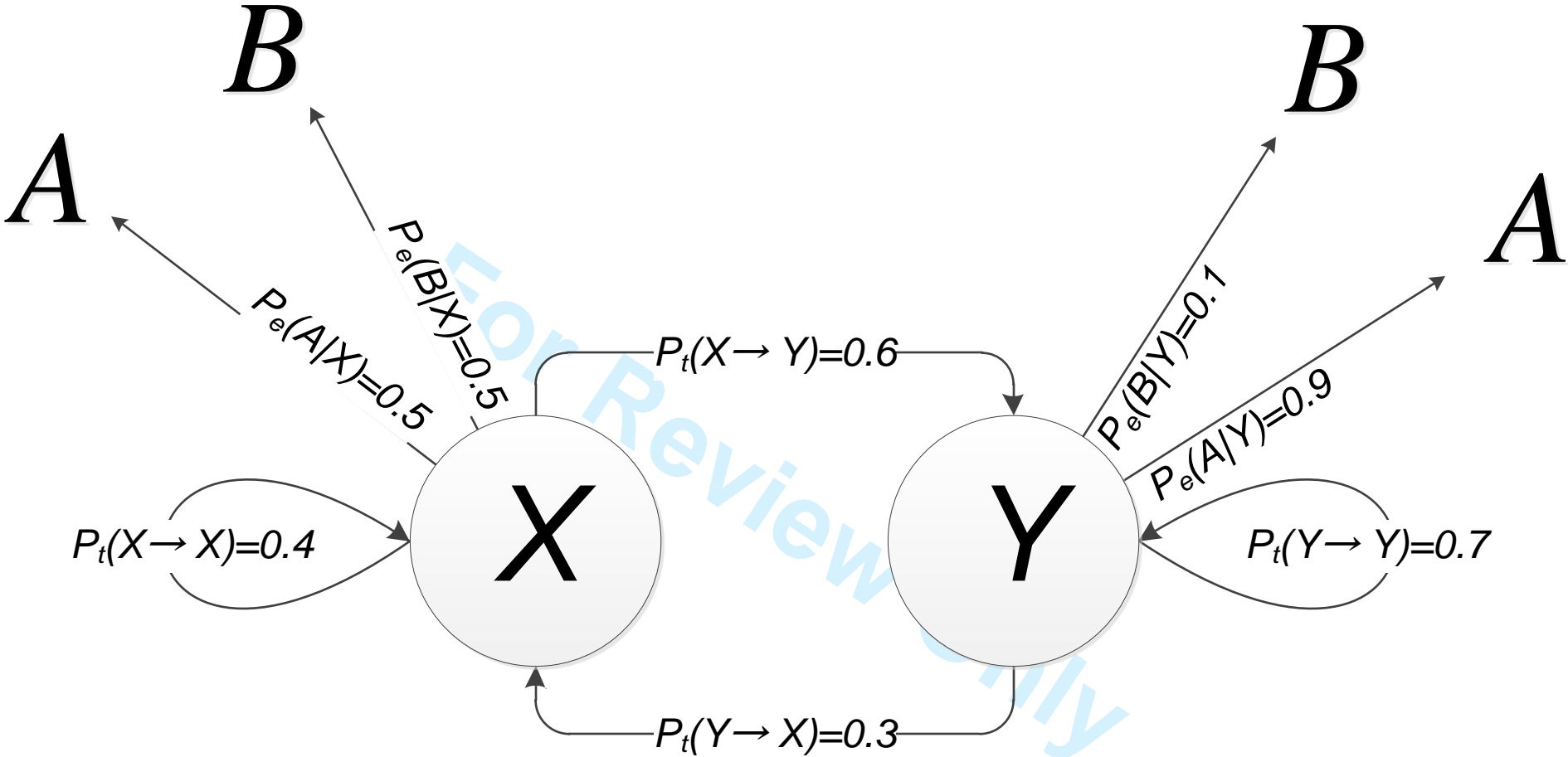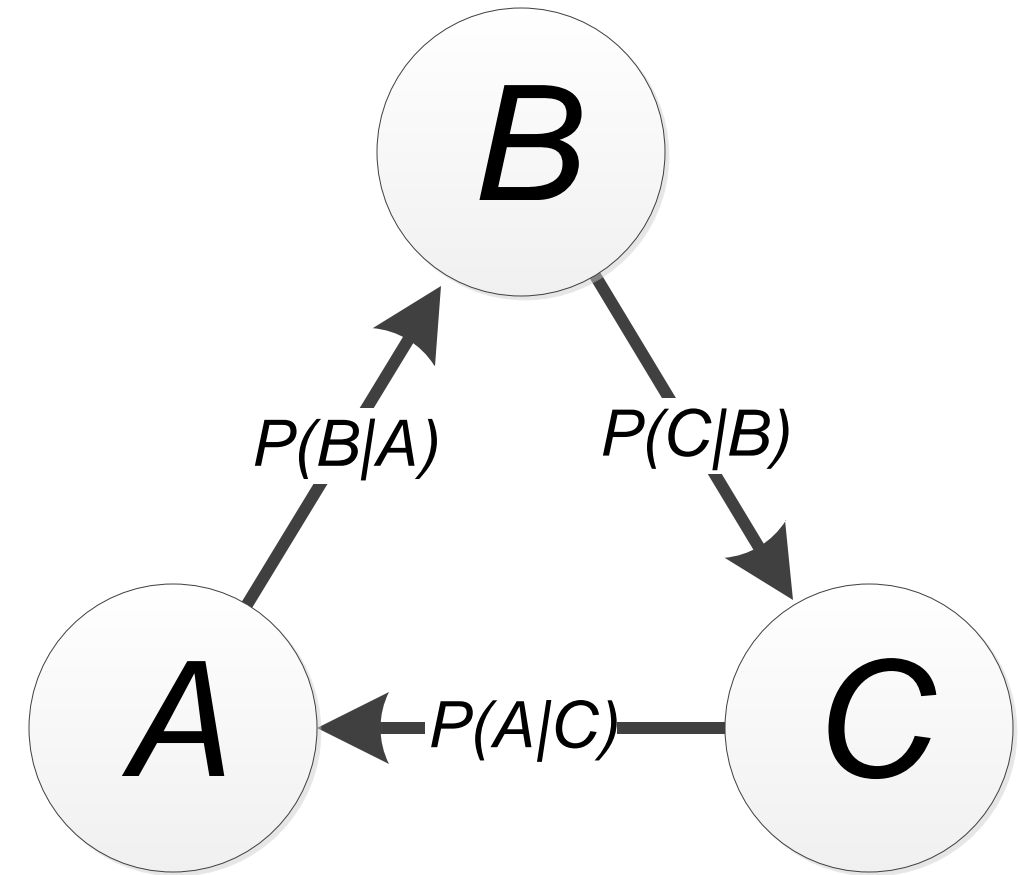| Type	Criterion	Example
Repetition	Single unit repeated more than once	Chickadee D-note mobbing call (Baker & Becker, 2002)
Diversity	A number of distinct units are present. Order is unimportant.	Birdsong repertoire size (Searcy, 1992)
Combination	Set of units has different information from each unit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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
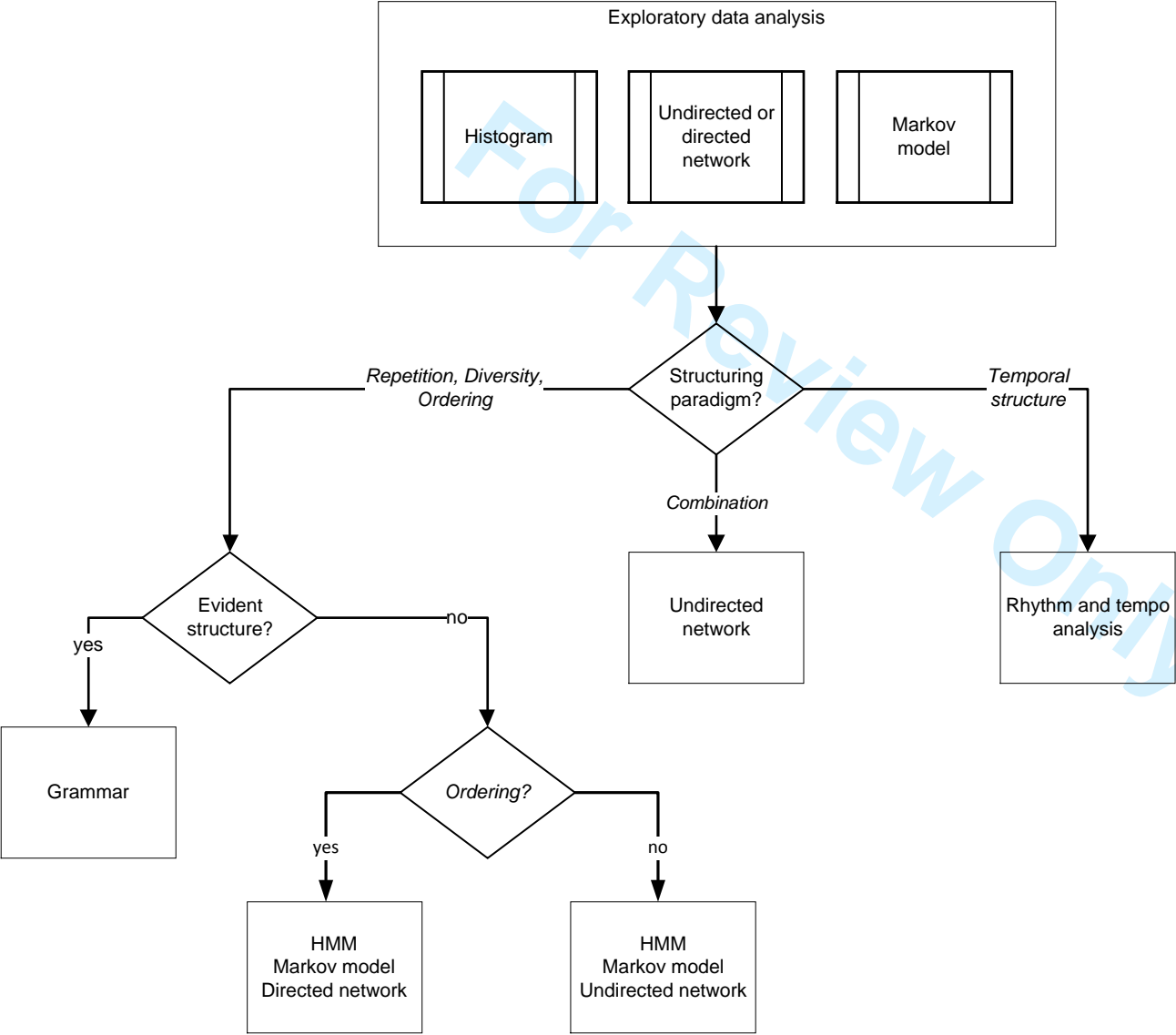
	A	*B*
P(X|AA)	0.2	0.8
P(X|AB)	0.6	0.4
P(X|BA)	0.3	0.7
P(X|BB)	0.9	0.1

P(B|AB)=0.4

P(B|BB)=0.1

P(B|BA)=0.7

P(A|AB)=0.6

P(A|BB)=0.9

P(B|AA)=0.8

P(A|AA)=0.2

P(A|BA)=0.3

$S \rightarrow b \mid \text{Song} \mid d$

$\text{Song} \rightarrow a \text{ Bseq Song} \mid \varepsilon$

$\text{Bseq} \rightarrow b \text{ Bs}$

$\text{Bs} \rightarrow b \text{ Bs} \mid \varepsilon$

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

$\Rightarrow a\,b\,b\,a\,\underline{\text{Bseq}}\text{ Song} \Rightarrow a\,b\,b\,a\,b\,\underline{\text{Bs}}\text{ Song} \Rightarrow a\,b\,b\,a\,b\,\underline{\text{Song}} \Rightarrow a\,b\,b\,a\,b$
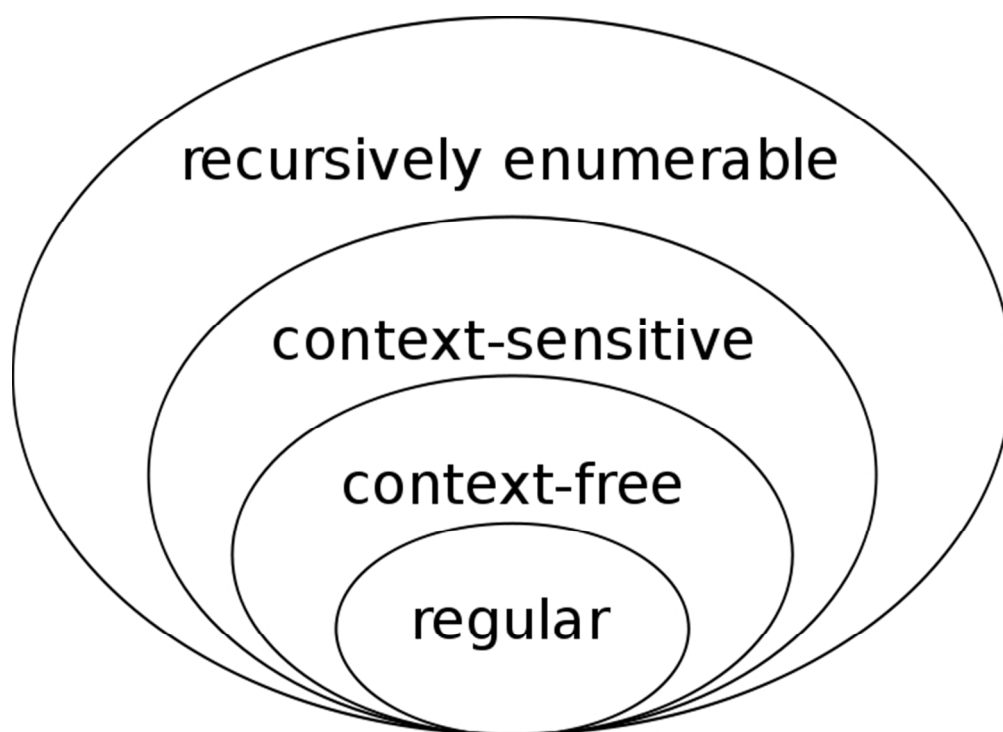
Figure 12. The classes of formal grammars known as the Chomsky hierarchy (Chomsky. 1957). Each class is a generalisation of the class it encloses, and is more complex than the enclosed classes. Image publicly available under the Creative Commons Attribution-Share Alike 3.0 Unported license.
https://commons.wikimedia.org/wiki/File:Wiki_inf_chomskeho_hierarchia.jpg
251x181mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```
┌─────────────────┐
│                 │
│ Data collection │
│                 │
└─────────────────┘
        │
        ▼
┌─────────────────┐
│                 │
│ Identifying     │
│ units           │
│                 │
└─────────────────┘
        │
        ▼
┌─────────────────┐
│ Characterising  │
│ sequence        │
└─────────────────┘
        │
        ▼
┌─────────────────┐
│                 │
│ Identifying     │
│ meaning         │
│                 │
└─────────────────┘
```

Data collection → Raw audio → Preprocessing

Data collection → Behavioural (context) data

Preprocessing → Filtering | Time-frequency analysis (e.g. spectrogram) | Time-series analysis

Division into units

Change in acoustic properties | Separated by silence | Series of sounds | Higher level of organisation

Select relevant units

Characterise sequence ← Extract features

Identify meaning

Repetition | Diversity | Composition

Ordering | Overlapping | Temporal

Clustering

Repertoire

Testable hypotheses and behavioural experiments

For Review Only