



Segmental Units in Nonhuman Animal Vocalization as a Window into Meaning, Structure, and the Evolution of Language

Dan C. Mann^{1*} and Marisa Hoeschele^{1,2}

¹Department of Cognitive Biology, University of Vienna

²Acoustics Research Institute, Austrian Academy of Sciences

*Corresponding author (Email: daniel.mann@univie.ac.at)

Citation – Mann, D.C., & Hoeschele, M. (2020). Segmental units in nonhuman animal vocalization as a window into meaning, structure, and the evolution of language. *Animal Behavior and Cognition*, 7(2), 151-158. doi: <https://doi.org/10.26451/abc.07.02.09.2020>

Abstract – Human vocalizations are made up of meaningless units or segments that are combined to create meaningful words and phrases. Jackendoff (1999) hypothesized that the ability of humans to combine segments together is necessitated by the fact that we need to express an almost limitless amount of symbolic or referential information that could occur in a different time or space. So far, there is very little evidence for this symbolic and referentiality meaning in animal vocalizations. Furthermore, segments have also rarely been identified in the animal kingdom, with units divided by intakes of breaths taken as the most fundamental. However, if we are to take Jackendoff’s hypothesis seriously, we must do more detailed analyses at the level of the segment (subunits within a single breath) in animal vocalizations. Here we discuss the current status of animal vocal communication and its relation to Jackendoff’s hypothesis. We propose further research into segmental units in animal vocalizations is a key next step to determining the evolution of human vocal behavior.

Keywords – Communication, Segments, Vocalizations, Meaning, Jackendoff’s hypothesis

The comparative method, comparing traits across various taxa, has been an essential tool in understanding complex behavior, particularly for systems that, like language or birdsong, do not leave fossil evidence. Comparisons of human linguistic capabilities to those of nonhuman animals have revealed that many animals possess components of language—even if no animal has full human-like language. Comparative research with nonhuman animals has made it clear that human language is built from a broad network of connected sub-processes and behaviors (Fitch, 2018). Conversely, the tools, methodologies, and models developed in human language research have greatly aided our ability to describe, measure, and quantify animal behavior (Doupe & Kuhl, 1999; Mol et al., 2017; Seyfarth et al., 1980). For instance, humans can distinguish sounds by modifying their vocal tract to affect resonance frequencies, known as “formants.” Formants are also present in the acoustic signal of many other species, the discovery of which led to a more complete understanding of how individuals can transmit information about their fitness (Reby et al., 2005).

One of the most famous examples of the relevance of the comparative method to acoustic communication is that of Robert Seyfarth, Dorothy Cheney, and Peter Marler’s (1980) work with vervet monkeys. The authors presented convincing evidence that nonhuman primate alarm calls could not be simply explained by arousal level or the internal state of an individual. The data showed that vervet

monkey alarm calls varied depending on the type of predator in the area; that is, vervet monkeys produce calls that are linked to concrete, external entities. Furthermore, the behavior of the receivers corresponded to the type of call given: vervet monkeys climbed into trees upon hearing the leopard alarm, looked down when hearing the snake alarm, and looked up when hearing the eagle alarm. Subsequent research by the authors found that primate calls likely differ in important ways from human words or expressions, namely that nonhuman primate calls are not used to change the mental state of other individuals (Cheney & Seyfarth, 1990). Nevertheless, the initial findings represented a crucial step in understanding the similarities and differences between human and nonhuman cognition and communication.

Subsequent comparative work from a wide range of species has revealed that many traits that were once believed to be unique to humans and unique to language (e.g., categorical perception, fast-mapping, descended larynx, etc.) could be found in nonhumans and, in some cases, were quite widespread (examples in Fitch, 2010). Importantly, many of these traits have been found in species that are not closely related to humans. For instance, vocal learning, or the acquisition of vocal signals through imitation, is an essential component of spoken human language but is also present in parrots, songbirds, hummingbirds, elephants, bats, pinnipeds, and cetaceans (Janik & Slater, 1997; Searcy & Nowicki, 2019). There is, however, little to no evidence for vocal learning in any nonhuman primate. Traits that are shared with closely related species—for humans, great apes—are interesting because they provide insights into ancestral forms, but traits that evolved independently in distinct lineages, like vocal learning, are equally interesting. Because these “analogous” traits are independent data points, we can use them to test hypotheses related to evolutionary function, adaptiveness, evolutionary constraints, and underlying mechanisms (Lattenkamp & Vernes, 2018; Nowicki & Searcy, 2014). For any analogous trait, there is the possibility that similarities are coincidental and have no deeper explanation, but detailed analyses of as many independent data points as possible allow us to discern whether this is the case.

How can we discover meaningful hypotheses about human language using a comparative method? Part of the answer, we argue, comes from studying humans from the perspective of a comparative researcher. Studying human language from a similar outside perspective as we study animal vocalizations can lead to insights that lead to testable hypotheses in other species. Here we discuss the example of Jackendoff’s hypothesis about the evolution of segments in human speech, which we describe from a comparative perspective in the next section. We then turn to the animal literature and a discussion of how we can use animal research to address Jackendoff’s ideas.

Studying Human Language from a Comparative Perspective – Jackendoff’s Hypothesis

A favorite rhetorical device of linguists is to invoke a hypothetical alien researcher of human language; for instance, Noam Chomsky stated that a “Martian scientist might reasonably conclude that there is a single human language, with differences only at the margins” (Chomsky, 2000, p. 7). On the one hand, Chomsky may be underselling linguistic diversity across human populations (Evans & Levinson, 2009). On the other hand, that there are at least universal tendencies across human languages is undeniable, particularly in the domain of sound systems (Blevins, 2004; Maddieson, 1984). In fact, the level of diversity found across human languages are constrained in part based on the physiological constraints of the human vocal apparatus. Although this point is obvious, it is important to recognize that human vocalizations all sound very similar in that they sound all human. To an alien researcher, and also potentially to other species, the variations across language types may be overshadowed by these species-specific features. Although other species can detect differences among human languages (e.g., elephants, *Loxodonta africana*, McComb et al., 2014; rats, *Rattus norvegicus*, Toro et al., 2003; cotton top tamarins, *Saguinus oedipus*, Ramus et al., 2000), most accounts of this ability are from experimental settings where the animals are trained to identify different languages or language types (e.g., Ramus et al., 2000; Toro et al., 2003). Rarely have we observed that animals naturally identify variations among human languages (but see McComb et al., 2014). At the same time, if we consider the vocalizations of other animals, we may well be guilty of the same problem: we focus on species-specific aspects of their sound, and not on

diversity. For example, guidebooks for bird songs have phonetic descriptions of the sound a particular bird makes, so that the species can be identified.

What parts of human speech are shared across languages? The study of linguistic sounds and their organization (phonology, phonetics, and phonotactics) is the domain of linguistics in which our knowledge of universal patterns, and their underlying mechanisms, is most advanced. For instance, extensive cross-linguistic typological research has revealed that, in spite of the great diversity of languages, all spoken languages have two broad classes of segments: plosives, transient bursts of energy (e.g., *p*, *d*, *k*), and vowels, periodic signals with clear harmonic structure that are typically made with little to no vocal tract obstruction (e.g., *i*, *u*, *a*; Hyman, 2008; Ladefoged & Maddieson, 1996; Lindblom & Maddieson, 1988; Maddieson, 1984). Furthermore, organization of these segment classes is asymmetrical, with plosive-vowel patterns, like *ka*, being near universal while the reverse pattern, *ak*, is much less common (related to the consonant-vowel preference and the Margin Hierarchy; Blevins, 1995; Breen & Pensalfini, 1999; Clements, 1990; Lowenstamm, 1996; Prince & Smolensky, 2002). Research at the level of the segment has informed questions related to innate grammatical constraints (Chomsky, 1965; Prince & Smolensky, 2002); the importance of exceptions to putative universals (Breen & Pensalfini, 1999); the role of misperception in shaping common sound patterns (Blevins, 2004; Ohala & Kawasaki, 1984); the timing, production, and coordination of articulatory gestures (Nam et al., 2009); and whether segments are distinct early on or emerge during development (Davis & MacNeilage, 1995; Giulivi et al., 2011; MacNeilage, 1998).

Crucially, if alien researchers were to treat spoken human language the same way that we animal communication researchers analyze animal vocalizations, they would likely miss many of the generalizations that have been so thoroughly researched in phonology. In acoustic animal communication, the most basic unit is typically a vocalization separated by periods of silence, a “breath group” or “syllable” (Kershenbaum et al., 2014). Yet, in spoken human language segments, words, phrases, and even sentences are often uttered as a continuous stream with minimal to no silent intervals. While some vocalizations like “how’s it going?” or “have a good night” are stereotyped and reoccurring, many acoustic strings are novel and may never be repeated. Only by investigating the units within a breath group do the building blocks of language become apparent.

These units, or “segments”, are typically meaningless and are divided by rapid transitions in the acoustic stream. For any language, the inventory of segments is limited, but even a small set of segments can be combined and rearranged to create an unbounded set of syllables and words. Some researchers, such as Jackendoff (1999), have suggested that human segment combinatorial abilities may have evolved to accommodate an expansion of referential and symbolic labels. These labels are defined as real world objects or actions, could be used outside of the presence of the referents, and were used with the intention to inform other individuals. Holistic phrases can support only a limited set of symbols before the acoustic signals begin to lose their perceptual distinctiveness. Combining and rearranging a small set of discrete units, “combinatorial phonology,” quickly permits the creation of an unlimited set of symbols (Jackendoff, 1999; Studdert-Kennedy, 1998).

Testing Jackendoff’s Hypothesis Using the Comparative Method

The use of segments as a productive building block of vocal behavior has the potential to be uniquely human. Considering Jackendoff’s hypothesis, there is a clear rationale for this potential uniqueness: rapid modulations are one strategy to cope with an increase in information transmission without overtaxing short-term memory. Combinatorial phonology permits an infinite set of labels, but human language is not just a set of labels. Humans combine labels to convey even more complex meaning, like expressing the relationship between concepts. Increasing the amount of information transmitted in the vocal signal will increase the duration of the signal. However, the signal cannot be infinitely long; at a certain point, memory limitations will affect the signal’s intelligibility. Pauses between each word, and especially between each segment, would greatly increase the length of the utterance. Rapid concatenation of segments, however, increases the rate of transfer and reduces the strain

on short-term memory (Liberman et al., 1967; Studdert-Kennedy, 1998). If animals do not use their vocalizations to transmit symbolic labels and relationships between labels, they do not face the same pressures to increase the transmission rate of perceptually distinct and discrete units (as opposed to fast trills where the same unit type is repeated).

If human segmental systems did, in fact, evolve due to the pressure of needing an unbounded symbolic inventory, we should expect vocal segments to be rare across the animal kingdom. An assessment of the animal vocal literature would suggest that this is the case. As stated above, in nonhumans, the breath group, often referred to as “note” or “syllable”, is usually treated as the most basic unit. Doupe and Kuhl (1999) refer to these units in songbirds as comparable to human phonetic units. The lack of analyses at sub-syllabic levels, though, may not be the result of rarity in animal vocal systems but rather a difficulty in analyzing units that have less clear divisions. Silence is typically an obvious boundary; transitions in acoustic parameters are less clear and require researchers to make difficult choices about the magnitude and quickness that constitute a potential unit. As such, segments present challenges for rigorous and reproducible research.

Even still, limited evidence for segments in nonhuman vocalizations exists. Banded mongoose close calls are composed of an initial noisy component followed by a harmonic component (Jansen et al., 2013). Jansen and colleagues found that the initial noisy segment was predictive of individual identity while the harmonic segment correlated with behavioral context; the authors state this is the first documentation of a nonhuman syllable containing discrete subunits in which each segment modifies the information of the whole syllable. Similar results have been noted in Campbell’s monkeys (Ouattara et al., 2009) and dingoes (Deaúx et al., 2016), where both species sometimes combine “syllables” together to create a multi-segment syllable with a novel meaning. In birdsong, terminology can vary from species to species, but units below the level of the syllable have been mentioned (Williams, 2004, calls them “elements”). Vocalizations that transition from harmonic to noisy within a breath group or have large frequency jumps have been described in a wide range of species, including species of nonhuman primates (Fitch et al., 2002), canids (Wilden et al., 1998), songbirds (Zollinger et al., 2008), frogs (Ryan & Guerra, 2014), and even fish (Rice et al., 2011). Because of the paucity of segmental analyses, it is unclear if these sub-syllabic variations are potential units or are an unmeaningful byproduct of biomechanical processes (see Berry et al., 1996; Fee et al., 1998; Fitch et al., 2002 for a discussion of non-linear phenomena in animal vocalizations). Furthermore, with the exception of the songbirds, the species for which sub-syllabic units have been referenced do not learn their vocalizations through the mimicry of a tutor or tutors. Even in many of the songbirds, the song and syllable inventories are set early in development.

While documentation of potential segments is rare in animals with more simple vocal systems, data is even more sparse in species whose vocal repertoires are large, complex, socially learned, and “generative;” that is, they can create novel units. However, these species are more likely to use segments to generate large repertoires and are more likely to be relevant for human phonology comparisons. Species, like budgerigars and grey catbirds for instance, have syllable repertoire sizes with no upper bound (Farabaugh et al., 1992; Kroodsmá et al., 1997). In budgerigars, these syllables have been described as non-stereotyped and non-repeating (Farabaugh et al., 1992). While the syllable has been described as the most basic unit in budgerigar warble song, a lack of stereotypy at the level of the syllable suggests a segmental analysis will find more repeating and stereotyped units. In our own recent work, we have shown that these syllables can be broken down into segments that behave in similar ways to human segments (Mann et al., unpublished).

Armed with this new perspective on budgerigar vocalizations, we can begin to examine other species to determine whether segments are more common than we thought in the animal kingdom.

Let us assume for a minute that the hypothesis proposed by Jackendoff (1999), that human combinatorial phonology evolved to accommodate a large symbolic inventory, is true. If we then also assume that animals do not use their vocalizations as symbolic or referential labels at the same level as humans (another assumption based, primarily, on a lack of evidence which we discuss later) – we should expect either: (1) combinatorial systems in animals to be practically non-existent or (2) animal combinatorial systems to be only superficially similar to human combinatorial phonology.

Expectation (1) does not seem to be supported by the data. Not only do combinatorial systems exist, generative systems have been documented for species of parrots, songbirds, and cetaceans (Engesser & Townsend, 2019; Rohrmeier et al., 2015). These species are able to create novel songs and/or syllables by rearranging and recombining more basic units. In songbirds, several phenomena that appear analogous to phenomena in human phonology have been documented, these include: coarticulation (Wohlgemuth et al., 2010), allophonic variation (Lachlan & Nowicki, 2015), phonemic contrast (Engesser et al., 2015), and phrase-final lengthening (Mann et al., unpublished; Tierney et al., 2011). These data suggest that the root of human phonology and the combinatorial systems of these species may be, at least at some level, analogous (i.e., “convergent evolution”). Therefore, they may have evolved to accommodate similar functional needs, suggesting that human phonology may not have evolved to support a sizable vocabulary, but perhaps evolved for non-linguistic functions (e.g., group cohesion, mate attraction, kin recognition, etc.).

However, expectation (2), that the combinatorial abilities found in animals may be largely unrelated to human segment concatenation, is better supported by data. For the most part, combinatorial phenomena in animals have been found at the level of the syllable/breath group, not the segment. The human voice can carry multiple information types, often simultaneously, much of which is unrelated to the linguistic signal (in fact, the linguistic signal need not be transmitted via the auditory channel as evidenced by sign languages). The arrangement of holistic (i.e., breath group) units could have evolved for one purpose while a segmental system later developed for a different purpose. If so, it could be that animal combinatoriality has parallels to human prosody (the melody and intonation of speech, see Filippi et al., 2019 for review of this topic) rather than human segments. As such, the prevalence and form of segmental systems in animal communication has relevance for our understanding of the evolution of combinatoriality, generativity, and vocal learning.

But if we go back to our assumptions, what if the assumption that animals do not convey symbolic or referential labels in their vocalizations as we defined above (or at least much closer than is known now) is false? Then, if the hypothesis proposed by Jackendoff (1999) is true, production of learned segments in animal vocalizations may be indicative of a species’ symbolic/referential abilities. After having segmented animal vocalizations, we can potentially better understand the function or even meaning structure of animal vocalizations. In assuming the syllable is the most basic unit, we could be missing out on vital information. One possibility, albeit a fairly optimistic possibility, is that segment analyses could reveal more complex labeling or even some form of symbolic reference in animals. Currently, no evidence exists to suggest that animals use their vocalizations to convey referential or symbolic labels (Cheney & Seyfarth, 1990; Fitch, 2016; Wheeler & Fischer, 2012), even though they may have the cognitive abilities that underlie symbolic concepts (Fitch, 2019; Seyfarth & Cheney, 2017). However, segmental analyses could reveal more complex structure in the vocal repertoires, which we could link to more complex meaning. This possibility is likely overly optimistic considering even experiments in a highly intelligent, social, vocal learning species like the bottlenose dolphin have failed to find evidence for vocalizations that transmit symbolic, referential meaning (Bastian et al., 1968). That being said, while we have made great advances in understanding animal behavior over the last few decades, most species are understudied and, even in well-studied taxa, our knowledge is often based on a few model species such as the zebra finch (Williams, 2004).

Conclusions

We are at an exciting point in the study of animal communication. Over the last 40 years, there have been great advances in understanding and clarifying the how and why animals use sound to communicate but many questions remained unanswered. We have more unique tools and perspectives to attack questions from multiple angles than ever before. We believe that we can push human-animal comparisons even further by attempting to analyze the level of the segment in animal vocal repertoires. We can further build on the work of Yip (2006) and investigate whether phonological phenomena exist in the signals of nonhuman animals. For example, do animals have classes of sounds (akin to human consonants, vowels,

fricatives, etc.) and rules that operate over those classes (like how English allows a vowel, but not a consonant, in a *h_d* frame: *had* but not *hld*)? And do the rules and classes vary by populations (like how English allows word initial *sp-* clusters but Spanish does not)? And can ambiguous signals affect how the acoustic system changes over time as it does with humans (Blevins, 2004)? Because segments have been so critical in understanding human language, we hope that this domain will open new lines of inquiry and allow us to answer questions that to date remain unsolved. In particular, segments could help us address complicated questions related to meaning in animal—human and nonhuman—vocal communication. The presence of segments in nonhumans does not necessitate referential or symbolic meaning in a nonhuman vocal system, just as Seyfarth et al.’s (1980) discovery that different acoustic signals corresponding to different behaviors in vervet monkeys did not necessitate referential meaning in primate communication. More likely, nonhuman segment systems would cast doubt on theories of human language evolution in which the phonetic/phonological systems evolved in order to accommodate referential meaning, theories such as in Jackendoff (1999).

References

- Bastian, J., Wall, C., & Anderson, C. L. (1968). Further investigation of the transmission of arbitrary environmental information between bottle-nose dolphins. *Naval Undersea Warfare Center, Technical Publication, 109*, 1–39.
- Berry, D. A., Herzel, H., Titze, I. R., & Story, B. H. (1996). Bifurcations in excised larynx experiments. *Journal of Voice, 10*, 129–138.
- Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Blackwell Publishing.
- Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.
- Breen, G., & Pensalfini, R. (1999). Arrernte: A language with no syllable onsets. *Linguistic Inquiry, 30*, 1–25.
- Cheney, D. L., & Seyfarth, R. M. (1990). Attending to behaviour versus attending to knowledge: Examining monkeys’ attribution of mental states. *Animal Behaviour, 40*, 742–753.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.
- Clements, N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 283–333). Cambridge University Press.
- Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*, 1199–1211.
- Deaúx, E. C., Allen, A. P., Clarke, J. A., & Charrier, I. (2016). Concatenation of “alert” and “identity” segments in dingoes’ alarm calls. *Scientific Reports, 6*, 1–9.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*, 567–631.
- Engesser, S., & Townsend, S. W. (2019). Combinatoricity in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*, e1493.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *The Behavioral and Brain Sciences, 32*, 429–494.
- Farabaugh, S. M., Brown, E. D., & Dooling, R. J. (1992). Analysis of warble song of the budgerigar: *Melopsittacus undulatus*. *Bioacoustics, 4*, 111–130.
- Fee, M. S., Shraiman, B., Pesaran, B., & Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature, 395*, 67–71.
- Filippi, P., Hoeschele, M., Spierings, M., & Bowling, D. L. (2019). Temporal modulation in speech, music, and animal vocal communication: Evidence of conserved function. *Annals of the New York Academy of Sciences, 1453*, 99–113.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.
- Fitch, W. T. (2016). Why formal semantics and primate communication make strange bedfellows. *Theoretical Linguistics, 42*, 97–109.
- Fitch, W. T. (2018). The biology and evolution of speech: A comparative perspective. *Annual Review of Linguistics, 4*, 255–279.

- Fitch, W. T. (2019). Animal cognition and the evolution of human language: Why we cannot focus solely on communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 375, 20190046.
- Fitch, W. T., Neubauer, J., & Herzog, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63, 407–418.
- Giulivi, S., Whalen, D. H., Goldstein, L. M., Nam, H., & Levitt, A. G. (2011). An articulatory phonology account of preferred consonant-vowel combinations. *Language Learning and Development*, 7, 202–225.
- Hyman, L. M. (2008). Universals in phonology. *Linguistic Review*, 25, 83–137.
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, 3, 272–279.
- Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, 26, 59–99.
- Jansen, D. A., Cant, M. A., & Manser, M. B. (2013). Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biology*, 10, 97.
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., ... Zamora-Gutierrez, V. (2014). Acoustic sequences in nonhuman animals: A tutorial review and prospectus. *Biological Reviews*, 91, 13–52.
- Kroodtsma, D. E., Houlihan, P. W., Falleon, P. A., & Wells, J. A. (1997). Song development by grey catbirds. *Animal Behaviour*, 54, 457–464.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Publishers.
- Lattenkamp, E. Z., & Vernes, S. C. (2018). Vocal learning: A language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences*, 21, 209–215.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 76, 431–461.
- Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. N. Li (Eds.), *Language, speech, and mind: Studies in honour of Victoria A. Fromkin* (pp. 62–78). Routledge.
- Lowenstamm, J. (1996). CV as the only syllable type. *Current Trends in Phonology: Models and Methods*, 2, 419–441.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *The Behavioral and Brain Sciences*, 21, 499–511.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge studies in speech science and communication. Cambridge University Press.
- McComb, K., Shannon, G., Sayialel, K. N., & Moss, C. (2014). Elephants can determine ethnicity, gender, and age from acoustic cues in human voices. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 5433–5438.
- Mol, C., Chen, A., Kager, R. W. J., & ter Haar, S. M. (2017). Prosody in birdsong: A review and perspective. *Neuroscience and Biobehavioral Reviews*, 81, 167–180.
- Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marisco, & I. Chitoran (Eds.), *Approaches to phonological complexity* (pp. 299–328). Mouton de Gruyter.
- Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, 28, 48–53.
- Ohala, J. J., & Kawasaki, H. (1984). Prosodic phonology and phonetics. *Phonology Yearbook*, 1, 113–127.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE*, 4(11), e7808.
- Prince, A., & Smolensky, P. (2002). *Optimality theory: Constraint interaction in generative grammar* (ROA). Blackwell Publishing Ltd.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288, 349–351.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, 272, 941–947.
- Rice, A. N., Land, B. R., & Bass, A. H. (2011). Nonlinear acoustic complexity in a fish “two-voice” system. *Proceedings of the Royal Society B: Biological Sciences*, 278, 3762–3768.
- Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140097.

- Ryan, M. J., & Guerra, M. A. (2014). The mechanism of sound production in tungara frogs and its role in sexual selection and speciation. *Current Opinion in Neurobiology*, *28*, 54–59.
- Searcy, W. A., & Nowicki, S. (2019). Birdsong learning, avian cognition and the evolution of language. *Animal Behaviour*, *151*, 217–227.
- Seyfarth, R. M., & Cheney, D. L. (2017). Precursors to language: Social cognition and pragmatic inference in primates. *Psychonomic Bulletin and Review*, *24*, 79–84.
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*, 801–803.
- Studdert-Kennedy, M. (1998). The particulate origins of language generativity: From syllable to gesture. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 202–221). Cambridge University Press.
- Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, *108*(37), 3–8.
- Toro, J. M., Trolalon, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, *6*, 131–136.
- Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*, 195–205.
- Wilden, I., Herzel, H., Peters, G., & Tembrock, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics*, *9*, 171–196.
- Williams, H. (2004). Birdsong and singing behavior. *Annals of the New York Academy of Sciences*, *1016*, 1–30.
- Wohlgemuth, M. J., Sober, S. J., & Brainard, M. S. (2010). Linked control of syllable sequence and phonology in birdsong. *Journal of Neuroscience*, *30*, 12936–12949.
- Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, *10*, 442–446.
- Zollinger, S. A., Riede, T., & Suthers, R. A. (2008). Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations. *The Journal of Experimental Biology*, *211*, 1978–1991.