7 A Road to a Better Understanding of Rhythms in Speech Using a Comparative Approach

LARA S. BURCHARDT AND SUSANNE FUCHS

7.1 Introduction

The production and perception of acoustic communication signals in different species, including humans, have intrigued scientists all over the world and from diverse research fields such as linguistics, biology, and psychology. Some questions are important to answer in all fields. One is about the temporal structure of acoustic communication, and how it is used to convey information. These temporal structures or rhythms can act on different levels, within a phrase or between phrases, and they can also help in determining phrase boundaries. Perspectives on these connected questions are manifold, and cross-talk between the three disciplines is often limited. In this chapter, we are advocating for more cross-talk using two prosodic markers connected to rhythms as examples to showcase the advantages of combining linguists' and biologists' knowledge about the respective phenomena.

Rhythms can be defined in various ways (for an overview, see Turk and Shattuck-Hufnagel, 2013). One example is given by Patel, where rhythm is the "systematic patterning of sounds in terms of timing, accent and grouping" (2008:96). Building on this, we consider rhythm as a nonrandom, ordered, predictable, and repeated alternation of different elements in a sequence. The motivation to use this definition is to be relatively independent of theoretical concepts in one or the other research domain. We think a broader perspective including humans and nonhuman animals is important because it allows us to better understand the underlying principles and may root rhythms of human language in evolution.

According to this definition, the building blocks of rhythm are elements in a sequence. Even if it is a challenge on its own (Fletcher, 2010), determining these elements in a known human language, that is, linguistic units such as syllables, words, and prosodic phrases, might be more straightforward than determining the building blocks of rhythm in nonhuman vocalizations. Among other approaches, rhythmic markers might be a way to tackle this issue in nonhuman animals. We will focus on two prosodic markers of rhythm, fundamental frequency (f0) declination and final lengthening. Both are frequently found in human communication and have also been reported in nonhuman communication.

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

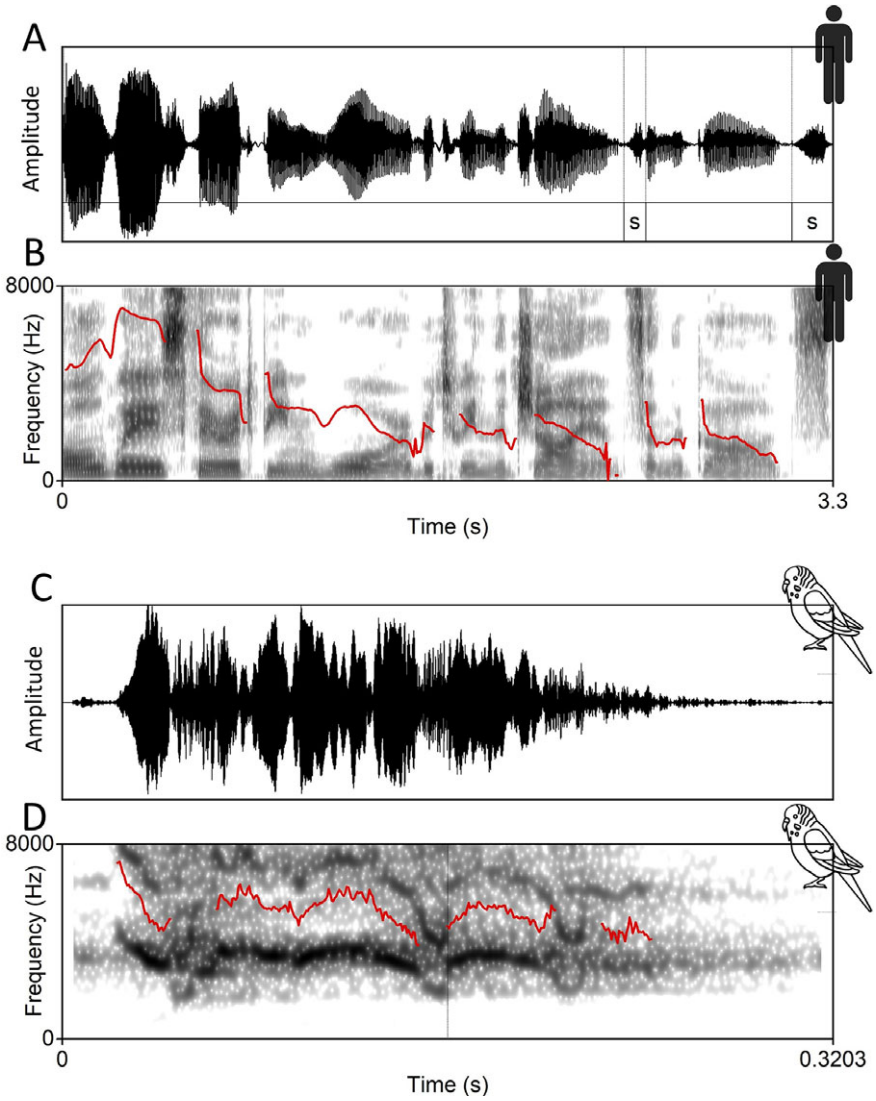


Figure 7.1 Explanation of f0 declination and final lenghthening in two examples.
A) Oscillogram of human language. The spoken text is: "Always there had been war between the giants and the gods." The duration of the "s" of "giants" and "gods" is annotated, and the final "s" is much longer, demonstrating final lengthening. B) The spectrogram for the same human sentence is shown with a solid line indicating the f0. A clear f0 decline is visible. C) An oscillogram of a budgerigar twittering. D) The corresponding spectrogram of the twittering. The f0 is shown as an extra solid line, and an f0 decline is visible.

Final lengthening refers to a phenomenon where a final or penultimate syllable of an utterance or prosodic phrase is produced for a longer duration than when the same is uttered within an utterance or prosodic phrase (Fletcher, 2010) (Figure 7.1A). Lengthening can signal the end of a unit, which varies nonlinearly with the strength of a boundary (Kentner et al.,

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

2023). It can co-occur with pauses and f0 lowering (Petrone et al., 2017). From a perceptual side, it is a phonetic signature that might help to mark the end of a unit (Schel et al., 2009).

F0 declination has been widely discussed in phonological (Ladd, 1988, 2008) and phonetic terms (Strik and Boves, 1995). We broadly define f0 declination as "the gradual decrease of f0 throughout an utterance" (Fuchs et al., 2015:35) to be as inclusive as possible concerning nonhuman animal communication (Figure 7.1A–D). F0 declination is common in statements as opposed to questions. From a phonetic perspective, it is calculated as a linear regression through f0 values in a given temporal window (e.g., 1–4 s in Yuan and Liberman, 2014:69) that often corresponds to interpausal units or annotated prosodic phrases. The linear regression slope is negative by definition (f0 decline) and flattens with utterance length (e.g., Cooper and Sorensen, 1981; Swerts et al., 1996; Fuchs et al., 2015).

Together, f0 declination and final lengthening can be used to create a sense of units in a rhythmic sequence, helping to signal boundaries between units and convey information about the length of a unit in the sequence itself that may be repeated over time to produce structured time events.

There is a lot of potential in studying this in a comparative way between species. Linguistics could help to answer long-standing questions in animal communication, for example using knowledge of human languages to determine phrase boundaries or meaningful units. The other way around, we can use knowledge about animal communication and the opportunities we might have when studying a wide variety of animal species with different cognitive abilities or physical constraints to solve debates in linguistics on how to delineate phenomena influenced by cognitive abilities or biophysical mechanisms. There might be universal underlying motor principles in humans and nonhuman animals that we can only find when studying and comparing both.

We can observe variations in pitch and timing in nonhuman animal communication similar to prosodic features in human speech (Briefer, 2012; Hotchkin and Parks, 2013; Filippi, 2016). Oftentimes, these changes are caused by physiological alterations (for example the emotional state can influence muscle tension [Briefer, 2012]). Different kinds of information can be conveyed by those changes, whether it be individual identity in birds (and many other species, e.g., Linhart et al., 2022) or context in primates (Crockford et al., 2018), among others.

There have been many recently published papers finding more and more similarities in human and nonhuman animal communication, for example finding that both penguins and gibbons also adhere to Zipf's Law of Brevity, where the most frequent elements in communication are shorter in duration (Favaro et al., 2020; Huang et al., 2020; Valente et al., 2021). In biology, more research is needed to explore the extent and complexity of prosodic features in nonhuman communication. Animal communication may resemble human prosody, but there are important differences as far as we know now: Many animal vocalizations are innate, not learned through acquisition, and may be less flexible and expressive than human speech (Janik and Slater, 2000; Tyack, 2019). At the same time, the spectrum between innate, adjusted, and learned vocalizations gives interesting options to study the prerequisites for prosodic phenomena.

To illustrate another field that would highly benefit from the comparative approach, we will focus on acoustic signals of humans and nonhuman animals using two concrete phenomena:

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

final lengthening and f0 declination. These two phonetic markers have been chosen because they are involved in determining the units that can form a rhythmic sequence. Final lengthening is a phonetic signature of the end of a phrase and signals a boundary, while the slope of f0 declination can signal the length of an entire phrase. Both are frequent in human communication, and they have been considered as being universal even if language-specific modifications may be found (Fletcher, 2010). There have been debates about the origin of prosodic phenomena in general, for example whether they rely on linguistic representations or general physical properties. Similar to other scientists (Pika et al., 2018; Matzinger and Fitch, 2021; Pouw and Fuchs, 2022; Hersh et al., 2023; Hoeschele et al., 2023), we strongly believe that more interdisciplinary cross-fertilization is needed to better understand what properties are shared among human and nonhuman animals in their rhythm communication. By comparing the use of prosodic features in human and nonhuman animal communication, we can get insights into their evolutionary origins and development of speech rhythms. We can also shed light on the cognitive and biological underpinnings of human language and communication. On a different level, the comparative approach can provide a valuable perspective on the diversity of prosodic features used across different languages and language families and therefore might raise new questions in more specific research areas. We will be able to better identify universal patterns and language-specific variations. The road toward such interdisciplinary exchange is not without obstacles, but the joint venture, in our case between a biologist and a linguist, can enhance the collection of species producing selected phenomena, can initiate new recording and open source databases (Hersh et al., 2023), and may reduce the human-centric view on the evolution of complex acoustic communication and the insider bias when doing comparative work (Hoeschele et al., 2023).

7.2 F0 Declination in Acoustic Signals: Human versus Nonhuman Animals

7.2.1 Characteristics of F0 Declination in Human Speech

F0 declination, the gradual decrease of f0 throughout an utterance, spans several words in human language and is a macro rhythm (rhythms with longer time units) in human speech. Strik and Boves (1995), among others, proposed to start analyzing time units larger than 1 second; otherwise, the calculation of the declination slope might be heavily affected by local f0 variations. F0 declination is not a rhythm in itself, but similar to final lengthening, it is a prosodic signature of a unit that can be repeated over time and form a rhythm.

While more cross-linguistic work on f0 declination using the same methodology is missing, it has been reported for various languages, such as Danish, Dutch, English, French, German, Greek, Japanese, and Spanish (see Fuchs et al., 2015, for an overview), so there is reason to believe it is a relatively robust phenomenon (Hauser and Fowler, 1992), at least in Indo-European languages. However, it is not mandatory and can be modulated; for example, f0 can rise at the end of a phrase (called continuation rise) to signal the speaker's motivation to continue talking, or it can rise signaling question intonation. There are also language specificities as well as other factors. For example, Lieberman et al. (1985) found more negative slopes in reading than in spontaneous speech.

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

A potential physiological origin of this phenomenon has been postulated. Lieberman (1967) measured subglottal pressure and f0 declination in three human participants reading declarative sentences and found a positive correlation between the two variables. He therefore suggested a potential origin in respiratory behavior, because respiration is a driving force of phonation. Others have argued that muscular tension in the vocal folds may be at the origin of f0 declination (e.g., Ohala, 1978) and that tensioning of the vocal folds is independent of respiration because the primary function of the larynx is to save lives and protect the lungs from foreign bodies, hence it must be independent and quick (Ohala, 1990). The challenge for or against one or the other argument or a mix of the two is that measuring subglottal pressure and laryngeal tension is very invasive, so empirical data is limited.

Apart from the physiological origin of f0 declination, cognitive processes seem plausible as well. Since the slope of the f0 declination is correlated with the length of the upcoming utterance, some anticipatory planning may be involved (Yuan and Liberman, 2014).

7.2.2 F0 Declination in Nonhuman Animal Communication

We occasionally find descriptions hinting at f0 declination in monkeys. For example, in a description of the vocalizations of the black and white colobus monkey, we find the following: "The final phrase is often deeper pitched than the others" (Marler, 1972:181). This refers to the alarm call "roar." This leads Schel et al. (2009) to hypothesize that this phenomenon is perceptually conspicuous and marks the end of the sequence. In colobus monkeys a roaring sequence consists of one or more roaring phrases, where a phrase is a basic unit, made up of ~15 "pulses," each with an average duration of 0.7 seconds, which makes a whole phrase around 10 seconds (Marler, 1972).

Confusingly, in the same species, the exact opposite was reported, at least for the black colobus monkey (*Colobus satanas*): "Initial phrases all decreased in pitch during delivery, the terminal phrases all increased" in pitch (Oates and Trocco, 1983:100). This could have different explanations: The phenomenon could be dependent on unknown parameters that have differed between the studies, or coincidentally the decline in final phrases was observed in a different colobus monkey species.

The most detailed study on f0 declination in nonhuman animals was conducted in vervet monkeys (*Cercopithecus aethiops*) and rhesus macaques (*Macaca mulatta*), two very common model species (Hauser and Fowler, 1992). For both species, vocal production also shows f0 declination, which is suggested to serve a similar communicative function as in human language (Hauser and Fowler, 1992). Under investigation were vocalizations uttered during aggressive interaction for vervet monkeys in the wild and an affiliative vocalization for wild rhesus macaques. For both species an almost linear decline in f0 could be shown over call bouts of two calls and of three calls. Furthermore, for two call bouts of vervet monkeys, a correlation between the duration of the bout and f0 decline could be shown, as expected from human language literature (e.g., Yuan and Liberman, 2014). No correlation could be found between the duration of the bout and the magnitude of the f0 decline in rhesus macaques. Another interesting thing to note is that the structured decline of f0 in vervet monkeys could only be seen in adults, not in juveniles, which could be explained by the fact that inter-call intervals in

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

juveniles are generally longer, indicating juveniles might take a breath between calls, in contrast to adults. This makes it even more interesting to study the phenomenon further.

Another instance of f0 declination being reported in monkeys stems from baboons, where f0 is also reported to decline within bouts of calling. This is described to be independent of rank and age in a species where f0 generally is highly correlated to these two parameters (Fischer et al., 2004), making it a more general phenomenon worthy of further investigation.

In birds, f0 decline was found in the vocalizations of the budgerigar (*Melopsittacus undulatus*). Here, "mean F0 measurements were lower for segments in syllable-final position when compared to medial segments" (Mann et al., 2021:6, Figure 4 caption).

7.2.3 Comparing F0 Declination in Human and Nonhuman Animal Communication

The number and detail of papers published on f0 declination are clearly more substantial in the linguistic domain, which shouldn't imply that this is an exclusively human phenomenon. Papers published on f0 declination in animal communication include primates and birds. They often lack empirical work on the underlying mechanisms (different for final lengthening – see below). While papers on animal communication are a stepping stone toward a broader perspective, they are mostly descriptive.

Whether or not f0 declination is primarily the result of a decrease in subglottal pressure, a reduction in laryngeal tension throughout an utterance, a marker of anticipatory planning of an utterance, or a mixture of these is still unclear. The invasiveness of recording data in favor of the first two explanations limits the empirical evidence in humans.

There may also be some challenges when comparing humans and nonhuman animals. The minimum length of an utterance that can be considered for calculating the f0 declination slope may have to be adjusted for various animal species, similar to the terminology used. For an untrained linguist, the term "syllable" that is used in animal communication may have a very different connotation than for a biologist, for whom it may be clear that this is an utterance between silent pauses. Joint venture investigations on humans and nonhuman animals might also reveal deeper insights because they differ in their respiratory, vocal, and cognitive repertoire.

7.3 Final Lengthening: Humans versus Nonhuman Animals

7.3.1 Final Lengthening in Human Speech

Phonetic studies in a variety of languages found that final lengthening is a reliable phonetic marker determining the end of a speech chunk and is very pronounced next to a following pause (e.g., Klatt, 1976; Edwards et al., 1991; for a review on languages, see Paschen et al., 2022). There have also been considerations that the lengthening of the final segment is part of the pause (e.g., Krivokapić et al., 2022) or can, in extreme cases, be produced instead of a pause in fast speech. Indeed, a pause is an important determiner of rhythm.

Final lengthening has been claimed to be universal in human language (Fletcher, 2010) with additional language specificities (e.g., Nakai et al., 2009). Since the term "universal" can be

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

ambiguous in meaning (Bickel, 2011), we refer to statistical universal here, which relies on robust statistical evidence across languages but also allows for exceptions. There is only recent empirical evidence, using the same methodology for 25 mostly understudied languages, that the lengthening of vowels is a statistically robust cross-linguistic phenomenon (Paschen et al., 2022). Language-specific variations, driven by phonological vowel length, were found as well. Sound-specific variations have also been reported, for example in Berkovits (1993) for Hebrew, who described stronger lengthening effects for final fricatives than stops. Paschen (2023) provided evidence for Lower Sorbian that lengthening occurs in vowels, sonorants, and fricatives, but not in stops. While these latter segmental influences may be specific to human language, they may also give some hints that continuous airstream mechanisms make final lengthening more likely.

The degree of lengthening has been extensively discussed. For example, the pi-gesture model (Byrd and Saltzman, 2003) proposes that the longer the segments, the closer they are to the boundary. Moreover, lengthening varies with the boundary type. Final segments at major boundaries, for example at the end of a sentence, may be longer than segments at phrase boundaries within a sentence. While language-specific variants exist, this does not exclude the assumption that the underlying principles are physical in nature but have been shaped in various ways by the properties of the sounds and the users of individual languages.

What are the underlying mechanisms that may cause final lengthening? Do we also find them in other behavior or other species?

7.3.2 Final Lengthening in Nonhuman Animal Communication

Final lengthening is getting increased attention in nonhuman animal acoustic studies. It was found in birds and primates. We can find it as well as f0 declination in the budgerigar, a vocal learning parrot. It was reported that segments at the end of vocalizations were more likely to be longer. In 14 adult budgerigars, it was found that segments in syllable-final positions are on average longer than medial segments (Mann et al., 2021). The budgerigar, thus, is the only nonhuman species so far where both final lengthening and f0 declination were observed. This is likely a sampling bias, where those phenomena have not been studied in many species. In another study on 80 different songbirds, the same could be shown: Song-final notes were significantly longer than nonfinal notes (Tierney et al., 2011). This is especially impressive as the analysis wasn't conducted per species but across all species, indicating this to be a generally observable phenomenon in songbirds. A more detailed analysis to find possible differences between families would be interesting. Both papers argue that final lengthening can be observed in these songbirds as well as in humans, because of similar motor constraints. Both humans and songbirds show high control of their vocal articulators with the possibility to rapidly adjust them during vocal production. Nevertheless, abrupt termination of these movements might be difficult, as opposed to a gradual relaxation and therefore slowing of articulators, resulting in final lengthening. This argument is further strengthened by the fact that budgerigar segments in particular are produced within a single breath (Tierney et al., 2011; Mann et al., 2021). It would be interesting to investigate respiratory kinematics and final lengthening in human speech production.

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

Final lengthening was also found in at least three different primate species: two crested gibbon species and the indri (Huang et al., 2020; Valente et al., 2021). In gibbons, this effect was found on two different structural levels, in vocal sequences and in bouts (where a vocal sequence is a short unit and a bout is made up of several sequences). The suggested reasons differ from those discussed in humans. The authors suggest a connection to males advertising their quality; that is, gibbons might be modulating the frequency of their calls rapidly at the end of sequences to advertise their individual quality to females as potential mating partners. An increase in frequency modulations, even though fast, would potentially lead to an increase in the duration of notes (Huang et al., 2020). If this is true, the observed phenomenon of final lengthening in crested gibbons would only be a byproduct of other processes.

7.3.3 Comparing Final Lengthening in Human and Nonhuman Animal Communication

There are different explanations for final lengthening in the literature. It has been understood as a general motor property rather than being innate. Tierney et al. (2011) attribute it to the energy efficiency of the underlying motor actions in humans and singing birds. As mentioned above, Huang et al. (2020) explain it as a byproduct rather than a phenomenon on its own. Matzinger and Fitch (2021) mention the possibility that the slowing down of articulators could be a result of a change from exhalation to inhalation; that is, respiratory dynamics and the occurrence of a breathing pause cause final lengthening. We think that it could be a plausible explanation for those species that produce segments within one breath, but in humans, final lengthening can also be found without a change from exhalation to inhalation. Nevertheless, the relation of one breath to one segment (or one phrase in human communication) may have been at the origin of human language evolution. In spontaneous interactive dialogues, more than 50% of all turns consisted of only one breathing cycle (Rochet-Capellan and Fuchs, 2014). Linguistic studies have mostly focused on language- and segment-specific properties in the implementation of final lengthening. Because the number of published papers on these specificities is so much more than in animal communication, one may implicitly assume that it is a phenomenon of human language.

Even if language variations exist, it does not exclude the possibility of an underlying motor principle. Humans and nonhuman animals clearly have motor constraints, but these may not be persistent under each and every situation, and all animals may also be able to compensate for it if needed. Monocausalities are rather rare in biological systems. For example, the available exhalation air in human speech may be physically constrained by the lung volume (vital capacity) of a speaker. In human and nonhuman acquisition, a high positive correlation between lung volume and utterance/vocalization length has been found (see the review in Fuchs and Rochet-Capellan, 2021). Evidence for a similarly strong correlation between utterance length and vital capacity in adults is missing, because humans with smaller lung capacities may adjust their laryngeal resistance and lose less expiratory air to compensate for their physical constraints.

All in all, evidence for final lengthening in nonhuman animals has changed the perspective that the phenomenon is a purely linguistic one to the notion that it might be grounded in general motor constraints.

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

7.4 Future Directions

7.4.1 A Roadmap

In the venture of finding the biological underpinnings of prosodic phenomena such as f0 declination and final lengthening through a comparative approach combining knowledge and advantages from studying human and nonhuman communication, there are several issues to overcome and steps to take. We will lay out a possible roadmap to achieve this here, with concrete steps.

1) *Establishing a common vocabulary and conceptual framework*: This touches on issues mentioned earlier, where terms such as "syllable" might be understood differently between linguists and biologists, but also within biology. Clear terminology and glossaries and clear but adjustable definitions and criteria for measuring and analyzing features are important steps.
2) *Establishing the necessary data basis*: We would need to develop a comprehensive database of vocalizations from a wide range of species to study these and other phenomena comparatively. A database should optimally include both natural or spontaneous and elicited vocalizations. Most importantly, meta-information about the social or ecological contexts is important to know, as well as potential age and sex.
3) *Identifying universal and species-specific patterns*: Once established, researchers can begin to identify robust patterns in the use of prosodic features across species, as well as species-specific variations, utilizing the database. Eventually, this can help to shed light on the evolutionary and ecological factors that shape the use of these features in different species with different needs and skills.
4) *Integrating insights from different disciplines*: With a comprehensive knowledge of theories, mechanisms, and constraints in linguistics, biology, psychology, and neuroscience from interdisciplinary collaborations, we can identify new research questions and generate novel insights into the biological and cognitive foundations of communication.

7.4.2 Prosodic Rhythm Markers and Respiration

A future endeavor toward a better understanding of prosodic markers of rhythm could be to investigate the interaction between respiration, phonation (voice quality), and anticipatory planning in humans and nonhuman animals. Respiration itself is a biological rhythm, and the duration of breathing cycles can constrain how long an animal can phonate. At the end of long sequences when lung volume is reduced, laryngeal adjustments may be required to continue phonation. While this has been modelled for humans (Zhang, 2016), it may also exist in nonhuman animals. The amount of air inhaled may further shape acoustic properties such as intensity and to some extent f0 (Watson et al., 2003).

Vocalizations in nonhuman primates such as screams or grunts may have specific prosodic features that can be shaped by respiratory control. Birds have more complex respiratory and laryngeal systems than humans, including not only lungs but several additional air sacs that are necessary during avian flight. Phonation is produced with a syrinx that is much closer to the

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

lungs than in humans, at the place where the trachea forks into the lungs. These physiological differences and control mechanisms are, on the one hand, a big challenge for comparative work; on the other hand, they may give us insights into which physiological or cognitive systems can produce these patterns and which features they have. Recent technological developments in thermography allow us to record breathing and acoustics even in free-ranging animals (Demartsev et al., 2022).

7.5 Conclusion

This chapter examined the intersection between biophysical and cognitive constraints on vocal communication. To that end, we investigated what is known about two concrete prosodic phenomena connected to rhythm in human language and nonhuman animal communication: final lengthening and f0 declination. Both prosodic phenomena have been found in different nonhuman animal species, mostly in birds but also in primates. We are sure that more examples will be found in different species. For final lengthening the evidence in nonhuman animals has changed the perspective that the phenomenon is a purely linguistic one to the notion that it might be grounded in general motor constraints. There is clearly an advantage when humans and nonhuman animals are taken into account, to find these explanations.

Box 7.1 Chapter Overview

Summary

Focusing on final lengthening and f0 declination as prosodic rhythm markers in human language and nonhuman animal communication, we explored the intersection of biophysical and cognitive factors. Examples from birds and primates challenge the notion of them being purely linguistic, suggesting a basis in general motor constraints. Further cross-species investigation promises additional insights.

Implications

F0 declination and final lengthening have been discussed as two properties demarcating the unit of a rhythm. Descriptions of rhythms in speech communication would benefit from the inclusion of nonhuman animal data for a better understanding of the underlying principles, and rooting rhythms of communication in evolution.

Gains

The evidence for f0 declination and final lengthening in nonhuman animals may change our human-centric perspective from purely linguistic phenomena to grounding it in general motor constraints, learned and innate behavior. A potential road toward further comparative analyses has been described.

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

REFERENCES

Berkovits, R., 1993. Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech* 36, 89–98. https://doi.org/10.1177/002383099303600105

Bickel, B., 2011. Statistical modeling of language universals. *Linguistic Typology* 15(2), 401–413. https://doi.org/10.1515/lity.2011.027

Briefer, E. F., 2012. Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology* 288, 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x

Byrd, D., and Saltzman, E., 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2

Cooper, W., and Sorensen, J., 1981. *Fundamental frequency in sentence production*. Springer Verlag.

Crockford, C., Gruber, T., and Zuberbühler, K., 2018. Chimpanzee quiet hoo variants differ according to context. *Royal Society Open Science* 5, 172066. https://doi.org/10.1098/rsos.172066

Demartsev, V., Manser, M. B., and Tattersall, G. J., 2022. Vocalization-associated respiration patterns: Thermography-based monitoring and detection of preparation for calling. *The Journal of Experimental Biology* 225, jeb243474. https://doi.org/10.1242/jeb.243474

Edwards, J., Beckman, M. E., and Fletcher, J., 1991. The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America* 89, 369–382. https://doi.org/10.1121/1.400674

Favaro, L., Gamba, M., Cresta, E., et al., 2020. Do penguins' vocal sequences conform to linguistic laws? *Biology Letters* 16, 20190589. https://doi.org/10.1098/rsbl.2019.0589

Filippi, P., 2016. Emotional and interactional prosody across animal communication systems: A comparative approach to the emergence of language. *Frontiers in Psychology* 7, 1393. https://doi.org/10.3389/fpsyg.2016.01393

Fischer, J., Kitchen, D. M., Seyfarth, R. M., and Cheney, D. L., 2004. Baboon loud calls advertise male quality: Acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology* 56, 140–148. https://doi.org/10.1007/s00265-003-0739-4

Fletcher, J., 2010. The prosody of speech: Timing and rhythm, in W. J. Hardcastle, J. Laver, and F. E. Gibbon (eds.), *The Handbook of Phonetic Sciences*, pp. 521–602. John Wiley & Sons, Ltd.

Fuchs, S., and Rochet-Capellan, A., 2021. The respiratory foundations of spoken language. *Annual Review of Linguistics* 7, 13–30. https://doi.org/10.1146/annurev-linguistics-031720-103907

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U. D., and Koenig, L. L., 2015. Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics* 52, 35–45. https://doi.org/10.1016/j.wocn.2015.04.002

Hauser, M. D., and Fowler, C. A., 1992. Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates. *The Journal of the Acoustical Society of America* 91, 363–369. https://doi.org/10.1121/1.402779

Hersh, T. A., Ravignani, A., and Burchardt, L. S., 2023. Robust rhythm reporting will advance ecological and evolutionary research. *Methods in Ecology and Evolution*, 14, 1398–1407. https://doi.org/10.1111/2041-210X.14118

Hoeschele, M., Wagner, B., and Mann, D. C., 2023. Lessons learned in animal acoustic cognition through comparisons with humans. *Animal Cognition* 26, 97–116. https://doi.org/10.1007/s10071-022-01735-0

Hotchkin, C., and Parks, S., 2013. The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems. *Biological Reviews* 88, 809–824. https://doi.org/10.1111/brv.12026

Huang, M., Ma, H., Ma, C., Garber, P. A., and Fan, P., 2020. Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: Insights into the origin of human language. *Animal Behaviour* 160, 145–155. https://doi.org/10.1016/j.anbehav.2019.11.017

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

Janik, V. M., and Slater, P. J. B., 2000. The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11. https://doi.org/10.1006/anbe.2000.1410

Kentner, G., Franz, I., Knoop, C. A., and Menninghaus, W., 2023. The final lengthening of pre-boundary syllables turns into final shortening as boundary strength levels increase. *Journal of Phonetics* 97, 101225. https://doi.org/10.1016/j.wocn.2023.101225

Klatt, D. H., 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59, 1208–1221. https://doi.org/10.1121/1.380986

Krivokapić, J., Styler, W., and Byrd, D., 2022. The role of speech planning in the articulation of pauses. *Journal of the Acoustical Society of America* 151, 402–413. https://doi.org/10.1121/10.0009279

Ladd, D. R., 1988. Declination "reset" and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America* 84, 530–544. https://doi.org/10.1121/1.396830

Ladd, D. R., 2008. *Intonational phonology*, 2nd ed. Cambridge University Press.

Lieberman, P., 1967. *Intonation, perception and language*. MIT Press.

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., and Miller, M., 1985. Measures of the sentence intonation of read and spontaneous speech in American English. *The Journal of the Acoustical Society of America* 77, 649–657. https://doi.org/10.1121/1.391883

Linhart, P., Mahamoud-Issa, M., Stowell, D., and Blumstein, D. T., 2022. The potential for acoustic individual identification in mammals. *Mammalian Biology* 102, 667–683. https://doi.org/10.1007/s42991-021-00222-2

Mann, D. C., Fitch, W. T., Tu, H.-W., and Hoeschele, M., 2021. Universal principles underlying segmental structures in parrot song and human speech. *Scientific Reports* 11, 776. https://doi.org/10.1038/s41598-020-80340-y

Marler, P., 1972. Vocalizations of East African monkeys II: Black and white colobus. *Behaviour* 42, 175–197. https://doi.org/10.1163/156853972X00266

Matzinger, T., and Fitch, W. T., 2021. Voice modulatory cues to structure across languages and species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 20200393. https://doi.org/10.1098/rstb.2020.0393

Nakai, S., Kunnari, S., Turk, A., Suomi, K., and Ylitalo, R., 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37, 29–45. https://doi.org/10.1016/j.wocn.2008.08.002

Oates, J. F., and Trocco, T. F., 1983. Taxonomy and phylogeny of black-and-white colobus monkeys: Inferences from an analysis of loud call variation. *Folia Primatologica* 40, 83–113. https://doi.org/10.1159/000156092

Ohala, J. J., 1978. The production of tone, in V. A. Fromkin (ed.), *Tone: A linguistic survey*, pp. 5–39. Academic Press.

Ohala, J. J., 1990. Respiratory activity in speech, in M. Hardcastle (ed.), *Speech production and speech modelling*, pp. 23–53. Kluwer.

Paschen, L., 2023. Final lengthening across various sound classes in a language documentation corpus of Lower Sorbian, in R. Skarnitzl and J. Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences: ICPhS 2023*, pp. 3252–3256. International Phonetic Association.

Paschen, L., Fuchs, S., and Seifart, F., 2022. Final lengthening and vowel length in 25 languages. *Journal of Phonetics* 94, 101179. https://doi.org/10.1016/j.wocn.2022.101179

Patel, A. D., 2008. *Music, language and the brain*. Oxford University Press.

Petrone, C., Truckenbrodt, H., Wellmann, C., et al., 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61, 71–92. https://doi.org/10.1016/j.wocn.2017.01.002

Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C., 2018. Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences* 285, 20180598. https://doi.org/10.1098/rspb.2018.0598

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

Pouw, W., and Fuchs, S., 2022. Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews* 141, 104836. https://doi.org/10.1016/j.neubiorev.2022.104836

Rochet-Capellan, A., and Fuchs, S., 2014. Take a breath and take the turn: How breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369, 20130399. https://doi.org/10.1098/rstb.2013.0399

Schel, A. M., Tranquilli, S., and Zuberbühler, K., 2009. The alarm call system of two species of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *Journal of Comparative Psychology* 123, 136–150. https://doi.org/10.1037/a0014280

Strik, H., and Boves, L., 1995. Downtrend in F0 and Psb. *Journal of Phonetics* 23, 203–220. https://doi.org/10.1016/S0095-4470(95)80043-3

Swerts, M., Strangert, E., and Heldner, M., 1996. F0 declination in read-aloud and spontaneous speech. *Proceedings of the Fourth International Conference on Spoken Language Processing*, ICSLP '96. Philadelphia, PA, pp. 1501–1504. https://doi.org/10.1109/ICSLP.1996.607901

Tierney, A. T., Russo, F. A., and Patel, A. D., 2011. The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences* 108, 15510–15515. https://doi.org/10.1073/pnas.1103882108

Turk, A., and Shattuck-Hufnagel, S., 2013. What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology* 4, 93–118. https://doi.org/10.1515/lp-2013-0005

Tyack, P. L., 2019. A taxonomy for vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375, 20180406. https://doi.org/10.1098/rstb.2018.0406

Valente, D., De Gregorio, C., Favaro, L., et al., 2021. Linguistic laws of brevity: Conformity in Indri indri. *Animal Cognition* 24, 897–906. https://doi.org/10.1007/s10071-021-01495-3

Watson, P. J., Ciccia, A. H., and Weismer, G., 2003. The relation of lung volume initiation to selected acoustic properties of speech. The *Journal of the Acoustical Society of America* 113, 2812–2819. https://doi.org/10.1121/1.1567279

Yuan, J., and Liberman, M., 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication* 65, 67–74. https://doi.org/10.1016/j.specom.2014.06.001

Zhang, Z., 2016. Respiratory laryngeal coordination in airflow conservation and reduction of respiratory effort of phonation. *Journal of Voice* 30, 760.e7–760.e13. https://doi.org/10.1016/j.jvoice.2015.09.015

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press

https://doi.org/10.1017/9781009295888.009 Published online by Cambridge University Press