



UvA-DARE (Digital Academic Repository)

Voice modulation: from origin and mechanism to social impact

Leongómez, J.D.; Pisanski, K.; Reby, D.; Sauter, D.; Lavan, N.; Perlman, M.; Varella Valentova, J.

DOI

[10.1098/rstb.2020.0386](https://doi.org/10.1098/rstb.2020.0386)

Publication date

2021

Document Version

Final published version

Published in

Philosophical Transactions of the Royal Society B: Biological Sciences

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

Leongómez, J. D., Pisanski, K., Reby, D., Sauter, D., Lavan, N., Perlman, M., & Varella Valentova, J. (2021). Voice modulation: from origin and mechanism to social impact. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1840), [20200386]. <https://doi.org/10.1098/rstb.2020.0386>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Introduction



Cite this article: Leongómez JD, Pisanski K, Reby D, Sauter D, Lavan N, Perlman M, Varella Valentova J. 2021 Voice modulation: from origin and mechanism to social impact. *Phil. Trans. R. Soc. B* **376**: 20200386. <https://doi.org/10.1098/rstb.2020.0386>

Accepted: 1 October 2021

One contribution of 11 to a theme issue ‘Voice modulation: from origin and mechanism to social impact (Part I)’.

Subject Areas:

behaviour, cognition, evolution, neuroscience

Keywords:

vocal control, voice perception, acoustic communication, evolution of speech, emotions, music

Authors for correspondence:

Juan David Leongómez
e-mail: jleongomez@unbosque.edu.co
Katarzyna Pisanski
e-mail: katarzyna.pisanski@cnrs.fr

†These authors contributed equally to this article.

Voice modulation: from origin and mechanism to social impact

Juan David Leongómez^{1,†}, Katarzyna Pisanski^{2,3,†}, David Reby², Disa Sauter⁴, Nadine Lavan⁵, Marcus Perlman⁶ and Jaroslava Varella Valentova⁷

¹Human Behaviour Laboratory (LACH), Faculty of Psychology, Universidad El Bosque, Bogota, DC 110121, Colombia

²Sensory Neuro-Ethology Laboratory (ENES), Neuroscience Research Centre of Lyon (CRNL), Jean Monnet University Saint-Etienne, Saint-Etienne, France

³CNRS – Centre National de la Recherche Scientifique, Laboratoire Dynamique du Langage, Université Lyon 2, Lyon, France

⁴Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

⁵Department of Biological and Experimental Psychology, Queen Mary, University of London, London, UK

⁶Department of English Language and Linguistics, University of Birmingham, Birmingham, UK

⁷Department of Experimental Psychology, Institute of Psychology, University of São Paulo, São Paulo 05508-030, Brazil

id JDL, 0000-0002-0092-6298; KP, 0000-0003-0992-2477; DR, 0000-0001-9261-1711; DS, 0000-0003-4872-0536; NL, 0000-0001-7569-0817; MP, 0000-0002-1269-3882; JVV, 0000-0002-2113-3385

Research on within-individual modulation of vocal cues is surprisingly scarce outside of human speech. Yet, voice modulation serves diverse functions in human and nonhuman nonverbal communication, from dynamically signalling motivation and emotion, to exaggerating physical traits such as body size and masculinity, to enabling song and musicality. The diversity of anatomical, neural, cognitive and behavioural adaptations necessary for the production and perception of voice modulation make it a critical target for research on the origins and functions of acoustic communication. This diversity also implicates voice modulation in numerous disciplines and technological applications. In this two-part theme issue comprising 21 articles from leading and emerging international researchers, we highlight the multidisciplinary nature of the voice sciences. Every article addresses at least two, if not several, critical topics: (i) development and mechanisms driving vocal control and modulation; (ii) cultural and other environmental factors affecting voice modulation; (iii) evolutionary origins and adaptive functions of vocal control including cross-species comparisons; (iv) social functions and real-world consequences of voice modulation; and (v) state-of-the-art in multidisciplinary methodologies and technologies in voice modulation research. With this collection of works, we aim to facilitate cross-talk across disciplines to further stimulate the burgeoning field of voice modulation.

This article is part of the theme issue ‘Voice modulation: from origin and mechanism to social impact (Part I)’.

1. Introduction

Vocal communication is widespread across tetrapod species. In many animals, including humans, the nonverbal properties of vocal cues can indicate both relatively static biological qualities of the vocalizer, such as sex, age and body size. Importantly, vocal cues can also dynamically signal motivations and emotions, such as aggression, fear, distress or pleasure. The vast diversity of functions and environments in which vocalizations are used has led to diverse anatomical, neural, cognitive and behavioural adaptations for their production and perception, including a capacity for volitional voice modulation in some species. However, while vocalizers in several species such as songbirds, cetaceans and pinnipeds have developed the ability to voluntarily alter their vocal output, existing research suggests that humans possess a remarkably

advanced capacity for voluntary vocal control—one which surpasses the abilities observed in other terrestrial mammals, including extant nonhuman primates. While vocal control is a necessary prerequisite for speech, it is also observed in human nonverbal vocalizations, such as conversational laughter, exaggerated roars of aggression or embellished cries of pain. This suggests that voice modulation plays an important role in both verbal and nonverbal communication and may often confer social and reproductive advantages.

Indeed, a large body of research has demonstrated the importance of nonverbal vocal parameters, such as voice pitch (the perceptual correlate of fundamental frequency, f_0) and vocal tract resonances (formant frequencies), in indicating biologically relevant qualities of vocalizers (e.g. sex, size, hormonal levels, dominance). This literature provides compelling evidence that the human vocal apparatus has been, at least partly, shaped by sexual selection, much like the vocal organs of other mammals. Studying variation between individuals in static vocal parameters has greatly advanced our understanding of their probable evolutionary origins and social functions. However, research on *within-individual voice modulation* has to date been surprisingly scarce, despite a promising interest and progressive rise in empirical research on vocal control. Also problematically most publications to date are scattered across thematically diverse specialist journals, without localized compendia of research. As a result, despite shared research interests, there has been very little cross-talk across disciplines in the voice sciences and many critical questions about the mechanisms, ontogeny, social functions and phylogenetic roots of human vocal behaviour remain unanswered.

The scarcity of research on within-individual voice modulation is particularly surprising. This behavioural phenomenon is present in a variety of animal species, is of interest to many disciplines (e.g. biology, psychology, ethology, anthropology, bioacoustics, communication studies, linguistics), and is a crucial predictor of social outcomes. Indeed, dynamic voice characteristics commonly affect the perceptions and behaviours of listeners. For instance, in our everyday (human) lives, voice features can elicit social stereotyping, including gender representations, that influence how listeners perceive the speaker. These perceptions can in turn affect important social decisions from who we vote for as a leader, hire for a job or choose as a romantic partner. As several studies in this issue show, the production and perception of human voice modulation can depend on social context as well as characteristics of the speaker and listener(s), and studying these interacting social factors is imperative to understand how voluntary (and sometimes deceptive) voice manipulation can influence interpersonal and social outcomes. Voice modulation also has implications for how listeners (human, nonhuman animal or machine) recognize speakers despite constant changes in their voice. Studying within-individual changes in nonverbal vocal parameters is thus necessary to develop effective voice recognition technology, automatic voice-based monitoring devices and voice resynthesis tools that currently depend too heavily on ostensibly stable speech parameters.

With this theme issue, we therefore aim to provide a comprehensive overview of current research on vocal communication, focusing on voice modulation. The issue serves to synthesize knowledge from multiple disciplines, while highlighting important open questions to direct ongoing

research and advance the voice sciences: When did vocal control originate in our ancestral lineage, and when and how does it emerge during ontogeny? What neural, anatomical and sociocultural factors shape its development and use in humans and other animals? What diverse functions does voice modulation serve, from nonverbal communication, to speech, to musicality? and how does human voice modulation influence the perceptions and behaviours of listeners, potentially to the benefit of the signaller, the receiver or both?

2. Topics and articles

This two-part issue, comprising 21 articles, represents a comprehensive and multidisciplinary synthesis of research and theory that seeks to: (i) establish the study of within-individual, in addition to between-individual, acoustic properties in nonverbal communication; (ii) integrate the empirical and theoretical literature; (iii) promote cross-talk across scientific disciplines including biology (anatomy, physiology, neurology), psychology (cognitive, developmental, cross-cultural, experimental, social), ethology (including primatology), anthropology, bioacoustics, communication studies, linguistics and the computer sciences, with the aim to establish a comparative framework in acoustic communication research; and finally, (iv) underscore the state-of-the-art in technological applications of voice research. For this reason, we have carefully selected contributions to cover five *critical topics* where further investigation is notably needed:

- (i) development and mechanisms driving vocal control and modulation;
- (ii) cultural and other environmental factors affecting voice modulation;
- (iii) evolutionary origins and adaptive functions of vocal control including cross-species comparisons;
- (iv) social functions and real-world consequences of voice modulation; and
- (v) state-of-the-art in multidisciplinary methodologies and technologies in voice modulation research.

Following our key aim to encourage cross-pollination between scientific areas and highlight the multidisciplinary nature of the field of voice modulation, both parts of this theme issue contain theoretical and empirical contributions addressing at least two, and often several, of the critical topics identified above. Because of this, and more importantly to highlight the crucial links across disciplines, the papers in this issue are *not* grouped by topic. Instead, we identify the multiple topics of each paper in brackets, after which each article is briefly described in the order in which it appears in the issue (which also corresponds to its citation number).

(a) Part 1

Both parts of this two-part issue begin by introducing the state-of-the-art in voice modulation research with three review papers. Hughes & Puts [1] (critical topics (ii), (iii) and (iv)) cover the current knowledge on human voice modulation in two fundamental and pervasive social contexts: mating and intrasexual competition. The authors show how altering acoustic parameters in speech, particularly voice pitch (f_0), can reveal biologically relevant qualities of the speaker and can influence the social perceptions of listeners.

It can also serve as a salient medium for signalling a person's dominance, attraction to another and relationship status, ultimately functioning to facilitate courtship success.

Taking a more mechanistic perspective, Belyk *et al.* [2] (critical topics (i), (iii) and (v)) offer a comprehensive description of the dual larynx motor networks hypothesis, describing the neural systems that underlie voice modulation or vocal production learning in the human brain. They review research that supports the existence of network-wide somatotopy, and elegantly highlight how mapping the neural systems of voice modulation can help researchers to uncover how humans evolved the ability to speak. The third and final review of Part 1 of this issue, by Matzinger & Fitch [3] (critical topics (i), (ii) and (iii)), centres on the physiological mechanisms structuring speech modulation across human languages. The authors demonstrate that identifying cues that are used similarly across many human languages (probably owing to a shared physiology in voice production and perception) can help indicate which of these cues result from physiological or basic cognitive constraints. Importantly, this approach may also help to pinpoint which cues may be employed more flexibly and are shaped by cultural evolution. This review is focused on human languages owing to a lack of much-needed data on nonhuman tetrapods, for which some promising candidates for future investigation of cues to structure in vocalizations are suggested.

The remaining seven contributions in Part 1 are empirical. Torres Borda *et al.* [4] (critical topics (ii) and (iii)) present evidence of the early emergence of vocal plasticity in an aquatic mammal, the infant harbour seal. Their results show that these pinniped pups, known to be capable of vocal learning in adulthood, can already modulate calls in response to noise at one to three weeks of age. Seal pups modify their calls when exposed to band-pass filtered noise that masks the typical pitch (f_0) range of conspecific calls. The results suggest that the ability to modulate vocalizations may arise early in development in mammals capable of vocal production learning.

From the perceiver's side, past studies have shown that changes in pitch, loudness and speed are linked to perceptions of emotion in vocal signals, but interestingly, also in instrumental music. Bedoya *et al.* [5] (critical topics (i) and (v)) take this one step further to show that voice-specific markers of emotion, such as vocal tremor, roughness and the sound of smiling, can be applied to musical stimuli to evoke analogous affective judgements. For example, musical sounds manipulated to feature the acoustic signature of a smiling voice are perceived as more positive, while musical sounds perceived to include more roughness, like the sounds of harsh screams, are perceived as relatively more negative. The authors show that perceptions of these emotions are similar across musicians and non-musicians, suggesting that the voice-specific acoustic manipulations tap into acoustic features that can be applied to potentially any carrier sounds, to be perceived as emotional. Contributing further to the link between voice and music, Yan *et al.* [6] (critical topics (i), (ii) and (iv)) provide a look into the prevalence of parental vocal music directed towards infants, showing that most parents sing to their young children every day. Indeed, despite the increasing prevalence of technology in infants' daily lives, singing appears to remain the primary musical activity of mother–infant dyads. These results are based on survey data collected from North American parents and a meta-analysis of research spanning three

decades. The authors also report that fathers sing less than do mothers and, as infants grow older, parents sing less. Nevertheless, the frequency of playing recorded music to infants does not change with children's age. This research complements a large body of literature on infant-directed speech or 'motherese', further implicating nonverbal voice modulation as a critical vocal behaviour in parent–offspring communication, from speech to singing.

Shifting from the social functions of soothing songs to agonistic threat displays, Kleisner *et al.* [7] (critical topics (ii), (iii) and (iv)) show that volitionally produced human aggressive vocalizations (roars) predict several times more variance in the actual physical strength of vocalizers than does speech. This was shown in two understudied African societies, urban-living Cameroonians and nomadic Hadza hunter–gatherers in Tanzania, complimenting earlier results obtained from European samples. Employing an innovative bottom-up, information-theoretic approach based on multimodal inference and averaging, this research provides novel support that human nonverbal vocalizations are homologous in form and function to those of other animals. Indeed, like other mammal species, humans produce roar-like vocalizations in competitive or combative contexts such as sports and warfare that the current study shows can exaggerate the physical strength of vocalizers, and thus may function to signal threat and formidability.

How is this capacity to control our voices, often on demand, mapped in the human brain? By simultaneously recording real-time changes in vocal anatomy and functional changes in the brains of trained singers and non-singer controls, Waters *et al.* [8] (critical topics (i), (iii), (iv) and (v)) explore the behavioural and neural correlates of laryngeal control, and their relationship to vocal expertise. Anatomical studies have shown that the human larynx receives innervation via direct connections from the primary motor cortex to the nucleus ambiguus. This pathway, sparser in other great apes and absent in monkeys, has been hypothesized to facilitate the speed and precision of laryngeal control needed for speech and song. Compared to controls, the results show that singers modulate their speech in a vocal size and pitch imitation task more accurately, and show stronger neural representations of larynx height within a region of the right dorsal somatosensory cortex. This research hints at a common neural basis for enhanced vocal control in speech and song, and a possible common neural substrate in the somatosensory cortex.

Given this remarkable precision and complexity of human vocal control, another key question is whether listeners can detect when others modulate their voices on purpose, and what effect that might have on how those speakers are perceived. Pinheiro *et al.* [9] (critical topics (iv) and (v)) test whether authenticity mediates the affective and social impressions that we form about others in two fundamental nonverbal vocalizations: crying and laughing. The authors show that listeners can often detect if someone is producing a laugh or cry volitionally rather than spontaneously. Importantly, vocalizers producing 'genuine' spontaneous laughs and cries are perceived as relatively more trustworthy and more aroused than those producing volitional vocalizations. This study shows that, despite the impressive control that humans have over vocal output, modulated vocalizations that do not sound authentic can alter listeners' social assessments, with potentially detrimental effects for the 'faker'.

In the final contribution of Part 1 of this issue, Winter *et al.* [10] (critical topics (ii), (iii) and (iv)) rethink the frequency

code. The frequency code hypothesis posits that the cross-modal association between low pitch and large size can explain a wide range of communicative phenomena, such as vocal modulation of pitch to signal dominance and deference, or to iconically depict small or large size. In a meta-analysis of speech production experiments across multiple languages (Korean, Japanese, Chinese, Catalan, Austrian German, German, Russian), the results show that speakers in these cultures lower, rather than raise, their voice pitch when speaking to an imagined more powerful superior compared to when speaking to an imagined friend or peer. In the light of these surprising results, the authors call for a pluripotent approach to the study of the frequency code and the interpretation of pitch modulation, underscoring that a speaker's pitch modulation is often interpreted differently depending on a wide range of contextual factors.

(b) Part 2

Like Part 1, the second part of this issue begins with three review papers. Bringing together current voice research in the neuro-behavioural sciences, Scott [11] (critical topics (i), (iv) and (v)) emphasizes a need to broaden the kinds of vocal production currently under popular study. Specifically, beyond speech, the study of voice modulation in singing and nonverbal vocalizations, and how this affects listeners' perceptions of speaker identity and social traits, promises to advance current knowledge on the neural underpinnings of vocal control. This is particularly relevant because the recruitment of right hemispheric neural networks for vocal control varies across voice modulation contexts. Following in this vein, Leongómez *et al.* [12] (critical topics (i), (ii) and (iii)) propose a bold new model for the evolution of musicality embedded in the context of human voice modulation. The authors argue that the capacity to process musical information was driven largely by its role in solidifying human social relationships, such as bonds between parents and offspring. In other words, musicality was selected for in complex social environments. Notably, by focusing on musicality, the proposed model allows for the possibility that musicality may predate music. Moreover, alongside evidence that vocal cues of expressivity also drive the perception of emotion in music [5] and that pitch modulation plays a role in human mating [1], the model illustrates that musicality is important beyond the realm of music, playing a role in domains like infant-directed speech [6,13], courtship and other language abilities.

In the third and final review paper of this issue, Bryant [13] (critical topics (ii) and (iii)) examines variation and potential universals in voice modulation across human cultures. The review focuses on three key communication contexts: vocal signalling of formidability/dominance, emotion, and infant-directed speech. Far beyond a literature review, the author underscores both the importance and the methodological or technological challenges faced by researchers studying human behaviour outside of western, educated, industrialized, rich, democratic (WEIRD) societies, including the great difficulty of adapting standardized experimental protocols to diverse cultures.

While most extant research in the human voice sciences still remains limited to WEIRD societies, there exists a growing trend towards studying under-represented and small-scale societies, as highlighted in this issue [7,14,15]. This is paralleled by an increase in large-scale international collaborations that combine data from diverse samples. In just such a study,

Ćwiek *et al.* [14] (critical topics (ii) and (iii)) present the biggest test of the *bouba-kiki* effect to date, showing that across 25 languages, most cultures associate the vocal sound *bouba* with rounded shapes and *kiki* with jagged shapes. Because a key question in the evolution and structure of language is whether the sounds of words are linked to their meaning, these results suggest that cross-modal correspondences between vocal sounds and shapes could have influenced the origins of certain words. In another cross-cultural study of laughter perception, Kamiloglu *et al.* [15] (critical topics (ii) and (iii)) compared data from Dutch and Japanese listeners. Consistent with the results of Pinheiro *et al.* [9], the authors show that listeners in both cultures judge spontaneous laughter as more positive than volitional laughter. Moreover, they show that listeners can identify whether laughs are produced by someone from their own culture, showing that human nonverbal vocal cues can encode cultural identity.

Cultural cues are important in human nonverbal communication, as are vocal indices of other socially relevant traits, including gender. In a voice imitation task involving 8–10-year-old boys and girls, Cartei *et al.* [16] (critical topics (i), (ii), (iii) and (iv)) show that children spontaneously masculinize or feminize their voices by lowering or raising their voice frequencies, respectively, depending on whether they are directing speech toward same-sex peers with stereotypically masculine (rugby) or feminine (ballet) interests. Strategic social voice modulation may thus emerge early in human development, but does it change as we become older? Tuomainen *et al.*'s [17] (critical topics (i), (ii) and (iv)) research suggests that it does, specifically in the context of making ourselves better understood. In a sample of individuals aged 8–80, the authors found that older people modulate their voices more than do younger people to overcome noise, and that people, regardless of age, alter their voices more in response to background speech than to other kinds of background noise.

This theme issue leaves little doubt that humans are capable of impressive voice modulation, strategically employed for a broad range of functions from improving speech intelligibility [17] to expressing biosocial traits, such as those related to dominance, threat and formidability [7,13,18]. How might such a complex capacity have arisen in the first place? Employing a comparative framework, the three penultimate papers of this issue provide insight into the origins and evolution of vocal control. Pisanski *et al.* [18] (critical topics (i), (iii) and (v)) explore the possibility that vocal modulation for body size exaggeration through the strategic manipulation of individual vocal tract resonances (formants) may have contributed to the origins of vocal complexity. While vocal tract elongation for size exaggeration has evolved independently in several vertebrate groups, the authors employ novel voice resynthesis technology paired with psychoacoustic playback experiments to show that smaller speech-like articulatory movements that shift only one or two formants in animal and human vocalizations can serve a similar yet less energetically costly size-exaggerating function. Their results point to a potential evolutionary pathway from formant modulation for size exaggeration, to formant modulation for vowel production and ultimately, for articulated speech.

Across the animal kingdom, vocal size exaggeration is indeed achieved through a variety of evolved behavioural or anatomical adaptations, sometimes resulting in large deviations from acoustic allometry, wherein animals sound larger (or smaller) than expected given their actual body size.

Ravignani & Garcia [19] (critical topics (i) and (iii)) ask whether such deviations can be predicted by the vocal learning abilities of the species. Using a comparative phylogenetic and acoustic analysis approach, their unexpected results show that multiple species belonging to clades showing vocal production learning do indeed deviate from allometric scaling, but in the opposite direction to that expected from size exaggeration mechanisms. Importantly, their approach may nevertheless allow researchers to identify species as potential vocal production learners, wherein vocal learning is an important prerequisite for voice modulation. Clearly, understanding commonalities in the acoustics of animal vocalizations, including comparisons between humans and our closest living primate relatives, can provide insight into how speech evolved. Grawunder *et al.* [20] (critical topics (i) and (iii)) show that the acoustic parameters of chimpanzee hoos, grunts, screams and barks are sufficient to discriminate the four call types, and that discrimination is similar when only using movements of vocal articulators, coded from video data. Their results show that while chimpanzees have functional use of a relatively wide vowel space, including [u] and [a], which shows considerable overlap with that of humans, they nevertheless do not produce high front vowel-like sounds like [i] that are common in human speech. The authors argue that chimpanzee vowel space use appears larger than that of monkey species examined to date, suggesting that expansion towards a human-like vowel space continued through hominoid evolution.

Closing this theme issue with a look towards the exciting but risky future of digital voice resynthesis, Guerouaou *et al.* [21] (critical topics (iv) and (v)) address the ethics of expressive vocal 'deep-fakes'. In an experimental ethics study, participants showed a high and broad acceptability of possible applications of expressive voice transformation technologies that alter the real vocal parameters of vocalizers, with only one general exception: when consent is lacking on the part of the vocalizer. Yet while the potential societal ramifications of voice resynthesis are not yet fully known, it is clear—as exemplified by several experiments in this issue [5,8,18]—that emerging voice resynthesis technologies are now giving researchers unprecedented experimental control over vocal manipulations, with the potential to revolutionize the future of voice sciences.

3. Concluding remarks

The 21 papers that comprise this two-part issue highlight the breadth and diversity of research on voice modulation.

Beyond extensive scientific applications within the human voice sciences and animal communication more broadly, research in this area also has crucial practical applications, from societal to technological. As the contributions to this two-part issue show, dynamic voice production and perception form a regular part of our everyday communication, whether physically 'voice-to-voice' or online and virtual.

To advance our understanding of vocal communication, from origin and mechanism to social impact, a comprehensive and panoptical view of the many phenomena involved in voice production and perception is essential. We believe that this necessarily must include the study of dynamic within-individual voice modulation, not least because vocal control is a necessary prerequisite for speech. Increasing dialogue and bridging the gap between disciplines is paramount; it is only through interdisciplinary collaboration that we will come to find a common terminology and definitions, ask the right questions, elaborate adequate and viable methodologies, and see the bigger picture to, ultimately, develop and test mature and robust hypotheses that encompass the wide range of complex phenomena involved in vocal modulation. In fact, by examining the biological and behavioural mechanisms and functions of vocal control, comparing across human populations and nonhuman animal species, researchers are already beginning to reveal the astonishing ubiquity and social consequences of within-individual modulation of vocal signals, thereby providing new insight into the evolution of speech, language, song and music.

Our goal is for this thematic issue to further inspire researchers from diverse fields to look beyond traditional disciplinary boundaries, helping to cement a long-lasting, interdisciplinary foundation for the field of voice modulation.

Data accessibility. This introductory article has no additional data.

Authors' contributions. J.D.L. and K.P. proposed the article concept and drafted the manuscript. All authors contributed towards editing and revising the final version.

Competing interests. The authors declare no competing interests.

Funding. J.D.L. was supported by Universidad El Bosque, Vice-rectory of Research (grant no. PCI.2015-8207). K.P. and D.R. were supported by the University of Lyon IDEXLYON project as part of the 'Programme Investissements d'Avenir' (ANR-16-IDEX-0005) to D.R. D.S. is supported by the ERC Starting (grant no. 714977). N.L. is supported by a Sir Henry Wellcome Fellowship (grant no. 220448/Z/20/Z).

Acknowledgements. We express our sincere gratitude to the authors and reviewers that contributed to this theme issue, and are extremely grateful to Editor Helen Eaton for her immense support throughout the entire process.

Guest Editor biographies



Juan David Leongómez is an associate professor and researcher at the Human Behaviour Laboratory (LACH), Faculty of Psychology, Universidad El Bosque in Bogota, Colombia. He holds a Bachelor's in Music Education from Universidad Pedagógica Nacional, Colombia, an MSc in Evolutionary Psychology from the University of Liverpool, UK, and a PhD from the University of Stirling, UK. He studies the evolutionary underpinnings of human behaviour, particularly those related to mate choice and the social effects of vocal signals, with a particular interest in the evolution of musicality. Juan David published among the first articles showing within-individual changes in voice pitch in response to the social status of the listener and has demonstrated strong effects of voice modulation on listeners in courtship contexts. He also supports the dissemination of open science practices and the use of opensource software and programming as tools to promote transparency and reproducibility, as well as to reduce national inequalities in production of, and access to, scientific knowledge.



Katarzyna (Kasia) Pisanski is a permanent CNRS researcher (*National Centre for Scientific Research*) affiliated with the University of Lyon and University Jean Monnet Saint Etienne, France, and working in collaboration with the University of Wrocław, Poland. Having obtained her PhD in 2014 from McMaster University, Canada, she now leads a successful research programme on human and non-human animal behaviour with a focus on acoustic communication. Kasia employs a multidisciplinary comparative and experimental approach to study the origins, ontogeny, mechanisms and evolved social functions of nonverbal voice production and perception. Her current research examines the roles of voice modulation in human nonverbal vocalizations (such as laughter and cries) and in speech, from shaping social interactions to providing unique insight into the evolution of acoustic communication in animals, including the *human animal*.



David Reby studies the structure, function and evolution of vocal signals in vertebrates. While well known for his seminal work on size communication and exaggeration in mammal vocal signals, David also leads a successful research programme on human voice modulation examining the development and voluntary expression of gender, and the functions of human nonverbal vocalizations. He has published over 100 papers in high-impact journals on the topic of animal and human vocal communication. He has acted as an Academic Editor for *PLoS One* for 14 years and has been Associate Editor for *Bioacoustics* since March 2020. David is a senior member of the *Institut Universitaire de France* and an elected member of the *French National Committee for Scientific Research*.



Disa Sauter is an associate professor of psychology at the University of Amsterdam, The Netherlands. She did her BSc in psychology and cognitive science at University College London in 2002, followed by a PhD in the same department in 2006. She did post-docs at Kings College London, Birkbeck College, and the Max Planck Institute for Psycholinguistics. She studies human vocal communication of emotion within the scope of her broad research programme on emotion. Disa has been Associate Editor for *Emotion Review* since 2014, is consulting editor at *Psychological Review*, and a member of the Editorial Boards of *Affective Science* and the *Journal of Nonverbal Behaviour*.



Nadine Lavan is currently a Sir Henry Wellcome Fellow at the Department of Biological and Experimental Psychology at Queen Mary University of London and obtained her PhD in psychology from the Royal Holloway University of London in 2017. Her research has examined the cognitive and neural mechanisms driving the perception of voices, including voice identity, voice imitation and vocal emotions. Nadine's current research is investigating how we perceive person characteristics, such as someone's age, regional origin and identity from their voice and their face.



Marcus Perlman's research investigates the evolution of language, with special interest in iconicity in speech, nonverbal vocalization and gesture. He has also conducted research on the volitional vocal control of nonhuman primates (the gorilla Koko), raising important questions about the role of vocal control in the evolution of language. He has been an Academic Editor for *PLoS One* for 3 years, and recently co-edited a special double issue for *Language and Cognition* on empirical approaches to studying iconicity in spoken and signed languages.



Jaroslava Varella Valentova is a professor at the Department of Experimental Psychology, University of Sao Paulo, Brazil, and obtained her PhD in anthropology at the Charles University, Czech Republic. Her research takes a unique cross-cultural approach to human nonverbal communication, focusing on its evolution, sociocultural functions and impacts, particularly in the context of sexual behaviour and sexual orientation. She has co-edited a book (*Handbook of Evolutionary Psychology*, 2018), guest edited a special issue in the *Human Ethology Bulletin* (2015) and she is guest editing a Research Topic about a 150 years' celebration of Darwin's book on human evolution and sexual selection (2022).

References

- Hughes SM, Puts DA. 2021 Vocal modulation in human mating and competition. *Phil. Trans. R. Soc. B* **376**, 20200388. (doi:10.1098/rstb.2020.0388)
- Belyk M, Eichert N, McGettigan C. 2021 A dual larynx motor networks hypothesis. *Phil. Trans. R. Soc. B* **376**, 20200392. (doi:10.1098/rstb.2020.0392)
- Matzinger T, Fitch WT. 2021 Voice modulatory cues to structure across languages and species. *Phil. Trans. R. Soc. B* **376**, 20200393. (doi:10.1098/rstb.2020.0393)
- Torres Borda L, Jadoul Y, Rasilo H, Salazar Casals A, Ravnignani A. 2021 Vocal plasticity in harbour seal pups. *Phil. Trans. R. Soc. B* **376**, 20200456. (doi:10.1098/rstb.2020.0456)
- Bedoya D, Arias P, Rachman L, Liuni M, Canonne C, Goupil L, Aucouturier J-J. 2021 Even violins can cry: specifically vocal emotional behaviours also drive the perception of emotions in non-vocal music. *Phil. Trans. R. Soc. B* **376**, 20200396. (doi:10.1098/rstb.2020.0396)
- Yan R, Jessani G, Spelke ES, de Villiers P, de Villiers J, Mehr SA. 2021 Across demographics and recent history, most parents sing to their infants and toddlers daily. *Phil. Trans. R. Soc. B* **376**, 20210089. (doi:10.1098/rstb.2021.0089)
- Kleisner K *et al.* 2021 Predicting strength from aggressive vocalizations versus speech in African bushland and urban communities. *Phil. Trans. R. Soc. B* **376**, 20200403. (doi:10.1098/rstb.2020.0403)
- Waters S, Kanber E, Lavan N, Belyk M, Carey D, Cartei V, Lally C, Miquel M, McGettigan C. 2021 Singers show enhanced performance and neural representation of vocal imitation. *Phil. Trans. R. Soc. B* **376**, 20200399. (doi:10.1098/rstb.2020.0399)
- Pinheiro AP, Anikin A, Conde T, Sarzedas J, Chen S, Scott SK, Lima CF. 2021 Emotional authenticity modulates affective and social trait inferences from voices. *Phil. Trans. R. Soc. B* **376**, 20200402. (doi:10.1098/rstb.2020.0402)
- Winter B, Oh GE, Hübscher I, Idemaru K, Brown L, Prieto P, Grawunder S. 2021 Rethinking the frequency code: a meta-analytic review of the role of acoustic body size in communicative phenomena. *Phil. Trans. R. Soc. B* **376**, 20200400. (doi:10.1098/rstb.2020.0400)
- Scott SK. 2021 The neural control of volitional vocal production: from speech to identity, from social meaning to song. *Phil. Trans. R. Soc. B* **376**, 20200395. (doi:10.1098/rstb.2020.0395)
- Leongómez JD, Havlicek J, Roberts SC. 2021 Musicality in human vocal communication: an evolutionary perspective. *Phil. Trans. R. Soc. B* **376**, 20200391. (doi:10.1098/rstb.2020.0391)
- Bryant GA. 2021 Vocal communication across cultures: theoretical and methodological issues. *Phil. Trans. R. Soc. B* **376**, 20200387. (doi:10.1098/rstb.2020.0387)
- Ćwiek A *et al.* 2021 The *bouba/kiki* effect is robust across cultures and writing systems. *Phil. Trans. R. Soc. B* **376**, 20200390. (doi:10.1098/rstb.2020.0390)
- Kamiloglu RG, Tanaka A, Scott SK, Sauter DA. 2021 Perception of group membership from spontaneous and volitional laughter. *Phil. Trans. R. Soc. B* **376**, 20200404. (doi:10.1098/rstb.2020.0404)
- Cartei V, Reby D, Garnham A, Oakhill J, Banerjee R. 2021 Peer audience effects on children's vocal masculinity and femininity. *Phil. Trans. R. Soc. B* **376**, 20200397. (doi:10.1098/rstb.2020.0397)
- Tuomainen O, Taschenberger L, Rosen S, Hazan V. 2021 Speech modifications in interactive speech: effects of age, sex and noise type. *Phil. Trans. R. Soc. B* **376**, 20200398. (doi:10.1098/rstb.2020.0398)
- Pisanski K, Anikin A, Reby D. 2021 Vocal size exaggeration may have contributed to the origins of vocalic complexity. *Phil. Trans. R. Soc. B* **376**, 20200401. (doi:10.1098/rstb.2020.0401)
- Ravnignani A, Garcia M. 2021 A cross-species framework to identify vocal learning abilities in mammals. *Phil. Trans. R. Soc. B* **376**, 20200394. (doi:10.1098/rstb.2020.0394)
- Grawunder S, Uomini N, Samuni L, Bortolato T, Girard-Buttoz C, Wittig RM, Crockford C. 2021 Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Phil. Trans. R. Soc. B* **376**, 20200455. (doi:10.1098/rstb.2020.0455)
- Gueraouaou N, Vaiva G, Aucouturier J-J. 2021 The shallow of your smile: the ethics of expressive vocal deep-fakes. *Phil. Trans. R. Soc. B* **376**, 20210083. (doi:10.1098/rstb.2021.0083)