



# A cross-species framework for classifying sound-movement couplings

Silvia Leonetti<sup>a,b,c,\*</sup>, Andrea Ravignani<sup>b,c,d,1</sup>, Wim Pouw<sup>e,\*\*,1</sup>

<sup>a</sup> Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, Turin 10123, Italy

<sup>b</sup> Department of Human Neurosciences, Sapienza University of Rome, Piazzale Aldo Moro 5, Rome 00185, Italy

<sup>c</sup> Comparative Bioacoustics Research Group, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, the Netherlands

<sup>d</sup> Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Aarhus C 8000, Denmark

<sup>e</sup> Donders Institute for Brain, Cognition, and Behavior, Radboud University Nijmegen, Houtlaan 4, Nijmegen 6525 XZ, the Netherlands

## ARTICLE INFO

### Keywords:

Comparative biology  
Evolutionary behavioral neuroscience  
Multimodal communication  
Vocalizations  
Biomechanics

## ABSTRACT

Sound and movement are entangled in animal communication. This is obviously true in the case of *sound-constituting* vibratory movements of biological structures which generate acoustic waves. A little less obvious is that other moving structures produce the energy required to sustain these vibrations. In many species, the respiratory system moves to generate the expiratory flow which powers the sound-constituting movements (*sound-powering movements*). The sound may acquire additional structure via upper tract movements, such as articulatory movements or head raising (*sound-filtering movements*). Some movements are not necessary for sound production, but when produced, impinge on the sound-producing process due to weak biomechanical coupling with body parts (e.g., respiratory system) that are necessary for sound production (*sound-impinging movements*). Animals also produce sounds *contingent* with movement, requiring neuro-physiological control regimes allowing to flexibly couple movements to a produced sound, or coupling movements to a perceived external sound (*sound-contingent movement*). Here, we compare and classify the variety of ways sound and movements are coupled in animal communication; our proposed framework should help structure previous and future studies on this topic.

## 1. Introduction

Laryngeal vocal folds in many mammals, including humans (*Homo sapiens*, henceforth humans), amphibians, and reptiles, are moved to produce sound. The syringeal vocal labia are governed by an analogous mechanism evolved in birds (Colafrancesco and Gridi-Papp, 2016; Fitch and Suthers, 2016; Herbst, 2016; Riede and Goller, 2010). The chest wall of penguins contracts vigorously, generating an air flow for their chest-vibrating vocalizations (Bustamante and Márquez, 1996; Favaro et al., 2015). Frogs, and many primates, such as siamangs (*Symphalangus syndactylus*), modulate frequency and/or amplitude of their vocal emission by inflating and deflating a vocal or laryngeal air sac (Burchardt et al., 2024; de Boer, 2009; Dudley and Rand, 1991; Riede et al., 2008). Complex movement of the larynx, beak, and neck of many birds, such as parrots, play a crucial role in the shaping of vocal emissions (Homerberger, 2017). Many marine mammals, such as northern elephant seals (*Mirounga angustirostris*), raise their heads to acoustically shape their deep-roar vocalizations (Frey and Gebler, 2010). Bats beat

their wings in synchrony with echo-localization vocal bursts (Lancaster et al., 1995). Chimpanzees (*Pan troglodytes*) accompany their pant hoots with vigorous whole-body movements (Tagliatela et al., 2015). Finally, humans dance in synchrony with instrumental and vocal music and in synchrony with others (Brown et al., 2006). The many cross-species examples of coupling of sound and movement suggest a special combination.

While there is already a rich literature concerned with how animals functionally combine ‘modalities’ in multimodal communication (Halfwerk et al., 2019; Partan and Marler, 1999; Pouw et al., 2021), none of those reviews single out and classify the special connection between sound and movement. The general understanding of multimodal communication is that different modes (like visual and sound production) are separate systems that are perceptually bound together to enhance or change the perceiver’s experience (e.g., Partan and Marler, 1999). However, movement and sound are often physically linked during their *production*, or originate from such links, creating a natural overlap between these ‘channels’ since they are produced together. Not

\* Corresponding author at: Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, Turin 10123, Italy.

\*\* Corresponding author.

E-mail addresses: [silvia.leonetti@unito.it](mailto:silvia.leonetti@unito.it) (S. Leonetti), [wim.pouw@donders.ru.nl](mailto:wim.pouw@donders.ru.nl) (W. Pouw).

<sup>1</sup> Co-senior authors.

all movement and sound are directly coupled of course, and some sound-movement couplings only exist in virtue of sustained support of the social-cultural and material environment (Heyes, 2018; Malafouris, 2016), such as musical instrument playing in humans.

How to make sense of what seems to be fundamentally different sound-movement couplings? This is the overarching topic of this paper. We will argue that movement-sound couplings can be sorted along a continuous classification of sound-movement relations. Between each conceptual class, no big conceptual break seems apparent, but when comparing the first and last classes, they seem to be a gulf apart. We will discuss the different sound-movement coupling classes in turn and in isolation while highlighting how they are combined with other classes. Communication therefore involves a nested set of sound-movement couplings (cf. Kelso and Tuller, 1984a). This forms a multi-class coalition of mechanically necessary, mechanically optional, and/or mechanically disconnected but primarily neurally coupled sounds and movement (see Fig. 1). As such, when it comes to communicative sound production, it is movement all the way down.

### 1.1. Overview

The review of the literature that we have conducted serves to exemplify the diversity of phenomena that fall under a particular class as proposed by our framework. The review is thereby by no means exhaustive, and no systematic search has been performed to serve such a goal. We do think that each class deserves its own systematic review, and we hope that our framework invites more targeted investigation of a particular sound-movement class, each of which can be the basis of a full-length systematic review. Thus the main goal of our review serves to exemplify the cross-species phenomena that fall under the sound-movement coupling classes that we propose.

Table 1 summarizes the classes of sound-movement coupling in the order reviewed here. Firstly, sound is often constituted by small oscillatory movements, i.e., vibrations, of biological structures (Section 2.1). Vibration needs to be generated by energy produced by sound-powering movements which enable and modulate sound production. For example, air flow is generated by movements of the respiratory systems, like the movement of the rib cage in mammals, or the movement of surrounding tissues of the air sacs in birds (Section 2.2). Other movements are used to filter the sound source, including articulatory movements and body postures, such as head raising, that dynamically modify the vocal tract (Section 2.3). A class of optional mechanical couplings connects sound production to peripheral bodily movements (Section 2.4). Some movement and sound couplings are primarily non-mechanical and more flexible than the previous classes (Section 2.5). Finally, movements can be coupled to external sounds (Section 2.6), and multiple external agents (Section 2.7). Fig. 1 exemplifies how all these 7 classes of sound-movement coupling can be present in a common human activity: joint singing.

## 2. The framework

### 2.1. Intrinsic sound-constituting movements

In this section, we overview key structures and mechanisms involved

in the production of sound-constituting movements in different species. The biophysics of sound production (and perception) very often<sup>1</sup> involves vibrating body parts of the producer (and the perceiver) of sound. As such, sound is a phenomenon that can be thought of as a system of coupling between sound-constitutive movements (sound source) and audition-constituting movements of the sound organs (e.g., cilia movements), both of which are physically coupled to the medium. As such, animals literally move to create motions in others.

#### 2.1.1. Sound-constituting movements through myoelastic aerodynamics (MEAD)

To entrain a medium to carry sound waves, terrestrial and semi-aquatic mammals (including humans), and many amphibians and reptiles, vibrate components of the larynx known as the vocal folds (Colafrancesco and Gridi-Papp, 2016; Fitch and Suthers, 2016; Herbst, 2016; Titze, 2008; Titze and Martin, 1998); for an overview see Ladich and Winkler (2017). The larynx shows much anatomical diversity across species, and it allows for the production of a variety of sounds, from infrasound up to ultrasound. In general, the larynx is a movable structure that is located at the top of the trachea; here phonation (Loucks et al., 2007; Raphael et al., 2007) occurs when an air flow generates a self-organized oscillatory movement upon meeting closed (adducted) and tensioned vocal folds. It is the laryngeal muscles that close (adduct) and open (abduct) the space between the vocal folds. The phonation process, at the level of vocal folds, is divided into two interdependent phases: “vocal fold posturing” and “vocal fold vibration” (Hunter et al., 2004). The first phase refers to slow deformations of the vocal cords (abduction/adduction and elongation) caused by the acting laryngeal muscles. These tonic muscle tensions occur recurrently, but not necessarily periodically, in human speech, occurring at frequencies lower than 10 Hz. The second phase concerns the oscillation of the vocal folds, characterized by small and relatively rapid deformations. This oscillation is caused by the air coming from the lungs, during the exhalation, which passes through the vocal folds. This vibration phase occurs at about 100 Hz or more in humans (Hunter et al., 2004).

As described in the MyoElastic AeroDynamic<sup>2</sup> (MEAD) theory (see Švec et al., 2023) these vibrations emerge or ‘self-organize’ from biophysical principles and do not require constant neuro-muscular control of vibrating motions of the vocal folds themselves (Herbst, 2016; Herbst et al., 2023). When energy is provided in the form of a sustained air flow, a pressure builds when meeting an obstructing barrier like the adducted vocal folds. Due to a pressure buildup below the glottis, the vocal folds are blown open, and the internal-external pressure gradient normalizes. As the pressure dissipates, the elastic restoring forces of the laryngeal tissues close the glottis again, and pressure builds up once again if expiratory air flow is sustained, restarting the cycle. This cyclical process of vocal fold oscillations is thus dependent on the powering air flow interacting with the elastic properties of the vocal folds and connective tissues, as well as the muscles (e.g., cricoarytenoid) controlling their tensioning (DeJonckere et al., 2024; Fitch and Suthers, 2016; Herbst, 2016; Herbst et al., 2023; Ladich and Winkler, 2017; Taylor et al., 2016; Titze, 2008; Zhang, 2016). Interestingly, the unusually low-frequency purrs observed in domestic cats (*Felis silvestris catus*) were always held to originate from active rather than passive vibrations. However, it has

<sup>1</sup> Whistling is a sound produced by the turbulent dynamics of air flows, rather than by vibrating organs like vocal folds. When whistling, air is forced through a small opening, such as the lips, creating a stream of air. This stream of air interacts with the edges of the opening, causing the air particles in the connected cavity to oscillate. This system functions as a Helmholtz oscillator, where the cavity of air acts like a spring-mass system. The oscillating air in the cavity generates sound waves in a manner similar to how vocal folds produce sound by vibration (or by making a spring-like motion), thus constituting sound.

<sup>2</sup> Note that Myo refers to muscle tissues.



**Fig. 1.** The 7 classes of sound-movement coupling: An example from joint singing in humans. Note. The figure illustrates how different classes of movements-sound couplings are combined in humans (Pexels, free license). In the figure, we use section numbering to refer to each class. 2.1) Intrinsic sound-constituting movements: Sound production is based on vocal-fold oscillations (small amplitude cyclical movements) (Wikipedia, free license). 2.2) Intrinsic sound-filtering movements: The air forced along the vocal cords makes them vibrate and this is delivered by movements of the surrounding structures of the lungs (e.g., rib cage). 2.3) Intrinsic sound-impinging movements: The articulatory movements of the jaw (as well as the tongue and lips) influence the frequency components of the sounds produced. Raising the head while singing alters the conformation of the vocal tract, therefore the frequency components of the sound emitted. 2.4) Intrinsic sound-powering movements: The limb movements, although not directly necessary for vocal production, can modulate the expiratory flow and therefore the vocal emission, through a weak limb-respiratory-vocal biomechanical coupling. 2.5) Intrinsic sound-contingent movements: The movement and posturing of the hand in relation to the linguistic meaning of the sounds, through neural coupling, plays a fundamental role in supporting the singers' semantic-emotional communication. Each singer reciprocally synchronizes the tempo of their vocal emissions with the tempo of their hands clapping. 2.6) Extrinsic sound-contingent movements: Each singer claps their hands entraining the rhythm of others' vocal emissions. 2.7) Mutual sound-contingent coupling: All singers coordinate both the tempo of their vocal emission and hand clapping with the tempo of others' vocal emission and hand clapping.

recently been shown that the same MEAD-based mechanisms play a role in purring (Herbst et al., 2023).

Vocal sound production in birds is also based on biophysical MEAD principles similar to laryngeal vocalizations; unlike mammals, however, birds' sound source is characterized by specialized syringeal membranes located at the branching of the trachea (Elemans et al., 2015; Goller and Larsen, 1997; Kriesell et al., 2020; Riede and Goller, 2010; Švec et al., 2023; Van den Berg, 1958). Among different bird families, there are substantial morphological and functional variations, in the cartilages, associated muscles and labia in the syrinx, responsible for typical vocal repertoires of each species. Interestingly, some bird taxa, such as songbirds, show a 'structurally bipartite' syrinx: its left and right sides constitute two independently controlled sound generators that can produce two different sounds (Brown and Riede, 2017; Riede and Goller, 2010; Suthers, 1990).

Even marine mammals produce sound in accordance with the MEAD theory (Madsen et al., 2023; Ravignani and Herbst, 2023). This similarity to land mammals is remarkable, given the different (re-)adaptations of their respiratory and sound-producing systems needed for underwater life. Among cetaceans, odontocetes, such as atlantic bottlenose dolphins (*Tursiops truncatus*) and killer whales (*Orcinus orca*), produce a wide range of sounds carried through self-sustaining vibrations of a head structure called a melon, blowing air through a nasal structure and making phonic lips in the nose vibrate (Madsen et al., 2023; Ravignani and Herbst, 2023). However, a recent study has found laryngeal structures, as a sound source, regulated by the same myoelastic-aerodynamic principles, in mysticetes, such as humpback whales (*Megaptera novaeangliae*; (Elemans et al., 2024). In other words, while some cetaceans and mysticetes phonate with a laryngeal MEAD mechanism, others evolved de-novo a non-laryngeal MEAD sound source. Also in the last group of fully-aquatic mammals, the sirenians (manatees and dugongs), the sound source is presumed to be the vocal

cords, located at the level of the larynx (Landrau-giovannetti et al., 2014) and once the sound is produced, it then resonates in the nasal region.

### 2.1.2. Sonating sound-constituting movements

Other mechanisms of non-vocal sound production exist, involving the vibration of a body part or external surface. These mechanisms are typically called "sonation", which includes: stridulation, aeroelastic flutter, and percussion (Clark, 2016; Clark and Feo, 2008).

Various lines of fish show an analogous gas-containing structure - the swim bladder. Sonic muscles, connected to the swim bladder directly deform it and act on it like a drum, where the walls of the swim bladder vibrate, generating a sound wave that propagates in the water (Bass and Chagnaud, 2012; Fine et al., 2002; Fine and Parmentier, 2015). Alternatively, the swim bladder can function as an underwater resonant bubble that passively radiates sounds produced by other organs such as the pharyngeal teeth and pectoral girdle (Barimo and Fine, 1998).

Stridulation is defined as the rubbing of one or more parts of the body against a rigid surface; such surface can be external or internal to the body. Stridulation is widespread in the animal kingdom (Fine and Parmentier, 2015). Here, we cite some explanatory examples. Among vertebrates, the streaked tenrec (*Hemicentetes semispinosus*) displays a characteristic stridulation behavior. The mid-dorsal region of this mammal contains keratinous quills. The step lines of these quills, when rubbed against each other, vibrate and produce a sound (Clark, 2016; Endo et al., 2010). In birds, the males of club-winged manakin (*Machaeropterus deliciosus*) move the wings high above their back, bringing them together to cause repeated collisions of the secondary wing feathers of both wings. One feather of the secondary wing of one side presents a structure, called the pick, that collides with the feather file on the secondary wing of the opposite side. This stridulation produces a loud tonal sound (Clark, 2016; Gómez-Bahamón et al., 2020).

**Table 1**  
Framework for the variety of sound-movement couplings.

Class & Definition	Explanation of definition	Examples
Intrinsic sound-constituting movements. <i>Movement (hereafter: M) is constitutive of sound (hereafter S), if the phenomenon of S would not exist without M.</i>	A vocal sound exists because the vibrations of the vocal folds create a propagating acoustic wave.	Humans (1) produce in-air <i>vocal sounds</i> because the vocal folds (in the larynx) oscillate. The oblique-lined tiger beetle produces sound by a file and scraper stridulation mechanism (2).
Intrinsic sound-powering movements. <i>M is powering S, if an instance of S is enabled by M to occur.</i>	Vocal sound-constituting movements require an energy source. This energy can be supplied by respiratory movements generating an air flow that powers vibrations.	In humans, during each respiratory phase, the respiratory system pushes air along the vocal folds (3). In the gulf coast toad, the air flow is pushed towards the vocal folds thanks to the cooperation of air sac movements and air-recycling systems, independently of breathing cycles (4).
Intrinsic sound-filtering movements. <i>M is filtering S, if an instance of S is modulated/attenuated in structure depending on M.</i>	In vocal sound production, articulatory movements of the body, and body posture changes, contribute to the modulation of emitted sound acoustic variables.	In humans, movements of the articulators (lips, tongue, jaws) during vocalization, alter particular frequency components of the sounds produced (5). The white-throated sparrow opens and closes the beak during vocal production. The position of the beak affects the acoustic properties of the sound, acting as an 'audio equalizer' for a sound generated elsewhere in the body (6).
Intrinsic sound-impinging movements <i>M is impinging on S, if another S–M relation is altered through mechanical interactions</i>	Movements that are not necessarily involved in the production of sound, but when produced affect the sound-producing process.	Humans couple movement of their upper limbs with sustained vocalizations such as in singing, recruiting a set of muscles that influences expiratory flow and therefore the amplitude and pitch of their voice (7). The greater spear-nosed bat couples ultrasound emission with locomotion and respiration movements (8).
Intrinsic sound-contingent movements. <i>M is contingent with S, if M covaries with S without mechanical necessity or interaction.</i>	Neural coupling between body parts that are not mechanically linked to move together.	In dialogue, humans temporally synchronize vocal emissions with semantically laden arm and finger postures guiding the interlocutor's attention (9). The superb lyrebird couples vocal emissions with leaps or wing movements through perceptual-motor neural links, during courtship displays (10).
Extrinsic sound-contingent movements. <i>M is contingent with S, if M covaries with S without mechanical necessity.</i>	Movements that couple with a unresponsive external sound.	Humans dance, coupling body movements to predictable external sound sequences (11). In an experimental setup, an eleonora cockatoo

**Table 1 (continued)**

Class & Definition	Explanation of definition	Examples
Mutual sound-contingent coupling. <i>A M sequence is mutually contingent on a S sequence if there is covariance between S and/or M sequences with reciprocal influence.</i>	Movement and sound couplings between external responsive agents, (and external unresponsive sounds).	synchronizes head movements to a series of external rhythmic sounds (12). Humans synchronize dance steps and singing voices in musical shows or choirs (13). The fiddler crab engages in "Mutual Social Entrainment", coordinating waving movements during courtship displays between individuals (14).

1. *H.sapiens*, (LaMar, n.d.); 2. *S.sasin*, (Clark et al., 2011); 3. *H. sapiens*, (Tong and Sataloff, 2022); 4. *Bufo valliceps*, (Gans and Maderson, 1973; Ladich and Winkler, 2017); 5. *H.sapiens*, (MacNeillage, 2010); 6. *Zonotrichia albicollis*, (Hoesel et al., 2000); 7. *H.sapiens*, (Pearson and Pouw, 2022; Pouw et al., 2020b; Werner et al., 2024); 8. *Phyllostomus hastatus*, (Suthers et al., 1972); 9. *H. sapiens*, (Feyereisen, 2017); 10. *Menura novaehollandiae*, (Dalziel et al., 2013); 11. *H. sapiens*, (Mårup et al., 2022); 12. *Cacatua galerita eleonora*, (Patel et al., 2009); 13. *H. sapiens*, (Müller et al., 2021); 14. *Uca annulipes*, (Backwell et al., 1998).

Stridulation is very common in invertebrates. A wide variety of insects show a common file and scraper stridulation mechanism (Masters, 1979).

Aeroelastic flutter displays have been extensively described in birds. Hummingbirds, especially, use their feathers to produce a communicative “trill”. This is a typical tonal sound, audible to the conspecifics, where the feathers vibrate at a frequency that depends on the morphology and stiffness of the feathers and the speed of the wingbeats (Clark et al., 2011; Clark and Feo, 2008).

Finally, many species show communicative percussive behaviors using body parts or external objects (including snapping with flippers, clapping and drumming using hands or feet). Such percussion produces communicative seismic and non-vocal sounds that propagate both in-air and through the substrate (Hill, 2001; Randall, 2001). Some mammals, such as rabbits and rodents, perform foot drumming (Giannoni et al., 1997; Rado et al., 1987; Randall, 2001). Non-human primates drum using their hands (Dufour et al., 2015) or on structures external to the body (Ravignani et al., 2013) or using artificial tools (Remedios et al., 2009). In a way, the rapid tongue clicks, which some bats (genus *Rousettus*) produce to echolocate is also a percussive behavior (Yovel et al., 2011); see Rojas et al. (2009) for similar approaches in human echolocation). Many amphibians use their body surfaces to produce vibrations. The impulsive force of their movements produces water surface waves that propagate in all directions (Narins, 1990). Birds, such as male red-bellied woodpeckers (*Melanerpes carolinus*) tap a substrate with their beaks (Garcia et al., 2020; Wilkins and Ritchison, 1999). Among aquatic mammals, male walruses produce intense percussion through collisions of their front flippers in the water (Larsen and Reichmuth, 2021), and harbor seals sonorously slap both the water and their own body (Kocsis et al., 2024). Finally, invertebrates such as jumping spiders drum too, as a signal to attract potential mates (Lim and Li, 2004).

2.1.3. Concluding remarks

Animals vocalize, sonate, stridulate, and beat and drum, making use of the resonant properties of body-internal or body-peripheral structures. Strikingly, despite all this variety, the MEAD principle reappears in a wide variety of convergently evolved mechanisms. Despite all the different mechanisms, - in the most general characterization possible - we can say that animals communicate acoustically by moving one another in some way.



## 2.2. Intrinsic sound-powering movements

The syrinx and larynx are considered *passive* vibrating organs. This is because other movements are often needed to provide the energy to induce such vibrations, for example in the form of an air flow that powers these vibrations. Such sound-powering movements are also important modulators of sound production as they control the intensity of the energy gradient and therefore the properties of the vibrations themselves. In this section, we examine the complex and coordinated series of sound-powering movements, focusing on respiratory movements (and thus ignoring other types of movements that can power vibrations, e.g., upper limb kinematics involved in hand clapping or drumming).

Respiratory muscles in land-dwelling animals originally evolved to generate an air flow to promote gas exchange, but are also often functioning as sound-powering movements. In terrestrial and semiaquatic mammals, lungs inflate and deflate during breathing, following the movement of the surrounding structures (rib cage, diaphragm and abdominal wall). The lungs move along with the surrounding structures because of a fluid-filled vacuum existing between the surface of the lungs and its surroundings. During each respiratory phase (exhalation and inhalation), the movement of the lungs decreases or increases the pressure of air in the lungs. An air pressure gradient is thereby created between the pressure of the external atmosphere and the lungs of the animal. The air then moves along this gradient which will equalize the internal-external air pressure, thereby dissipating the energy gradient. As discussed in [Section 2.1](#), vibrations responsible for sound production are dependent on this air flow.

The sound production process is a chained transfer of energy, from lung movement (mechanical energy) to an air pressure gradient (aerodynamic energy), to vibration (mechanical energy), to another air pressure gradient known as sound waves (acoustic energy). Interestingly, in human phonation, the rate of conversion of aerodynamic energy into an audible sound, defined as "glottis efficiency", is rather low, of the order of 0.0001–1 %. Although more in-depth studies are needed to better define "global vocal efficiency", it seems that the human body may be inefficient as a voice producer ([Titze, 1992](#); [Zhang, 2016](#)).

While the sound-powering movements may be produced by the respiratory muscles, there is a clear coordinative change from passive respiration to speaking. Quiet breathing entails rhythmic inspiratory/expiratory intervals further characterized by the use of a low range of the tidal volume of the lungs ([Seikel et al., 2023](#)). Conversely, quick but deeper inhalations, followed by a long expiratory phase, power speaking. This expiratory phase features small and fast-scale deviations of increased or decreased expiratory drive that deviate from this general slow expiratory phase; such sudden small increases in expiratory drive can affect intensity and the fundamental frequency in speech prosody, for example during emphasis of a particular word or syllable ([Petroni et al., 2017](#); [Pouw et al., 2020a](#)). During speech, subtle modulations of expiratory flow are required to control the intensity and other prosodic variables ([Petroni et al., 2017](#)), and respiration shows distinct dynamics in performing fricatives, plosives, and other key phonological units. Multiple nested respiratory movements, therefore, underlie speaking ([Fuchs et al., 2019](#); [Fuchs and Rochet-Capellan, 2021](#); [Sundberg et al., 1993](#)); they do not simply provide a steady air flow, but are an important part of the linguistic work cycle and modulate speech information structure ([Fuchs and Rochet-Capellan, 2021](#)).

Since speech uses more tidal volume, the rib cage's elastic recoil becomes stronger ([Seikel et al., 2023](#)). This means more complex neuromuscular control, involving many secondary respiratory muscles around the trunk ([Seikel et al., 2023](#)). Speech can therefore be seen as changing the stable routines of the respiratory system as it dramatically alters its dynamics for speech to function. Such increased respiratory flexibility has been suggested to explain why only anatomically modern humans and Neanderthals have been found to have unusually enlarged vertebral canals at the level where nerves innervate the lungs, which has

been held to be evidence for increased neuromuscular control for complex vocal productions ([MacLarnon and Hewitt, 1999](#)).

Primates also show a certain degree of control of air flow during phonation and this has an impact on acoustic variables. Squirrel monkeys (*Saimiri sciureus*) selectively modulate degree, latency and amplitude of expiratory movements producing a vast vocal repertoire. Specifically, the tensioning of the abdominal muscles, and consequently the subglottic pressures, influence amplitude and frequency of vocalizations ([Häusler, 2000](#)).

In birds, coordinated control of respiratory movements plays a crucial role in ensuring sound production. In canaries and many other bird species, songs consist of sequences of "notes", each produced in conjunction with expiratory "mini-breaths" ([Calder, 1970](#)). The respiratory system of birds has several specializations that increase the efficiency of sound production. In addition to paired lungs, birds have a series of posterior and anterior air sacs that work as bellows, expanding and contracting to ensure controlled movement of gas across the exchange surfaces of the lungs ([Akester, 1960](#); [Duncker, 1974](#)); for an overview see [Farmer \(2015\)](#); [Maina \(2002\)](#). Contraction of the abdominal muscles leads to compression of the abdominal air pocket. This compression forces the movement of air, which travels through the respiratory tract, meets the membranes of the syringe, causing them to vibrate.

There are also sound-powering movements that are not related to respiratory functioning. Fully aquatic mammals and amphibians have evolved sound production mechanisms based on air flow but independent of breathing. In addition to the lungs, air sacs and air-recycling systems cooperate to push air across the sound source, independently of breathing cycles, allowing these species to vocalize while avoiding trips to the surface ([Ladich and Winkler, 2017](#)).

### 2.2.1. Concluding remarks

In this section have added a second layer of nested movements that contribute to sound production in animals. Across species, several strategies have been developed to improve vocal efficiency, (defined as the amount of energy that is transferred from the air flow to the sound source). To this end, in many species, the respiratory and the vocal system cooperate and overlap their structures, functions and mechanisms. Specifically, respiratory-related muscles must act in synergy with laryngeal muscles that tune the sound-constituting vibrations. The role of the respiratory system as a sound-powering system is a prime example of exaptation, where traits originally selected for a particular function start to enhance fitness of the animal due to another function ([Gould and Vrba, 1982](#); [Larson, 2013](#)). The staggering variety of evolutionary innovations for sound-powering movement suggests that powering should not be understood as an on/off condition. Sound-powering movements are complex modulators of acoustic features that determine biological sound structures since they are continuously coupled to sound-constituting movements.

## 2.3. Intrinsic sound-filtering movements

We have examined the complex coordination of vibrating organs with the movements that power and modulate them. A third set of movements and postures affect the vocal sound production process through filtering. Some of these sound-filtering movements are articulatory and motor adaptations specifically developed to shape the vocal tract, while others are physical activities that may serve other functions such as visual signaling too, but that constrain or have been "extra-purposed" to serve vocalizations as well.

### 2.3.1. Sound-filtering through upper vocal tract body part movements

In humans, and in many species of terrestrial and aquatic vertebrates, the source-filter theory explains the processes and movements that affect acoustic variables after the energy delivered by sound-powering movements have been converted into vocal sound.

According to the source-filter theory, the mechanism of vocal sound production is the result of a two-phase process (Chiba and Kajiyama, 1941). The first phase occurs at the level of the source (e.g., larynx) where a sound is produced. In this phase, the vocal sound has a number of features e.g., fundamental frequency, harmonics, noise components, degree of periodicity and intensity. During the second phase, the vocal sound then radiates through the vocal tract. The vocal tract acts as a filter by attenuating some frequency components of the source sound and amplifying others. It is the configuration of the vocal tract (e.g., determined by laryngeal movement), further shaped by the position and movement of the articulators (e.g., lips, tongue, jaw, teeth) that influences the acoustic properties of the vocal sound emitted (Browman and Goldstein, 1992; Tokuda, 2021). The coordination of these different sound-filtering movements is a complicated affair, with varying ideas about which articulatory movements came online first in the human lineage (MacNeilage, 2010). Importantly, the articulatory degrees of freedom, natural frequencies and neuro-muscular stabilities of all these articulatory movements determine to an important degree the structure of human speech. For example, most energy in the speech acoustic signal found across all spoken languages is characterized by a 2–8 Hz quasi-rhythmicity (Chandrasekaran et al., 2009), associated with jaw motion (MacNeilage, 2010; Poeppel and Assaneo, 2020). Similar orofacial-phonation coupling rhythms in this ‘theta’ range were observed in marmosets (*Callithrix jacchus*) (Risueno-Segovia and Hage, 2020).

In songbirds, there are a variety of sound-filtering movements that contribute to their vocal productions. Such movements affect the supra-syringeal vocal tract shaping and include movement of the tongue and beak (Hoese et al., 2000). Kinematic analysis and behavioral studies have shown that dynamic movements of the beak and the degree of beak opening in songbirds affect the acoustic frequencies and, therefore the harmonic content of the vocal sounds produced. For example, white bellbirds (*Procnias albus*), before singing, inflate their throat by swallowing air and opening their beaks to their maximum degrees, thus regulating the resonance properties of the vocal tract (Podos and Cohn-Haft, 2019). However, some birds, such as doves, vocalize with a closed beak (Hoese et al., 2000; Homberger, 2017; Riede et al., 2016, 2004). In the expiratory phase of their vocalizations, the beak and the nostrils are closed and the air is directed through the upper esophagus causing its inflation. This structure acts like a radiating chamber attenuating high-frequency sounds, whereby the favored lower-frequency sound is radiated from the inflated structure and cutaneous tissue (Beckers et al., 2003; Riede et al., 2016).

But also less obvious sound-filtering movements are produced by birds. In parrots, the movements of the lingual, laryngeal, cranial and mandibular apparatuses and the movement of the neck collaborate in a complex way, allowing a means of production of vocal and consonant sounds that sound highly similar to humans (Homberger, 2017). The ability to produce such complex movements lies in the highly mobile skull system (a so-called kinetic skull). Indeed, the ability of some species of birds to produce a rich panoply of formant and rhythmic modulations does not seem to relate to syringeal complexity but rather to the degree of controlled mobility of the larynx, the glottal lips and hyoid structures (Homberger, 2017).

### 2.3.2. Sound-filtering through air sac movements

In many species, the vocal tract is extended by another morphological structure that can act as a radiating chamber during vocal production: an air sac. Several mammalian lineages have evolved a variety of laryngeal air sacs which, as a lateral expansion of the vocal tract, have been found to constrain sound production (Burchardt et al., 2024; Frey et al., 2007; Nishimura, 2020), though their function across and within species is a matter of ongoing discussion (e.g., see (Burchardt et al., 2024; Dunn, 2018; Nishimura, 2020; Perlman and Salmi, 2017)). We mention here only *some* relevant examples. Male reindeer (*Rangifer tarandus*) use the stylohyoid and thyroepiglottic muscles to inflate the

air sac during the rutting calls. Siamang (*Symphalangus syndactylus*) uses their laryngeal air sac to amplify lower frequency components relative to the higher components in their loud booming and barking calls during singing (de Boer, 2009; Mott, 1924). In this regard, (Burchardt et al., 2024) obtained in a small dataset a correlation between air sac volume and increased energy at the lower frequencies relative to higher ones, in line with earlier acoustic modeling (de Boer, 2009; Frey et al., 2007; Riede and Titze, 2008). Note though, that the function of air sacs remains somewhat of a mystery, and there are plenty of alternative hypotheses. For example, the presence of air sacs might allow for the production of longer and faster sequences at shorter intervals, avoiding the risk of hyperventilation (Hewitt et al., 2002). Or they might produce an additional air pressure that increases expiratory flow, thereby amplifying calls (Mott, 1924).

### 2.3.3. Sound-filtering through head posture movement

Another clearly visible sound-filtering movement is head movements. In many vertebrate species, including humans, head posturing affects vocal acoustics through supra-laryngeal tract shaping (Flory, 2016; Miller et al., 2014). Other terrestrial and semi-aquatic mammals, such as red deer (*Cervus elaphus*), mongolian gazelle (*Procapra gutturosa*) and northern elephant seals (*Mirovanga angustirostris*) show a typical “roaring posture”: the neck is extended and the head elevated upwards in such a way as to lengthen the vocal tract. The result is a deep vocalization characterized by a decrease in the formant frequencies (Frey and Gebler, 2010). Formant frequencies arise from the acoustic resonances of the vocal tract air column. Formants are defined as accumulations of acoustic energy estimated from the spectral envelope of a signal (Aalto et al., 2018).

Among birds, king penguins, while calling, raise their beak upwards and stretch their neck to their maximum extension. This particular posture allows the sound to travel farther avoiding being shielded by the bodies of the individuals in the large colonies where they live (Kriesell et al., 2020). This modulation of sound propagation thereby promotes individual recognition of mate and chick calls. Roosters have a less mobile laryngeal system, which raises questions about how complex consonant-vowel like sounds of “cock-a-doodle-doo” may be produced without such means of vocal tract shaping. Homberger (2017) hypothesizes, these consonant-vowel like sounds are achieved because of the retroflexion of the cervical vertebral column through head raising. Interestingly, it seems that when some morphology limits vocal production, such degrees of freedom may be obtained through other vocal-tract shape modulation in the form of head-raising movements.

### 2.3.4. Concluding remarks

In this section, we have examined the interdependent movements involving the upper vocal tract and their effects on the sound produced. The types of sound-relating movements considered until now (e.g., head raising, air sac inflation, etc.) also are powerful visual cues. This firstly means that vocalization can become informative about bodily activity and states, similar to how certain sound qualities may inform about body size. Secondly, in line with the concept of exaptation (defined in the section: “2.2.1 Concluding remarks”), this means that sound-filtering movements, previously shaped by natural selection for sound production, may be repurposed or “extra-purposed”. The sound-filtering movements can then assume new or additional functional roles that add new selective pressures, for instance, as part of visual communication (Starnberger et al., 2014). Thus, some signals that are traditionally seen as visual communication may have originated as sound-filtering movements. Though of course, this may go the other direction too: signals evolved for visual communication, which had accidental (filtering) effects on sound-production, may have become exapted or extra-purposed for sound-filtering too.

## 2.4. Intrinsic sound-impinging movements

The movements we discussed so far are those recruited solely for the purpose of producing or modulating sounds. However, as soon as sound-relating movements become visible or sensed in some other domain, such behaviors are open for further exaptation/repurposing. Since the original function of sound-movement coupling may remain next to the newly acquired function, we think “extra-purposing” is a more appropriate term. Our emphasis on “extra-purposing” becomes especially clear in the current section, where sound-impinging movements are considered: movements that serve multiple functions simultaneously in the context of communication. In this section, we overview a set of movements that physically impinge sound production, while they are not physically required for sound production in any way. These movement-sound couplings come from relatively weak biomechanical links between the locomotor, respiratory, and vocal systems. The coordination among these systems involves new muscle groups and establishes new synergies (Latash, 2008), characterized by multiple coordinated stabilities across interconnected elements. These new stabilities enable a sound production that may exhibit different characteristics compared to vocalization produced without sound-impinging movements.

### 2.4.1. Sound-impinging movements through respiratory-vocal coupling

The coupling between whole-body movement, respiration, and vocal communication has been described in many vertebrates. Human adults exhibit coupling between movement, respiration, and vocalization. They perform pulse-like upper limb movements known as “beat gestures” (for an overview see Pouw and Fuchs, 2022). These pulse-like movements have been shown to modulate expiratory flow (Werner et al., 2024) and, consequently, vocal production (Pouw et al., 2020b). They are also directly related to the degree of muscle tensioning around the trunk, which is associated with postural control and arm movements (Hodges and Gandevia, 2000; Pouw et al., 2023). The biomechanics behind the movement-respiration-vocalization coupling, therefore, helps explain why the voice’s amplitude can change with forceful gestural arm movements (see Pouw and Fuchs, 2022), such as in the case of singing (Pearson and Pouw, 2022). Biomechanics also explains why unique individuals with profound language or motor production issues largely maintain gesture-vocal synchrony, as obtained with aphasia (Jenkins and Pouw, 2023) or deafferentation (Pouw et al., 2020b).

Potentially limb-respiratory-vocal coupling can also provide a reason for why infants as young as 9 months of age can synchronize their gestures with vocal utterances that are in transition to more mature speech (Borjon et al., 2024; Ejiri, 1998; Ejiri and Masataka, 2001). For example, Borjon and colleagues (2024) have recently shown that infants aged between 9 and 24 months move their hands and heads concurrently and synchronously during spontaneous vocalizations. Some further evidence shows that language learning may be related to crawling and walking development (Borjon et al., 2024; Walle and Campos, 2014). Pouw and Fuchs (2022) suggest on the basis of a cross-species review that the coordination between an infant’s limb movements and vocalizations during development may be influenced by the biomechanical connections within the respiratory-vocal system and whole-body movements.

Among mammals, bats perform movement-respiratory-vocal coupling during the production of ultrasonic echolocation sounds: Beating of the wings for locomotion synchronizes with the respiratory-vocal cycles in a 1:1 or polyrhythmic fashion. This synchronization arises from a functional alignment (or *synergy*) of weak biomechanical influences of wing-powered flight on respiration with the respiratory flow needed for vocalization (Lancaster et al., 1995). The abdominal muscles, which usually provide primary power for ultrasonic production, are *in flight* attenuated as another muscle now delivers the primary drive. In flight the large pectoral muscles that power flight now also simultaneously support respiratory cycling (Lancaster et al., 1995; Pouw

and Fuchs, 2022; Thomas and Suthers, 1972).

Importantly, in this case, the constraints of biomechanics do not necessitate the synchronization of vocal pulses with wingbeats. For example, *Eptesicus fuscus* adjusts the number of vocal pulses per wingbeat when evading objects, requiring denser sampling with echolocation, which clearly overrides any energetic optimizations that exist when coupling sound-producing movements with vocalization (Ghose et al., 2006). Ultimately, it is the context that determines what is optimal overall.

Adult rats, during social interaction, coordinate ultrasonic vocalizations and locomotion with extraordinary temporal precision: within a few hundred milliseconds vocal production starts and ends together with the onset and offset of the movement (Laplagne and Elías Costa, 2016). Furthermore, a movement-respiratory-vocal link is observed in gerbils (*Meriones unguiculatus*). They saltate and emit ultrasound when their forelimbs touch the ground after a hop, or during their copulation movements (Blumberg, 1992). Touching the ground, after the hop, an expiratory flow is produced which coincides with a laryngeal constriction that helps stiffen the thorax. Together this results in a high-frequency sound production (Blumberg, 1992).

Similarly to rodents, also some birds seem to couple hops to vocalizations (Berg et al., 2019a). White-crested laughing thrushes (*Garrulax leucolophus*) increase vocalization durations during synchronized hopping as compared to sitting (Chinkangsadarn, 2012; Laplagne and Elías Costa, 2016). Forelimb coupling with vocalization is evident in birds too, showing respiratory-vocal coupling in flight (see Berg et al., 2019a). But also in communicative displays such coupling is observed. The brown-headed cowbird (*Molothrus ater*) engages in vigorous multimodal courtship displays, consisting of sequences of postural changes and vocalizations emitted in synchrony with the wingbeats (Cooper and Goller, 2004). However, the displays are highly variable as the wing movements vary in intensity and the vocal production may be omitted entirely. Importantly, these courtship vocalizations require the activity of specific respiratory abdominal muscles. While moving the wings this respiratory muscle activity decreases, even though the pressure of the air sac remains constant. Like bats, cowbirds exhibit a sophisticated movement-respiratory-vocal synergy; unlike bats, cowbirds’ coupling is deployed for visual display rather than locomotion.

Movement-respiratory-vocal coupling may guide vocal development (Berg et al., 2013). Chicks of green-rumped parrotlets (*Forpus passerinus*), seem to couple the different uses of their wings in and outside of the nest with their vocal development stages. Initially, vocalizations show a characteristic development from early, premature, to mature vocalization closely synchronized with the motoric development from passive, to exploratory wing-flapping, to flight-supporting wingbeats, respectively. According to these researchers, wing use modulates expiratory pressures, providing the right biophysical milieu for further vocal-respiratory skills to develop (Berg et al., 2013). This might associated with development in zebra finches: Inhibiting the motoric development of flight, by applying reversible wing-clipping or reducing cage size, also inhibits a range of markers of vocal learning influencing the neural plasticity involved in vocal learning (Liu et al., 2022).

### 2.4.2. Other sound-impinging movements

In the overview of the current class we focus on movements that are under control of the animal, and when produced, impinge on the respiratory-vocal system. But there are many other examples of where movements might impinge on some key element of the sound-producing process (Orlikoff, 2008; Vorperian et al., 2015), some of which are not under direct control of the animal. Most striking perhaps, are studies showing how heartbeats can couple directly with the voice. This was first found in humans in the 80’s (Orlikoff and Baken, 1989). During stationary phonation of a vowel (such as “a”), there is a systematic coupling between the variation of the heart rate and the variation of the periodic jitter in the fundamental frequency, as well as the intensity, of vocal production. How the heartbeat relates to vocalization can either be



explained by variations in the blood volume in the vocal folds or because the heart mechanically pushes against the lungs (Orlikoff, 1990; Orlikoff and Baken, 1989). That heartbeat phases can be retrieved from vocal acoustics in humans has been further shown in machine learning research (James, 2015; Mesleh et al., 2012). However, the exact biomechanical connection between heartbeats and vocalization is unknown and this research seems to have been discontinued in phonetics research in humans. We also have not found exactly similar findings in comparative bioacoustics, aside from research showing how heart rate is correlated with acoustic features of vocalization, see e.g., Stewart et al., (2015) for a study on prairie voles (*Microtus ochrogaster*). Such findings are often explained according to the Polyvagal Theory - in terms of a shared arousal-sensitive neural network that modulates cardiac and vocal activity (Porges and Lewis, 2009). It is possible, however, that heart rates couple rhythmically and *mechanically* with vocalizations in non-human animals, as it seems to do in sustained vocalizations in humans.

#### 2.4.3. Concluding remarks

The sound-impinging movements we have overviewed here show an interesting characteristic of being “optional”; this is likely because the biomechanical constraints are rather weak. That is, the bat is not obligated to couple wingbeats with vocalizations because performing a wingbeat increases the expiratory drive by only so much. Such drive could be counteracted with compensatory inspiratory muscle activations. In a similar vein, the cowbird can and does stop vocalizing altogether during vigorous wing movements which likely destabilizes vocalization (Cooper and Goller, 2004).

The flexibility of aligning or counteracting biomechanics is likely dependent on the coupling strength and the neurophysiological control that is available to the animal. Interestingly, while non-vocal learning South American bird species scale their wing beats with vocalization durations suggesting a biomechanical constraint, *vocal learners* tend to show less of such scaling (Berg et al., 2019b). This makes us wonder whether such vocal learners are more flexible in how to negotiate biomechanical constraints at any moment in time given what is optimal within a wider context.

Finally, we can wonder whether the coordination of sound production with movements that impinge sound may also become a communicative display itself. For instance, for an animal to be able to stably vocalize while counteracting effects on sound-impinging movements, or vocalizing in synchrony by aligning sound-impinging movements, entails a signal production that is potentially costly to sustain and only suitable individuals can bear such costs (see for a discussion on costly signaling Penn and Számadó, 2020). Sound-impinging movements as communicative displays may therefore become a reliable indicator of fitness itself and possibly come under sexual selection, as is possibly the case in courtship displays of cowbirds (Cooper and Goller, 2004; see also Soma and Shibata, 2023).

#### 2.5. Intrinsic sound-contingent movement

In previous sections, we have discussed sound-impinging movements that are not strictly necessary for sound production, but when invoked, they need to be negotiated or incorporated in some way with the sound-producing process. In this section we go one step further: we describe the coordination of vocalizations and body movements that are not (clearly) mechanically linked, but they are coupled by the organism. These *intrinsic sound-contingent movements* are therefore special in that they imply a certain degree of “mechanical arbitrariness” and increased neurophysiological control.

In humans, speaking may co-occur with gestures that pattern for the sake of directing someone’s attention. The ability to meaningfully time complex semiotically formed motor-routines cannot of course be solely or primarily explained by biomechanical constraints (Holler, 2022). Neurobiological and behavioral studies are needed to elucidate this

sound-movement coupling, and the field of gesture studies is ordained with understanding how semiotic movements align with semiotic sounds (Feyereisen, 2017). In this section, however, we will mainly consider more basic underlying timing principles of neural coordination of sound and movement.

The timing of simple vocalizations and finger movements - which is difficult to imagine being biomechanically coupled (but see Silva et al., 2007) - still naturally synchronizes (Kelso and Tuller, 1984b; Parrell et al., 2014; Treffner and Peter, 2002; Zelic et al., 2015). In such research, it is generally found that in-phase or out-of-phase synchronization are two stable attracting modes of coupling that are more generally found when humans produce two rhythmic outputs, such as flexing and extending the left and right finger rhythmically (Turvey, 1990). Vocal-hand coordination can be more complex than simple 1:1 phase-coupling however, as rhythmic behavior can involve poly-rhythmic coupling as well (Zelic et al., 2015). We can speak therefore of an attractor landscape, and it turns out that this landscape seems similarly structured by more general principles that do not only apply to vocal-motor coupling. Similar to other types of movements, vocal-hand coupling follows a highly predictable hierarchy of modes as described by the so-called Farey tree and Arnold Tongues (Pikovsky et al., 2001). This hierarchy dictates that polyrhythmic coupling is more stable for simple integer ratios 1:2 as compared to more complex ratios 3/4, and even more unstable than 4/9 (Zelic et al., 2015). Since there are no obvious mechanical links between the left and right index-fingers, or between the voice and hand, we should wonder what constrains two rhythmic outputs to show such structural patterns.

The answer has generally been sought in more general principles of coupled oscillator systems that self-organize into stable modes of interaction (Pikovsky et al., 2001). This is because the coordination stabilities described above are seen for a variety of coupled systems, including people swinging in a rocking-chair together (Richardson et al., 2007) or even non-living coupled oscillators like metronomes or lasers (Pikovsky et al., 2001). The key take-home here is that while coordination of vocalization and movement may at times be “mechanically arbitrary” they may still be constrained in their coordination in non-arbitrary ways as dictated by principles of coupled oscillators. In the end, why the coordination happens as it does, also boils down to physical constraints, but these constraints arise out of fundamental physical interactions that arise in coupled oscillating systems as diverse as neurons or lasers.

Next to such general self-organizing principles that govern contingent coupling between voice and hand, it has additionally been argued that the hand and articulation are very strongly interconnected because their neural hubs are anatomically close (Iverson and Thelen, 1999). Neurophysiological studies have demonstrated that the corticospinal pathways that regulate human hand muscle movements exhibit increased excitability during speech compared to a resting condition (Miyata and Kudo, 2014; Tokimura et al., 1996). The neurobiological coupling of sound and limb movement might also explain the ease of perceptual binding of *perceived* sound and movement. For example, the effectiveness of multimodal messages in human face-to-face communication has been demonstrated as it increases processing speed in different communicative contexts as compared to unimodal messages (Holler and Levinson, 2019); for an overview see (Trujillo and Holler, 2023).

Many animal species communicate by integrating and synchronizing different sensory modalities (Elias et al., 2003; Lewis et al., 2001; Todt and Fiebelkorn, 1980). Among vertebrates, the production of a laryngeal/syringeal communicative sound or vibration is often associated with movement in several multimodal displays; multimodality is generally used to improve signal quality in social contexts (Burchardt et al., 2019; Dalziel et al., 2013; Hasson and Hamilton, 1997; Marler, 1967; Partan and Marler, 2005). Among terrestrial mammals, great apes produce multimodal displays in the context of agonistic interaction. They assert their social status and intimidate competitors by producing



complex and loud hoots combined with rapid and exaggerated locomotion activity, jumping from branch to branch, shaking and throwing objects and beating hands and feet against the buttress roots of trees (Geissmann, 2000; Goodall, 1986; Müller and Anzenberger, 2002; Partan and Marler, 2005; Soldati et al., 2022). Similar phenomena have been described in marine mammals: during aggressive interaction, bottlenose dolphins (*Tursiops truncatus*) show an elaborated underwater behavior characterized by a visual display including jaw clapping and tail slapping that co-occur with the emission of rhythmic vocal bouts (Herzing, 2015).

Among birds, the male red-winged blackbirds (*Agelaius phoeniceus*) defend their territory with multimodal displays aimed at neighboring conspecifics: they couple advertising songs with a visual display characterized by the spread of wings and erect epaulets (Fusani et al., 2014; Peek, 1972). Female red-winged blackbirds exhibit aggressive and territorial-marking behavior that is associated with a particular type of male song, by fanning their tails and flapping their wings. Here the association of two sensory modalities has an evident pragmatic role: when the same visual display is associated with a different type of song, the behavior assumes a different function, related to maintaining the pair bond or courtship (Beletsky, 1982; Partan and Marler, 2005). Sound-movement coupling can also support social bonding functions or, during courtship displays, increase mating success in birds (Mitoyen et al., 2019; Ota et al., 2015). Spotted palm thrushes (*Cichladusa guttata*) show a sophisticated performance of antiphonal duets, aimed at maintaining stable pair bonds. The male in a pair attains the partner's attention by accompanying the vocal duet with a precise rhythmic movement of tail and wings. The emission of each note is coupled by a cycle of raising the wings above the head, during which the feathers are unfurled. The cycle ends when the wings close in conjunction with the end of a note emission (Todd and Fiebelkorn, 1980). Lyrebird courtship display is characterized by a multimodal dance performance in which four types of songs (with different durations and amplitude) are coordinated with their own specific choreographies that include precise sequences of wings, legs, and tail movement (Dalziel et al., 2013). Java Sparrows (*Lonchura oryzivora*) similarly show a coordination of courtship song notes and hopping (Soma and Shibata, 2023). Male lance-tailed manakin (*Chiroxiphia lanceolata*) jump between perches emitting pip calls in synchrony with the end of the jump during his courtship rituals (DuVal, 2007), though this might also have biomechanical origins.

Fish too show likely sound-contingent movements in multimodal courtship displays. For instance, damselfish (*Dascyllus albisella*) males advertise their nest and convey information about their body size; they produce vocal signals called “chirps” simultaneously with visual displays such as conspicuous jumps rising in the water column and then rapidly swimming down (Lobel and Mann, 1995; Myrberg et al., 1986).

### 2.5.1. Concluding remarks

This section provided examples of sound-movement couplings with no obvious biomechanical connection supporting those couplings: Animals seem to connect functionally peripheral bodily systems through neural entrainment. Future research may show that these links are governed by general neural dispositions that can be self-organizing too, where two oscillating systems show stable modes. This is also the case of humans performing vocal-movement coupling as well as inter-movement coupling. In other cases, there might be a complex sequential sound-movement coupling that cannot simply be understood via emergent domain-general principles of self-organization (cf. Pikovsky et al., 2001).

### 2.6. Extrinsic sound-contingent movement

In the previous section, the movement and sounds produced both originate from the organism and are thus accessible and fully within its control. Here we discuss how movements may become coupled to

external sound patterns; sounds need to be expected or anticipated for movement coupling to occur. We discuss whether the ability to couple movements with external sounds is widespread among species.

We define external sound-contingent movements as the capacity to couple body movements to an external auditory sequence; in its simplest form, the external sound sequence is isochronous and fully predictable, consisting of a series of evenly paced units (Kotz et al., 2018; Merker et al., 2009). In auditory cognitive neuroscience, they call this audio-motor entrainment and it is held to be a key building block of characteristically human cognition, deployed in behaviors such as music perception, dance, and speaking (Fink et al., 2021; Phillips-Silver et al., 2010; Richter and Ostovar, 2016).

However, the ability to produce and perceive certain rhythmic patterns, only develops during human ontogenesis. The human fetus produces spontaneous rhythmic patterns related to physiology, rather than cognition, such as breathing, hiccups, and sucking. The fetus is also consistently exposed to external rhythmic sounds, such as the mother's footsteps and vocalizations, and this experience is believed to enhance the development of neurocognitive substrates for music, dance, and language (Larsson et al., 2019; Provasi et al., 2014). Newborns show sensory-motor synchronization to external sounds, but only within a narrow time range close to their spontaneous motor rhythm (Provasi and Bobin-Bègue, 2003). Only around 18 months do infants show the ability to synchronize their movement with external rhythms even away from their natural rate of movement (Rocha and Mareschal, 2017). During the development the accuracy of synchronization at wide ranges of intervals increases: adults can easily synchronize to metric beats, between 94 and 176 beats per minute (Patel and Iversen, 2014; van Noorden and Moelants, 1999; Zentner and Eerola, 2010); See also Repp and Su (2013) for a review of empirical findings in sensorimotor synchronization).

The ability to rhythmically entrain body movements to a sound pattern may partly depend on the natural frequencies of the particular body part (Mårup et al., 2022). A critical example of this comes from research on coordination dynamics given the biomechanical link between voice and body movement, showing that the voice carries information about rhythmic upper limb movement phases and tempo (Pouw et al., 2022). Beat perception could then be understood as a matter of neural systems resonating to the structure in the sound (see also Bayne and Williams, 2023; Large and Jones, 1999a) - a non-linear composite process where a sound structure interacts with perceptual-motor neural systems that have a particular response function because they tend to operate on temporal scales relevant for the functions that those perception and action systems generally perform (e.g., bodily movement).

The capacity of beat perception and synchronization in non-human animals is a recent area of empirical work (Kotz et al., 2018; Ravignani et al., 2017; Wilson and Cook, 2016). Different degrees of audiomotor entrainment abilities have been found in a few species, such as in several parrot species and Asian elephants, for which there is also behavioral evidence of vocal learning abilities (Cook et al., 2013; Hasegawa et al., 2011; Schachner et al., 2009; Stoeger et al., 2012). For instance, cognitive experiments have shown that sulphur-crested cockatoos (*Cacatua galerita*) adapt head movements to a series of experimentally modified auditory rhythmic stimuli (Patel et al., 2009). Finding audiomotor entrainment abilities in vocal learning species could be explained by the “vocal learning and rhythmic synchronization hypothesis” which states that the capacity of beat perception and synchronization is a byproduct of the ability to produce complex vocal signals learned through imitation (Janik and Slater, 1997; Patel and Iversen, 2014) i.e., vocal learning ability; (Janik and Slater, 1997; Patel and Iversen, 2014). This hypothesis thus proposes that the rhythmic synchronization ability originates in the neural circuitry for complex vocal learning. Vocal learning species could become excellent study models for understanding the evolution of human speech and music (Patel et al., 2009). If vocal learning in birds and mammals involves

neural circuits that have common evolutionary bases and if the ability to synchronize to an external beat is related to the neural circuitry for complex vocal learning, then the biological mechanisms underlying audiomotor entrainment could also have a shared evolution (Fitch and Jarvis, 2013; Patel and Iversen, 2014).

However, entrainment abilities to external rhythmic stimuli have been found also in less vocally flexible animal species. A California sea lion, trained through reinforcement, is able to entrain head bobbing to an external auditory rhythm and to generalize this behavior to different tempo and musical stimuli (Cook et al., 2013). This behavioral evidence suggests that the capacity of beat perception and synchronization could not strictly depend on the ability for vocal mimicry. Therefore, this trait may be more common in the animal kingdom than previously thought (Cook et al., 2013).

#### 2.6.1. Concluding remark

In this section, we have focused on how animals move to external sounds, which we have called extrinsic sound-contingent movement. We learn that the coupling of the movement to the sound seems to be *contingent* on the resonant properties of the sensorimotor system (Large and Jones, 1999b). An effective comparative approach to test the capacity to perceive a certain rhythmic structure in the sound and entrain it, requires tailoring experiments to the range of tempi that a specific species can entertain. Indeed, more ecologically relevant species-specific experimental conditions would open the way for new evidence of audiomotor entrainment across animal taxa.

#### 2.7. Mutual sound-contingent movements

As a final step, we shortly discuss how movements and sound may become mutually coupled. We call this mutual sound-contingent movement coupling, and it can consist of sound-contingent movement coupling between agents who produce sounds, or who are regulated by an external sound signal as well (e.g., music). This coupling requires an appreciation of reciprocal causation in interaction (Anichini et al., 2023; De Jaegher et al., 2010) and adapted neurophysiological mechanisms that guide perception-action couplings. This capacity therefore goes beyond the coordination of sound and movement with an unresponsive external sound signal (Knoblich et al., 2011), or beyond the capacity of an individual to couple two rhythms that originate endogenously.

Humans show mutual sound-contingent movements in interactive social behaviors (Keller et al., 2014) such as dancing (Bigand et al., 2024a), playing instruments together, with a conductor, or in improvisation (Eerola et al., 2017; Wöllner, 2018) or when singing in a choir (see Fig. 1). Choir-singing and other such collective vocal music making synchronizes whole-body movement and vocalization while being coupled to the movements of the conductor or other musicians (Müller et al., 2018). Mutual sound-contingent coupling is also apparent in taking turns in conversation based on multimodal perceptions of suitable turn transitions (ter Bekke et al., 2024; Trujillo and Holler, 2023).

In our reading of the literature, mutual sound-contingent movements are not often deconstructed into key mechanisms that can be attributed to sustain what is a more holistic non-decomposable coordination. However, a recent study investigated how the movement coupling between agents is regulated in collective dance by deconstructing movement motifs present in the data at several levels using dimensionality reduction techniques (Bigand et al., 2024a). To dance together, two independent factors were found to be important when coupling to an external sound: shared sensory input (music) and interpersonal adaptation to the movements (via visual contact). Interestingly, the anteroposterior parts of the body (such as the head) synchronize to the music while the lateral parts (such as the arms) synchronize through visual contact. This demonstrates the human ability to assign simultaneous but distinct synchronies to each spatial axis of the body and to each of its components (Bigand et al., 2024a). While similar analysis awaits for sound-contingent movement coupling with vocalizations, this study

shows that there are ways to simplify the very complex multimodal and multi-agent entanglements.

What is the ontogenetic basis for engaging in mutually coupled sound-contingent movements? Likely this is the socially embedded musical interactions practiced in childhood. These occur in the form of temporal entrainment, i.e., the ability to perceive external sequences and move in time, such as during clapping games and nursery rhymes. But this also is regulated by affective interaction: forms of emotional communication between newborn-caregiver, such as the typical slow, regular, and repetitive language directed at the child, exaggerated prosody, accompanied by synchronization of movement and gestures (Provasi and Bobin-Bègue, 2003; Repp, 2005). More elaborate synchronization capabilities gradually mature with development. Studies that have investigated the ontogeny of the ability to synchronize body movements with music and with a musical partner have shown that between 5 months and 5 years, this ability is not yet refined. Improvements in precise phase synchronization are observed later in childhood or near adolescence (Eerola et al., 2006; Kirschner and Tomasello, 2009; Zentner and Eerola, 2010) for a review and discussion of the ontogeny of these capabilities see (Phillips-Silver and Keller, 2012).

Do we find anything close to mutual sound-contingent movements in non-human animals? Many species have been found to show ecologically relevant mutual synchronization of acoustic or visual displays although involving only one modality and a lower degree of complexity than the one we see in humans (Chauvigné et al., 2014; Greenfield, 1994; Phillips-Silver et al., 2010). Groups of male fiddler crabs (*Uca annulipes*) synchronize waving movements of their claws during courtship displays. It seems that, to attract females, males compete to emit their signal before their neighbors, and this may have the incidental effect of synchronizing the signal (Backwell et al., 1998). Using bioluminescent visual signals to communicate, males of some firefly species exhibit inter-individual synchronization of flashing in the tropical night (Buck, 1988). This shows an inter-individual ability to coordinate using rhythmic visual signals. Acoustic synchronization between conspecifics has been described in many taxa (Greenfield, 1994). Frogs simultaneously emit advertisement calls to attract females (Klump and Gerhardt, 1992), and males of many species of insects synchronize their chirps in chorusing (Greenfield, 1994; Otte, 1992). Other examples of chorusing have been found in young wolves (Coscia et al., 1991) and dolphins (Janik and Sayigh, 2013).

We think it is certainly possible that in some of these collective unimodal emergent synchronizations that have been studied, the coupling may be more multimodal than previously thought, involving not only synchronous emission of sounds but also regulations of such emissions due to visual information of near neighbors that are part of the collective. For instance, a more detailed analysis of the underwater behavior of Atlantic spotted dolphins (*Stenella frontalis*) and bottlenose dolphins (*Tursiops truncatus*) revealed the capacity to synchronize vocal and physical displays in these species. They emit different types of synchronized vocalizations, combined with intra- and interspecific aggressive behaviors, and furthermore, they coordinate body postures and physical movements with the rhythm of the vocalizations (Herzing, 2015).

#### 2.7.1. Concluding remark

Here, we have discussed how organisms couple the movement of body parts and the vocal emissions to the movements of others and external sounds so that they are simultaneous, such as the capacity to create perception-action loops on multiple levels. This skill, overall, entails a framing that cannot be adequately addressed on the level of the individual anymore and that turns out to have great relevance in human social-ecological niche, such as shown in, chanting, dancing, and dialoguing. This does not mean that this is the most complex of behavior, however. Emergent mutually coupled movements or sound productions are abundant in non-human animals, and it is unclear whether the neurophysiological regulatory system needed for multi-modal versus

more unimodal perception-action loops is a difference in degree or a difference in kind (Greenfield and Roizen, 1993). Nevertheless, clear mutual sound-contingent movements performed predictively seem to be rare in other species. Future studies in this direction could employ more sophisticated tools to firstly assess whether there are more rhythms that are being regulated, and if so, whether they are differently regulated as is recently shown in mutually sound-contingent movements of dance (Bigand et al., 2024b).

### 3. Discussion

#### 3.1. Overview

As discussed through our review, movement and sound can be coupled in animals (including humans) in many ways. We proposed a framework to classify all these different kinds of sound-movement couplings. We started with sound-constituting movements, such as the vibrating movements of body parts, like the vocal folds or external surfaces, which transfer mechanical or aerodynamic energy to acoustic energy (sound waves). We then overviewed sound-powering movements such as respiratory actions, which provide the energy for vibrations to be sustained and modulated. Subsequently, we considered the sound-filtering movements that shape the vocal tract, studying how sound is modified thanks to the different movements and postures of the larynx, neck, head, and articulatory musculoskeletal components. We then considered a range of sound-impinging movements: Movements that are optionally implicated with vocal production due to weak biomechanical constraints. The final sections considered those movements and sounds that are contingently coordinated: in this case the movements coupled with sound that show no clear mechanical link with the sound-producing process. The sound-contingent movements can be produced within an organism, relative to a non-responsive external sound; Alternatively, there are sound-movement contingencies that arise between mutually coupled agents. We hope our framework will be equally helpful for those venturing into comparative bioacoustics and multimodal communication. Specifically, we identified a disconnect between, on the one hand, describing movements as directly causally implicated with sound production and, on the other hand, understanding separate “auditory” and “visual” signals integrated by the perceiver. Instead, our classification emphasizes a rather continuous and entangled movement-sound coupling regime, bridging both extremes.

#### 3.2. Pitfalls and prospects

Our proposed framework is by no means complete. For instance, we have not overviewed detailed musculo-skeletal actions that make things move in each class or that have more secondary roles in facilitating such prime movements, e.g., muscles and other connective tissues that can act as for example brakes, stabilizers, tuners, struts (Profeta and Turvey, 2018; Turvey and Fonseca, 2014). Additionally, our proposed framework is incomplete as it does not discuss the possible differentiable adaptive functions of the different classes of sound-movement coupling. What is sound-movement coupling good for in the ecological niche of the animal? Instead, our framework has focused on building a roadmap to characterize the nuances in connections which movement and sound display when coupled. We also invite more in-depth research on the interrelationships of different sound-movement couplings. Indeed, different sound-movement coupling classes are often combined and themselves coordinated. For example, breathing and articulatory movements coordinate in humans: singers and beatboxers strategically time the adduction of the vocal folds to produce different kinds of voicing onsets (e.g., hard onset, staccato onset, breathy onset; Sundberg, 1993; Sundberg et al., 1993). Many such inter-relations exist between the sound-movement coupling classes; we hope a more refined language which dissects multi-class composite behaviors can help understand the entangled relationship between them.

Our overview of movement and sound coupling is far from exhaustive in some other ways too. Movement and sound can be coupled and regulated by even more multisensory control loops. Critically, locomotor movement self-stimulates the visual field (Gibson, 1966), and all body movements will affect sensations of body posture and orientation (vestibular and proprioceptive information) in a closed-loop like fashion. In fact, some have argued that the vestibular system is a key bodily sense that allows for controlling synchronized movement-sound coupling (Coulson and Momsen, 2023). Thus, while we have focused on movement and sound coupling, those couplings themselves are nested in a multisensory milieu of self- and other stimulation (Godfrey-Smith, 2020).

#### 3.3. Key challenges for neuroscience and behavioral research in sound-movement coupling

What lessons can we draw that could guide future research and theory in this domain? Theoretically, we started out with several classes of sound-movement couplings where the physiology of movement is in some way mechanically implicated with the sound production process. In these cases, it is clear that the nervous system does not have to organize and couple body parts *de novo* - the brain does not need to be educated about how to vibrate the vocal folds, it just needs to be able to regulate the vibrations by discovering that under different embodied conditions (e.g., adducting the vocal folds, and expiring) the body will respond in a certain way. In this sense, biomechanics provide natural covarying information between sound and movement that the nervous system can learn through discovery and ultimately can use to support communication. MacNeilage (2010) famously says in this regard that “we get the biomechanics for free” (p. 295). Or so it seems. Biomechanical constraints actually are the result of adaptations in the species that were painstakingly “earned” in the arena of natural selection.

Fultot and colleagues (2019) argue that there is a reason why large multicellular organisms like mammals have not evolved (into) highly specialized bio-morphologies that optimize a single function but tend to favor flexibility: Extreme morphological adaptation would reduce the degrees of freedom of these potential animals so dramatically that they would be poorly adapted for many other tasks. This is where more complex neural systems with bodies with high degrees of freedom come in. Over different situations soliciting different adaptive behaviors, neural systems can dynamically coordinate musculoskeletal systems in such a way as to become an efficient foraging device, a courtship device, and whatever else is required for the survival of the species (Fultot et al., 2019). Having a flexible way to organize body parts in task-specific devices is what Fultot and colleagues understand as the proper function of neural systems, which obviates the need for bodies to *physically* morph over evolutionary time into highly specialized task-specific devices. Complex neural systems are useless of course if the body plan and its dispositions do not have a potential fit with aspects of the environment, i.e., a particular physical morphology is needed to be able to ‘functionally’ morph into a variety of task-specific devices.

An organism can also modify its environment. For example, the mole cricket (*Gryllotalpa vineae*) has a very loud sonification (Turner, 2009). This is because they dig a burrow for supporting the acoustic physics of amplification of their calls. Further, an orangutan can *flexibly organize its body* to build its own resonators on the fly, by cupping the hand in front of their mouths to shape their vocalization (de Boer et al., 2015; Hardus et al., 2009). Now consider humans, who have been transducing their movements into sounds in all kinds of ways (e.g., instrumentally, with joint movements to sounds in piano-playing; biomechanically, e.g., beatboxing). These abilities are partly derived from humans having created a particularly strong social and cultural environment that can act as another type of environment for learning certain sound and movement couplings (Falands et al., 2023; Phillips-Silver and Trainor, 2005). Such early social-cultural constraints likely were an important further driver for the development of nervous systems adapted for



learning a particular sound-movement coupling (Heyes, 2018). The result is a nervous system which naturally learns to couple sound and movement with a degree of flexibility that is quite unique in the animal kingdom; auditory-motor entrainment and speaking, is a likely product of this flexibility.

The challenge we think is to gain a deeper understanding of the adaptive trade-offs for when communicative feats seem to be a product of a morphological specialization and/or a (socially-scaffolded) neuro-physiological sound-movement assemblies. The “calculus” of these trade-offs will need to relate costs and benefits of particular adaptations on the level of mechanisms, ontogenesis, and evolution (Tinbergen, 1963).

## Funding

SL was supported by Independent Max Planck Research Group Leader Funding to AR and by the European Union (ERC, TOHR, 101041885). WP is funded by a VENI grant (VI.Veni 0.201 G.047), DFG VICOM grant (PO2841/1-1), and supported by Language in Interaction (024.001.006). AR is supported by the European Union (ERC, TOHR, 101041885) and the HFSP research grant (RGP0019/2022). The Comparative Bioacoustics Group is supported by Max Planck Independent Research Group Leader funding to AR. Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF117).

## Declaration of Competing Interest

We have no competing interests to declare.

## Data availability

No data was used for the research described in the article.

## References

- Aalto, D., Malinen, J., & Vainio, M. (2018). Formants. In *Oxford Research Encyclopedia of Linguistics*.
- Akester, A.R., 1960. The comparative anatomy of the respiratory pathways in the domestic fowl (*Gallus domesticus*), pigeon (*Columba livia*) and domestic duck (*Anas platyrhynchos*). *J. Anat.* 94, 487–505.
- Anichini, M., de Reus, K., Hersh, T.A., Valente, D., Salazar-Casals, A., Berry, C., Keller, P. E., Ravnigani, A., 2023. Measuring rhythms of vocal interactions: a proof of principle in harbour seal pups. *Philos. Trans. R. Soc. B Biol. Sci.* 378, 20210477. <https://doi.org/10.1098/rstb.2021.0477>.
- Backwell, P., Jennions, M., Passmore, N., Christy, J., 1998. Synchronized courtship in fiddler crabs. *Nature* 391, 31–32. <https://doi.org/10.1038/34076>.
- Barimo, J.F., Fine, M.L., 1998. Relationship of swim-bladder shape to the directionality pattern of underwater sound in the oyster toadfish. *Can. J. Zool.* 76, 134–143. <https://doi.org/10.1139/z97-160>.
- Bass, A.H., Chagnaud, B.P., 2012. Shared developmental and evolutionary origins for neural basis of vocal-acoustic and pectoral-gestural signaling. *Proc. Natl. Acad. Sci.* 109, 10677–10684. <https://doi.org/10.1073/pnas.1201886109>.
- Bayne, T., Williams, I., 2023. In search of the beat. *Mind Lang.* 38, 907–924. <https://doi.org/10.1111/mila.12439>.
- Beckers, G.J.L., Suthers, R.A., Cate, C. ten, 2003. Mechanisms of frequency and amplitude modulation in ring dove song. *J. Exp. Biol.* 206, 1833–1843. <https://doi.org/10.1242/jeb.00364>.
- Beletsky, L.D., 1982. Vocalizations of female northern orioles. *Condor* 84, 445–447. <https://doi.org/10.2307/1367454>.
- Berg, K.S., Beissinger, S.R., Bradbury, J.W., 2013. Factors shaping the ontogeny of vocal signals in a wild parrot. *J. Exp. Biol.* 216, 338–345. <https://doi.org/10.1242/jeb.073502>.
- Berg, K.S., Delgado, S., Mata-Betancourt, A., 2019a. Phylogenetic and kinematic constraints on avian flight signals. *Proc. R. Soc. B Biol. Sci.* 286, 20191083. <https://doi.org/10.1098/rspb.2019.1083>.
- Berg, K.S., Delgado, S., Mata-Betancourt, A., 2019b. Phylogenetic and kinematic constraints on avian flight signals. *Proc. R. Soc. B Biol. Sci.* 286, 20191083. <https://doi.org/10.1098/rspb.2019.1083>.
- Bigand, F., Bianco, R., Abalde, S.F., Novembre, G., 2024a. The geometry of interpersonal synchrony in human dance. *Curr. Biol.* <https://doi.org/10.1016/j.cub.2024.05.055>.
- Bigand, F., Bianco, R., Abalde, S.F., Novembre, G., 2024b. The geometry of interpersonal synchrony in human dance. *Curr. Biol.* <https://doi.org/10.1016/j.cub.2024.05.055>.
- Blumberg, M.S., 1992. Rodent ultrasonic short calls: locomotion, biomechanics, and communication. *J. Comp. Psychol.* 106, 360–365. <https://doi.org/10.1037/0735-7036.106.4.360>.
- Borjon, J.I., Abney, D.H., Yu, C., Smith, L.B., 2024. Infant vocal productions coincide with body movements. *Dev. Sci.* 27, e13491. <https://doi.org/10.1111/desc.13491>.
- Browman, C.P., Goldstein, L., 1992. Articulatory phonology: an overview. *Phonetica* 49, 155–180. <https://doi.org/10.1159/000261913>.
- Brown, S., Martinez, M.J., Parsons, L.M., 2006. The neural basis of human dance. *Cereb. Cortex* 16, 1157–1167. <https://doi.org/10.1093/cercor/bhj057>.
- Brown, C., Riede, T., 2017. *Comparative Bioacoustics: An Overview*. Bentham Science Publishers.
- Buck, J., 1988. Synchronous rhythmic flashing of fireflies. II. *Q. Rev. Biol.* 63, 265–289. <https://doi.org/10.1086/415929>.
- Burchardt, L.S., Norton, P., Behr, O., Scharff, C., Knörnschild, M., 2019. General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccolaryx bilineata*. *R. Soc. Open Sci.* 6, 181076. <https://doi.org/10.1098/rsos.181076>.
- Burchardt, L.S., Sande, Y., van de, Kehy, M., Gamba, M., Ravnigani, A., Pouw, W., 2024. A toolkit for the dynamic study of air sacs in siamang and other elastic circular structures. *PLOS Comput. Biol.*
- Bustamante, J., Márquez, R., 1996. Vocalizations of the chinstrap penguin *pygoscelis antarctica*. *Colonia Waterbirds* 19, 101–110. <https://doi.org/10.2307/1521812>.
- Calder, W.A., 1970. Respiration during song in the canary (*Serinus canaria*). *Comp. Biochem. Physiol.* 32, 251–258. [https://doi.org/10.1016/0010-406X\(70\)90938-2](https://doi.org/10.1016/0010-406X(70)90938-2).
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., Ghazanfar, A.A., 2009. The natural statistics of audiovisual speech. *PLOS Comput. Biol.* 5, e1000436. <https://doi.org/10.1371/journal.pcbi.1000436>.
- Chauvigné, L.A.S., Gitau, K.M., Brown, S., 2014. The neural basis of audiomotor entrainment: an ALE meta-analysis. *Front. Hum. Neurosci.* 8.
- Chiba, T., Kajiyama, M., 1941. *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan.
- Chinkangsadarn, 2012. Singing Behavior of White-Crested Laughing-Thrush (*Garrulus leucolophus*). [WWW Document]. URL (<https://www.thaiscience.info/Journals/Article/SJST/10890420.pdf>) (accessed 7.9.24).
- Clark, C.J., 2016. Locomotion-Induced Sounds and Sonations: Mechanisms, Communication Function, and Relationship with Behavior. In: Suthers, R.A., Fitch, W.T., Fay, R.R., Popper, A.N. (Eds.), *Vertebrate Sound Production and Acoustic Communication*, Springer Handbook of Auditory Research. Springer International Publishing, Cham, pp. 83–117. [https://doi.org/10.1007/978-3-319-27721-9\\_4](https://doi.org/10.1007/978-3-319-27721-9_4).
- Clark, C.J., Elias, D.O., Prum, R.O., 2011. Aeroelastic flutter produces hummingbird feather songs. *Science* 333, 1430–1433. <https://doi.org/10.1126/science.1205222>.
- Clark, C.J., Feo, T.J., 2008. The Anna's hummingbird chirps with its tail: a new mechanism of sonation in birds. *Proc. R. Soc. B Biol. Sci.* 275, 955–962. <https://doi.org/10.1098/rspb.2007.1619>.
- Colafrancesco, K.C., Gridi-Papp, M., 2016. Vocal Sound Production and Acoustic Communication in Amphibians and Reptiles. In: Suthers, R.A., Fitch, W.T., Fay, R.R., Popper, A.N. (Eds.), *Vertebrate Sound Production and Acoustic Communication*. Springer International Publishing, Cham, pp. 51–82. [https://doi.org/10.1007/978-3-319-27721-9\\_3](https://doi.org/10.1007/978-3-319-27721-9_3).
- Cook, P., Rouse, A., Wilson, M., Reichmuth, C., 2013. A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* 127, 412–427. <https://doi.org/10.1037/a0032345>.
- Cooper, B.G., Goller, F., 2004. Multimodal signals: enhancement and constraint of song motor patterns by visual display. *Science* 303, 544–546. <https://doi.org/10.1126/science.1091099>.
- Coscia, E.M., Phillips, D.P., Fentress, J.C., 1991. Spectral analysis of neonatal wolf canis lupus vocalizations. *Bioacoustics* 3, 275–293. <https://doi.org/10.1080/09524622.1991.9753190>.
- Coulson, S., Momen, J., 2023. Discourse and communicative pragmatics. In: *APA Handbook of Neuropsychology, Volume 1: Neurobehavioral Disorders and Conditions: Accepted Science and Open Questions*, Vol. 1, APA Handbooks in Psychology®. American Psychological Association, Washington, DC, US, pp. 43–58. <https://doi.org/10.1037/0000307-003>.
- Dalziel, A.H., Peters, R.A., Cockburn, A., Dorland, A.D., Maisey, A.C., Magrath, R.D., 2013. Dance choreography is coordinated with song repertoire in a complex avian display. *Curr. Biol.* 23, 1132–1135. <https://doi.org/10.1016/j.cub.2013.05.018>.
- de Boer, B., 2009. Acoustic analysis of primate air sacs and their effect on vocalization. *J. Acoust. Soc. Am.* 126, 3329–3343. <https://doi.org/10.1121/1.3257544>.
- de Boer, B., Wich, S.A., Hardus, M.E., Lameira, A.R., 2015. Acoustic models of orangutan hand-assisted alarm calls. *J. Exp. Biol.* 218, 907–914. <https://doi.org/10.1242/jeb.110577>.
- De Jaegher, H., Di Paolo, E., Gallagher, S., 2010. Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. <https://doi.org/10.1016/j.tics.2010.06.009>.
- DeJonckere, P.H., Lebacqz, J., DeJonckere, P.H., Lebacqz, J., 2024. The Physics of the Human Vocal Folds as a Biological Oscillator. In: *New Insights on Oscillators and Their Applications to Engineering and Science*. IntechOpen. <https://doi.org/10.5772/intechopen.113958>.
- Dudley, R., Rand, A.S., 1991. Sound production and vocal sac inflation in the túngara frog, *physalaemus pustulosus* (Leptodactylidae). *Copeia* 1991, 460–470. <https://doi.org/10.2307/1446594>.
- Dufour, V., Poulin, N., Charlotte Curé, Sterck, E.H.M., 2015. Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. *Sci. Rep.* 5, 11320. <https://doi.org/10.1038/srep11320>.
- Duncker, H.-R., 1974. Structure of the avian respiratory tract. *Respir. Physiol.* 22, 1–19. [https://doi.org/10.1016/0034-5687\(74\)90044-9](https://doi.org/10.1016/0034-5687(74)90044-9).
- Dunn, J.C., 2018. Sexual selection and the loss of laryngeal air sacs during the evolution of speech. *Anthropol. Sci.* 126, 29–34. <https://doi.org/10.1537/ase.180309>.



- DuVal, E.H., 2007. Cooperative display and lekking behavior of the lance-tailed manakin (*Chiroxiphia lanceolata*). *Auk* 124, 1168–1185. <https://doi.org/10.1093/auk/124.4.1168>.
- Eerola, T., Jakubowski, K., Moran, N., Keller, P.E., Clayton, M., 2017. Shared periodic performer movements coordinate interactions in duo improvisations. *R. Soc. Open Sci.* 5, 171520. <https://doi.org/10.1098/rsos.171520>.
- Eerola, T., Luck, G., Toivainen, P., 2006. An investigation of pre-schoolers' corporeal synchronization with music. In: *Proceedings of the 9th International Conference on Music Perception and Cognition*. Alma Mater Studiorum University of Bologna, Bologna, Italy, pp. 472–476.
- Ejiri, K., 1998. Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica* 55, 226–237. <https://doi.org/10.1159/000028434>.
- Ejiri, K., Masataka, N., 2001. Co-occurrences of preverbal vocal behavior and motor action in early infancy. *Dev. Sci.* 4, 40–48. <https://doi.org/10.1111/1467-7687.00147>.
- Elemans, C.P.H., Jiang, W., Jensen, M.H., Pichler, H., Mussman, B.R., Nattestad, J., Wahlberg, M., Zheng, X., Xue, Q., Fitch, W.T., 2024. Evolutionary novelties underlie sound production in baleen whales. *Nature* 627, 123–129. <https://doi.org/10.1038/s41586-024-07080-1>.
- Elemans, C.P.H., Rasmussen, J.H., Herbst, C.T., Düring, D.N., Zollinger, S.A., Brumm, H., Srivastava, K., Svane, N., Ding, M., Larsen, O.N., Sober, S.J., Švec, J.G., 2015. Universal mechanisms of sound production and control in birds and mammals. *Nat. Commun.* 6, 8978. <https://doi.org/10.1038/ncomms9978>.
- Elias, D.O., Mason, A.C., Maddison, W.P., Hoy, R.R., 2003. Seismic signals in a courting male jumping spider (*Araneae:Salticidae*). *J. Exp. Biol.* 206, 4029–4039. <https://doi.org/10.1242/jeb.00634>.
- Endo, H., Kobayashi, D., Kimura, J., Rakotondraparany, F., Matsui, A., Yonezawa, T., Shinohara, A., Hasegawa, M., 2010. A quill vibrating mechanism for a sounding apparatus in the streaked tenrec (*Hemicentetes semispinosus*). *PLoS One* 27, 427–432. <https://doi.org/10.2198/zsj.27.427>.
- Falandays, J.B., Kaaronen, R.O., Moser, C., Rorot, W., Tan, J., Varma, V., Williams, T., Youngblood, M., 2023. All intelligence is collective intelligence. *J. Multiscale Neurosci.* 2, 169–191. <https://doi.org/10.56280/1564736810>.
- Farmer, C.G., 2015. The evolution of unidirectional pulmonary airflow. *Physiology* 30, 260–272. <https://doi.org/10.1152/physiol.00056.2014>.
- Favaro, L., Gamba, M., Alfieri, C., Pessani, D., McElligott, A.G., 2015. Vocal individuality cues in the African penguin (*Spheniscus demersus*): a source-filter theory approach. *Sci. Rep.* 5, 17255. <https://doi.org/10.1038/srep17255>.
- Feyereisen, P., 2017. *The Cognitive Psychology of Speech-Related Gesture*. Routledge.
- Fine, M.L., Malloy, K.K.L., King, C.B., Mitchell, S.L., Cameron, T.M., 2002. Sound generation by the toadfish swimbladder. *Bioacoustics* 12, 220–222. <https://doi.org/10.1080/09524622.2002.9753701>.
- Fine, M.L., Parmentier, E., 2015. Mechanisms of Fish Sound Production. In: Ladich, F. (Ed.), *Sound Communication in Fishes, Animal Signals and Communication*. Springer, Vienna, pp. 77–126. [https://doi.org/10.1007/978-3-7091-1846-7\\_3](https://doi.org/10.1007/978-3-7091-1846-7_3).
- Fink, B., Blasing, B., Ravnigani, A., Shackelford, T.K., 2021. Evolution and functions of human dance. *Evol. Hum. Behav.* 42, 351–360. <https://doi.org/10.1016/j.evolhumbehav.2021.01.003>.
- Fitch, W.T., Jarvis, E.D., 2013. Birdsong and Other Animal Models for Human Speech, Song, and Vocal Learning. <https://doi.org/10.7551/mitpress/9548.003.0028>.
- Fitch, W.T., Suthers, R.A., 2016. Vertebrate Vocal Production: An Introductory Overview. In: Suthers, R.A., Fitch, W.T., Fay, R.R., Popper, A.N. (Eds.), *Vertebrate Sound Production and Acoustic Communication*. Springer International Publishing, Cham, pp. 1–18. [https://doi.org/10.1007/978-3-319-27721-9\\_1](https://doi.org/10.1007/978-3-319-27721-9_1).
- Flory, Y., 2016. The influence of head and body postures on the acoustic speech signal. *Int. J. Speech, Lang. Law* 23, 141–146. <https://doi.org/10.1558/ijsll.v23i1.26677>.
- Frey, R., Gebler, A., 2010. Chapter 10.3 Mechanisms and evolution of roaring-like vocalization in mammals. In: Brudzynski, S.M. (Ed.), *Handbook of Behavioral Neuroscience, Handbook of Mammalian Vocalization*. Elsevier, pp. 439–450. <https://doi.org/10.1016/B978-0-12-374593-4.00040-1>.
- Frey, R., Gebler, A., Fritsch, G., Nygrén, K., Weissengruber, G.E., 2007. Nordic rattle: the hoarse vocalization and the inflatable laryngeal air sac of reindeer (*Rangifer tarandus*). *J. Anat.* 210, 131–159. <https://doi.org/10.1111/j.1469-7580.2006.00684.x>.
- Fuchs, S., Rochet-Capellan, A., 2021. The respiratory foundations of spoken language. *Annu. Rev. Linguist.* 7, 13–30. <https://doi.org/10.1146/annurev-linguistics-031720-103907>.
- Fuchs, S., Savin, E., Solt, S., Ebert, C., Krifka, M., 2019. Antonym adjective pairs and prosodic iconicity: evidence from letter replications in an English blogger corpus. *Linguist. Vanguard* 5. <https://doi.org/10.1515/lingvan-2018-0017>.
- Fultot, M., Adrian Frazier, P., Turvey, M.T., Carello, C., 2019. What are nervous systems for? *Ecol. Psychol.* 31, 218–234. <https://doi.org/10.1080/10407413.2019.1615205>.
- Fusani, L., Donaldson, Z., London, S.E., Fuxjager, M.J., Schlenger, B.A., 2014. Expression of androgen receptor in the brain of a sub-oscine bird with an elaborate courtship display. *Neurosci. Lett.* 578, 61–65. <https://doi.org/10.1016/j.neulet.2014.06.028>.
- Gans, C., Maderson, P.F.A., 1973. Sound producing mechanisms in recent reptiles: review and comment. *Am. Zool.* 13, 1195–1203. <https://doi.org/10.1093/icb/13.4.1195>.
- Garcia, M., Theunissen, F., Sèbe, F., Clavel, J., Ravnigani, A., Marin-Cudraz, T., Fuchs, J., Mathevon, N., 2020. Evolution of communication signals and information during species radiation. *Nat. Commun.* 11, 4970. <https://doi.org/10.1038/s41467-020-18772-3>.
- Geissmann, T., 2000. *Gibbon songs and human music from an evolutionary perspective*. In: *The Origins of Music*. The MIT Press, Cambridge, MA, US, pp. 103–123.
- Ghose, K., Horiuchi, T.K., Krishnaprasad, P.S., Moss, C.F., 2006. Echolocating bats use a nearly time-optimal strategy to intercept prey. *PLOS Biol.* 4, e108. <https://doi.org/10.1371/journal.pbio.0040108>.
- Giannoni, S.M., Márquez, R., Borghi, C.E., 1997. Airborne and substrate-borne communications of *Microtus (Terricola) gerbei* and *M. (T.) duodecimcostatus*. *Acta Theriol.* 42, 123–141. <https://doi.org/10.4098/AT.arch.97-16>.
- Gibson, J.J., 1966. *The senses considered as perceptual systems*. Houghton Mifflin, Oxford, England.
- Godfrey-Smith, P., 2020. *Metazoa: Animal Minds and the Birth of Consciousness*. HarperCollins UK.
- Goller, F., Larsen, O.N., 1997. A new mechanism of sound generation in songbirds. *Proc. Natl. Acad. Sci.* 94, 14787–14791. <https://doi.org/10.1073/pnas.94.26.14787>.
- Gómez-Bahamón, V., Tuero, D.T., Castaño, M.I., Jahn, A.E., Bates, J.M., Clark, C.J., 2020. Sonations in Migratory and Non-migratory Fork-tailed Flycatchers (*Tyrannus savana*). *Integr. Comp. Biol.* 60, 1147–1159. <https://doi.org/10.1093/icb/icaa115>.
- Goodall, J., 1986. Social rejection, exclusion, and shunning among the Gambia chimpanzees. *Ethol. Sociobiol.* 7, 227–236. [https://doi.org/10.1016/0162-3095\(86\)90050-6](https://doi.org/10.1016/0162-3095(86)90050-6).
- Gould, S.J., Vrba, E.S., 1982. Exaptation—a missing term in the science of form. *Paleobiology* 8, 4–15. <https://doi.org/10.1017/S0094837300004310>.
- Greenfield, M.D., 1994. Synchronous and alternating choruses in insects and anurans: common mechanisms and diverse functions. *Am. Zool.* 34, 605–615. <https://doi.org/10.1093/icb/34.6.605>.
- Greenfield, M.D., Roizen, I., 1993. Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature* 364, 618–620. <https://doi.org/10.1038/364618a0>.
- Halfwerk, W., Varkevisser, J., Simon, R., Mendoza, E., Scharff, C., Riebel, K., 2019. Toward testing for multimodal perception of mating signals. *Front. Ecol. Evol.* 7.
- Hardus, M.E., Lameira, A.R., Van Schaik, C.P., Wich, S.A., 2009. Tool use in wild orangutans modifies sound production: a functionally deceptive innovation? *Proc. R. Soc. B Biol. Sci.* 276, 3689–3694. <https://doi.org/10.1098/rspb.2009.1027>.
- Hasegawa, A., Okanoya, K., Hasegawa, T., Seki, Y., 2011. Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Sci. Rep.* 1, 120. <https://doi.org/10.1038/srep00120>.
- Hasson, O., Hamilton, W.D., 1997. Amplifiers and the handicap principle in sexual selection: a different emphasis. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 235, 383–406. <https://doi.org/10.1098/rspb.1989.0006>.
- Häusler, U., 2000. Vocalization-correlated respiratory movements in the squirrel monkey. *J. Acoust. Soc. Am.* 108, 1443–1450. <https://doi.org/10.1121/1.1289209>.
- Herbst, C.T., 2016. Biophysics of Vocal Production in Mammals. In: Suthers, R.A., Fitch, W.T., Fay, R.R., Popper, A.N. (Eds.), *Vertebrate Sound Production and Acoustic Communication*. Springer International Publishing, Cham, pp. 159–189. [https://doi.org/10.1007/978-3-319-27721-9\\_6](https://doi.org/10.1007/978-3-319-27721-9_6).
- Herbst, C.T., Elemans, C.P.H., Tokuda, I.T., Chatzioannou, V., Švec, J.G., 2023. Dynamic system coupling in voice production. *J. Voice*. <https://doi.org/10.1016/j.jvoice.2022.10.004>.
- Herzing, D., 2015. Synchronous and rhythmic vocalizations and correlated underwater behavior of free-ranging atlantic spotted dolphins (*Stenella frontalis*) and bottlenose dolphins (*Tursiops truncatus*) in the Bahamas. *ABC* 2, 14–29. <https://doi.org/10.12966/abc.02.02.2015>.
- Hewitt, G., MacLarnon, A., Jones, K.E., 2002. The functions of laryngeal air sacs in primates: a new hypothesis. *IJFP* 73, 70–94. <https://doi.org/10.1159/000064786>.
- Heyes, C., 2018. *Cognitive Gadgets: The Cultural Evolution of Thinking*. Harvard University Press.
- Hill, P.S.M., 2001. Vibration and animal communication: a review. *Am. Zool.* 41, 1135–1142. <https://doi.org/10.1093/icb/41.5.1135>.
- Hodges, P.W., Gandevia, S.C., 2000. Changes in intra-abdominal pressure during postural and respiratory activation of the human diaphragm. *J. Appl. Physiol.* 89, 967–976. <https://doi.org/10.1152/jappl.2000.89.3.967>.
- Hoes, W.J., Podos, J., Boettcher, N.C., Nowicki, S., 2000. Vocal tract function in birdsong production: experimental manipulation of beak movements. *J. Exp. Biol.* 203, 1845–1855. <https://doi.org/10.1242/jeb.203.12.1845>.
- Holler, J., 2022. Visual bodily signals as core devices for coordinating minds in interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 377, 20210094. <https://doi.org/10.1098/rstb.2021.0094>.
- Holler, J., Levinson, S.C., 2019. Multimodal language processing in human communication. *Trends Cogn. Sci.* 23, 639–652. <https://doi.org/10.1016/j.tics.2019.05.006>.
- Homberger, D.G., 2017. Functional-morphological diversity and complexity of the larynx as a basis for vocal complexity in birds: analogies to human speech, 577.8–577.8 FASEB J. 31. [https://doi.org/10.1096/fasebj.31.1\\_supplement.577.8](https://doi.org/10.1096/fasebj.31.1_supplement.577.8).
- Iverson, J.M., Thelen, E., 1999. Hand, mouth and brain. The dynamic emergence of speech and gesture. *J. Conscious. Stud.* 6, 19–40.
- James, A.P., 2015. Heart rate monitoring using human speech spectral features. *Hum. Cent. Comput. Inf. Sci.* 5, 33. <https://doi.org/10.1186/s13673-015-0052-z>.
- Janik, V.M., Sayigh, L.S., 2013. Communication in bottlenose dolphins: 50 years of signature whistle research. *J. Comp. Physiol. A* 199, 479–489. <https://doi.org/10.1007/s00359-013-0817-7>.
- Janik, V.M., Slater, P.J.B., 1997. Vocal Learning in Mammals. In: Slater, P.J.B., Rosenblatt, J.S., Snowdon, C.T., Milinski, M. (Eds.), *Advances in the Study of Behavior*. Academic Press, pp. 59–99. [https://doi.org/10.1016/S0065-3454\(08\)60377-0](https://doi.org/10.1016/S0065-3454(08)60377-0).
- Jenkins, T., Pouw, W., 2023. Gesture-speech coupling in persons with aphasia: a kinematic-acoustic analysis. *J. Exp. Psychol. Gen.* 152, 1469–1483. <https://doi.org/10.1037/xge0001346>.

- Keller, P.E., Novembre, G., Hove, M.J., 2014. Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130394. <https://doi.org/10.1098/rstb.2013.0394>.
- Kelso, J.A., Tuller, B., 1984a. Converging evidence in support of common dynamical principles for speech and movement coordination. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 246, R928–R935. <https://doi.org/10.1152/ajpregu.1984.246.6.R928>.
- Kelso, J.A., Tuller, B., 1984b. Converging evidence in support of common dynamical principles for speech and movement coordination. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 246, R928–R935. <https://doi.org/10.1152/ajpregu.1984.246.6.R928>.
- Kirschner, S., Tomasello, M., 2009. Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. <https://doi.org/10.1016/j.jecp.2008.07.005>.
- Klump, G.M., Gerhardt, H.C., 1992. Mechanisms and Function of Call-Timing in Male-Male Interactions in Frogs. In: McGregor, P.K. (Ed.), *Playback and Studies of Animal Communication*. Springer US, Boston, MA, pp. 153–174. [https://doi.org/10.1007/978-1-4757-6203-7\\_11](https://doi.org/10.1007/978-1-4757-6203-7_11).
- Knoblich, G., Butterfill, S., Sebanz, N., 2011. Chapter three - Psychological Research on Joint Action: Theory and Data. In: Ross, B.H. (Ed.), *Psychology of Learning and Motivation, Advances in Research and Theory*. Academic Press, pp. 59–101. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>.
- Kocsis, K., Duengen, D., Jadoul, Y., Ravignani, A., 2024. Harbour seals use rhythmic percussive signalling in interaction and display. *Animal Behaviour* 207, 223–234.
- Kotz, S.A., Ravignani, A., Fitch, W.T., 2018. The evolution of rhythm processing. *Trends Cogn. Sci.* 22, 896–910. <https://doi.org/10.1016/j.tics.2018.08.002>.
- Kriesel, H.J., Le Bohec, C., Cerwenka, A.F., Hertel, M., Robin, J.-P., Ruthensteiner, B., Gahr, M., Aubin, T., Düring, D.N., 2020. Vocal tract anatomy of king penguins: morphological traits of two-voiced sound production. *Front Zool.* 17, 5. <https://doi.org/10.1186/s12983-020-0351-8>.
- Ladich, F., Winkler, H., 2017. Acoustic communication in terrestrial and aquatic vertebrates. *J. Exp. Biol.* 220, 2306–2317. <https://doi.org/10.1242/jeb.132944>.
- LaMar, M.D., n.d. Human acoustics: From vocal chords to inner ear (Ph.D.).
- Lancaster, W.C., Henson, O.W., Jr, Keating, A.W., 1995. Respiratory muscle activity in relation to vocalization in flying bats. *J. Exp. Biol.* 198, 175–191. <https://doi.org/10.1242/jeb.198.1.175>.
- Landrau-giovannetti, N., Mignucci-giannoni, A.A., Reidenberg, J.S., 2014. Acoustical and anatomical determination of sound production and transmission in West Indian (*Trichechus manatus*) and Amazonian (*T. inunguis*) Manatees. *Anat. Rec.* 297, 1896–1907. <https://doi.org/10.1002/ar.22993>.
- Laplagne, D.A., Elias Costa, M., 2016. Rats synchronize locomotion with ultrasonic vocalizations at the subsecond time scale. *Front. Behav. Neurosci.* 10.
- Large, E.W., Jones, M.R., 1999a. The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>.
- Large, E.W., Jones, M.R., 1999b. The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>.
- Larsen, O.N., Reichmuth, C., 2021. Walrus produce intense impulse sounds by clap-induced cavitation during breeding displays. *Royal Society open science* 8 (6), 210197.
- Larson, V., 2013. A rose blooms in the winter: the tradition of the hortus conclusus and its significance as a devotional emblem. *Dialog* 52, 303–312. <https://doi.org/10.1111/dial.12066>.
- Larsson, M., Richter, J., Ravignani, A., 2019. Bipodal steps in the development of rhythmic behavior in humans, 2059204319892617 *Music Sci.* 2. <https://doi.org/10.1177/2059204319892617>.
- Latash, M.L., 2008. *Synergy*. Oxford University Press.
- Lewis, E.R., Narins, P.M., Cortopassi, K.A., Yamada, W.M., Poinar, E.H., Moore, S.W., Yu, X., 2001. Do male white-lipped frogs use seismic signals for intraspecific communication? *Am. Zool.* 41, 1185–1199. <https://doi.org/10.1093/icb/41.5.1185>.
- Lim, M., Li, D., 2004. Courtship and male-male agonistic behaviour of *Cosmophasis umbratica* Simon, an ornate jumping spider (Araneae: Salticidae) from Singapore. *Raffles Bull. Zool.* 52.
- Liu, W., Landstrom, M., Cealie, M., MacKillop, I., 2022. A juvenile locomotor program promotes vocal learning in zebra finches. *Commun. Biol.* 5, 1–9. <https://doi.org/10.1038/s42003-022-03533-3>.
- Lobel, P.S., Mann, D.A., 1995. Spawning sounds of the damselfish, *dascyllus albisella* (pomacentridae), and relationship to male size. *Bioacoustics* 6, 187–198. <https://doi.org/10.1080/09524622.1995.9753289>.
- Loucks, T.M.J., Poletto, C.J., Simonyan, K., Reynolds, C.L., Ludlow, C.L., 2007. Human brain activation during phonation and exhalation: common volitional control for two upper airway functions. *NeuroImage* 36, 131–143. <https://doi.org/10.1016/j.neuroimage.2007.01.049>.
- MacLarnon, A.M., Hewitt, G.P., 1999. The evolution of human speech: the role of enhanced breathing control. *Am. J. Phys. Anthropol.* 109, 341–363. [https://doi.org/10.1002/\(SICI\)1096-8644\(199907\)109:3<341::AID-AJPA5>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2).
- MacNeillage, P.F., 2010. *The Origin of Speech*. Oxford University Press.
- Madsen, P.T., Siebert, U., Elemans, C.P.H., 2023. Toothed whales use distinct vocal registers for echolocation and communication. *Science* 379, 928–933. <https://doi.org/10.1126/science.adc9570>.
- Maina, J.N., 2002. Structure, function and evolution of the gas exchangers: comparative perspectives. *J. Anat.* 201, 281–304. <https://doi.org/10.1046/j.1469-7580.2002.00099.x>.
- Malafouris, L., 2016. *How Things Shape the Mind* [WWW Document]. MIT Press. <https://mitpress.mit.edu/9780262528924/how-things-shape-the-mind/> (accessed 4.3.24).
- Marler, P., 1967. Animal communication signals. *Science* 157, 769–774. <https://doi.org/10.1126/science.157.3790.769>.
- Mårup, S.H., Möller, C., Vuust, P., 2022. Coordination of voice, hands and feet in rhythm and beat performance. *Sci. Rep.* 12, 8046. <https://doi.org/10.1038/s41598-022-11783-8>.
- Masters, W.M., 1979. Insect disturbance stridulation: its defensive role. *Behav. Ecol. Sociobiol.* 5, 187–200.
- Merker, B.H., Madison, G.S., Eckerdal, P., 2009. On the role and origin of isochrony in human rhythmic entrainment. Special Issue on “The Rhythmic Brain” *Cortex* 45, 4–17. <https://doi.org/10.1016/j.cortex.2008.06.011>.
- Mesleh, A., Skopin, D., Baglikov, S., Quteishat, A., 2012. Heart rate extraction from vowel speech signals. *J. Comput. Sci. Technol.* 27, 1243–1251. <https://doi.org/10.1007/s11390-012-1300-6>.
- Miller, N.A., Gregory, J.S., Aspden, R.M., Stollery, P.J., Gilbert, F.J., 2014. Using active shape modeling based on MRI to study morphologic and pitch-related functional changes affecting vocal structures and the airway. *J. Voice* 28, 554–564. <https://doi.org/10.1016/j.jvoice.2013.12.002>.
- Mitoyen, C., Quigley, C., Fusani, L., 2019. Evolution and function of multimodal courtship displays. *Ethology* 125, 503–515. <https://doi.org/10.1111/eth.12882>.
- Miyata, K., Kudo, K., 2014. Mutual stabilization of rhythmic vocalization and whole-body movement. *PLOS ONE* 9, e115495. <https://doi.org/10.1371/journal.pone.0115495>.
- Mott, F., 1924. A study by serial sections of the structure of the larynx of hylobates syndactylus (Siamang Gibbon). *Proc. Zool. Soc. Lond.* 94, 1161–1170. <https://doi.org/10.1111/j.1096-3642.1924.tb03336.x>.
- Müller, A.E., Anzenberger, G., 2002. Duetting in the titi monkey *callicebus cupreus*: structure, pair specificity and development of duets. *Folia Primatol.* 73, 104–115. <https://doi.org/10.1159/000064788>.
- Müller, V., Delius, J.A.M., Lindenberg, U., 2018. Complex networks emerging during choir singing. *Ann. N. Y. Acad. Sci.* 1431, 85–101. <https://doi.org/10.1111/nyas.13940>.
- Müller, V., Ohström, K.-R.P., Lindenberg, U., 2021. Interactive brains, social minds: neural and physiological mechanisms of interpersonal action coordination. *Neurosci. Biobehav. Rev.* 128, 661–677. <https://doi.org/10.1016/j.neubiorev.2021.07.017>.
- Myrberg, A.A., Mohler, M., Catala, J.D., 1986. Sound production by males of a coral reef fish (*Pomacentrus partitus*): its significance to females. *Anim. Behav.* 34, 913–923. [https://doi.org/10.1016/S0003-3472\(86\)80077-X](https://doi.org/10.1016/S0003-3472(86)80077-X).
- Narins, P.M., 1990. Seismic communication in anuran amphibians. *BioScience* 40, 268–274. <https://doi.org/10.2307/1311263>.
- Nishimura, T., 2020. Primate Vocal Anatomy and Physiology: Similarities and Differences Between Humans and Nonhuman Primates. In: Masataka, N. (Ed.), *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism*. Springer, Singapore, pp. 25–53. [https://doi.org/10.1007/978-981-15-4250-3\\_2](https://doi.org/10.1007/978-981-15-4250-3_2).
- Orlikoff, R.F., 1990. Heartbeat-related fundamental frequency and amplitude variation in healthy young and elderly male voices. *J. Voice* 4, 322–328. [https://doi.org/10.1016/S0892-1997\(05\)80049-9](https://doi.org/10.1016/S0892-1997(05)80049-9).
- Orlikoff, R.F., 2008. Voice production during a weightlifting and support task. *Folia Phoniatr. Et. Logop.* 60, 188–194. <https://doi.org/10.1159/000128277>.
- Orlikoff, R.F., Baken, R.J., 1989. The effect of the heartbeat on vocal fundamental frequency perturbation. *J. Speech Hear Res.* 32, 576–582. <https://doi.org/10.1044/jshr.3203.576>.
- Ota, N., Gahr, M., Soma, M., 2015. Tap dancing birds: the multimodal mutual courtship display of males and females in a socially monogamous songbird. *Sci. Rep.* 5, 16614. <https://doi.org/10.1038/srep16614>.
- Otte, D., 1992. Evolution of Cricket Songs. *J. Orthoptera Res.* 25–49. <https://doi.org/10.2307/3503559>.
- Parrell, B., Goldstein, L., Lee, S., Byrd, D., 2014. Spatiotemporal coupling between speech and manual motor actions. *J. Phon.* 42, 1–11. <https://doi.org/10.1016/j.wocn.2013.11.002>.
- Partan, S., Marler, P., 1999. Communication goes multimodal. *Science* 283, 1272–1273. <https://doi.org/10.1126/science.283.5406.1272>.
- Partan, S.R., Marler, P., 2005. Issues in the classification of multimodal communication signals. *Am. Nat.* 166, 231–245. <https://doi.org/10.1086/431246>.
- Patel, A.D., Iversen, J.R., Bregman, M.R., Schulz, I., 2009. Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830. <https://doi.org/10.1016/j.cub.2009.03.038>.
- Patel, A.D., Iversen, J.R., 2014. The evolutionary neuroscience of musical beat perception: the action simulation for auditory prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* 8.
- Pearson, L., Pouw, W., 2022. Gesture-vocal coupling in Karnatak music performance: a neuro-bodily distributed aesthetic entanglement. *Ann. N. Y. Acad. Sci.* 1515, 219–236. <https://doi.org/10.1111/nyas.14806>.
- Peek, F.W., 1972. An experimental study of the territorial function of vocal and visual display in the male red-winged blackbird (*Agelaius phoeniceus*). *Anim. Behav.* 20, 112–118. [https://doi.org/10.1016/S0003-3472\(72\)80180-5](https://doi.org/10.1016/S0003-3472(72)80180-5).
- Penn, D.J., Számadó, S., 2020. The handicap principle: how an erroneous hypothesis became a scientific principle. *Biol. Rev. Camb. Philos. Soc.* 95, 267–290. <https://doi.org/10.1111/brv.12563>.
- Perlman, M., Salmi, R., 2017. Gorillas may use their laryngeal air sacs for whinny-type vocalizations and male display. *J. Lang. Evol.* 2, 126–140. <https://doi.org/10.1093/jole/lzx012>.

- Petrone, C., Fuchs, S., Koenig, L.L., 2017. Relations among subglottal pressure, breathing, and acoustic parameters of sentence-level prominence in German. *J. Acoust. Soc. Am.* 141, 1715–1725. <https://doi.org/10.1121/1.4976073>.
- Phillips-Silver, J., Aktipis, C.A., Bryant, G., 2010. The ecology of entrainment: foundations of coordinated rhythmic movement. *Music Percept.* 28, 3–14. <https://doi.org/10.1525/mp.2010.28.1.3>.
- Phillips-Silver, J., Keller, P., 2012. Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6. <https://doi.org/10.3389/fnhum.2012.00026>.
- Phillips-Silver, J., Trainor, L.J., 2005. Feeling the beat: movement influences infant rhythm perception, 1430–1430 *Science* 308. <https://doi.org/10.1126/science.1110922>.
- Pikovsky, A., Kurths, J., Rosenblum, M., 2001. *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press, Cambridge, Mass.
- Podos, J., Cohn-Haft, M., 2019. Extremely loud mating songs at close range in white bellbirds. *Curr. Biol.* 29, R1068–R1069. <https://doi.org/10.1016/j.cub.2019.09.028>.
- Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* 21, 322–334. <https://doi.org/10.1038/s41583-020-0304-4>.
- Porges, S.W., Lewis, G.F., 2009. The polyvagal hypothesis: common mechanisms mediating autonomic regulation, vocalizations, and listening. *Handb. Mamm. vocalizations: Integr. Neurosci. Approach* 255–264.
- Pouw, W., Burchardt, L.S., Selen, L., 2023. Postural Muscular Eff. Up. Limb Mov. Voicing. <https://doi.org/10.1101/2023.03.08.531710>.
- Pouw, W., Fuchs, S., 2022. Origins of vocal-entangled gesture. *Neurosci. Biobehav. Rev.* 141, 104836. <https://doi.org/10.1016/j.neubiorev.2022.104836>.
- Pouw, W., Harrison, S.J., Dixon, J.A., 2022. The importance of visual control and biomechanics in the regulation of gesture-speech synchrony for an individual deprived of proprioceptive feedback of body position. *Sci. Rep.* 12, 14775. <https://doi.org/10.1038/s41598-022-18300-x>.
- Pouw, W., Harrison, S.J., Esteve-Gibert, N., Dixon, J.A., 2020a. Energy flows in gesture-speech physics: the respiratory-vocal system and its coupling with hand gestures. *J. Acoust. Soc. Am.* 148, 1231–1247. <https://doi.org/10.1121/10.0001730>.
- Pouw, W., Paxton, A., Harrison, S.J., Dixon, J.A., 2020b. Acoustic information about upper limb movement in voicing. *PNAS*. <https://doi.org/10.1073/pnas.2004163117>.
- Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R.S., Wiggins, G.A., 2021. Multilevel rhythms in multimodal communication. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 20200334. <https://doi.org/10.1098/rstb.2020.0334>.
- Profeta, V.L.S., Turvey, M.T., 2018. Bernstein's levels of movement construction: a contemporary perspective. *Hum. Mov. Sci.* 57, 111–133. <https://doi.org/10.1016/j.humov.2017.11.013>.
- Provasi, J., Bobin-Bègue, A., 2003. Spontaneous motor tempo and rhythmical synchronisation in 2½- and 4-year-old children. *Int. J. Behav. Dev.* 27, 220–231. <https://doi.org/10.1080/01650250244000290>.
- Provasi, J., Doyère, V., Zélati, P.S., Kieffer, V., Perdry, H., El Massioui, N., Brown, B.L., Dellatolas, G., Grill, J., Droit-Volet, S., 2014. Disrupted sensorimotor synchronization, but intact rhythm discrimination, in children treated for a cerebellar medulloblastoma. *Res. Dev. Disabil.* 35, 2053–2068. <https://doi.org/10.1016/j.ridd.2014.04.024>.
- Rado, R., Levi, N., Hauser, H., Witcher, J., Alder, N., Intrator, N., Wollberg, Z., Terkel, J., 1987. Seismic signalling as a means of communication in a subterranean mammal. *Anim. Behav.* 35, 1249–1251. [https://doi.org/10.1016/S0003-3472\(87\)80183-5](https://doi.org/10.1016/S0003-3472(87)80183-5).
- Randall, J.A., 2001. Evolution and function of drumming as communication in mammals. *Am. Zool.* 41, 1143–1156. <https://doi.org/10.1093/icb/41.5.1143>.
- Raphael, L.J., Borden, G.J., Harris, K.S., 2007. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Lippincott Williams & Wilkins.
- Ravignani, A., Herbst, C.T., 2023. Voices in the ocean. *Science* 379, 881–882. <https://doi.org/10.1126/science.adg5256>.
- Ravignani, A., Honing, H., Kotz, S.A., 2017. The evolution of rhythm cognition: timing in music and speech. *Frontiers in human neuroscience* 11, 303.
- Ravignani, A., Olivera, V.M., Gingras, B., Hofer, R., Hernández, C.R., Sonnweber, R.-S., Fitch, W.T., 2013. Primate drum kit: a system for studying acoustic pattern production by non-human primates using acceleration and strain sensors. *Sensors* 13, 9790–9820. <https://doi.org/10.3390/s130809790>.
- Remedios, R., Logothetis, N.K., Kayser, C., 2009. Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc. Natl. Acad. Sci.* 106, 18010–18015. <https://doi.org/10.1073/pnas.0909756106>.
- Repp, B.H., 2005. Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. <https://doi.org/10.3758/BF03206433>.
- Repp, B.H., Su, Y.-H., 2013. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* 20, 403–452. <https://doi.org/10.3758/s13423-012-0371-2>.
- Richardson, M.J., Marsh, K.L., Isenhower, R.W., Goodman, J.R.L., Schmidt, R.C., 2007. Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. <https://doi.org/10.1016/j.humov.2007.07.002>.
- Richter, J., Ostovar, R., 2016. “It Don’t Mean a Thing if It Ain’t Got that Swing” – an alternative concept for understanding the evolution of dance and music in human beings. *Front. Hum. Neurosci.* 10.
- Riede, T., Beckers, G.J.L., Blevins, W., Suthers, R.A., 2004. Inflation of the esophagus and vocal tract filtering in ring doves. *J. Exp. Biol.* 207, 4025–4036. <https://doi.org/10.1242/jeb.01256>.
- Riede, T., Eliason, C.M., Miller, E.H., Goller, F., Clarke, J.A., 2016. Coos, booms, and hoots: The evolution of closed-mouth vocal behavior in birds. *Evolution* 70, 1734–1746. <https://doi.org/10.1111/evo.12988>.
- Riede, T., Goller, F., 2010. Peripheral mechanisms for vocal production in birds – differences and similarities to human speech and singing (Special Issue on Language and Birdsong). *Brain Lang.* 115, 69–80. <https://doi.org/10.1016/j.bandl.2009.11.003>.
- Riede, T., Titze, I.R., 2008. Vocal fold elasticity of the Rocky Mountain elk (*Cervus elaphus nelsoni*) – producing high fundamental frequency vocalization with a very long vocal fold. *J. Exp. Biol.* 211, 2144–2154. <https://doi.org/10.1242/jeb.017004>.
- Riede, T., Tokuda, I.T., Munger, J.B., Thomson, S.L., 2008. Mammalian laryngeal air sacs add variability to the vocal tract impedance: physical and computational modeling. *J. Acoust. Soc. Am.* 124, 634–647. <https://doi.org/10.1121/1.2924125>.
- Risueno-Segovia, C., Hage, S.R., 2020. Theta synchronization of phonatory and articulatory systems in marmoset monkey vocal production. *Curr. Biol.* 30, 4276–4283.e3. <https://doi.org/10.1016/j.cub.2020.08.019>.
- Rocha, S., Mareschal, D., 2017. Getting into the groove: the development of tempo-flexibility between 10 and 18 months of age. *Infancy* 22, 540–551. <https://doi.org/10.1111/inf.12169>.
- Rojas, J.A.M., Hermosilla, J.A., Montero, R.S., Espi, P.L.L., 2009. Physical analysis of several organic signals for human echolocation: oral vacuum pulses. *Acta Acust. U Acust.* 95, 325–330. <https://doi.org/10.3813/AAA.918155>.
- Schachner, A., Brady, T.F., Pepperberg, I.M., Hauser, M.D., 2009. Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836. <https://doi.org/10.1016/j.cub.2009.03.061>.
- Seikel, J.A., Drumright, D.G., Hudock, D.J., 2023. *Anatomy & physiology for speech, language, and hearing*. Plural Publishing.
- Silva, P., Moreno, M., Mancini, M., Fonseca, S., Turvey, M.T., 2007. Steady-state stress at one hand magnifies the amplitude, stiffness, and non-linearity of oscillatory behavior at the other hand. *Neuroscience letters* 429 (1), 64–68.
- Soldati, A., Fedurek, P., Dezechache, G., Call, J., Zuberbühler, K., 2022. Audience sensitivity in chimpanzee display pant hoots. *Anim. Behav.* 190, 23–40. <https://doi.org/10.1016/j.anbehav.2022.05.010>.
- Soma, M., Shibata, M., 2023. Dancing in Singing Songbirds: Choreography in Java Sparrows. In: Seki, Y. (Ed.), *Acoustic Communication in Animals: From Insect Wingbeats to Human Music* (Bioacoustics Series Vol.1. Springer Nature, Singapore, pp. 95–111. [https://doi.org/10.1007/978-981-99-0831-8\\_6](https://doi.org/10.1007/978-981-99-0831-8_6).
- Starnberger, I., Preininger, D., Hödl, W., 2014. The anuran vocal sac: a tool for multimodal signalling. *Anim. Behav.* 97, 281–288. <https://doi.org/10.1016/j.anbehav.2014.07.027>.
- Stewart, A.M., Lewis, G.F., Yee, J.R., Kenkel, W.M., Davila, M.I., Sue Carter, C., Porges, S.W., 2015. Acoustic features of prairie vole (*Microtus ochrogaster*) ultrasonic vocalizations covary with heart rate. *Physiol. Behav.* 138, 94–100. <https://doi.org/10.1016/j.physbeh.2014.10.011>.
- Stoeger, A.S., Mietchen, D., Oh, S., de Silva, S., Herbst, C.T., Kwon, S., Fitch, W.T., 2012. An asian elephant imitates human speech. *Curr. Biol.* 22, 2144–2148. <https://doi.org/10.1016/j.cub.2012.09.022>.
- Sundberg, J., 1993. BREATHING BEHAVIOR DURING SINGING.
- Sundberg, J., Titze, I., Scherer, R., 1993. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. The Voice Foundation's 22nd Annual Symposium. *J. Voice* 7, 15–29. [https://doi.org/10.1016/S0892-1997\(05\)80108-0](https://doi.org/10.1016/S0892-1997(05)80108-0).
- Suthers, R.A., 1990. Contributions to birdsong from the left and right sides of the intact syrinx. *Nature* 347, 473–477. <https://doi.org/10.1038/347473a0>.
- Suthers, R.A., Thomas, S.P., Suthers, B.J., 1972. Respiration, wing-beat and ultrasonic pulse emission in an echo-locating bat. *J. Exp. Biol.* 56, 37–48. <https://doi.org/10.1242/jeb.56.1.37>.
- Švec, J.G., Schutte, H.K., Chen, C.J., Titze, I.R., 2023. Integrative insights into the myoelectric-aerodynamic theory and acoustics of phonation. scientific tribute to donald G. Miller. *J. Voice* 37, 305–313. <https://doi.org/10.1016/j.jvoice.2021.01.023>.
- Tagliatela, J.P., Russell, J.L., Pope, S.M., Morton, T., Bogart, S., Reamer, L.A., Schapiro, J.J., Hopkins, W.D., 2015. Multimodal communication in chimpanzees. *Am. J. Primatol.* 77, 1143–1148. <https://doi.org/10.1002/ajp.22449>.
- Taylor, A.M., Charlton, B.D., Reby, D., 2016. Vocal Production by Terrestrial Mammals: Source, Filter, and Function. In: Suthers, R.A., Fitch, W.T., Fay, R.R., Popper, A.N. (Eds.), *Vertebrate Sound Production and Acoustic Communication*. Springer International Publishing, Cham, pp. 229–259. [https://doi.org/10.1007/978-3-319-27721-9\\_8](https://doi.org/10.1007/978-3-319-27721-9_8).
- ter Bekke, M., Levinson, S.C., van Otterdijk, L., Kühn, M., Holler, J., 2024. Visual bodily signals and conversational context benefit the anticipation of turn ends. *Cognition* 248, 105806. <https://doi.org/10.1016/j.cognition.2024.105806>.
- Thomas, S.P., Suthers, R.A., 1972. The physiology and energetics of bat flight. *J. Exp. Biol.* 57, 317–335. <https://doi.org/10.1242/jeb.57.2.317>.
- Tinbergen, N., 1963. On aims and methods of Ethology. *Z. f. uR. Tierpsychol.* 20, 410–433. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>.
- Titze, I.R., 1992. Vocal efficiency. *J. Voice* 6, 135–138. [https://doi.org/10.1016/S0892-1997\(05\)80127-4](https://doi.org/10.1016/S0892-1997(05)80127-4).
- Titze, I.R., 2008. The human instrument. *Sci. Am.* 298, 94–101.
- Titze, I.R., Martin, D.W., 1998. Principles of voice production. *J. Acoust. Soc. Am.* 104, 1148. <https://doi.org/10.1121/1.424266>.
- Todt, D., Fiebelkorn, A., 1980. Display, timing and function of wing movements accompanying antiphonal duets of *cichlasda guttata*. *Behaviour* 72, 82–105. <https://doi.org/10.1163/156853980X00069>.
- Tokimura, H., Asakura, T., Tokimura, Y., Oliviero, A., Rothwell, J.C., 1996. Speech-induced changes in corticospinal excitability. *Ann. Neurol.* 40, 628–634. <https://doi.org/10.1002/ana.410400413>.
- Tokuda, I., 2021. The source-filter theory of speech. *Oxf. Res. Encycl. Linguist.* <https://doi.org/10.1093/acrefore/9780199384655.013.894>.



- Tong, J.Y., Sataloff, R.T., 2022. Respiratory function and voice: the role for airflow measures. *J. Voice* 36, 542–553. <https://doi.org/10.1016/j.jvoice.2020.07.019>.
- Treffner, P., Peter, M., 2002. Intentional and attentional dynamics of speech–hand coordination (Current issues in motor control and coordination). *Hum. Mov. Sci.* 21, 641–697. [https://doi.org/10.1016/S0167-9457\(02\)00178-1](https://doi.org/10.1016/S0167-9457(02)00178-1).
- Trujillo, J.P., Holler, J., 2023. Interactionally embedded gestalt principles of multimodal human communication. *Perspect. Psychol. Sci.*, 17456916221141422 <https://doi.org/10.1177/17456916221141422>.
- Turner, J.S., 2009. *The Extended Organism: The Physiology of Animal-Built Structures*. Harvard University Press.
- Turvey, M.T., 1990. Coordination. *Am. Psychol.* 45, 938–953. <https://doi.org/10.1037/0003-066X.45.8.938>.
- Turvey, M.T., Fonseca, S.T., 2014. The medium of haptic perception: a tensegrity hypothesis. *J. Mot. Behav.* 46, 143–187. <https://doi.org/10.1080/00222895.2013.798252>.
- Van den Berg, J., 1958. Myoelastic-aerodynamic theory of voice production. *J. Speech Hear. Res.* 1, 227–244. <https://doi.org/10.1044/jshr.0103.227>.
- van Noorden, L., Moelants, D., 1999. Resonance in the perception of musical pulse. *J. N. Music Res.* 28, 43–66. <https://doi.org/10.1076/jnmr.28.1.43.3122>.
- Vorperian, H.K., Kurtzweil, S.L., Fourakis, M., Kent, R.D., Tillman, K.K., Austin, D., 2015. Effect of body position on vocal tract acoustics: acoustic pharyngometry and vowel formants. *J. Acoust. Soc. Am.* 138, 833–845. <https://doi.org/10.1121/1.4926563>.
- Walle, E.A., Campos, J.J., 2014. Infant language development is related to the acquisition of walking. *Dev. Psychol.* 50, 336–348. <https://doi.org/10.1037/a0033238>.
- Werner, R., Selen, L., Pouw, W., 2024. Arm movements increase acoustic markers of expiratory flow, in: *Proceedings of the International Conference of Speech Prosody (SpeechProsody2024)*. Leiden. <https://doi.org/10.1101/2024.01.08.574611>.
- Wilkins, H.D., Ritchison, G., 1999. Drumming and tapping by red-bellied woodpeckers: description and possible causation (Descripción y Posible Causa del Tamborileo y Picoteo Lento en *Melanerpes carolinus*). *J. Field Ornithol.* 70, 578–586.
- Wilson, M., Cook, P.F., 2016. Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic bulletin & review* 23, 1647–1659.
- Wöllner, C., 2018. Call and response: Musical and bodily interactions in jazz improvisation duos. *Musica Sci.* <https://doi.org/10.1177/1029864918772004>.
- Yovel, Y., Geva-Sagiv, M., Ulanovsky, N., 2011. Click-based echolocation in bats: not so primitive after all. *J. Comp. Physiol. A* 197, 515–530. <https://doi.org/10.1007/s00359-011-0639-4>.
- Zelic, G., Kim, J., Davis, C., 2015. Articulatory constraints on spontaneous entrainment between speech and manual gesture. *Hum. Mov. Sci.* 42, 232–245. <https://doi.org/10.1016/j.humov.2015.05.009>.
- Zentner, M., Eerola, T., 2010. Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci.* 107, 5768–5773. <https://doi.org/10.1073/pnas.1000121107>.
- Zhang, Z., 2016. Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *J. Acoust. Soc. Am.* 139, 1493–1507. <https://doi.org/10.1121/1.4944754>.