



Origins of vocal-entangled gesture

Wim Pouw^{a,*}, Susanne Fuchs^b

^a Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, the Netherlands

^b Leibniz Center General Linguistics, Berlin, Germany

ARTICLE INFO

Keywords:

Beat gestures
Manual gesture
Biomechanics
Locomotor-respiratory coupling
Cross-species comparison
Evolutionary biology
Multimodal prosody
Ontogeny
Phylogeny

ABSTRACT

Gestures during speaking are typically understood in a representational framework: they represent absent or distal states of affairs by means of pointing, resemblance, or symbolic replacement. However, humans also gesture along with the rhythm of speaking, which is amenable to a non-representational perspective. Such a perspective centers on the phenomenon of vocal-entangled gestures and builds on evidence showing that when an upper limb with a certain mass decelerates/accelerates sufficiently, it yields impulses on the body that cascade in various ways into the respiratory–vocal system. It entails a physical entanglement between body motions, respiration, and vocal activities. It is shown that vocal-entangled gestures are realized in infant vocal–motor babbling before any representational use of gesture develops. Similarly, an overview is given of vocal-entangled processes in non-human animals. They can frequently be found in rats, bats, birds, and a range of other species that developed even earlier in the phylogenetic tree. Thus, the origins of human gesture lie in biomechanics, emerging early in ontogeny and running deep in phylogeny.

1. General introduction

Co-speech gestures are a special class of upper-limb movements that occur during speaking and have been referred to as a “window on the mind” (Goldin-Meadow, 2003). They have become a pivotal topic in anthropology, language evolution, (psycho-)linguistics, psychology, neuroscience, and cognitive science (Feyereisen, 2017; Goldin-Meadow and Brentari, 2017; Kendon, 2004; McNeill, 2000). Gestures as windows onto the mind allude to their referential qualities. They can, for example, bring a state of affairs to the fore by pointing to a distal object or depicting an absent one. Gestures also serve as pragmatic signs regulating conversation (Holler and Levinson, 2019). The varied ways gestures can fulfill a representational function have been the primary topic of concern for researchers studying gesture (de Ruiter, 2007; Feyereisen, 2017; Goldin-Meadow and Brentari, 2017; Hostetter and Alibali, 2008; Kendon, 2004; Novack and Goldin-Meadow, 2017).

However, humans also gesture in a way that is best not characterized as representational. A non-representational approach to gesture promotes a distributed source of meaning that is co-reliant on bodily dynamics (Cuffari, 2012; Darwin, 1948; LeCron Foster, 1992; Merleau-Ponty, 1945; Morgenstern et al., 2021; Paolo et al., 2018; Rączaszek-Leonardi and Kelso, 2008; Sheets-Johnstone, 2011). Consider

that salient and rapid changes in manual movement (i.e., pulses) are often temporally synchronized with vocalizations (McNeill, 1992). Such gestures are frequently understood to “beat” (McNeill, 1992) or “rhythmically pulse” (Leonard and Cummins, 2011) with prosodic aspects of speaking (Casasanto and Jasmin, 2010; Efron et al., 1972; McNeill, 2005; Wagner et al., 2014). Even when co-speech gestures represent through depiction or pointing, they often still retain this overlying vocal-synchronized quality (Wagner et al., 2014; Ginosar et al., 2019). We refer to this phenomenon as vocal-entangled gesture and we will argue that such gestures have communicative potential by serving as an index for some embodied state of affairs rather than as a representation of purely mental content.

The purpose of this review is to show that vocal-entanglement is a fundamental aspect of gesture – it emerges early in ontogeny and runs deep in phylogeny. The importance of vocal-entangled gesture is rarely, if at all, mentioned in influential evolutionary accounts of gesture (cf. Arbib, 2005; Corballis, 2002; Gärdenfors, 2017; Kendon, 2017; Levinson and Holler, 2014; Sterelny, 2012; Tomasello, 2008; Zwyciszynski et al., 2018). When vocal-entangled gestures defined as “beat” gestures are mentioned, they are held to emerge 1.5 million (!) years later than the use of representational gestures in the hominin lineage (Fröhlich et al., 2019). The common view that beat gestures are complex human devices

* Corresponding author.

E-mail addresses: wim.pouw@donders.ru.nl (W. Pouw), fuchs@leibniz-zas.de (S. Fuchs).

¹ Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

that were invented after some basic language was in place is comparable to approaches in ontogeny. There it is suggested that infants only learn to use beat gestures after 33 months of age, which is 18 (!) months later than the latest emergence of iconic representational gesturing (Esteve-Gibert and Guellai, 2018; Capone and McGregor, 2004; Iverson and Thelen, 1999). In comprehensive theories about the cognitive and linguistic functions of gestures, the vocal-entangled aspect of gesture is often not mentioned at all, or is only mentioned as an afterthought by admitting that beat gestures are not within the scope of the theory (Goldin-Meadow and Brentari, 2017; Hostetter and Alibali, 2019; Kita and Özyürek, 2003; Krauss et al., 2000; Murgiano et al., 2021; Novack and Goldin-Meadow, 2017; Pouw et al., 2014), with some exceptions (McNeill, 2005; Iverson and Thelen, 1999; Rusiewicz and Esteve-Gibert, 2018). Thus, the vocal-entangled nature of gesture is frequently neglected in theoretical accounts of multimodal communication, no matter what the explanatory goal (phylogeny, ontogeny, mechanistic).

We can only speculate as to why this is. Firstly, the study of representational aspects of gesture over non-representational ones is promoted due to the historical makeup of the field of gesture studies, which is concerned with semiotics and cognition (Feyereisen, 2017; Kendon, 2004), but not with movement and biomechanics. That the pulse quality of gesture does not figure into phylogenetic accounts of gesture may have a different reason: gestures are often (implicitly) burdened with explaining how the *faculty of language* emerged (Fitch, 2010). Gestures then figure in heavy-duty arguments a) that the development of language was kick-started by early pantomimic gesture practices, as they provided the first instances of the capacity of displaced reference, and b) that the manual system was the first communicative system that allowed for some intentional control (Arbib, 2005; Corballis, 2002; Gärdenfors, 2017; Sterelny, 2012). As it is admittedly difficult to propose that a proto-version of “beat” gestures can kick-start language proper, the non-representational pulsing aspect of gesture never figures in heavy-duty arguments about the (gestural) origins of language.

Finally, definitions that are in use by researchers can sometimes obscure the continuities in behavior within and between species. In gesture studies, the more restrictive term beat gesture requires it to fulfill some pragmatic function within the discourse (McNeill, 1992). Pulsing movements that are often synchronized with vocalization in infants and non-human animals are therefore not deemed to be a beat gesture. The strict definition has thereby fortified the idea that representationhood is ontogenetically and phylogenetically primary to the faculty of gesture.

But isn't it likely that more sophisticated cognitive skills such as representation are preceded by and arise out of not-yet representational skills? We aim to show that the cognitive modesty of vocal-entangled gestures can contribute to the grounding of human gesture in a much broader cross-species and gradually developing history (Darwin, 1863; DeSilva, 2021) that connects bipedalism, respiratory-motor coupling, and the complexification of the respiratory-vocal system, a preadaptation to the faculty of speech. This phylogenetic history is mirrored (MacNeilage, 2010) in infant vocal-motor babbling, and indeed much earlier than an infant can refer to absent states of affairs. This account broadens and deepens the potential origins of human gesture.

To unpack these issues, **Section 2** overviews the phenomenon of human multimodal prosody, linking gestures with vocal properties of speech. **Section 3** introduces an alternative perspective that grounds multimodal prosody in biomechanics. It highlights how the different systems are *physically connected* – that the pulse quality of gestures cascades onto the respiratory-vocal system. In **Section 4**, we show how human infants explore gesture-speech biomechanics in vocal-motor babbling. Moving to phylogeny in **Section 5**, we ground gesture-speech biomechanics in the broader cross-species phenomenon of locomotor-respiratory coupling (LRc), which is then related to locomotor-respiratory-vocal coupling (LRVc). Then we show that phylogenetic theories of gestural communication are amenable to a revision that takes into account physical aspects of gesture. We conclude in **Section 6** that

connections between gesture, respiration, and vocalization emerge early in acquisition and run deep in phylogeny, inviting new research directions.

2. Human multimodal prosody

Despite the relative absence of vocal-entangled gesture in grand theories of why humans gesture, there is an active area of research in gesture studies that focuses on the pervasiveness of vocal-entangled gestures and their role in speech production and perception (Bosker and Peeters, 2021; Wagner et al., 2014), sometimes called “Multimodal Prosody.” In corpus linguistics, multimodal prosody is frequently assessed by human annotators comparing the moment of a sudden halt of a gesture stroke (the “gesture apex”) with moments of pitch accents during unscripted speech (Im and Baumann, 2020; Loehr, 2012; McClave, 1998; Mendoza-Denton and Jannedy, 2011). Such research has found that expressive moments of gesture often co-occur with sharp rises in fundamental frequency (FO) and speech intensity typical of prosodically relevant contrasts in speech. Experimental studies show that gesture and speech flexibly couple and robustly synchronize when either activity is perturbed. For example, a gesture slows down if speech is slowed down (Kelso et al., 1983; Pouw and Dixon, 2019; Stoltmann and Fuchs, 2017; Rusiewicz et al., 2013), and speech does the same with gesture if the movement must readjust due to a disruption of visual feedback (Chu and Hagoort, 2014). Indeed, gesture–speech coupling is bidirectional (Dohen and Roustan, 2017; Krahmer and Swerts, 2007; Krivokapić et al., 2017; Parrell et al., 2014; Trujillo et al., 2021). Even pointing – a referential act – aligns with the pitch accented syllable, such that saying “PApa” or “paPA” will attract the gesture apex to synchronize with the utterance of the emphasized syllable (Esteve-Gibert and Prieto, 2013; Rochet-Capellan et al., 2008).

Although it is well known that gesturing and speech style vary considerably from person to person, there is some cross-modal correspondence that listeners seem to detect by listening to speech or seeing gesture. When perceiving a pulsing gesture desynchronized with pitch accented speech, a type of bottom-up surprise effect occurs as an event-related readiness potential known as an N400 (Morett et al., 2020). The voice is even perceived contextually relative to the presence or absence of a pulsing gesture. Listeners perceiving speech with a pulsing gesture hear a lexical stress on the pulse-aligned syllable (Bosker and Peeters, 2021), much like how lips can inform about speech sounds (Fowler, 1986; McGurk and MacDonald, 1976; Pouw and Dixon, 2022). Further, computational approaches have shown that gestures can be synthesized from speech acoustics alone (Ginosar et al., 2019; Alexanderson et al., 2020). Recurrent Neural Networks can be trained on complex associations of a speaker's speech acoustics and gestural motions. After such training, they can be successfully employed to produce very natural-looking synthetic pulsing gestures based on novel speech acoustics from that person (see, e.g., <https://www.youtube.com/watch?v=xzTE5sobpFY>). While the (non-linear) relations that tie gestural movements with prosody of speech remain unknown, the success of deep neural networks provides proof that the acoustic signal is informative enough to reconstruct pulsing gestures that are entangled with speech – as if reconstructing a key (gesture) from a lock (speech).

The coordinative stabilities that hold between manual gesture and speech prosody require an explanation. According to some views, gesture–speech coupling is considered to be controlled top-down, governed by dedicated brain modules (de Ruiter, 2000; Krauss et al., 2000; Feyereisen, 2017). The alternative view proposed here is that multimodal prosody originates bottom-up from physically coupled systems that are further neurally regulated top-down in a loop-like fashion, e.g., by the constraints of a specific prosodic system of a given language. In the next section, we describe how gestures are physically linked to the respiratory and laryngeal systems.

3. Gesture-speech biomechanics

3.1. Upper-limb-respiratory links

How can gestures reach the vocal system? Biomechanics research shows that rapid upper-limb movements can interact with thoracic muscle systems, including those involved in respiration. Specifically, upper-limb movements recruit a whole ensemble of muscles (see Fig. 1), such as trunk/core muscles (transversus abdominus; Hedges and Richardson, 1997) that are of key importance to expiratory control. Even the diaphragm, a primary respiratory muscle, contracts (more forcefully) so as to maintain postural integrity during accelerative upper-limb movements (Hedges and Gandevia, 2000a; b), thereby changing intra-abdominal pressures. Pelvic floor muscles also contract in a similar anticipatory fashion to peak accelerations of arm movements, thereby increasing expiratory flow (Hedges et al., 2007). These muscle actions are peripheral to the focal muscle activations that drive upper-limb movement. They are called anticipatory and reactionary postural adjustments (APAs) as they counteract the destabilizing forces away from the body's center of mass of upper-limb movements. APAs are generally produced about 30–50 ms before and after the initiation of destabilizing limb movements (Aruin and Latash, 1995) and are part of a reflex-like system co-regulated by spinal cord and cerebellar neural feedback loops (Colnaghi et al., 2017; Latash, 2008).

The consequences of rapid upper-limb motion for respiration in turn have consequences for speech production. Yet, in speech research, the focus on respiration has been mostly dedicated to two muscle groups: 1) the diaphragm and 2) the internal and external intercostalis muscles. The diaphragm separates the abdominal cavity from the ribcage. Its contraction goes hand in hand with active inhalation. The external intercostalis muscles between the ribs are additional muscles of inhalation, which elevate and expand the ribcage. Exhalation can be passive, for example when the diaphragm relaxes or after deep inhalation where elastic recoil of the ribs is activated. Active exhalation will involve activation of the inner intercostalis muscles located between the ribs and some abdominal muscles. All these muscles have different sizes and work on different temporal scales. The diaphragm drives the overall slow breathing rhythm, with breathing cycles being on average around

4 s for speech. Faster adaptations of the respiratory system are made possible by the intercostalis muscles.

A very important conceptual point is that the lungs themselves do not include muscles, but they can be flexibly expanded or compressed. How does that work? The lungs are surrounded by a thin membrane, the pleural visceralis, and the rib cage is surrounded by another membrane, the pleural parietalis. Both constitute the pleura (see Fig. 2). The small gap between these two membranes is filled with liquid so that both are connected but can slide along each other (Perkins et al., 1986). Since the lungs are so tightly connected to the ribcage, *any motion affecting the ribcage, like moving the arm during gesturing, can have a (small) effect on lung volume and subglottal pressure*. However, the respiratory-muscular context of speech is often studied when subjects are not producing any gross body movements, and (consequently) there has only been a focus on the inner intercostalis muscles and their role in prominent syllable

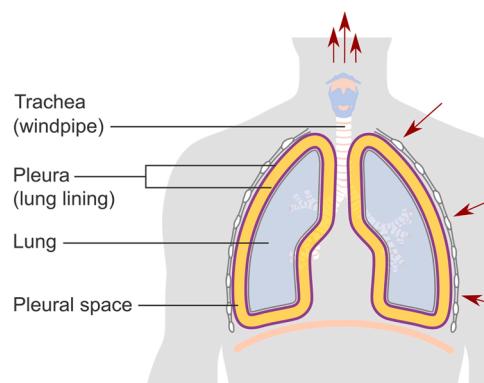


Fig. 2. Any movements of the ribcage can induce changes in lung volume. The lungs are passive and surrounded by a thin membrane, the pleural visceralis, and the rib cage is surrounded by the pleural parietalis. These two membranes are connected by the fluid-filled pleural space, which acts as a vacuum. This figure was retrieved and modified from Wikipedia and originally published under a CC BY-SA 4.0 license. Note that the pleural space depicted here is larger than in reality; it has been enlarged for visualization purposes.

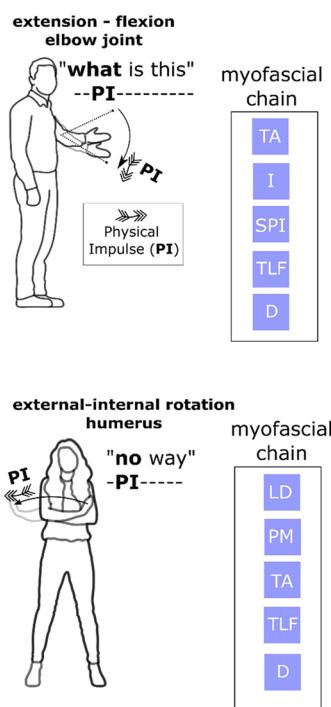
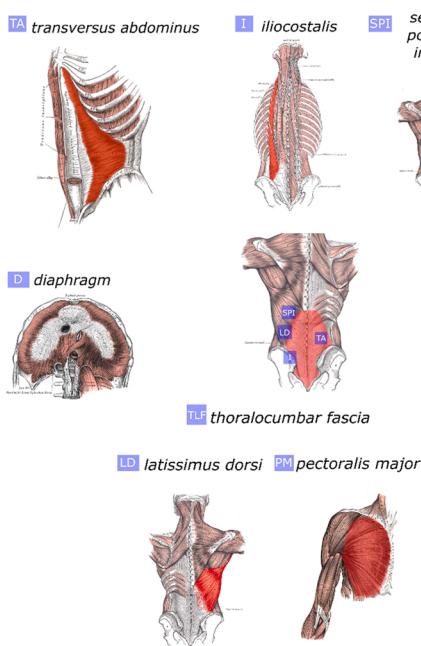


Fig. 1. Example of respiratory-related muscle systems involved in anticipatory and reactionary postural adjustments, (APAs) to upper-limb movements. On the right, we have two example gestures with a particular physical impulse on a syllable. On the left there are potential muscle groups recruited for focal action or in anticipatory/reactive fashion to maintain postural stability at the moment of a physical impulse (PI). For example, for external-internal rotation of the humerus, the pectoralis major and latissimus dorsi are focal muscles that activate and stabilize acceleration and deceleration, which will constrain the rib cage (Basmajian and de Luca, 1985; Cerqueira and Garbelline, 1999; Hoit et al., 1990). Such an action will also recruit APAs that can affect respiratory functions, such as the transversus abdominus (Hedges and Richardson, 1997) and even the diaphragm, which is shown in transverse perspective here (Hedges and Gandevia, 2000a; b). Furthermore, these muscle systems interconnect with a complex net of connective tissues called the thoralocumbar fascia (Willard et al., 2012) (the myofascial chain units are presented in arbitrary order). This figure was retrieved from Pouw and colleagues, reused and adapted with permission by the authors.

production (for an overview, see Petrone et al., 2017; Fuchs et al., 2019).

It is important to further note that the respiratory flow is modulated by the larynx – an upper valve on top of the trachea connecting to the lungs, regulating how much air can escape during exhalation or stream in during inhalation. Its primary function is to protect the lungs from foreign bodies, as well as to stabilize the thorax by holding one's breath during vigorous movement, and it is of course crucial for speaking and phonation. Changes in the stiffness of the vocal folds and the degree of glottal opening also impact lung volume and subglottal pressure.

Since the upper-limb system is close to the respiratory–vocal system, biomechanical interactions are likely to arise (Klein and Codd, 2010), as exemplified in Fig. 1. Indeed, people with respiratory problems (Chronic Obstructive Lung Disease: COPD) recruit shoulder and elbow muscles to assist in breathing during exercise (Dourado et al., 2006). COPD patients' inspiratory and expiratory capacities correlate with upper-limb muscle strength, suggesting some interdependence between these systems (Liu et al., 2019). COPD patients also coordinate breathing and upper-limb movement so as not to hamper respiratory functioning (Dolmäge et al., 2013). Given this line of research, Liu and colleagues (2019) go as far as to conclude that when primary respiratory drivers are impaired, "muscles around the shoulder and distal muscles around the elbow are accessory respiratory muscles" (p. 2032).

Anatomical facts can explain why upper-limb movements would distribute their mechanical effects over the thoracic musculoskeletal system (see Fig. 3). When inspecting the shoulder girdle, one can see that it is suspended by the clavicle (Levin, 1997). The clavicle is ill-poised to withstand any vertical loading from the upper limbs (e.g., while carrying your groceries) as it is horizontally oriented. Indeed, in humans, the clavicle bone breaks often. However, the reason that humans do not always break the clavicle when moving our upper limbs is that the shoulder girdle forms a tent-like structure, with compressive bone elements, which are interconnected by tensile elements in the form of muscles and connective tissue called fascia (Moccia et al., 2016).

The web-like suspension mechanism that stabilizes the scapula and shoulder joint is a prime example of a (bio-)tensegrity structure (Levin, 1997, 2006). Its architecture enables the tensile and flexible elements to exercise a pre-stress on the system, thereby allowing it to be perturbed and deform without collapse. When pushing or pulling on an element of the tensegrity system, forces are absorbed in a distributed and non-linear way over the elements of the system (Ingber, 2008; Silva et al., 2007; Turvey and Fonseca, 2014). What tensegrity thus implies is that actions happening locally can reverberate more globally, which is a helpful biological design feature when coordinating multiple musculoskeletal units in a synergetic way (Profeta and Turvey, 2018).

Thus, gestures are actually whole-body coordinations performed by pre-stressed musculoskeletal chains that directly constrain the thoracic region and thereby respiratory–vocal control. In what follows, we move towards more direct evidence for the gesture–speech biomechanics thesis: *when an upper-limb segment with a certain mass (or multiple segments with a certain combined mass) sufficiently accelerates or decelerates, it yields physical impulses on the musculoskeletal system, the cascading mechanical effects of which will constrain respiratory–vocal activity.*

3.2. Upper-limb–respiratory–vocal biomechanics

In singing pedagogy, it is established wisdom that posture and peripheral muscle tensions are important for stable voice production (McCoy, 2012). Singers are sometimes even advised to tense the buttocks to control voicing (see, e.g., <https://www.youtube.com/watch?v=bYKdW32-rxM>). Common wisdom aside, there is considerable evidence that bodily postures affect singing (Cardoso et al., 2019; Longo et al., 2020; Miller et al., 2012b, 2014; Pettersen et al., 2005; Pettersen and Westgaard, 2004, 2005; Pettersen, 2006). Piano playing modifies back and shoulder positioning (e.g., extrarotated shoulder), which correlates with decreased energy at the harmonic formant attributable to changes in upper airway flow (Longo et al., 2020). The pectoralis major, an arm-rotating muscle, is active during phonation in expert and non-expert singers, which helps to counteract upper thorax inflation, thereby guiding expiratory control (Pettersen, 2006). Indeed, as Pettersen (2006) highlights, expiring during phonation is a dynamic affair recruiting different muscles at different phases to maintain the same stable subglottal pressure.

What evidence is there that *upper-limb* movements affect vocalization and speech biomechanically? A direct encounter with gesture–speech biomechanics was observed in an experiment where participants needed to vocalize the vowel /ə/ (as in cinema) at a steady-state using one expiratory flow. These vocalizations were performed as steadily as possible while rhythmically moving the upper limbs, versus not moving at all, while seated or standing. Posture was manipulated as it is known that anticipatory postural adjustments during upper-limb movements are much more forceful when standing as opposed to a more stable sitting posture (Cordo and Nashner, 1982). Upper-limb movement type was manipulated by increasing the physical impulse that a gesture produced by putting different masses in motion (wrist < arm < two-arm movement involves increasing masses in motion). Participants were instructed to repeatedly make a sudden halt at the maximum extension of the movement at 80 beats per minute, thereby creating a rhythmic physical impulse (or beat) coinciding with the deceleration of the hand.

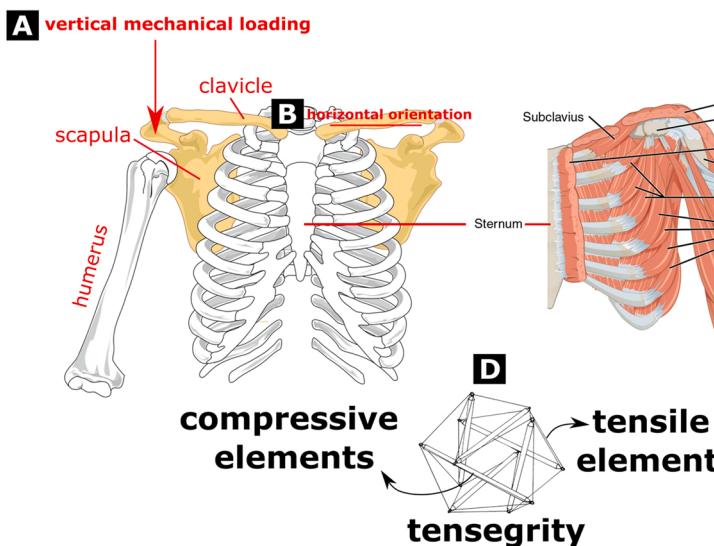


Fig. 3. The lungs passively follow the ribcage by biomechanical necessity. Bone (compressive elements) and muscle (tension elements) structures of the shoulder and thoracic regions are shown. On the left, it is shown that if there is some vertical loading (A) on the clavicle due to, for example, carrying your groceries, this loading will potentially transfer vertical forces onto the clavicle. The entire shoulder system is suspended on the sternum via an almost horizontally oriented clavicle (B). The scapula has a crucial stabilizing function (Handling et al., 2010) for upper-limb movement and muscles attach to it that also wrap around the rib cage (C: serratus anterior); it has been likened to the center of a wheel, with the muscles acting as the tensile spokes (Levin, 1997). This design architecture, where tensile and compressive elements form a pre-stressed system that allows focal forces to distribute globally, is referred to as a tensegrity structure, of which the tetrahedron is an idealized example (D). These figures were retrieved and modified from Wikipedia and originally published under a CC BY-SA 4.0 license.

The findings showed that during upper-limb movements, there were peaks in the fundamental frequency and the envelope of the intensity (amplitude envelope) of phonation, even more so for higher mass arm vs. wrist motions (see Fig. 4). These peaks occurred at moments where there were abrupt changes in vertical velocity at the maximum extension (i.e., peaks in decelerations). No such acoustic peaks were present when participants vocalized without movements. Moreover, these acoustic peaks were more pronounced when participants were standing vs. sitting, confirming the importance of posture-stabilizing muscle adjustment.

Several follow up studies provided additional evidence for the gesture–speech biomechanics thesis, extending its effects to more naturalistic speech (Pouw et al., 2020a, 2020c, 2020d). When the movement frequency was guided at a slower or faster pace via feedback from the motion tracking system, then vocal acoustics also oscillated at those slower or faster frequencies (Pouw et al., 2020d). In another study (Pouw et al., 2020c) participants wore a respiration belt that tracks chest kinematics. They were asked to repeatedly utter mono-syllables with an initial plosive consonant (/pa/) at 1-second intervals while moving their upper limbs in different timing relations with their vocalizations. When the physical impulse of a wrist or arm movement (peak in deceleration) was timed synchronously with the utterance, intensity was higher for that vocalization. Vocalization intensity was lower when the movement impulse was timed in alternation with the utterance. This shows that it is not movement per se, but the physical impulses produced by movements that are driving acoustic effects. It was further observed that chest-circumference changes occurred during vocalizations, but such changes were amplified when moving the upper limbs. Both intensity and F0 were positively related to such respiratory kinematic changes,

providing direct evidence for the mediating role of respiration in the effect of gesture on speech.

In a subsequent study (Pouw et al., 2020a) participants produced meaningful full-sentence speech while not moving or moving the wrist or arm at 80 beats per minute. It was replicated that at moments of physical impulses of wrist and arm movements, the intensity and F0 of speech peaked. Again, for higher mass arm vs. lower mass wrist movements, these effects were more extreme. Thus, although organizing meaningful speech brings with it major constraints on prosody due to syntactic and semantic layers of expression, the physical impulses of upper-limb movements are still detectable in such complex speech-vocalization acoustics.

Of course, detecting statistically reliable effects of gesture loadings in vocalizers' acoustics due to mechanical loading of upper-limb movements does not necessitate that those acoustic changes are detectable or meaningful for listeners and thus for communication. Therefore, in a follow-up study (Pouw et al., 2020d), listeners were asked to synchronize with movements of vocalizers, whom they could not see but only hear. The vocalizers produced steady-state vocalizations while rhythmically moving their wrist or arm at different tempos. It was found that listeners could synchronize their own wrist and arm movements with the vocalizers' arm and wrist motions. An astonishing feat is that listeners not only detected the frequency (i.e., rhythm) but also the phasing of the movement. They grasped that peaks in vocalizations indicate that the movement was changing its velocity during the extension–flexion motion. This allowed them to synchronize in-phase with the vocalizers. These findings mean that there is quite detailed information about movement in the vocalization (see Fig. 4).

There have been several applied extensions of the basic

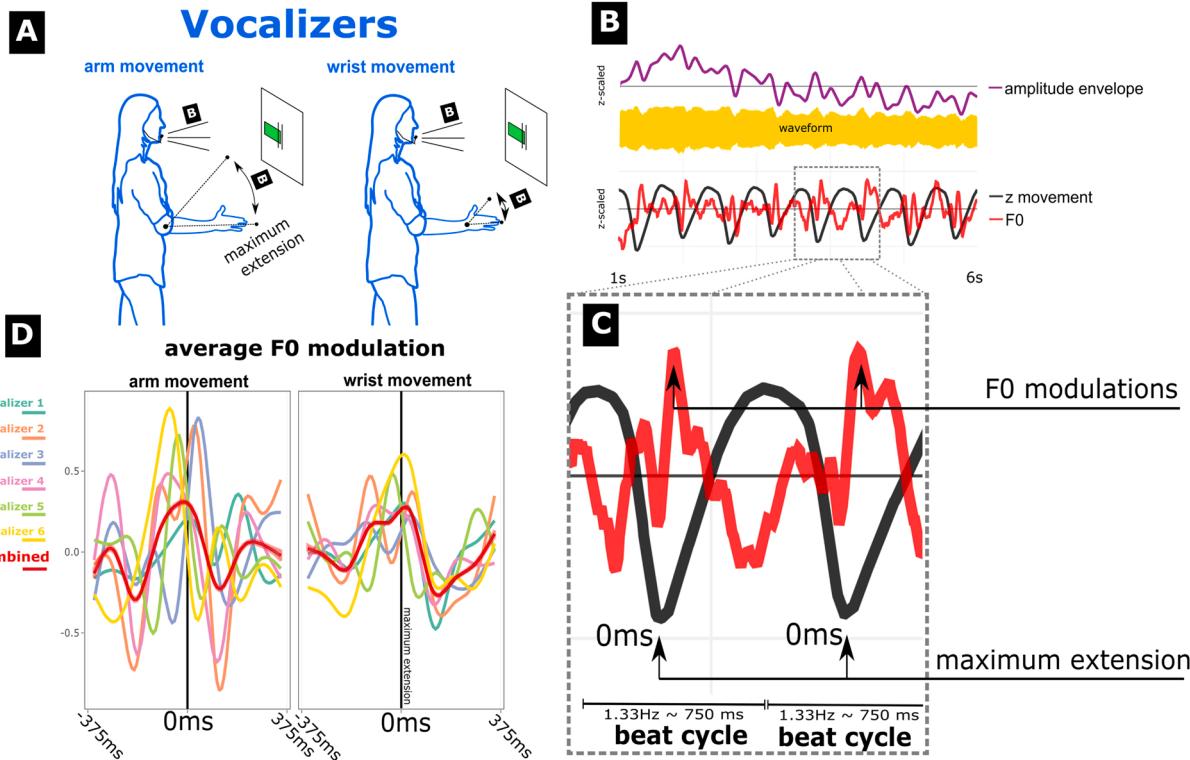


Fig. 4. Example results for gesture–speech biomechanics: this figure has been adapted with permission by the authors (Pouw et al., 2020d) and can serve as a summary of a line of research referred to as “gesture–speech biomechanics” (Pouw et al., 2019a; b, 2020a,b,c,d). When vocalizers are continuously moving their hands around the elbow or wrist joint (A) while vocalizing within only one breath cycle, unintended vocal inflections occur during the maximum extension moment when the hand suddenly stops, i.e., when there is a “pulse” in the movement (B). The inflections are positive peaks in vocal amplitude, as well as in the fundamental frequency (c), here shown in z-scaled units. When people are performing arm movement, the acoustic changes are more extreme as compared to wrist movement (here shown for six vocalizers). However, manual impulses have small (about 1–8 Hz) but perceptually apparent effects on vocal acoustics (D) for both wrist and arm movements (Pouw et al., 2020d). Further research in this program has revealed a modulating role of posture, respiration, and acceleration rate and has extended such effects to mono-syllabic utterances, fluent speech, and singing.

gesture–speech biomechanics research. In professional vocal performers of a south-Indian musical style, vocal FO is found to be most strongly coupled to accelerations of their hands, suggesting that the original link between manual forces and vocalization can be aesthetically repurposed in music (Pearson and Pouw, 2022). Similarly, even in persons with severe motor (deafferentation: Pouw et al., 2020b) and language pathologies (aphasia: Jenkins and Pouw, 2022), evidence has been found for the coupling of gesture acceleration and vocal acoustics. Thus the basic connection between gesture acceleration and speech acoustics seems to be maintained when any of the modalities are severely compromised, or when the interaction with the modalities is heavily enculturated.

Based on the review of studies in this section we conclude that the ultimate net effect of the specific flexion–extension pulsing movements we have studied is an increase in expiratory flow during vocalization. However, interactions of the upper-limb system with the respiratory–vocal system are likely multitudes more complex. This is because different upper-limb movements can recruit different accessory respiratory muscles involved in expiratory but also inspiratory control. Indeed, even for the studied flexion–extension arm motions during standing, other muscles (e.g., iliocostalis) are tensing for postural control that are involved in *inspiratory* action, which would on its own result in a drop in sub-glottal pressures when vocalizing. The astonishing complexity of respiratory interactions can be further appreciated in that the actions of primary respiratory muscles depend on the context of accessory muscle actions: “[if] contraction of the neck muscles fixes the first two ribs, the lateral parts of the intercostals can increase rib cage volume. If, however, abdominal muscles fix the most caudal rib, contraction of the same muscles would have the opposite effect” (p. 171 in Perry et al., 2010).

It is further possible that head gestures, which are also known to couple with prosodic aspects of speech (Esteve-Gibert et al., 2021; Liu et al., 2020), have their own subtle physical effects on speech production. Thus, while we will not review work that suggests biomechanical speech interactions for head movements (Anegawa et al., 2008; Honda et al., 1999; Liu et al., 2020; Miller et al., 2012a; b), we do want to mention that such “head gestures” are potentially amenable to a similar approach to the one we adopt here (see, e.g., Scarr and Harrison, 2017).

We have reviewed ample research that denies a clear-cut divide between manual gesture and the vocal system. Gesturing interacts with speaking via the respiratory system. However, humans must be able to counteract any constraints of gesture, or alternatively they can also allow such effects to align with the reaching of particular vocal targets. In both modes, gesture and speech are coordinated. This means that an effect of gesture on acoustics is not obligatory (Cravotta et al., 2021) as there is flexibility to counteract gesture-related perturbations with other (laryngeal or respiratory) muscle tensions. In addition to active counteraction, gesture–speech biomechanical constraints may also be deliberately amplified or simply aligned with a vocal target (e.g., Pearson and Pouw, 2022). In this way, neural–cognitive control loops that are shaped by culture and language regulate gesture–speech biomechanical interactions. Note that this of course does not explain why some individuals do or do not gesture, nor does it explain how different languages and cultures may lead one to avoid or co-opt these supposed biomechanical stabilities that arise out of moving and speaking at the same time. It certainly also does not address gesturing or signing in a sign language. The biomechanical account promoted here only serves as a mechanistic explanation of why vocal-entangled gestures exist at all in humans.

In the next section we question the assumption in developmental psychology that the coordination of speech and gesture is the last multimodal competence to emerge. After all, it seems that gesture and voice are biomechanically connected from the get-go.

4. Ontogeny: vocal–motor babbling as an exploration of biomechanics

There is a consensus that “beat gestures” occur at around 33–48 months of age (for a review, see Esteve-Gibert and Guellai, 2018), much later than the onset of pointing, which occurs at around 15 months at the very latest, and also still much later than iconic gestures, which emerge by 26 months at the latest (Colonnesi et al., 2010; Ozcaliskan and Goldin-Meadow, 2011). Yet infants move their upper limbs in a pulsing way with their vocalization *before anything as sophisticated as deixis or iconicity gets off the ground*. Notably, the pulsing quality also present in “beat gestures” is already vigorously used at 9 months of age in the form of vocal–motor babbling. Here we will give an overview of the developmental context of gesture-vocal babbling and how and how it might relate to the physical link between gesture and vocalization (also see Fig. 5).

The production of well-formed syllable cycles is an important developmental milestone for the infant (Iverson et al., 2007; MacNeilage, 1998), as mandibular oscillations are the key basis for generating the 3–8 Hz theta structure that generally characterizes human spoken languages (Poeppel and Assaneo, 2020). However, before any syllable cycle gets off the ground, the infant needs to gain the flexibility to vocalize. This flexibility is something that infants seem to naturally develop as they discover the degrees of freedom of vocalization and their social effects (for an example, see https://www.youtube.com/watch?t=88&v=_JmA2CIUvUY&feature=youtu.be). We might think of crying as reflecting such exploration, but it is not the primary way for infants to explore their vocal abilities. Rather, newborn infants engage in grunts (McCune, 2021) as well as the production of ‘vocants’ or vowel-like vocalizations (Oller et al., 2019). Grunts occur due to forceful air pressure increases leading to short and sudden bouts of airflow escaping through an otherwise closed glottis (and without supraglottal constrictions). McCune (2021) suggests that grunts are initially involuntary but are at a certain moment (around 6–9 months) intentionally co-opted for social employment by the infant. Vocants can also be flexibly employed, even with neutral affect, which stands in stark contrast to the functional rigidness of crying, which primarily signals distress (Oller et al., 2013). Further vocalizations have a particular developmental trajectory (Koopmans-van Beinum and Stelt, 1986), with interrupted phonations that require more complex laryngeal–respiratory coordination arising around 5 months, followed by babbling that adds an articulatory coordinative component (around 6–10 months).

Recently it was shown that vocalizations are about 35 times more abundant in infant communication than the use of gestures, especially in early developmental phases of about 4 months. However, in later phases (7 and 11 months), vocalizations are reduced and gestures increased, but still ending with about 2.5 times more vocalizations than gestures (Burkhardt-Reed et al., 2021). Though not central to their findings, the authors also found that “non-social” gestures including *rhythmic* manual movements reached a prominent peak at 7 months of age, *being the most abundant type of gesture*, which is critical for our current concerns. Based on this general divergence on the number of gestures occurring relative to vocalizations during the first year, Burkhardt-Reed et al. (p. 11 in 2021) conclude that vocalization, not gesture, is the primary means of communication for infants. They further suggest this has evolutionary implications, since “if language indeed originated from gestural use, gestural activity should have occurred to a far greater extent [...].”

While gestures may not be more primary, there is an interesting well-replicated phenomenon that directs us to the possibility that gestures guide speech-like vocal development. After the initial period of ‘vocating,’ more speech-like utterances emerge in the form of canonical babbling. Canonical babbling occurs from 6 to 10 months of age and is a critical period where the mandibular cycle is used (MacNeilage and Davis, 2000) in the production of reduplicated consonant-vowel syllable chains, like /baba/ or /mama/ (Ejiri and Masataka, 2001). In a longitudinal study of infants who were still developing into the canonical

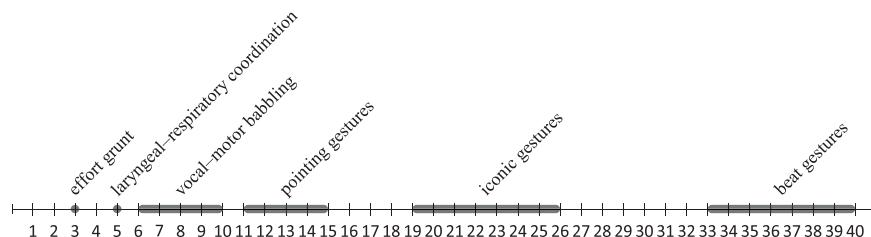


Fig. 5. Notable reported milestones in literature reviewed below on a developmental timeline (in months).

babbling stage, Ejiri (1998) found that infants suddenly increased the number of times they rhythmically shook a rattle during free play sessions. This increase in rattling co-occurred with increases in rhythmic manual (but not leg) actions. Importantly, a sharp increase in rhythmical actions in the exact same period wherein infants rapidly increase their canonical babbling has now been reported in several studies (Burkhardt-Reed et al., 2021; Cobo-Lewis et al., 1996; Eilers et al., 1993; Iverson and Wozniak, 2007; Locke et al., 1995; Thelen, 1979). Further, when babbling is delayed, rhythmic manual behavior and postural stability are also reduced (Iverson and Wozniak, 2007). Ejiri (1998) found that during the peak babbling stage, vocalization occurred during rhythmic manual movement as much as 75% of the time. Up to the month where canonical babbling emerged, infants were just as likely to shake a rattle that made no sound as they were to shake sound-producing rattles, indicating that the rhythmic actions were not produced for sound production per se.

Further, rhythmic actions in infants increase in prevalence during vocalizations preceding and around the canonical babbling onset at 6–7 months of development (as opposed to other actions, e.g., handling, mouthing), but then decrease from months 8–11 (Ejiri and Masataka, 2001). When acoustically analyzing the vocalizations that occur or do not occur with rhythmic actions, longer syllable durations and formant transitioning are found when rhythmic actions are present. Such acoustic markers are associated with more adult-like syllable productions. This suggests that the coordination of mostly manual (Ejiri, 1998; Iverson, 2010) rhythmic actions with vocalization affords novel vocal-articulatory stabilities for the infant. Though co-occurrence does not mean causation, these findings at least show that rhythmic manual movements precede the onset of canonical babbling (Iverson and Fagan, 2004).

So rhythmic action and canonical babbling rapidly emerge in succession, with (co-vocal) rhythmic movements emerging first. However, such rhythmic manual actions suddenly decrease. Why is this? The developmental psychologist Esther Thelen (Thelen, 1979) would insist that this is because the productive developmental work of rhythmic movements is completed at some point in development. Then the infant moves on to exploring new sensorimotor regularities: “behaviors that may originate as ‘by-products’, so to speak, of normal maturation, may be used opportunistically by infants to serve a variety of functions for which more complex behavior is as yet unavailable” (p. 713–714 in Thelen, 1979).

A dynamical systems account inspires the current perspective on the developmental emergence of gesture and speech coupling (Iverson, 2010; Iverson and Thelen, 1999, 2003). In this account, it is held that initial non-communicative hand-mouth coalitions in, for example, feeding are the ontogenetic drivers of manual–speech coupling in communication. Iverson and Thelen review a host of neurological evidence that upper-limb and speech production areas are closely located neurally (e.g., in the supplementary motor cortex) and may share common neural systems that are dedicated to timing, such that “patterns of co-activation may be influenced by a common precise timing mechanism in the lateral perisylvian cortex” (Iverson and Thelen, 1999, p. 22). To explain this gradual synchronization of manual action and speech utterances, Iverson and Thelen (1999, p. 35) suggest that “as

words are practiced, they are able to activate the gesture system sufficiently to form synchronous couplings, and thus the two motor systems become” entrained. Synchronization is thus assumed to depend on the emergence of the first meaningful words.

However, even before the flexible production of the first meaningful words, infants from 11 to 19 months of age synchronize gesture and vocalization (Esteve-Gibert and Prieto, 2014; Murillo et al., 2018). No reliable differences are found in adult-like gesture–speech synchrony across age, despite differences in gesture frequencies and speech abilities (Esteve-Gibert and Prieto, 2014). This is further corroborated and extended by a longitudinal study showing that 9-month-old infants show evident gesture–vocal synchrony and further increase this synchrony over the following 9 months (Murillo et al., 2018).

The following revision of dynamical systems accounts of gesture development by Iverson and Thelen (1999) is therefore needed: During the early development of the canonical babbling phase, there are exploratory vocal productions that co-occur with physical impulses during concurrent exploratory body movements. Such physical impulses occur by chance and affect vocal production through respiratory interactions, which can solicit the joint neural regulation of manual and vocal systems. As infants start to discover the regularities of this physics, they may start to exploit upper-limb–respiratory–vocal stabilities for social effects. In fact, rhythmic interactions are often solicited and responded to by caregivers from very early on (Moreno-Núñez et al., 2017). It is in this way that gesture–vocal coupling can be further discovered and reinforced through social support, similar to how grunts start to become employed communicatively by infants (McCune, 2021). Thereby manual rhythmic co-vocal gestures gradually become physically and intentionally stabilized at 6–10 months of age and are part of a wider developing physical context (e.g., posture development, respiratory modulation skills) that is supportive of the production of well-formed reduplicated syllables in canonical babbling (Ejiri and Masataka, 2001). At some point the developmental work of gesture–vocal biomechanics has been saturated, and the infant might even decouple such systems, as is observed in the sudden decrease in co-vocal rhythmic movements, all in a continued search for novel articulatory–vocal degrees of freedom, which can now be explored without manual movement. Gesture–vocal decoupling should then be seen as a removal of a developmental crutch, as the infant now has more locally controlled vocalizing abilities, which are employed to produce the first meaningful words.

The vocal acoustics of human infants have been shown to covary with the supine or upright posture of the infant (Lin and Green, 2007), which has been related to upright posture physically optimizing respiratory drive for vocalization (LeBarton and Iverson, 2016). There hides an evolutionary logic in this notion that acoustics are informative about the body: “in addition to broadcasting respiratory events, the vocal apparatus could evolve the capacity to express respiratory pressure (via modulation of sound amplitude) or muscle tension (via modulation of pitch), and so on” (Tchernichovski and Oller, 2016, p. 422). It can thus be adaptive for peripheral bodily conditions to imprint on the voice, for example, in the case where an infant is expressing physiological–emotional–psychological well-being to the caretaker. Therefore, vocal–motor babbling imbues the utterances with a bodily

component that is a particularly potent communicative invitation to the parent (Moreno-Núñez et al., 2021; Murillo et al., 2021; Mehr et al., 2021). In the infant's successful multimodal solicitation of social attention from the parent, this may then, over time, shape a context for new communicative abilities to develop (e.g., more complex deictic gestures).

To conclude, we have reviewed a suite of studies showing that infants peak at about nine months in coordinating rhythmic manual movements with vocalizations. Currently, gesture studies are conservative about what a “beat” gesture is, but given the above, they might become more lenient as to what counts as its precursor, namely vocal-entangled gesture in the form of vocal–motor babbling. This would mean accepting that *vocal-entangled gesture is already present before any iconic or deictic gesture emerges in development*. Next we will propose a similar revision to evolutionary thinking about the origins of gesture.

5. Phylogeny

The literature on the evolution of human communication is daunting in breadth, and it seems that every possible angle to approach the subject has been taken (for an excellent overview, see Fitch, 2010). In broad strokes, theorists oppose one another based on whether music, speech, and/or manual and whole-body gesture were the first forms of (proto-)language. Further discussions consider the putative adaptive functions of these particular modes of communication and their required preadaptations.

Even within similar-minded camps there are considerable differences in opinion, e.g., diverging as much as 5.5 million years as to when intentional vocal control developed (cf. Fröhlich et al., 2019; Levinson and Holler, 2014). It is difficult to foresee at what scientific juncture any perspective will become more viable than the other. To increase the viability of the phenomenon of vocal-entangled gesture as a basic evolutionary feat, we therefore aim to show that the pulse quality of gesture has a role to play in any particular evolutionary rendering of spoken human communication. For this to be convincing, we also need to show that the vocal-entangled gesture is shared among non-human animals from branches of the phylogenetic tree that split from that of humans much longer ago than those of any species that has so far been shown to be related to the human faculty of gesture.

5.1. Locomotor–respiratory coupling

Vocal-entangled gesture can be traced back to elementary beginnings: it is rooted in the phenomenon of locomotor–respiratory coupling (LRC). The LRC literature shows that a varied number of animal species “time” their respiratory cycles with locomotion (Bramble and Carrier, 1983). During moments of peak mechanical loading of the forelimbs on the thoracic region, the expiratory phase of the respiratory cycle is initiated. “Timing” deserves scare quotes as it is an emergent stability between coupled articulatory systems, rather than a centrally controlled top-down process that does the timing. Namely, the mechanical loading on the thoracic region mechanically increases expiratory flow, which disposes the respiratory cycles to align with those constraints.

The biomechanics of locomotor–respiratory coupling is described in detail in galloping horses: they start inspiration shortly after the leading forelimb leaves the ground, instigating accelerated forward translation while increasing abdominal volume. This acceleration displaces the visceral organs more backward, acting like the cocking of a piston. When decelerating during forelimb ground contact, this “visceral piston” moves forward against the diaphragm, reducing abdominal and thoracic volume, which leads to sudden increases in lung pressure, thereby forcing exhalation to synchronize with forelimb strides (Bramble, 1989; see also Lafontaine et al., 1996). It is only during more active locomotion in quadrupeds that locomotor–respiratory coupling becomes so strong that one stride cycle synchronizes with one respiratory cycle (i.e., 1:1 coupling; Boggs, 2002).

Of course, in different species the precise biomechanics may be different, but similar tendencies for 1:1 or polyrhythmic (e.g., 1:2) coupling, especially in fast-paced locomotion patterning, have been shown in for example rabbits (*Oryctolagus cuniculus*; Simons, 1999), domestic cats (Iscoe, 1981), wallabies (*Macropus eugenii* Desmarest; Baudinette et al., 1987), and rhinoceroses (*Ceratotherium simum*; Young et al., 1992), as well as in avian species (i.e., wingbeat–respiratory coupling) such as geese (*Branta canadensis*; Funk et al., 1992) and bats (e.g., *Phyllostomus hastatus*; Suthers et al., 1972; see also Carpenter, 1986) – indeed, it is clear that LRC is a pervasive phenomenon for tetrapod species in general (Boggs, 2002). Even in humans, there is still mechanical loading of foot landings (i.e., lower limbs) that can constrain the diaphragm, changing inter-abdominal pressures and thereby affecting respiratory cycling. But such constraints are obviously less powerful; rather a weak coupling is observed in the form of simple integer ratio locomotor–respiratory coupling of 2:1 or 4:1, and sometimes complete decoupling of locomotion and respiration is also observed (Banzett et al., 1992; McDermott et al., 2003; O'Halloran et al., 2012; Perségl et al., 1991). When the upper limbs are used for locomotion, LRC is more apparent: humans in wheelchairs show respiratory coupling with their upper-limb cycling in a polyrhythmic fashion (Amazeen et al., 1992; see also Ebert et al., 2000).

With LRC a deep connection can be made between body movements and respiratory–(vocal) coupling and its first emergence in early land-dwelling quadruped tetrapods, thus going back at least 300 million years by very conservative estimates (Clack, 2006). Of course, some intermediate steps are needed to complete the connection between vocal-entangled gesture and LRC. In the following subsection, we will provide an overview of locomotor–respiratory–vocal coupling (LRVC) in animals, a natural extension of LRC that can be seen as a precursor to vocal-entangled gesture.

5.2. Locomotor–respiratory–vocal connections

As already mentioned, bats (e.g., *Phyllostomus hastatus*) tend to synchronize their respiratory cycles in a 1:1 fashion with their wing beats (Suthers et al., 1972; Wong and Waters, 2001), much like horses, which synchronize their expiration with forelimb mechanical loading (Bramble and Carrier, 1983). However, of crucial survival importance in bats is that they need to echo-vocalize to perceive the environment, especially during locomotion. In line with earlier findings that respiration coordinates with wingbeats, it has also been found that concurrent echo-vocalization bursts synchronize with wingbeats (Suthers et al., 1972; Lancaster et al., 1995).

If we look closely at bat LRVC, we find that echo-vocalization pulses unfold during moments of downward to upward wingbeat transitions at variable 1:1, 1:2, or 1:3 coupling ratios (Suthers et al., 1972). Biomechanical research on bats (*Pteronotus parnellii*) has shown that when bats vocalize while being stationary, the muscles in the abdominal wall are synchronized with echo-location vocalizations, with no consistent muscle activation of the upper thoracic region (pectoralis). However, for echo-vocalization during flight, the abdominal muscles are synchronized with vocalizations and the large muscles that power flight (pectoralis) show this synchronization (see Fig. 6). This is because the muscles that power flight also affect respiratory functioning. During downward wingbeats, energetic potential is transferred onto the thoracic region that increases expiratory flow that would otherwise be generated by the contraction of the abdominal muscles. The bats are using their resources efficiently by organizing coordinative structures (Kugler et al., 1980) that can yield similar solutions with different sets of muscles. The abdominal muscles need to work much less vigorously during flight, as the expiratory drive is in part delivered by the wingbeat-related muscle activity of the pectoralis.

Biomechanical stabilities can be flexibly dealt with. Brown-headed cowbirds (*Molothrus ater*) variably vocalize during more or less vigorous wingbeat displays (Cooper and Goller, 2004), due to the

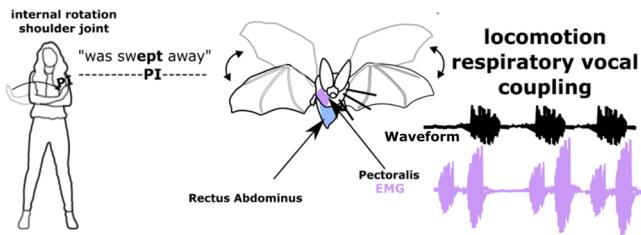


Fig. 6. This figure draws attention to the continuity of gesture–speech biomechanics as reviewed in Section 3, and the phenomenon of locomotor–respiratory–vocal coupling, here schematically exemplified for a flying bat (Lancaster et al., 1995): humans synchronize (1.3 Hz) their arm movements with F0 and intensity, while bats synchronize (10 Hz) their wingbeats (and muscle contractions measurable by ElectrMyoGraphy or EMG) with echo-location bursts. This figure reuses and adapts material from Pouw et al. (2021), with permission from the authors.

resultant biomechanical constraints on their air sacs. In moderately vigorous displays, reduced muscle activation is found of a primary respiratory driver of vocalization without wing displays. This diminished respiratory effort while nonetheless maintaining air sac pressure (Cooper and Goller, 2004) provides suggestive evidence for a biomechanical cooperation with wingbeat muscle activity that helps to maintain air sac pressure, which is otherwise regulated by different sets of muscles. During very vigorous wing displays, these birds do stop vocalizing, as this coordination is simply too unstable to maintain at those intensities (Cooper and Goller, 2004). These findings in cowbirds can be further related to Lancaster et al.'s (1995) observation that bats are not obligated to perform locomotor–respiratory–vocal coupling – they can echo-vocalize at any moment of the wingbeat cycle (Moss et al., 2006). This resonates with a point that bears repeating: the constraints of biomechanics are always present. They may push an animal to do things one way rather than another way. But there is always flexibility to overcome constraints and *to do something in a different way* or to stop evoking such constraints altogether.

Rodents too tend to “time” their high-arousal and copulation vocalizations with rhythmic hind- and forelimb motions (Blumberg, 1992). For example, during hopping locomotion, gerbils (*Meriones unguiculatus*) jump up with all fours leaving the ground, but upon return, when the forelimbs hit the ground, a vocalization is emitted (Blumberg, 1992; for a similar observation in dogs, see Bramble and Carrier, 1983). Vocalizing can actually be biomechanically adaptive. It involves a constriction of the larynx, which then tunes the thoracic stiffness through lung pressure increases relative to a fully dilated glottis (Blumberg, 1992). Thus, increased subglottal pressure through laryngeal control is a way to change the pre-stress characteristics of what we above called a tensile integrity structure, optimizing absorption of mechanical shocks of the forelimbs during ground contact. This notion of thoracic stabilization through glottal constriction can be related to findings showing that grunts improve human throwing velocities in baseball and forearm velocities in tennis players (O'Connell et al., 2014; Tammany et al., 2021), and even to grunts in infants exerting some physical effort, which then are co-opted for communicative signaling (McCune, 2021; Raine et al., 2017).

Findings on vocal development in green-rumped parrotlets (*Forpus passerinus*; Berg et al., 2013) bear a striking similarity to vocal–motor babbling in infants as reviewed in Section 4. Like bats, adult green-rumped parrotlets tend to emit contact calls during the downward wingbeats during flight, typical of LRVC. Nestling young parrotlets go through a critical canonical stage of vocal development where there is a sudden decrease in call segment duration and an increase in the fundamental frequency characteristic of adult-like call structure. Such decreased call duration becomes more like the structure of adult contact calls and suddenly emerges “shortly after [parrotlets] began using wings to aid locomotor displacements inside the nest cavity” (p. 344). These

wing beats in the nest are still of a lower intensity than those in flight, and this might explain why parrots do not reach adult call production before flight. It is only when young parrots take their first virgin flights that their calls become fully adult-like. Berg et al. (2013) explain the relation between wingbeats and vocal development as follows: “the high levels of expiratory pressure generated by wing-powered flight might then provide individuals with a fortuitous and energetically inexpensive way of amplifying their high frequency calls” (p. 344). This research on parrots is remarkably relatable to vocal development in human infants, where canonical syllable-like babbling rapidly follows upon the emergence of manual rhythmic movements (Iverson and Fagan, 2004). Vocal development is then scaffolded by latent biomechanical limb–respiratory–vocal stabilities, which are actualized at a particular stage of the bodily development (see also Zhang and Ghazanfar, 2018).

To conclude, it is important to explicate the homology between locomotor-respiratory(-vocal) coupling and gesture-speech biomechanics. Locomotor-respiratory(-vocal) coupling emerges because musculoskeletal elements (e.g., in the pectoral region) that produce locomotor actions lead to respiratory action, thereby modulating inspiration/expiration at certain accelerative moments during locomotion. Such modulations can either be counteracted or can be functionally incorporated in the form of aligning breathing cycles and vocalization with pectoral limb cycles. The mechanism that drives locomotion-respiratory-vocal coupling is therefore homologous to gesture-speech biomechanics: The production of a pectoral limb movement for a pulsing gesture leads to inspiratory/expiratory modulations during accelerative moments of the gesture, which is frequently incorporated in the production of prosodic speech.

5.3. Bipedalism and upper-limb–respiratory–vocal connections

Based on the literature on locomotor–respiratory(-vocal) coupling in extant species reviewed above, we assume that the pectoral limb system must have constrained the evolution of the respiratory system in animals, and vice versa (Carrier et al., 1984; Klein and Codd, 2010; Perry et al., 2010). Accordingly, Klein and Codd (2010) argue that “[t]he large diversity of [ventilatory] structures and mechanisms may be attributed, at least partially, to overcoming constraints between locomotion and breathing” (p. 526). Different solutions have evolved from this original constraint. In the monitor lizards of the genus *Veranus*, the intercostal muscles are intensively used for their accordion-like “axial” locomotion, which has large effects on rib cage mobility. Intercostal muscles that regulate rib cage movement are therefore less efficiently employable for respiration. As a likely evolutionary antidote, a supplemental respiratory mechanism is present in the form of “gular” pumping by an inflatable air sac reservoir that pumps air into the lungs to maintain aerobic metabolism (Owerkowicz et al., 1999). In crocodilians (*Alligator mississippiensis*), who also move with axial contractions, a distinctive diaphragm morphology developed, where this muscle functions as a type of piston that allows the decoupling of locomotion and respiratory cycles altogether (Farmer and Carrier, 2000). These examples then contrast with other quadruped mammals where respiratory cycles are more strictly coupled in a 1:1 fashion with locomotion (Bramble and Carrier, 1983).

Another optimizing solution that has been proposed to prevent rigid locomotor-respiratory coupling, such as the crocodile diaphragm, is through bipedalism (Carrier et al., 1984). In strict bipeds, the respiratory system is freed from locomotor-related perturbations from the forelimbs onto the thoracic region, since the now-promoted *upper* limbs do not carry the body during locomotion. Bipedalism then allows for more variable relationships between locomotion and respiration, and explains why locomotor-respiratory coupling is mostly lost in bipedal humans (as reviewed above). Carrier et al. (1984) suggest that energetic benefits arise from the independence of locomotion perturbations and respiratory processes that have had unique adaptive advantages in humans. For example, this can explain why humans are excellent

long-distance runners, as the decoupling allows slower respiratory cycles, which are energetically more optimal than respiratory cycles that would couple with the fast locomotion strides.

This is not to say that locomotion is entirely independent of respiration in humans, as already mentioned. But it is much more flexible. And indeed, in other bipedal locomoting species such as non-avian birds, more flexible and weak polyrhythmic (e.g., 3:1) locomotion-respiratory coupling is observed (Boggs, 2002; Nassar et al., 2001). It is this homology with non-avian birds that Carrier and colleagues (1984) emphasize when arguing that bipedalism has advantages for flexibility in respiratory control.

In humans the respiratory system was not the only system freed through bipedalism from interactions with locomotion. The pectoral limb system was freed for non-locomotion action too. Thus, the pectoral limbs could now move for different and more varied prehensile purposes, while still constraining the respiratory system. *This re-purposing of the upper-limb system for more complex interactions with the (social) environment also means that the respiratory system itself became biomechanically impacted by new manual intentionality and diverse perturbations that were simply not elicited before bipedalism.*

Extending this to vocal control is an obvious next step. A complex respiratory system must be in place to vocalize in varied ways. MacLarnon and Hewitt (1999) have shown that the level of innervation required for advanced speech respiratory control seems to be uniquely anatomically accommodated in humans as well as Neanderthals, but not in other hominid members of earlier descent, such as *Homo ergaster* or the *australopithecines*. Specifically, the vertebral canal, which houses the spinal cord, is expanded at the level of the thorax for us late hominids, suggesting a higher density of innervation (a thicker bundle of nerves) to serve the more complex respiratory-control needs for speech (MacLarnon and Hewitt, 1999). Though not mentioned by MacLarnon and Hewitt (1999), the nerves dedicated for upper-limb control also run through exactly this level of the vertebral canal. MacLarnon and Hewitt (1999) reflect on the fact that postural control challenges or respiratory liberation from locomotion due to bipedalism cannot explain an increased thoracic innervation, as *Homo ergaster* was bipedal but did not have increased innervation as compared to Neanderthals and humans. Thus, we cannot simply go from bipedalism to increased respiratory control as a preadaptation to speech; something else is needed to complete the connection.

Bipedalism entails a diversification of manual behavior. For example, carrying something in one hand with a tonic muscle tensioning and grasping things with another with a varied distribution of muscle tensioning goes beyond purely simple isochronous and symmetric oscillatory modes such as in locomotion. It is possible that given the biomechanical connections already reviewed, new manual feats drove a complexification of respiratory-vocal functions too. We can further appreciate that “[f]rom an engineering perspective, bipedalism is a ridiculous answer to the need for locomotion, posing problems akin to balancing an apple on top of a moving pencil” (Walker and Shipman, 1996, p. 199). We should add to this pencil two pairs of chained chaotic pendula in the form of new and varied upper-limb perception-action routines that push and pull away from the center of mass. Thus, as opposed to bipedalism in and of itself, new manual behaviors that arose because of bipedalism must also have driven the complexification of the whole thoracic and upper-limb muscle system, and this might have played a role in constructing a respiratory niche for spoken language to eventually flourish (MacLarnon and Hewitt, 1999).

5.4. Revisions to any theory about the evolution of gesture

At this juncture we discuss what the ancient limb-respiratory-vocal connections might mean for phylogenetic stories. How did vocal-entangled gesture evolve in humans?

In gesture-first accounts there is a recurring argument that the flexibility of how gestural communication is employed in non-human apes

such as bonobos and chimpanzees stands in stark contrast to the putative inflexible use of *vocalizations* in these species (Corballis, 2002; Tomasello, 2008). This supposed flexibility in the manual domain is used to argue for the primacy of a gestural communication system, serving as the early preadaptation to kick-start the intentional signaling needed for language in any modality. However, this view can be brought into harmony with our account. Once bipedal, the upper-limb systems gained novel possibilities for interacting with the (social) environment. New levels of intentionality were elicited by the (social) environment. These intentional capabilities of the upper-limb movement system were then co-opted by the respiratory-vocal system.

This line of reasoning follows arguments that speech or music competencies in humans might be parasitic on capabilities in the motor domain (Ghazanfar, 2013; Larsson, 2014; Larsson et al., 2019; Michelletta et al., 2013; Ravignani et al., 2019). Relatedly, social vocalizations in bats have been suggested to structure in a way that seems derived from the respiratory-locomotor-vocal stabilities during echo-location in flight (Burchardt et al., 2019). For humans, we can then surmise that respiratory-vocal competencies derive from the manual system through natural biomechanical interactions between these systems.

Specifically, producing particular quasi-rhythmic and variable contrasts in the intensity and F0 of vocalizations could be achieved through physical impulses of the upper limbs that were more flexibly employable in this varied way. Thus, if we hold that the respiratory-vocal system was not able to intentionally produce sharp vocalic contrasts within one breath in some structured way (e.g., quasi-rhythmically), the intentionally skilled upper-limb system could then aid in the production of such contrasts through biomechanical entrainment, similar to what can happen in ontogeny as we have reviewed for green-rumped parrotlets and human infants.

This biomechanical scaffolding can help to explain one of the identified weaknesses of gesture-first theories: why did the vocal-articulatory system take over as a dominant mode of communicating? In the current line of reasoning, it is because the apprentice (the putative unstable and inflexible vocal system) came to replace the master (the putative intentionally skilled upper-limb system). Yet this could only have happened after some manual (en)training, which provided the scaffolding for the vocal-articulatory system to become dominant, in turn reaping the adaptive advantages of communication in this modality that have been identified by a host of scholars (e.g., communicating in the dark or while the hands are busy; for an overview of such benefits, see Fitch, 2010).

As an intermediate conclusion: if a theory holds that the manual system offered some early cognitive, intentional, or semantic flexibilities that provided a scaffold for the development of the vocal system (Arbib, 2005; Donald, 1991; Zywickzynski et al., 2018), then the physical origins of vocal-entangled gesture provide an important existential argument for these systems to co-adapt. This answers to what we might call the “entanglement challenge” to unimodal-first theories: *once a vocal or gesture modality is linguistically in place, how is it that gesture and vocalization became entangled in the human species?* This challenge is arguably poorly met in gesture-first (Arbib, 2005; Corballis, 2002) or gesture-dominant (Donald, 1991; Zywickzynski et al., 2018) theories of the evolution of human communication, despite perhaps offering good reasons why gestures were dominant first before transitioning to the other modality (see, e.g., Gentilucci and Corballis, 2006).

Another possible reconciliation is apparent in our view on vocal-entangled gesture and evolutionary accounts that argue for vocal and gestural co-evolution (Kendon, 2017; McNeill, 2012), as the very distinction between upper-limb movement, respiration, and vocalization fades away when appreciating their biomechanics. Biomechanics is thus an existential basis for gesture-speech co-evolution. For this merging of views to succeed, it would require extant accounts to shift focus onto non-representational aspects of gesture (Kendon, 2017). This shift in focus would even apply to the views on gesture evolution of David McNeill, who does not seem to consider the pulsing-quality of

gesture central as an initial driver in connecting the evolution of gesture and speech (McNeill, 2012).

There are several types of merges possible between our views and gesture–speech co-evolution views. A superficial merge entails that biomechanics was a byproduct that emerged because using both gesture and vocalization simultaneously was the most productive solution for flexibly representing thoughts in a combined visual-auditory format (Goldin-Meadow and Brentari, 2017). Thus, it is possible to hold that manual gestures proliferated because they are representationally potent in a way that complements vocalization and that gesture–speech biomechanics was a so-called spandrel (Gould and Lewontin, 1979).

Another more completed merge of co-evolution accounts with the biomechanical one would entail that the coupling of vocalization and upper limbs itself had adaptive functions. Then the proliferation of gesture and speech today should be (in part) rooted in those adaptive functions.

If a vocalization is an unavoidable by-product of other bodily events, if the occurrence of the vocalization predicts a behavioral or physiological state in the caller, and if a conspecific can detect and react to the vocalization in a way that benefits the caller or the listener, then the vocalization is likely to generate the establishment of and become fixed within a communicatory system. Furthermore with time, the incidental vocalization can become ritualized and perhaps be emitted independently of other physical constraints. (Blumberg, 1992, p. 364)

Thus the early adaptive function of vocal-entangled gesture may lie in its indexical communicative potential. As mentioned above, the infant–caretaker communication is valuable for the caretaker (and the infant) to gain (and emit) information about the peripheral bodily states. Much like how rodents have been found to emit high-frequency calls that are informative about abdominal muscle contractions which are a physiological response to being cold (Blumberg and Sokoloff, 2001; see also Zhang and Ghazanfar, 2016), human gesture-entangled vocalization may inform about the broader physical state of the infant. Thus infants' vocal-motor babbling could be indexical for the gradients of peripheral muscle tensioning that are co-informative of the infants current emotional states. In adulthood we can further imagine that indexical gesture–vocal coupling can function to support “social grooming” (Dunbar, 2016; Savage et al., 2020) – a way to “forge[s] and reinforce[s] affiliative inter-individual relationships … by synchronizing and harmonizing the moods, emotions, actions or perspectives of two or more individuals”. Gesture–vocal coupling would play an important role in such processes. Namely, humans can acoustically detect the phasing of upper-limb movements of vocalizers (Pouw et al., 2020d), and even from just hearing each other speak, bodily sways can become coordinated (Shockley et al., 2007, 2003). Thus, voices are not abstract sounds perceived at a distance; they are produced by bodies and are perceived accordingly in relation to their wider physiological constituents (Pisanski et al., 2016). Social vocal-entangled gesturing then can then be seen as a potential for synchronizing bodily physiology through sound.

Gesture–vocal entanglement can be related to the adaptive functions of indexical cues, and thereby also the derived adaptive functions of “faking” or “representing” such indexical cues. As is well known in non-human animals, acoustics is indexically informative about body size (Fitch and Kelley, 2000; Ghazanfar et al., 2007; Raine et al., 2019; Reby et al., 2005; Wright et al., 2021). Some species might even exaggerate such qualities (de Boer et al., 2015; Hardus et al., 2009). We deem it possible that tensioning of the upper limbs may have similarly played a role in early aggressive or territorial vocal displaying, and could play a more general role in emotional expression today. Tensioning gestures provide potential information about action-readiness, bodily activity (Pouw et al., 2020d), and emotional states, much like how chest beats are informative about body size in gorillas (Wright et al., 2021), and much like how tensioning of the vocal folds (and heightened FO) is itself deeply connected to (and thus an indexical cue for) emotional states

(Bolinger, 1986). It is not that producing a gesture allows someone to signal their internal state of anger, rather it is by producing this high-tensioned bodily impulse that they enact that anger (Merleau-Ponty, 1945). As such, gesturing as vocal-coupled bodily tension is a “credible signal” about an emotional-embodied state of affairs (a putative origin of music; Mehr et al., 2021; also see Zahavi and Zahavi, 1999).

To conclude, we have inserted vocal-entangled gesture in several accounts about language origins. While we are not committed to any particular evolutionary rendering, we are convinced that vocal-entangled gestures fit in any of these accounts. We will now discuss two critical pieces of the puzzle that need to be put in place for a completed account of the origins of vocal-entangled gestures: What about our closest living ape relatives? And what about the brain?

5.5. What about non-human apes?

What about other non-human apes, such as chimpanzees and bonobos, and the singing hylobatids (gibbons and siamangs)? Especially great apes are central to current mainstream discussions about the origins of gesture and speech (Fröhlich et al., 2019), yet they have not figured at all in the current account. As Ravignani and Kotz (2020) comment on the gesture–speech biomechanics thesis: “evidence from other primates suggests we evolved movement precision and dexterity … before direct corticobulbar projections for enhanced laryngeal control (Simonyan, 2014). Why, then, have all other dexterous primates not evolved similar capacities for embodied vocal control?” (p. 23223).

Firstly, there is very little research on truly multimodal aspects of primate communication (Slocombe et al., 2011; Liebal et al., 2022), in contrast to unimodal vocal and unimodal gesture research. As a consequence, little is known about bodily–vocal coordination, or even about the coordination between respiratory cycles and arboreal locomotion. Though it has been reported that some non-human primates restrict their sound production to one continuous unit per inhalation or exhalation (MacLarnon and Hewitt, 1999; Marler and Hobbett, 1975), it is unclear whether vocalization entrains to peripheral body parts. Even more generally we do not know whether rhythms are shared in the manual, respiratory, and vocal domains, there being however some exceptional studies that have pursued interests that are moving in this direction (Dufour et al., 2015; Perlman et al., 2012; Lameira et al., 2019; Perlman and Salmi, 2017).

Secondly, recall that some animals such as monitor lizards and crocodiles have developed a bodily morphology that circumvents or counteracts the biomechanical constraints that exist between locomotion and respiratory dynamics in many tetrapod species (Arcadi et al., 1998; Klein and Codd, 2010; Perry et al., 2010). It is interesting in this regard that all non-human primate species have several types of air sacs



Fig. 7. A siamang (*Symphalangus syndactylus*) vocalizing while suspending. This species produces advertisement calls during vigorous brachiating, sometimes referred to as locomotion displays (Geissmann and Orgeldinger, 2000) or what might turn out to be vocal-entangled locomotion. This image is copyrighted under creative commons CC-BY-2 (photo by Su Neko) Wikipedia.

(Ybarra, 1995), a respiratory-related morphological feature that humans no longer possess (see Fig. 7.). The function of these air sacs is under debate (Boer, 2012; Fitch et al., 2016; Hauser et al., 2002; Negus, 1929, p. 96–120). Suggestions range from “relatively functionless” (Harrison, 1995), for multimodal displaying (Perlman and Salmi, 2017), or to avoid over-oxygenation by “rebreathing” stored CO₂-rich air that was expired previously from the lungs (Maclarnon and Hewitt, 2004; Negus, 1929) to “vocal amplification,” possibly serving to exaggerate body size (Fitch and Hauser, 2002; Hewitt et al., 2002). The sound amplification through air sacs lies in the resonances that arise due to such cavities, but possibly also in the passive elastic recoil of the air sacs, which provides a potent energy reservoir to support airflow (Negus, 1929; Boer, 2012; Maclarnon and Hewitt, 2004). Thus, based on such reasoning, air sacs may have even served to overcome physiological challenges for the respiratory (-vocal) system during locomotion. Importantly, in our translated reading of Hayama (1996), but also Mott (1924), it is argued that air sacs function to stabilize the thorax during brachiation by modulating the explosive expiratory pressures generated by the upper-limb system on the thorax. Air sacs in primates have thus been envisaged as a counter-adaptation to locomotor-respiratory biomechanical constraints.

If – analogous to buccal air sac pumping in the monitor lizards – ape air sacs are a morphological adaption that enables a decoupling (or less rigid coupling) between respiratory(-vocal) and locomotion dynamics, then this would indicate entirely different constraints in respiratory-limb-vocal dynamics as compared to humans. We therefore conclude that we simply know too little at present to problematize the lack of vocal-entangled gesture in non-human primates; however the problem could disappear if we can gain a better understanding of a) air sac functioning in relation to locomotor-respiratory coupling, and b) potential gesture-vocal coupling in apes in general (Liebal et al., 2022).

5.6. What about the brain? Neural implications for limb-respiratory-vocal connections

The coordination stabilities latent in the bio-architecture of the skeleton, muscles, and connective tissues will evolutionarily elicit active neural regulation and optimization (Boggs, 2002). For example, when an external force passively moves the upper limbs, respiration cycles still increase in frequency, suggesting a role for reafferent feedback about muscle and connective tissue deformations (Waisbren et al., 1990). Further, sudden thoracic compression has reflex-like effects on laryngeal closure in humans, suggesting a neural regulatory link between those systems (Baer, 1979; Je, 1973). In rats, fore- and hindlimb central pattern generators (CPGs) and reafferent feedback modulate respiratory CPGs (Gal et al., 2020; Giraudin et al., 2008, 2012), from which can be concluded: “pathways from both the forelimbs and hindlimbs have direct access to the medullary respiratory centers, thereby providing the substrate for coupling the ensemble of spinal respiratory outputs (including phrenic, intercostal and abdominal motoneurons) to ongoing locomotory movements” (Giraudin et al., 2008, p. 2634). Further, the size of human cerebellar regions correlates with the frequency of beat gestures during speaking (Bernard et al., 2015). It turns out that these cerebellar regions are also a key neural regulator of anticipatory postural adjustments (Coffman et al., 2011; Colnaghi et al., 2017), in line with the current perspective that relates gesture, posture, respiration, vocalization, and bipedalism (Pouw et al., 2019a).

In retracing the phylogenetic origins of gesture through the comparative analysis of neurophysiology, Bass and Chagnaud (2012) conclude that the connection between voice and upper-limb control may be traced back to pectoral-fin acoustic signaling in ray-finned (teleost) fish. They comprehensively review work on the role of the caudal hindbrain in teleost fish in the control of muscles for both locomotion and sonic signaling, concluding that the pectoral system has been neutrally co-regulated with sonic signaling from very early, thus even before the emergence of terrestrial quadrupeds.

To conclude, evolutionarily ancient neural regulatory systems in the brain and spinal cord tend to exploit and tune biomechanical redundancies (Damm et al., 2020; Pouw et al., 2021; Jékely et al., 2021; cf. Fitch et al., 2016). On this matter, Terrance Deacon (Deacon, 1998, ch. 8) suggests that vocal neural control tends to be greater for animal species that mechanically entangle respiratory-vocal systems with peripheral skeletal muscle systems (e.g., wingbeats in birds). This is because such peripheral muscle systems tend to have different control hierarchies that are less reflexive and automatic in nature than the (originally) more viscerally controlled vocal system. Thus natural interactions between limb musculoskeletal activity, respiration, and vocalization can then lead to new neural control interdependencies that can support the complexification of vocal control. It is therefore important not to think of our account as “only about the body”; it naturally invokes neural implications.

6. Conclusion

- (1) The question of why vocal-entangled gestures emerged in human communication has been neglected despite the well-developed research program known as multimodal prosody, which shows that in humans, gesture and speech tend to be coupled. We provide evidence that gestures can affect the respiratory-vocal system through the physical impulses that they generate, leading to a cascading effect on subglottal pressure and the acoustic output during vocalization. The connection between motion, respiration, and vocalization brings the new perspective that vocal-motor babbling in early infancy is developmentally productive for gesture-vocal biomechanics. We argue that contrary to current thinking, vocal-entangled hand movements are the first basic gesture to emerge in infants. Vocal-entangled gestures have been shown in light of the phylogeny and the emergence of bipedalism to be connected by a development from locomotor-respiratory coupling to locomotor-respiratory-vocal coupling. We conclude that no matter which popular view of the origins of human communication one follows, the relation between pulsing movements, respiration, and vocalization provides a potent avenue to a much-needed incremental account (Darwin, 1863; Deacon, 1998; Fitch, 2010) of why gestures are entangled with vocalization in humans. Our approach has several advantages: a) it puts the emergence of human gesture back into the realm of ontogeny and roots it in phylogeny, b) it connects the body and the mind, and c) it is in line with biological perspectives that view intentional systems as emergent upon aligning more basic processes into adaptive ways (Deacon, 2011; Juarrero, 1999).
- (2) A challenge arising from our views is how to relate non-representational pulsing gestures to pointing, iconic, or symbolic gestures. Vocal-entangled gestures function indexically by informing others about bodily tensions and their psychological correlates through the performance of a holistic multimodal display. Through gaining new manual skills, for example tool making, such evolutionarily ancient tensioning gestures at some point may have become more strategically employable than simply signaling embodied states such as anger, readiness to fight, or bodily distress in the case of an infant. The strategic reuse and flexibilization of indexical cueing, we would surmise, are a preadaptation to more advanced representation through iconicity.
- (3) Our views suggest several new lines of research. Firstly, more empirical evidence is needed about bodily gestures and their recruitment of different muscles that affect expiration and inspiration (cf. Lancaster et al., 1995), i.e., we need to survey potentially functional gesture-respiratory-vocal muscle synergies (Latash, 2008) that potentially support reaching communicative targets in ways that are more energetically efficient than unimodal productions (e.g., Lancaster et al., 1995; Gillooly and

Ophir, 2010). We further see the importance of biomechanics research in vocal-motor babbling in infants, and developmental research that relates babbling/gesture rates to vocal development as possibly mediated by differences in parental, cultural, or other (e.g., sex) developmental factors. As mentioned, the role of gesture as tensioning bodily actions as it relates to emotional experience and expression is in need of further exploration. These new directions of research aligns with a shift towards a more dynamic monitoring of the role of gesture in communication. For example, there is research showing that when people cannot gesture, there is not a clear impact on acoustics when averaging over the speech acoustics during gesture versus under passive conditions (Cravotta et al., 2019, 2021; Hoetjes et al., 2014). At the same time, researchers in machine learning are reporting successes in predicting when a gesture will occur based on ongoing acoustics during utterances (Yunus et al., 2020). For this research to be reconciled, we need to accept that humans can reach highly similar acoustic targets with different bodily *means*, and therefore we should dynamically monitor those bodily means (Lancaster et al., 1995). Humans can speak, vocalize, or sing without gesturing, and with minimal whole-body participation. But if gesturing is present, the process of speaking, vocalizing, or singing must unfold in a different way (even when the acoustics are nearly indistinguishable in the end if we bin and average them; see, e.g., Pouw et al., 2020a for a comparison between static and dynamic analyses of gesture–speech coupling).

- (4) Finally, a thorough cross-species research program is needed on vocalization and locomotor-respiratory coupling, including in non-human primates. Non-human primates are heavily dependent on their pectoral limbs for locomotion such as during brachiating, yet they are surprisingly the only set of species for which we know nothing about locomotor-respiratory(-vocal) coupling. In extension, there needs to be more research on multimodal aspects of primate communication, which requires a coordination of still disparate gesture versus acoustic research programs on primates (Slocombe et al., 2011; Liebal et al., 2022). Interdisciplinary efforts may further overcome methodological challenges by sharing data in open access databases for animal communication, using video-based motion tracking, and using machine learning algorithms for source separation of multiple voices or filtering background noise.
- (5) From our emphasis on vocal-entangled gesture, a more general reflection on extant theories of multimodal communication emerges. Researchers and theorists may have been looking *through* gestures too much, looking for signs of something that is occluded. This search reflects a pre-theoretical commitment to gestures as carriers of representational content and has hitherto lost its connection to bodily processes. We argue that vocal-entangled gestures are bodily utterances which can have consequences for respiration and vocalization. They are present in ontogeny, have deep roots in phylogeny, and have natural communicative significance.

Conflicts of Interest

We have no conflicts of interest to report.

Data availability

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

Acknowledgments

Wim Pouw and Susanne Fuchs would like to thank James A. Dixon, Steven J. Harrison, Lisette de Jonge-Hoekstra, and Alexandra Paxton, and the Center for the Ecological Study of Perception and Action for

their contributions to the development of these ideas. The authors also would like to thank Asli Ozyurek, Teresa Bejarano Fernández, James B. Harrod, for discussing with us aspects of this paper. We thank Jenny Michlich for first pointing us to cross-species connections way back when, digitally scanning and sharing pages from her hard-copy of Negus (1929), and for proofreading the manuscript. We also thank Azaria Boots for being a second proofreader, and Kirsten Brock for proofreading the revised version. We thank Pilar Prieto, Nuria Esteve Gibert, and Pascal Perrier for reading an early synopsis of the current paper, Nicola Miller for reading and commenting on a preliminary draft, and Aude Noiray, Dan Dediu, Hans-Rutger Bosker, Eva Murillo, Jared L. Tagliatela, Lara Burchardt, Laura L. Koenig, Marcin Włodarczak, Mohammad Ali Nazari, Simone Falk, and Natalie Awad Schwob for providing feedback on a first completed draft of this paper. We sincerely thank two anonymous reviewers and one signing reviewer (Asif Ghazanfar) for their valuable comments on a previous (even more lengthy) version of this manuscript.

Funding

This project has been funded by a VENI grant (VI.Veni 0.201G.047) awarded by the Dutch Research Council (NWO) to Wim Pouw (PI) and the Salammbo grant (FFU791/8-1) awarded by ANR-DFG to Susanne Fuchs (PI). Wim Pouw would further like to acknowledge financial support via a Donders Fellowship awarded by the Donders Center for Brain, Cognition and Behavior, and the Language in Interaction consortium project ‘Communicative Alignment in Brain & Behaviour’ (CABB).

References

- Alexanderson, S., Henter, G.E., Kucherenko, T., Beskow, J., 2020. Style-controllable speech-driven gesture synthesis using normalising flows. *Comput. Graph. Forum (Print.)* 39 (2), 487–496.
- Amazeen, P., Amazeen, E.L., Beek, P.J., 1992. Locomotor-respiratory coupling during wheelchair propulsion. *J. Appl. Physiol.* 4 (72), 1357–1379.
- Anegawa, E., Tsuyama, H., Kusukawa, J., 2008. Lateral cephalometric analysis of the pharyngeal airway space affected by head posture. *Int. J. Oral. Maxillofac. Surg.* 37 (9), 805–809.
- Arbib, M.A., 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28 (2), 105–124.
- Arcadi, A.C., Robert, D., Boesch, C., 1998. Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* 39 (4), 505–518.
- Aruin, A.S., Latash, M.L., 1995. Directional specificity of postural muscles in feed-forward postural reactions during fast voluntary arm movements. *Exp. Brain Res.* 103 (2), 323–332.
- Baer, T., 1979. Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *J. Acoust. Soc. Am.* 65 (5), 1271–1275.
- Banzett, R.B., Mead, J., Reid, M.B., Topulos, G.P., 1992. Locomotion in men has no appreciable mechanical effect on breathing. *J. Appl. Physiol.* 72 (5), 1922–1926.
- Basmajian, J.V., de Luca, C.J., 1985. *Muscles Alive: Their Functions Revealed by Electromyography*. Williams & Wilkins, Baltimore.
- 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. USA* 109, 10677–10684.
- Baudinette, R.V., Gannon, B.J., Runciman, W.B., Wells, S., Love, J.B., 1987. Do cardiorespiratory frequencies show entrainment with hopping in the tammar wallaby? *J. Exp. Biol.* 129 (1), 251–263.
- Berg, K., Beissinger, S., Bradbury, J., 2013. Factors shaping the ontogeny of vocal signals in a wild parrot. *J. Exp. Biol.* 216 (2), 338–345.
- Bernard, J.A., B Millman, Z., Mittal, V.A., 2015. Beat and metaphoric gestures are differentially associated with regional cerebellar and cortical volumes. *Hum. Brain Mapp.* 36 (10), 4016–4030.
- Bernstein, N., 1967. *The Co-ordination and Regulations of Movements*. Pergamon Press, Oxford, UK.
- Blumberg, M., 1992. Rodent ultrasonic short calls: Locomotion, biomechanics, and communication. *J. Comp. Psychol.* 106 (4), 360–365.
- Blumberg, M.S., Sokoloff, G., 2001. Do infant rats cry? *Psychol. Rev.* 108 (1), 83–95.
- 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 (6), 907–914.
- Boer, B.D., 2012. Air sacs and vocal fold vibration: Implications for evolution of speech. *Theor. Et Hist. Sci.* 9 (0), 13–28.
- Boogs, D.F., 2002. Interactions between locomotion and ventilation in tetrapods. *Comp. Biochem. Physiol. Part A: Mol. Integr. Physiol.* 133 (2), 269–288.
- Bolinger, D., 1986. *Intonation and Its Parts: Melody in Spoken English*. Stanford University Press.

- Bosker, H.R., Peeters, D., 2021. Beat gestures influence which speech sounds you hear. *Proc. Royal Soc. B* 288 (1943), 1–9. <https://doi.org/10.1098/rspb.2020.2419>.
- Bramble, D., Carrier, D.R., 1983. Running and breathing in mammals. *Science* 219 (4582), 251–256.
- Bramble, D.M., 1989. Axial-appendicular dynamics and the integration of breathing and gait in mammals. *Am. Zool.* 29, 171–186.
- Burchardt, L.S., Norton, P., Behr, O., Scharff, C., Knörnschild, M., 2019. General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*. *R. Soc. Open Sci.* 6 (1), 181076.
- Burkhardt-Reed, M.M., Long, H.L., Bowman, D.D., Bene, E.R., Oller, D.K., 2021. The origin of language and relative roles of voice and gesture in early communication development. *Infant Behav. Dev.* 65, 101648.
- Capone, N.C., McGregor, K.K., 2004. Gesture development. *J. Speech, Lang., Hear. Res.* 47 (1), 173–186.
- Cardoso, R., Lumini-Oliveira, J., Meneses, R.F., 2019. Associations between posture, voice, and dysphonia: A systematic review. *J. Voice* 33(1), 124 e1–124, e12.
- Carpenter, R.E., 1986. Flight physiology of intermediate-sized fruit bats (Pteropodidae). *J. Exp. Biol.* 120 (1), 79–103.
- Carrier, D.R., Kapoor, A.K., Kimura, T., Nickels, M.K., Satwanti, Scott, E.C., So, J.K., Trinkaus, E., 1984. The energetic paradox of human running and hominid evolution [and Comments and Reply]. *Curr. Anthropol.* 25 (4), 483–495.
- Casasanto, D., Jasmin, K., 2010. Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLOS ONE* 5 (7), e11805.
- Cerdeira, E.P., Garbelline, D., 1999. Electromyographic study of the pectoralis major, serratus anterior and external oblique muscles during respiratory activity in humans. *Electromogr. Clin. Neurophysiol.* 39 (3), 131–137.
- Chu, M., Hagoort, P., 2014. Synchronization of speech and gesture: Evidence for interaction in action. *J. Exp. Psychol.: Gen.* 143, 3.
- Clack, J.A., 2006. The emergence of early tetrapods. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* 232 (2), 167–189.
- Cobo-Lewis, A.B., Oller, D.K., Lynch, M.P., Levine, S.L., 1996. Relations of motor and vocal milestones in typically developing infants and infants with Down syndrome. *Am. J. Ment. Retard.* 100 (5), 456–467.
- Coffman, K.A., Dum, R.P., Strick, P.L., 2011. Cerebellar vermis is a target of projections from the motor areas in the cerebral cortex. *Proc. Natl. Acad. Sci.* 108 (38), 16068–16073.
- Colnaghi, S., Honeine, J.-L., Sozzi, S., Schieppati, M., 2017. Body sway increases after functional inactivation of the Cerebellar Vermis by cTBS. *Cerebellum* 16 (1), 1–14.
- Colonnese, C., Stams, G.J.J.M., Koster, I., Noom, M.J., 2010. The relation between pointing and language development: A meta-analysis. *Dev. Rev.* 30 (4), 352–366.
- Cooper, B.G., Goller, F., 2004. Multimodal signals: Enhancement and constraint of song motor patterns by visual display. *Science* 303 (5657), 544–546.
- Corballis, M.C., 2002. From Hand to Mouth: The Origins of Language. Princeton University Press, Princeton, NJ.
- Cordo, P.J., Nashner, L.M., 1982. Properties of postural adjustments associated with rapid arm movements. *J. Neurophysiol.* 47 (2), 287–302.
- Cravotta, A., Busà, M.G., Prieto, P., 2019. Effects of encouraging the use of gestures on speech. *J. Speech, Lang., Hear. Res.* 62 (9), 3204–3219.
- Cravotta, A., Prieto, P., Busà, M.G., 2021. Exploring the effects of restraining the use of gestures on narrative speech. *Speech Commun.* 135, 25–36.
- Cuffari, E., 2012. Gestural sense-making: hand gestures as intersubjective linguistic enactments. *Phenomenol. Cogn. Sci.* 11 (4), 599–622.
- Damm, L., Varoqui, D., De Cock, V.C., Dalla Bella, S., Bardy, B., 2020. Why do we move to the beat? A multi-scale approach, from physical principles to brain dynamics. *Neurosci. Biobehav. Rev.* 112, 553–584.
- Darwin, C., 1863. The Descent of Man, and Selection in Relation to Sex. Princeton University Press, Princeton.
- Darwin, C., 1948. The Expression of the Emotions in Man and Animals. University of Chicago Press.
- Deacon, T.W., 1998. The Symbolic Species: The Co-Evolution of Language and the Brain. W.W. Norton.
- Deacon, T.W., 2011. Incomplete nature: How mind emerged from matter. WW Norton & Company.
- DeSilva, J., 2021. A Most Interesting Problem: What Darwin's Descent of Man Got Right and Wrong about Human Evolution. Princeton University Press.
- Dohen, M., Roustan, B., 2017. Co-production of speech and pointing gestures in clear and perturbed interactive tasks: Multimodal designation strategies. *Interspeech 2017*. Stockholm, Sweden, pp. 1–5.
- Dolmage, T.E., Janaudis-Ferreira, T., Hill, K., Price, S., Brooks, D., Goldstein, R.S., 2013. Arm elevation and coordinated breathing strategies in patients with COPD. *Chest* 144 (1), 128–135.
- Donald, M., 1991. Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition. Harvard University Press, Boston.
- Dourado, V.Z., Antunes, L.C. d O., Tanni, S.E., de Paiva, S.A.R., Padovani, C.R., Godoy, I., 2006. Relationship of upper-limb and thoracic muscle strength to 6-min walk distance in COPD patients. *Chest* 129 (3), 551–557.
- 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 (1), 11320.
- Dunbar, R., 2016. Human Evolution: Our Brains and Behavior. Oxford University Press, Oxford.
- Ebert, D., Raßler, B., Heftner, H., 2000. Coordination between breathing and forearm movements during sinusoidal tracking. *Eur. J. Appl. Physiol.* 81 (4), 288–296.
- Efron, D., Efron, J.M., Veen, S.V., 1972. Gesture, Race and Culture: A Tentative Study of the Spatio-Temporal and "Linguistic" Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living under Similar as Well as Different Environmental Conditions. Mouton.
- Eilers, R.E., Oller, D.K., Levine, S., Basinger, D., Lynch, M.P., Urbano, R., 1993. The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. *Infant Behav. Dev.* 16 (3), 297–315.
- Ejiri, K., 1998. Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica* 55 (4), 226–237.
- Ejiri, K., Masataka, N., 2001. Co-occurrences of preverbal vocal behavior and motor action in early infancy. *Dev. Sci.* 4 (1), 40–48.
- Esteve-Gibert, N., Guellai, B., 2018. Prosody in the auditory and visual domains: A developmental perspective. *Front. Psychol.* 9 (338), 1–10.
- Esteve-Gibert, N., Prieto, P., 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech, Lang., Hear. Res.* 56 (3), 850–864.
- Esteve-Gibert, N., Prieto, P., 2014. Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun.* 57, 301–316.
- Esteve-Gibert, N., Lovenbruck, H., Dohen, M., D'Imperio, M., 2021. Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech. *Dev. Sci.*, e13154.
- Farmer, C.G., Carrier, D.R., 2000. Ventilation and gas exchange during treadmill locomotion in the American alligator (*Alligator mississippiensis*). *J. Exp. Biol.* 203 (Pt 11), 1671–1678.
- Feyereisen, P., 2017. The Cognitive Psychology of Speech-Related Gesture. Routledge, New York.
- Fitch, T., Hauser, M.D., 2002. Acoustic communication. In: Simmons, A., Fay, R.R., Popper, A.N. (Eds.), *Unpacking "Honesty": Vertebrate Vocal Production and the Evolution of Acoustic Signals*. Springer Handbook of Auditory Research. Springer, New York.
- Fitch, W.T., 2010. The Evolution of Language. Cambridge University Press.
- Fitch, W.T., Kelley, J.P., 2000. Perception of vocal tract resonances by whooping cranes (*Grus americana*). *Ethology* 106 (6), 559–574.
- Fitch, W.T., de Boer, B., Mathur, N., Ghazanfar, A.A., 2016. Monkey vocal tracts are speech-ready. *Sci. Adv.* 2 (12), e1600723.
- Fowler, C.A., 1986. An event approach to the study of speech perception from a direct-realist perspective. *J. Phon.* 14 (1), 3–28.
- Fröhlich, M., Sievers, C., Townsend, S.W., Gruber, T., van Schaik, C.P., 2019. Multimodal communication and language origins: Integrating gestures and vocalizations. *Biol. Rev.* 94 (5), 1809–1829.
- Fuchs, S., Koenig, L.L., Petrone, C., 2019. Exploring the source of short-term variations in respiratory data. *J. Acoust. Soc. Am.* 145 (1), EL66–EL71.
- Funk, G.D., Milson, W.K., Steeves, J.D., 1992. Coordination of wingbeat and respiration in the Canada goose. I. Passive wing flapping. *J. Appl. Physiol.* 73 (3), 1014–1024.
- Gal, J.-P.L., Colnot, E., Cardoit, L., Bacqué-Cazenave, J., Thoby-Brisson, M., Juvin, L., Morin, D., 2020. Modulation of respiratory network activity by forelimb and hindlimb locomotor generators. *Eur. J. Neurosci.* 52 (4), 3181–3195.
- Gärdenfors, P., 2017. Demonstration and pantomime in the evolution of teaching. *Front. Psychol.* 8.
- Geissmann, T., Orgeldinger, M., 2000. The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*. *Anim. Behav.* 60 (6), 805–809.
- Gentilucci, M., Corballis, M.C., 2006. From manual gesture to speech: A gradual transition. *Neurosci. Biobehav. Rev.* 30 (7), 949–960.
- Ghazanfar, A.A., Turesson, H.K., Maier, J.X., van Dinther, R., Patterson, R.D., Logothetis, N.K., 2007. Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr. Biol.* 17 (5–2), 425–430.
- Gillooly, J.F., Ophir, A.G., 2010. The energetic basis of acoustic communication. *Proc. R. Soc. B: Biol. Sci.* 277 (1686), 1325–1331.
- Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., Malik, J., 2019. Learning individual styles of conversational gesture. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 3497–3506.
- Giraudin, A., Cabirol-Pol, M.-J., Simmers, J., Morin, D., 2008. Intercostal and abdominal respiratory motoneurons in the neonatal rat spinal cord: Spatiotemporal organization and responses to limb afferent stimulation. *J. Neurophysiol.* 99 (5), 2626–2640.
- Giraudin, A., Le Bon-Jégo, M., Cabirol, M.-J., Simmers, J., Morin, D., 2012. Spinal and pontine relay pathways mediating respiratory rhythm entrainment by limb proprioceptive inputs in the neonatal rat. *J. Neurosci.* 32 (34), 11841–11853.
- Goldin-Meadow, S., 2003. Hearing Gesture: How Our Hands Help Us Think. Harvard University Press.
- Goldin-Meadow, S., Brentari, D., 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40, e46.
- Gould, S.J., Lewontin, R.C., 1979. The spandrels of San Marco and the Panglossian paradigm. *Proc. R. Soc. B* 205, 581–598.
- Handling, M.A., Curtis, A.S., Miller, S.L., 2010. The origin of the long head of the triceps: A cadaveric study. *J. Shoulder Elb. Surg.* 19 (1), 69–72.
- Hardus, M.E., Lameira, A.R., Schaik, C.S., Wich, S.A., 2009. Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proc. R. Soc. B: Biol. Sci.* 276 (1673), 3689–3694.
- Harrison, D., 1995. The Anatomy and Physiology of the Mamalian Larynx. Cambridge University Press, Cambridge.
- Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298 (5598), 1569–1579.
- Hayama, S., 1996. The origin of the completely closed glottis. Why does not the monkey fall from a tree? *Primate Res.* 12 (2), 179–206. Retrieved from. https://www.jstage.jst.go.jp/article/psj1985/12/2/12_2_179/_article-char/ja/.

- Hewitt, G., MacLarnon, A., Jones, K.E., 2002. The functions of laryngeal air sacs in primates: A new hypothesis. *Folia Primatol.* 73 (2–3), 70–94.
- Hodges, P.W., Gandevia, S.C., 2000a. Activation of the human diaphragm during a repetitive postural task. *J. Physiol.* 522 (Pt 1), 165–175.
- Hodges, P.W., Gandevia, S.C., 2000b. Changes in intra-abdominal pressure during postural and respiratory activation of the human diaphragm. *J. Appl. Physiol.* 89 (3), 967–976.
- Hodges, P.W., Richardson, C.A., 1997. Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Exp. Brain Res.* 114 (2), 362–370.
- Hodges, P.W., Sapsford, R., Pengel, L.H.M., 2007. Postural and respiratory functions of the pelvic floor muscles. *Neurorol. Urodyn.* 26 (3), 362–371.
- Hoetjes, M., Krahmer, E., Swerts, M., 2014. Does our speech change when we cannot gesture? *Speech Commun.* 57, 257–267.
- Hoit, J.D., Banzett, R.B., Brown, R., Loring, S.H., 1990. Speech breathing in individuals with cervical spinal cord injury. *J. Speech, Lang., Hear. Res.* 33 (4), 798–807.
- Holler, J., Levinson, S.C., 2019. Multimodal language processing in human communication. *Trends Cogn. Sci.* 23 (8), 639–652.
- Honda, K., Hirai, H., Masaki, S., Shimada, Y., 1999. Role of vertical larynx movement and cervical lordosis in FO control. *Lang. Speech* 42 (Pt 4), 401–411.
- Hostetter, A.B., Alibali, M.W., 2008. Visible embodiment: Gestures as simulated action. *Psychon. Bull. Rev.* 15 (3), 495–514.
- Hostetter, A.B., Alibali, M.W., 2019. Gesture as simulated action: Revisiting the framework. *Psychon. Bull. Rev.* 26 (3), 721–752.
- Im, S., Baumann, S., 2020. Probabilistic relation between co-speech gestures, pitch accents and information status. *Proc. Ling. Soc. Am.* 5 (1), 685–697.
- Ingbret, D.W., 2008. Tensegrity and mechanotransduction. *J. Bodyw. Mov. Ther.* 12 (3), 198–200.
- Iscoe, S., 1981. Respiratory and stepping frequencies in conscious exercising cats. *J. Appl. Physiol.* 51 (4), 835–839.
- Iverson, J.M., 2010. Multimodality in infancy: Vocal-motor and speech-gesture coordinations in typical and atypical development. *Enfance* 3 (3), 257–274.
- Iverson, J.M., Fagan, M.K., 2004. Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Dev.* 75 (4), 1053–1066.
- Iverson, J.M., Thelen, E., 1999. Hand, mouth and brain: The dynamic emergence of speech and gesture. *J. Conscious. Stud.* 6 (11–12), 19–40.
- Iverson, J.M., Thelen, E., 2003. The hand leads the mouth in ontogenesis too. *Behav. Brain Sci.* 26 (2), 225–226.
- Iverson, J.M., Wozniak, R.H., 2007. Variation in vocal-motor development in infant siblings of children with autism. *J. Autism. Dev. Disord.* 37 (1), 158–170.
- Iverson, J.M., Hall, A.J., Nickel, L., Wozniak, R.H., 2007. The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain Lang.* 101 (3), 198–207.
- Je, R., 1973. Extra-segmental reflexes derived from intercostal afferents: Phrenic and laryngeal responses. *J. Physiol.* 233, 1.
- Jékely, G., Keijzer, F., Godfrey-Smith, P., 2015. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1684), 1–9. <https://doi.org/10.1098/rstb.2015.0181>.
- Jenkins, T., Pouw, W., 2022. A kinematic-acoustic analysis of gesture-speech coupling in aphasia. Preprint. <https://doi.org/10.31234/osf.io/457nd>.
- Juarro, A., 1999. *Dynamics in Action: Intentional Behavior as a Complex System*. MIT Press.
- Kelso, J.A.S., Tuller, B., Harris, K., 1983. A "Dynamic Pattern" perspective on the control and coordination of movement. In: MacNeilage, P. (Ed.), *The Production of Speech*. Springer-Verlag, Berlin, pp. 137–173.
- Kendon, A., 2004. Gesture: Visible Action as Utterance. Cambridge University Press.
- Kendon, A., 2017. Reflections on the "gesture-first" hypothesis of language origins. *Psychon. Bull. Rev.* 24 (1), 163–170.
- Kita, S., Özürek, A., 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48 (1), 16–32.
- Klein, W., Codd, J.R., 2010. Breathing and locomotion: Comparative anatomy, morphology and function. *Respir. Physiol. Neurobiol.* 173, S26–S32.
- Koopmans-van Beinum, F.J., Stelt, J.M., 1986. Early stages in the development of speech movements. In: Lindblom, B., Zetterström, R. (Eds.), *Precursors of Early Speech*. Palgrave Macmillan, London, pp. 37–50.
- Krahmer, E., Swerts, M., 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57 (3), 396–414.
- Krauss, R.M., Chen, Y., Gottesman, R.F., 2000. Lexical gestures and lexical access: A process model. In: McNeill, D. (Ed.), *Language and Gesture*, Language Culture and Cognition. Cambridge University Press, Cambridge, pp. 261–283.
- Krivokapić, J., Tiede, M.K., Tyrone, M.E., 2017. A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Lab. Phonol.* 8 (1), 3.
- Kugler, P.N., Scott Kelso, J.A., Turvey, M.T., 1980. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In: Stelmach, G.E., Requin, J. (Eds.), *Advances in Psychology*, volume 1 of *Tutorials in Motor Behavior*. North-Holland, pp. 3–47.
- Lafontuna, C.L., Reinach, E., Saibene, F., 1996. The effects of locomotor-respiratory coupling on the pattern of breathing in horses. *J. Physiol.* 492, 587–596.
- Lameira, A.R., Eerola, T., Ravignani, A., 2019. Coupled whole-body rhythmic entrainment between two chimpanzees. *Sci. Rep.* 9 (1), 18914.
- Lancaster, W.C., Henson, O.W., Keating, A.W., 1995. Respiratory muscle activity in relation to vocalization in flying bats. *J. Exp. Biol.* 198 (1), 175–191.
- Larsson, M., 2014. Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Anim. Cogn.* 17 (1), 1–14.
- Larsson, M., Richter, J., Ravignani, A., 2019. Bipedal steps in the development of rhythmic behavior in humans. *Music Sci.* 2, 205920431982617.
- Latash, M.L., 2008. Synergy. Oxford University Press.
- LeBarton, E.S., Iverson, J.M., 2016. Associations between gross motor and communicative development in at-risk infants. *Infant Behav. Dev.* 44, 59–67.
- LeCron Foster, M., 1992. Body process in the evolution of language. In: Sheets-Johnstone, M. (Ed.), *Giving the Body Its Due*. State University of New York Press, New York, pp. 208–233.
- Leonard, T., Cummins, F., 2011. The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* 26 (10), 1457–1471.
- Levin, S.M., 1997. Putting the shoulder to the wheel: A new biomechanical model for the shoulder girdle. *Biomed. Sci. Instrum.* 33, 412–417.
- Levin, S.M., 2006. Tensegrity: The new biomechanics. In: Hutson, M., Ellis, R. (Eds.), *Textbook of Musculoskeletal Medicine*. Oxford University Press, Oxford, England, pp. 69–80.
- Levinson, S.C., Holler, J., 2014. The origin of human multi-modal communication. *Philos. Trans. R. Soc. B: Biol. Sci.* 369 (1651), 20130302.
- Liebal, K., Slocombe, K.E., Waller, B.M., 2022. The language void 10 years on: Multimodal primate communication research is still uncommon. *Ethol. Ecol. Evol.* 0 (0), 1–14.
- Lin, H.-C., Green, J.A., 2007. Effects of posture on newborn crying. *Infancy* 11 (2), 175–189.
- Liu, X., Li, P., Wang, Z., Lu, Y., Li, N., Xiao, L., Duan, H., Wang, Z., Li, J., Shan, C., Wu, W., 2019. Evaluation of isokinetic muscle strength of upper limb and the relationship with pulmonary function and respiratory muscle strength in stable COPD patients. *Int. J. Chronic Obstr. Pulm. Dis.* 14, 2027–2036.
- Liu, Y., Shamei, A., Chow, U.Y., Soo, R., Pineda Mora, G., de Boer, G., Gick, B., 2020. F0-related head movement in blind versus sighted speakers. *J. Acoust. Soc. Am.* 148 (2), EL190–EL194.
- Locke, J.L., Bekken, K.E., McMinnlarson, L., Wein, D., 1995. Emergent control of manual and vocal-motor activity in relation to the development of speech. *Brain Lang.* 51 (3), 498–508.
- Loehr, D.P., 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol.* 3 (1), 71–89.
- Longo, L., Di Stadio, A., Ralli, M., Marinucci, I., Ruoppolo, G., Dipietro, L., de Vincenzi, M., Greco, A., 2020. Voice parameter changes in professional musicians-singers singing with and without an instrument: The effect of body posture. *Folia Phoniatr. Et. Logop.* 72 (4), 309–315.
- MacLarnon, A., Hewitt, G., 2004. Increased breathing control: Another factor in the evolution of human language. *Evolut. Anthropol.: Issues, N., Rev.* 13 (5), 181–197.
- MacLarnon, A.M., Hewitt, G.P., 1999. The evolution of human speech: The role of enhanced breathing control. *Am. J. Phys. Anthropol.* 109 (3), 341–363.
- MacNeilage, P.F., 1998. The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21 (4), 499–511.
- MacNeilage, P.F., 2010. *The Origin of Speech*. Oxford University Press.
- MacNeilage, P.F., Davis, B.L., 2000. On the origin of internal structure of word forms. *Science* 288 (5465), 527–531.
- Marler, P., Hobbett, L., 1975. Individuality in a long-range vocalization of wild chimpanzees. *Z. für Tierpsychol.* 38 (1), 97–109.
- McClave, E., 1998. Pitch and manual gestures. *J. Psycholinguist. Res.* 27 (2), 69–89.
- McCoy, S.J., 2012. *Your Voice: An Inside View*. Inside View Press.
- Mccune, L., 2021. Laryngeal vocalizations: Grunts as a gateway to language? *Lingua* 264, 103161.
- McDermott, W.J., Van Emmerik, R.E.A., Hamill, J., 2003. Running training and adaptive strategies of locomotor-respiratory coordination. *Eur. J. Appl. Physiol.* 89 (5), 435–444.
- McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. *Nature* 264 (5588), 746–748.
- McNeill, D., 1992. *Hand and Mind: What Gestures Reveal about Thought*. University Of Chicago Press, Chicago.
- McNeill, D., 2000. *Language and Gesture*. Cambridge University Press.
- McNeill, D., 2005. *Gesture and Thought*. The University of Chicago Press, Chicago.
- McNeill, D., 2012. *How Language Began: Gesture and Speech in Human Evolution*. Cambridge University Press.
- Mehr, S.A., Krasnow, M.M., Bryant, G.A., Hagen, E.H., 2021. Origins of music in credible signaling. *Behav. Brain Sci.* 44, 1–41.
- Mendoza-Denton, N., Jannedy, S., 2011. Semiotic layering through gesture and intonation: A case study of complementary and supplementary multimodality in political speech. *J. Engl. Linguist.* 39 (3), 265–299.
- Merleau-Ponty, M., 1945. *Phenomenology of Perception*. Routledge.
- Micheletta, J., Engelhardt, A., Matthews, L., Agil, M., Waller, B.M., 2013. Multicomponent and multimodal lipsmacking in crested macaques (*Macaca nigra*): Lipsmacking behavior in crested macaques. *Am. J. Primatol.* 75 (7), 763–773.
- Miller, N.A., Gregory, J.S., Semple, S.I.K., Aspden, R.M., Stollery, P.J., Gilbert, F.J., 2012b. Relationships between vocal structures, the airway, and craniocervical posture Investigated using magnetic resonance imaging. *J. Voice* 26 (1), 102–109.
- Miller, N.A., Gregory, J.S., Semple, S.I.K., Aspden, R.M., Stollery, P.J., Gilbert, F.J., 2012a. The effects of humming and pitch on craniofacial and craniocervical morphology measured using MRI. *J. Voice* 26 (1), 90–101.
- 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 (5), 554–564.

- Moccia, D., Nackashi, A.A., Schilling, R., Ward, P.J., 2016. Fascial bundles of the infraspinatus fascia: Anatomy, function, and clinical considerations. *J. Anat.* 228 (1), 176–183.
- Moreno-Núñez, A., Rodríguez, C., Del Olmo, M.J., 2017. Rhythmic ostensive gestures: How adults facilitate infants' entrance into early triadic interactions. *Infant Behav. Dev.* 49, 168–181.
- Moreno-Núñez, A., Murillo, E., Casla, M., Rujas, I., 2021. The multimodality of infant's rhythmic movements as a modulator of the interaction with their caregivers. *Infant Behav. Dev.* 65, 101645.
- Morett, L.M., Landi, N., Irwin, J., McPartland, J.C., 2020. N400 amplitude, latency, and variability reflect temporal integration of beat gesture and pitch accent during language processing. *Brain Res.* 1747, 147059.
- Morgenstern, A., Chévreuil, L., Blondel, M., Vincent, C., Thomas, C., Boutet, D., 2021. "Of thee I sing": An opening to Dominique Boutet's kinesiological approach to gesture. *Lang. Modalities* 1, 3–16.
- Moss, C.F., Bohn, K., Gilkerson, H., Surlykke, A., 2006. Active listening for spatial orientation in a complex auditory scene. *PLOS Biol.* 4 (4), e79.
- Mott, F., 1924. A study by serial sections of the structure of the larynx of Hylobates syndactylus (Siamese gibbon). In: Proceedings of the Zoological Society of London, Vol. 94. Blackwell Publishing Ltd., Oxford, UK, pp. 1161–1170.
- Murgiano, M., Motamed, Y., Vigliocco, G., 2021. Situating language in the real-world: The role of multimodal iconicity and indexicality. *J. Cogn.* 4, 1–38.
- Murillo, E., Ortega, C., Otones, A., Rujas, I., Casla, M., 2018. Changes in the synchrony of multimodal communication in early language development. *J. Speech, Lang., Hear. Res.* 61 (9), 2235–2245.
- Murillo, E., Montero, I., Casla, M., 2021. Contribution of gular pumping to lung ventilation in monitor lizards. *Front. Psychol.* 12, 616812.
- Nassar, P.N., Jackson, A.C., Carrier, D.R., 2001. Entrainment of the natural frequencies of running and breathing in guinea fowl (*Numida meleagris*). *J. Exp. Biol.* 204 (9), 1641–1651.
- Negus, V.E., 1929. The Mechanism of the Larynx. Wm. Heinemann (medical books) Limited, London, pp. 96–120.
- Novack, M.A., Goldin-Meadow, S., 2017. Gesture as representational action: A paper about function. *Psychon. Bull. Rev.* 24 (3), 652–665.
- O'Connell, D.G., Hinman, M.R., Hearne, K.F., Michael, Z.S., Nixon, S.L., 2014. The effects of "grunting" on serve and forehand velocities in collegiate tennis players. *J. Strength Cond. Res.* 28 (12), 3469–3475.
- O'Halloran, J., Hamill, J., McDermott, W.J., Remelius, J.G., Van Emmerik, R.E.A., 2012. Locomotor-respiratory coupling patterns and oxygen consumption during walking above and below preferred stride frequency. *Eur. J. Appl. Physiol.* 112 (3), 929–940.
- Oller, D.K., Budner, E.H., Ramsdell, H.L., Warlaumont, A.S., Chorna, L., Bakeman, R., 2013. Functional flexibility of infant vocalization and the emergence of language. *Proc. Natl. Acad. Sci.* 110 (16), 6318–6323.
- Oller, D.K., Caskey, M., Yoo, H., Bene, E.R., Jhang, Y., Lee, C.-C., Bowman, D.D., Long, H. L., Budner, E.H., Vohr, B., 2019. Preterm and full term infant vocalization and the origin of language. *Sci. Rep.* 9 (1), 14734.
- Owerkowicz, T., Farmer, C.G., Hicks, J.W., Brainerd, E.L., 1999. Contribution of gular pumping to lung ventilation in monitor lizards. *Science* 284 (5420), 1661–1663.
- Ozcaliskan, S., Goldin-Meadow, S., 2011. Is there an iconic gesture spurt at 26 months? In: Stam, G., Ishino, M. (Eds.), *Integrating Gestures: The Interdisciplinary Nature of Gesture*. Gesture Studies. John Benjamins Publishing Company, Amsterdam, Netherlands, pp. 163–174.
- Paolo, E.A.D., Cuffari, E.C., Jaeger, H.D., 2018. Linguistic Bodies: The Continuity between Life and Language. MIT Press.
- Parrell, B., Goldstein, L., Lee, S., Byrd, D., 2014. Spatiotemporal coupling between speech and manual motor actions. *J. Phon.* 42, 1–11.
- Pearson, L., Pouw, W., 2022. Gesture-vocal coupling in music performance: A kinematic-acoustic analysis of a south Indian vocal music corpus. *Ann. N. Y. Acad. Sci.*
- Perkins, W.H., Perkins, W.H., Kent, R.D., 1986. Functional Anatomy of Speech, Language, and Hearing: A Primer. Allyn and Bacon.
- Perlman, M., Salmi, R., 2017. Gorillas may use their laryngeal air sacs for whinny-type vocalizations and male display. *J. Lang. Evol.* 2 (2), 126–140.
- Perlman, M., Patterson, F.G., Cohn, R.H., 2012. The human-fostered gorilla Koko shows breath control in play with wind instruments. *Biolinguistics* 6 (3–4), 433–444.
- Perry, S.F., Similowski, T., Klein, W., Codd, J.R., 2010. The evolutionary origin of the mammalian diaphragm. *Respir. Physiol. Neurobiol.* 171 (1), 1–16.
- Perségl, L., Jordan, M., Viala, D., 1991. Evidence for the entrainment of breathing by locomotor pattern in human. *J. De. Physiol.* 85 (1), 38–43.
- 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 (3), 1715–1725.
- Pettersen, V., 2006. Preliminary findings on the classical singer's use of the pectoralis major muscle. *Folia Phoniatr. Et. Logop.* 58 (6), 427–439.
- Pettersen, V., Westgaard, R.H., 2004. Muscle activity in professional classical singing: A study on muscles in the shoulder, neck and trunk. *Logop. Phoniatr. Vocology* 29 (2), 56–65.
- Pettersen, V., Westgaard, R.H., 2005. The activity patterns of neck muscles in professional classical singing. *J. Voice: Off. J. Voice Found.* 19 (2), 238–251.
- Pettersen, V., Bjørkøy, K., Torp, H., Westgaard, R.H., 2005. Neck and shoulder muscle activity and thorax movement in singing and speaking tasks with variation in vocal loudness and pitch. *J. Voice: Off. J. Voice Found.* 19 (4), 623–634.
- Pisanski, K., Cartei, V., McGettigan, C., Raine, J., Reby, D., 2016. Voice modulation: A window into the origins of human vocal control? *Trends Cogn. Sci.* 20 (4), 304–318.
- Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* 1–13. <https://doi.org/10.1038/s41583-020-0304-4>.
- Pouw, W., Dixon, J.A., 2019. Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cogn. Sci.* 43 (3), e12721.
- Pouw, W., Harrison, S.J., Dixon, J.A., 2020b. 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.* <https://doi.org/10.1038/s41598-022-18300-x> (In press).
- Pouw, W., de Nooijer, J.A., van Gog, T., Zwaan, R.A., Paas, F., 2014. Toward a more embedded/extended perspective on the cognitive function of gestures. *Front. Psychol.* 5 (359), 1–14.
- Pouw, W., Dixon, J.A., 2022. What you hear and see specifies the perception of a limb-respiratory-vocal act. *Proceedings of the Royal Society B* 289 (1979), 1026. <https://doi.org/10.1098/rspb.2022.1026>.
- Pouw, W., Harrison, S.J., Dixon, J.A., 2019a. Gesture-speech physics: The biomechanical basis of the emergence of gesture-speech synchrony. *J. Exp. Psychol.: Gen.* 149, 391–404.
- Pouw, W., Paxton, A., Harrison, S.J., Dixon, J.A., 2019b. Acoustic specification of upper limb movement in voicing. *Proceedings of the 6th Meeting of Gesture and Speech in Interaction*. Universitätsbibliothek Paderborn, Paderborn, Germany, pp. 75–80.
- Pouw, W., de Jonge-Hoekstra, L., Harrison, S.J., Paxton, A., Dixon, J.A., 2020a. Gesture-speech physics in fluent speech and rhythmic upper limb movements. *Ann. N. Y. Acad. Sci.* 1491 (1), 89–105.
- Pouw, W., Harrison, S.J., Esteve-Gibert, N., Dixon, J.A., 2020c. Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *J. Acoust. Soc. Am.* 148 (3), 1231–1247.
- Pouw, W., Paxton, A., Harrison, S.J., Dixon, J.A., 2020d. Acoustic information about upper limb movement in voicing. *Proc. Natl. Acad. Sci.* 117 (12), 11364–11367.
- Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R., Wiggins, G., 2021. Multilevel rhythms in multimodal communication. *Philos. Trans. R. Soc. B: Biol. Sci.* 376, 1835.
- Profeta, V.L.S., Turvey, M.T., 2018. Bernstein's levels of movement construction: A contemporary perspective. *Hum. Mov. Sci.* 57, 111–133.
- Raine, J., Pisanski, K., Reby, D., 2017. Tennis grunts communicate acoustic cues to sex and contest outcome. *Anim. Behav.* 130, 47–55.
- Rączaszek-Leonardi, J., Kelso, S., 2008. Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *N. Ideas Psychol.* 26 (2), 193–207.
- Raine, J., Pisanski, K., Bond, R., Simmer, J., Reby, D., 2019. Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLOS ONE* 14 (3), e0213034.
- Ravignani, A., Kotz, S.A., 2020. Breathing, voice, and synchronized movement. *Proc. Natl. Acad. Sci.* 117 (38), 23223–23224.
- Ravignani, A., Dalla Bella, S., Falk, S., Kello, C.T., Noriega, F., Kotz, S., 2019. Rhythm in speech and animal vocalizations: A cross-species perspective. *Ann. N. Y. Acad. Sci.* 1453 (1), 79–98.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W.T., Clutton-Brock, T., 2005. Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proc. R. Soc. B: Biol. Sci.* 272 (1566), 941–947.
- Rochet-Capellán, A., Laboissière, R., Galván, A., Schwartz, J., 2008. The speech focus position effect on jaw–finger coordination in a pointing task. *J. Speech, Lang., Hear. Res.* 51 (6), 1507–1521.
- Ruiter, D., Peter, J., 2007. Postcards from the mind: The relationship between speech, imagistic gesture and thought. *Gesture* 7 (1), 21–38.
- de Ruiter, J.P., 2000. The production of gesture and speech. In: McNeill, D. (Ed.), *Language and Gesture*, Language Culture and Cognition. Cambridge University Press, Cambridge, pp. 284–311.
- Rusiewicz, H.L., Esteve-Gibert, N., 2018. Set in time. In: Prieto, P., Esteve-Gibert, N. (Eds.), *The Development of Prosody in First Language Acquisition*. John Benjamins Publishing Company, p. 103.
- Rusiewicz, H.L., Susan, S., Iverson, J., Szuminsky, N., 2013. Effects of prosody and position on the timing of deictic gestures. *J. Speech, Lang., Hear. Res.* 56 (2), 458–470.
- Savage, P.E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., Fitch, W.T., 2020. Music as a coevolved system for social bonding. *Behav. Brain Sci.* 44, 1–36.
- Scarr, G., Harrison, H., 2017. Examining the temporo-mandibular joint from a biotensegrity perspective: A change in thinking. *J. Appl. Biomed.* 15 (1), 55–62.
- Sheets-Johnstone, M., 2011. *The Primacy of Movement*. John Benjamins, Amsterdam.
- Shockley, K., Santana, M.-V., Fowler, C.A., 2003. Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29 (2), 326–332.
- Shockley, K., Baker, A.A., Richardson, M.J., Fowler, C.A., 2007. Articulatory constraints on interpersonal postural coordination. *J. Exp. Psychol.: Hum. Percept. Perform.* 33 (1), 201–208.
- 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. *Neurosci. Lett.* 429 (1), 64–68.
- Simons, R.S., 1999. Running, breathing and visceral motion in the domestic rabbit (*Oryctolagus cuniculus*): Testing visceral displacement hypotheses. *J. Exp. Biol.* 202 (5), 563–577.
- Simonyan, K., 2014. The laryngeal motor cortex: Its organization and connectivity. *Curr. Opin. Neurobiol.* 28, 15–21.
- Slocombe, K.E., Waller, B.M., Liebal, K., 2011. The language void: The need for multimodality in primate communication research. *Anim. Behav.* 81 (5), 919–924.
- Sterelny, K., 2012. Language, gesture, skill: The co-evolutionary foundations of language. *Philos. Trans. R. Soc. B: Biol. Sci.* 367 (1599), 2141–2151.
- Stoltmann, K., Fuchs, S., 2017. Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate. *J. Multimodal Commun. Stud.* 4 (1–2), 63–68.

- 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 (1), 37–48.
- Tammany, J.E., O'Connell, D.G., Latham, S.E., Rogers, J.A., Sugar, T.S., 2021. The effect of grunting on overhead throwing velocity in collegiate baseball pitchers. *Int. J. Sports Sci. Coach.* 16 (5), 1111–1116.
- Tchernichovski, O., Oller, D.K., 2016. Vocal development: How marmoset infants express their feelings. *Curr. Biol.* 26 (10), R422–R424.
- Thelen, E., 1979. Rhythmic stereotypes in normal human infants. *Anim. Behav.* 27, 699–715.
- Tomasello, M., 2008. The Origins of Human Communication. MIT Press, Cambridge, MA.
- Trujillo, J.P., Özürek, A., Holler, J., Drijvers, L., 2021. Speakers exhibit a multimodal Lombard effect in noise. *Sci. Rep.* 11 (1), 16721.
- Turvey, M.T., Fonseca, S.T., 2014. The medium of haptic perception: A tensegrity hypothesis. *J. Mot. Behav.* 46 (3), 143–187.
- Wagner, P., Malisz, Z., Kopp, S., 2014. Gesture and speech in interaction: An overview. *Speech Commun.* 57, 209–232.
- Waisbren, S.J., Whiting, C.S., Nadel, E.R., 1990. Effects of passive limb movement on pulmonary ventilation. *Yale J. Biol. Med.* 63 (6), 549–556.
- Walker, A., Shipman, P., 1996. The Wisdom of the Bones: In Search of Human Origins. Knopf.
- Willard, F.H., Vleeming, A., Schuenke, M.D., Danneels, L., Schleip, R., 2012. The thoracolumbar fascia: Anatomy, function and clinical considerations. *J. Anat.* 221 (6), 507–536.
- Wong, J.G., Waters, D.A., 2001. The synchronisation of signal emission with wingbeat during the approach phase in soprano pipistrelles (*Pipistrellus pygmaeus*). *J. Exp. Biol.* 204 (Pt 3), 575–583.
- Wright, E., Grawunder, S., Ndayishimiye, E., Galbany, J., McFarlin, S.C., Stoinski, T.S., Robbins, M.M., 2021. Chest beats as an honest signal of body size in male mountain gorillas (*Gorilla beringei beringei*). *Sci. Rep.* 11 (1), 6879.
- Ybarra, M.A.S., 1995. A comparative approach to the non-human primate vocal tract: Implications for sound production. In: Zimmermann, E., Newman, J.D., Jürgens, U. (Eds.), Current Topics in Primate Vocal Communication. Springer, US, Boston, MA, pp. 185–198.
- Young, I.S., Warren, R.D., Altringham, J.D., 1992. Some properties of the mammalian locomotory and respiratory systems in relation to body mass. *J. Exp. Biol.* 164 (1), 283–294.
- Yunus, F., Clavel, C., Pelachaud, C., 2020. Sequence-to-sequence predictive model: From prosody to communicative gestures. *arXiv* 2008, 07643.
- Zahavi, A., Zahavi, A., 1999. The Handicap Principle: A Missing Piece of Darwin's Puzzle. Oxford University Press.
- Zhang, Y.S., Ghazanfar, A.A., 2016. Perinatally influenced autonomic system fluctuations drive infant vocal sequences. *Curr. Biol.* 26 (10), 1249–1260.
- Zhang, Y.S., Ghazanfar, A.A., 2018. Vocal development through morphological computation. *PLOS Biol.* 16 (2), e2003933.
- Zywickzynski, P., Wacewicz, S., Sibierska, M., 2018. Defining pantomime for language evolution research. *Topoi* 37 (2), 307–318.

Glossary

beat gesture: A communicative hand movement with a pulse quality that is timed with a speech segment and has a pragmatic function.

constraint: Some factor that expands, limits, or transforms an organism's degrees of freedom (Bernstein, 1967; Kugler et al., 1980). A constraint is therefore not only a limiting factor, but it can also mean that new opportunities become available. For example, gravity constrains an organism to jump at a certain height, but it also constrains it in a way that enables running as a stable mode of locomotion (Kugler et al., 1980).

gesture: A bodily movement or posture that is communicative in potential. A body movement functions communicatively when perceivers are sensitive to the producer's bodily movement/posture in relation to some state of affairs. These signals can convey information about a certain state of affairs through (loose) covariance relations (e.g., a circle-tracing body movement iconically covaries with objects with circular edges such as a soccer ball, and this body movement is a gesture if a perceiver is potentially sensitive to this covariance).

impulse: The change in momentum of an object (or body segment). The momentum of a segment is determined by the mass of the segment (M) * velocity of the segment (V). Since the mass of a body segment can be treated as a constant, impulse effectively equates to a change in the velocity of the segment ΔV (i.e., acceleration, or A), which, when multiplied by the mass, yields a quantity of force (F), given that $F = M * A$. The physical impulse (J), then, is the force exerted over some time window (w) starting from some time point (t0): $J = \int_{t0}^{t0+w} F dt$

physical impulse: The cascading (mechanical) effects of a local impulse of a body segment on wider musculoskeletal activity.

pulse: A rapid change in the amplitude of a signal.