accepted as Kadavá, Š., Ćwiek, A., & Fuchs, S. (2026). Communicating from head to toe: The physiological processes in multimodal language. In J. A. Bateman & C.-I. Tseng (Eds), The Handbook of Linguistics and Multimodality. Wiley/Blackwell.

Communicating from head to toe: the physiological processes in multimodal language

Šárka Kadavá[1,2,3], Aleksandra Ćwiek[1,4], Susanne Fuchs[1]

[1] Leibniz-Centre General Linguistics, Berlin, Germany

[2] University of Göttingen, Germany

[3] Radboud University, Nijmegen, Netherlands

[4] Centre for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland

Abstract

This chapter examines how physiological processes shape multimodal language production and perception across the entire body. Drawing on interdisciplinary research from phonetics, motor control, and movement science, we demonstrate that language is fundamentally physical, constrained and enabled by anatomical structures from head to toe. We show how the head's extrinsic laryngeal muscles create mechanical linkages between head movement and vocal production, how upper limb gestures influence speech through the tensegrity architecture of the torso affecting respiratory systems, and how lower body posture provides the foundational stability enabling upper body communication. These physiological interdependencies reveal that communicative behavior emerges from integrated bodily systems rather than isolated articulators. We argue for developing holistic models that incorporate breathing, postural control, and biomechanical constraints as central organizing principles of multimodal communication.

Keywords: physiology, multimodal communication, breathing, posture, multimodal prosody, gesture-speech coordination, biomechanical constraints

1. Introduction to physiological features in multimodal language

What is language? Everyone will agree that language is central to human existence, and yet if we ask this question to members of a scientific community, the answer will heavily depend on the tradition they were brought up in. In 1954, Morris Halle wrote: "*Real languages are not minimal redundancy codes invented by scholars fascinated by the powers of algebra, but social institutions serving fundamental needs of living people in a real world. [In trying to understand] how human beings communicate by means of language, it is impossible for us to discount physical considerations, [i.e.] the facts of physics and physiology*." (Halle 1954: 79–80, cited in Ohala, 1978, p. 5)

In line with this account, this chapter draws on the linguistic tradition that explores how patterns found in languages can be derived from principles which are independent from language itself but related to the physical properties of the body that shapes and constrains the way we talk (Ohala and Solé, 2008). Why specifically this tradition? Because language is movement and involves coordinated actions of the whole body. In the following sections, we review interdisciplinary research on the biological, physical, and cognitive foundations of multimodal communication, drawing on movement science, motor control, phonetics, and phonology.

Starting with a simple illustration of how important physical properties actually are, you may remember the tale of the *Three Little Pigs*. Each pig builds a house to protect itself from a wolf. The first pig chooses grass, a flexible but fragile material, and the wolf easily blows the house away, so that this pig runs to the next one. The second pig relies on wood, which demands more effort from the wolf before it finally gives away. The two pigs are now running to the third who constructed a house of stone. No matter how vigorously the wolf attacks, the stone house withstands every attempt, and the pigs remain safe.

The moral is obvious: the outcome depends on the physical properties of the materials. Grass, with its flexibility, cannot withstand external stress. Wood offers more strength but is still vulnerable. Stone combines mass and durability, giving it the capacity to resist destructive forces. If the pigs had been confronted by an earthquake instead, the story would unfold differently: the more flexible houses of grass and wood would have had an advantage over the rigid but brittle stone.

Such properties are not only decisive for objects such as houses, but also for living entities. Organisms have specific material characteristics of their own structures, from the hardness of bones to the elasticity of skin and muscle. Imagine the potential movement of an animal without bones, like an octopus, in comparison to a mammal with bones, such as a cat. And just as a movement can be influenced by the presence or absence of bones, physical features are crucial for what kind of sounds we produce.

Some studies have dealt with physical constraints in speech production and focused particularly on the hard structure of the vocal tract, especially the jaw-teeth complex and the palate shape (e.g., Brunner, Fuchs and Perrier, 2009; Weirich and Fuchs, 2013; Blasi *et al.*, 2019; Dediu, Janssen and Moisik, 2019). In a comprehensive study, Blasi and colleagues (2019) observed that labiodental consonants such as /f/ and /v/ are often absent in languages of hunter-gatherer communities. Their edge-to-edge bite configuration is adapted to heavy diets, which makes the production of labiodentals less likely and effortful. In contrast, modern populations eat a softer diet and have developed an overbite. Since labiodentals require the lower lip to be placed on the upper incisors, which is exactly the characteristic of an overbite, this arrangement led to more frequent productions of labiodentals, ultimately incorporating them into phonological inventories of the world's languages.

The palate is another constraint: it limits the tongue's movement, but it also allows the tongue to produce shapes that it would be unable to produce without this upper boundary, such as /s/ (Stone, 1991). The individual steepness of the alveolar ridge affects the way speakers realize phonemic contrasts at the alveolar-postalveolar place of articulation (Weirich and Fuchs, 2013). The curvature of the palate in the coronal plane constrains how much variability speakers produce (Brunner, Fuchs and Perrier, 2009). Dediu *et al.* (2021) proposed that if we look into historical changes in vocal tract morphology, in particular the hard structures that were fossilized, we may be able to reconstruct inter-individual variation, sound change, and model the potential functions these vocal tract forms may be able to produce.

Breathing constitutes another physiological process that is intrinsically linked to speech. Fuchs and Rochet-Capellan (2021) demonstrated that breathing patterns are integral to speech planning and execution, with respiratory cycles providing the temporal scaffolding around which utterances are organized. In other words, the possible duration of a breath constrains the length of an utterance. The robust finding that speech intervals cluster around two seconds across diverse languages and cultures (Burchardt, Fuchs and Paschen, 2025) may reflect this respiratory constraint operating at the most basic level of linguistic organization.

These examples demonstrate the importance of physical properties in speech. But why should we stop here? Language is multimodal and is communicated from the head to the toe. Thus, in this chapter, we will argue that physical constraints and affordances of the whole body deserve more systematic attention in our understanding of multimodal communication.

2. Multimodal phenomena from head to toe linking to physiological features

In what follows, we collect vast evidence that what we call language extends far beyond speech. In the spirit of the popular children's song *Head, Shoulders, Knees and Toes*, we start at the top with the head (*Section 2.1*), continue through the upper body (*Section 2.2*) to land at the roots of our bodies, the legs and toes (*Section 2.3*). In each section, we first introduce a selected body of literature on how the concrete body part participates in communication and social interaction, focusing on the functional component. After that, we review evidence showing that the various functions are not arbitrary but emerge from anatomical and physiological mechanisms.

2.1. Phenomena of the head

2.1.1. How does the head participate in language?

The head houses the articulatory organs for speech production as well as the sensory organs for vision and hearing. It is also where many of the visual cues for meaning emerge. Listeners are, in turn, attending to these signals, facilitating efficient integration of conveyed messages. For instance, Paris et al. (2013) reported that visual speech (e.g., information about lip movements) presented before auditory speech speeds up behavioral response to stimuli when compared to an auditory-only condition. A well-known demonstration of how integration of auditory and visual signals interact is the McGurk effect, where seeing a speaker producing the syllable (e.g., /ga/) while hearing another (e.g., /ba/) leads to the perception of a third, fused syllable (e.g., /da/) (McGurk and McDonald, 1976).

Communicative head movements include mainly head nods and shakes. Head nodding is strongly associated with agreement and acknowledgement. In Japanese, for instance, nods occur in about 42% of listener responses, functioning as a primary backchannel mechanism that sustains conversational flow and prevents problematic silences (Maynard, 1986; Kita and Ide, 2007). Head shakes, conversely, are widely used to signal negation. Cross-cultural preference in expressing "no" with a head shake has been documented across Asia, Africa, Europe, and the Americas (Bross, 2020). Exceptions exist, such as in Bulgaria and southern Albania, where head shaking conventionally means "yes," demonstrating how culture-specific practices can diverge from widespread patterns (Darwin, 1872; Kendon, 2002; Bross, 2020).

Beyond nodding and shaking, the head systematically contributes to prosody. Head movements align with pitch accents, stress, and boundaries, forming what has been called *visual prosody*. In noisy conditions, listeners identify syllables more accurately when natural head movements are present compared to conditions where they are suppressed (Munhall *et al.*, 2004). For example, lexical tones in Mandarin can be identified visually with above-chance accuracy after minimal training, based on the head and neck movements that accompany tone production (Chen and Massaro, 2008). A study of French prosody revealed that speakers modulate head nod kinematics (i.e., range, duration and stiffness) as a form of prosodic enhancement, with differences between two observed strategies becoming more pronounced under stronger focus conditions (Carignan *et al.*, 2024). Further, Pagel *et al.* (2023) reported that head movements increase in amplitude and speed during clear or loud speech, mirroring greater articulatory effort in the vocal tract.

The head is also the locus for orofacial expressions, which interact closely with vocal production. Eyebrow raising marks prosodic prominence, focus, or interrogativity across various languages and contexts (Krahmer and Swerts, 2004; Borràs-Comes *et al.*, 2014; Nota, Trujillo and Holler, 2021). Orofacial expressions can also act as a compensatory strategy when communication channels are compromised:

when whispering, speakers produce more pronounced orofacial expressions (e.g., eyebrow movements and eye openings), compensating for the lack of fundamental frequency information (Żygis and Fuchs, 2023; Sardhaei, Żygis and Sharifzadeh, 2024).

Head involvement in communication develops early. Infants begin to gain control over their heads within the first few months of life, a crucial step that later enables them to engage in joint attention by following an adult's gaze and head orientation (Morales, Mundy and Rojas, 1998). Even before they can produce speech sounds, they use head movements in communicative exchanges. In the second half of their first year, rudimentary head nods and shakes emerge, which are later developed to communicate assent and dissent (Kettner and Carpendale, 2013). Research also shows that 9- to 24-month-old infants produce head movements that are tightly coupled in time with vocalizations (Borjon *et al.*, 2024). In turn, caregivers modify their head movement patterns when interacting with infants (Shepard, Spence and Sasson, 2012). This parallels the acoustic modifications in infant-directed speech (Cristia, 2013), suggesting coordinated adaptations across multiple communicative channels.

2.1.2. How is the head linked to physiology?

The human head comprises approximately 8% of total body mass. The adoption of bipedalism positioned this substantial mass atop a vertical, flexible cervical spine – creating what Walker and Shipman (1997, p. 199) famously described as an engineering challenge "akin to balancing an apple on top of a moving pencil."

What creates a direct physiological link between head movement and vocal production is the network of extrinsic laryngeal muscles, commonly known as the *strap muscles* which include the sternohyoid, sternothyroid, thyrohyoid, and omohyoid. The larynx is suspended in the neck by this web of muscles, some of which connect directly to the skull and jaw, creating a mechanical linkage system where changes in head posture can alter tension on the vocal apparatus (Erickson, Baer and Harris, 1982; Honda, 2008).

An upward head rotation directly influences vocal fold tension through the engagement of both intrinsic and extrinsic laryngeal muscles (Hwan Hong *et al.*, 1997; Honda, 2008). The cricothyroid muscle, among the muscles central to *f*0 control through its regulation of vocal fold tension, works in an antagonistic relationship with the strap muscles (Erickson, Baer and Harris, 1982; Titze, 1994). Research using *in vivo canine* models demonstrates that contraction of different strap muscles produces predictable changes in *f*0: sternohyoid and sternothyroid muscle activation corresponds to raised *f*0 and vocal intensity, while thyrohyoid muscle activation produces the opposite effect (Hwan Hong *et al.*, 1997). The intrinsic laryngeal muscles must coordinate precisely with these extrinsic systems and air pressure differential to control vocal fold characteristics (Zhang, 2016). In an MRI study, Miller and colleagues (2012) observed that raising pitch from a low to a high note is accompanied by systematic postural and structural adjustments, including increased craniocervical angles and elevation of the hyoid and larynx.

Recent experimental evidence further corroborates the existence of this physiological coupling. Munhall *et al.* (2004) found that speaker head movement during speech production explains 63% of the variance in *f*0 using sentence-by-sentence analysis, with head elevation correlating with *f*0 increases and lowering with *f*0 decreases. Liu *et al.* (2020) extended this work through a comparison of congenitally blind and sighted speakers, finding positive correlations between *f*0 and vertical head movement in both groups, with stronger correlations in blind speakers compared to sighted speakers. Importantly, sighted speakers displayed higher head movement per semitone ratios, suggesting that while the physiological coupling operates independently of visual input, sighted speakers enhance their head movements for visual signalling. A more recent motion capture and acoustic analyses study demonstrated that head rotation angle is a robust predictor of *f*0, with upward head rotation producing systematic increases in fundamental frequency (Ćwiek and Fuchs, 2025). The study reveals approximately a 0.85 Hz increase in *f*0 for every degree of upward head rotation.

The neurobiological basis for these head–voice connections may extend beyond biomechanics: the Sensorimotor Account of Multimodal Prosody (SAMP; Momsen, 2024) proposes that the vestibular system (i.e., the inner ear organs that detect head movement and help maintain balance) plays a crucial role in speech by integrating acoustic prosodic features with inertial motion signals from co-speech movements across the entire body. This framework suggests that when we speak, the vestibular system processes not only how our head moves but also the vibrational energy produced by vocalization itself.

Physiological links can also be found between speech and facial expressions. Smiling represents one of the most studied facial gestures, affecting both speech production and perception. Smile changes fundamental and formant frequencies (raises *f*0 and *f*1 but decreases *f*2), increases the lip spread, and can lower larynx position (Tartter, 1980; Ohala, 2005; Fagel, 2010). Crucially, listeners are able to identify smiled speech from acoustic cues alone (Tartter, 1980), and that even when the speech is whispered (Tartter and Braun, 1994).

The head's role in communication reveals an entanglement between physiology and language. Through extrinsic laryngeal muscles linking skull and vocal apparatus, head movements influence voice production – a coupling that both speakers and listeners exploit for communicative purposes.

2.2. Phenomena of the upper body

2.2.1. How does the upper body participate in language?

Hands and torso, two phenomena elaborated in this section, are another example of the whole body participating in communication. Worldwide, sign and tactile languages are used in deaf and deaf-blind communities. The way in which the upper limbs and torso participate in signing practices has been documented in the field of sign linguistics, particularly in sign phonology (e.g., Brentari, 1999). As sign languages are addressed in *Chapter 4*, this chapter focuses primarily on spoken languages and refers to sign languages where necessary.

In spoken languages, people often accompany their speech with various hand gestures in face-to-face conversation (McNeill, 1992; Kendon, 2004), but also when interacting over the telephone (Bavelas *et al.*, 2008). What usually distinguishes them from, say, goal-oriented manual behaviour, such as grasping a cup and bringing it to the mouth to drink, is that they are produced with communicative intention. You can, for instance, do a very similar movement of bringing your hand shaped as if holding a cup to your mouth to communicate to someone that you would like to drink something.

Experimental evidence indeed suggests that there is something special about communicating hands. These gestures are larger, produced with greater vertical amplitude, and have more complex movement patterns compared to hand movements produced during actions (Trujillo *et al.*, 2018). Moreover, gestures tend to be tightly coupled with the shared understanding between interlocutors. Among others, Gerwing and Bawelas (2004) showed that when only one participant has access to information about an object, their manual expressions towards the interlocutor are more elaborate and informative. Similarly, pragmatic context shapes gesture use and form. Pointing gestures with different purposes have been found to move differently, especially in terms of speed and duration of the extension (Raghavan, Raviv and Peeters, 2023).

Gestures may serve several functions, both for their producers and perceivers. It has been argued that they reflect externalized mental content (Goldin-Meadow *et al.*, 2001; Hostetter and Alibali, 2008, 2019) and that thanks to their imagistic nature, they offer a secondary mode of thinking (Kita, 2000), freeing up cognitive resources by, for instance, supporting the processing of visuo-spatial information. In this line of research, gesturing has been shown to propel learning of advanced concepts (Rueckert *et al.*, 2017), problem solving (Kirk and Lewis, 2017), spatial processing (Chu and Kita, 2011), second language (Gullberg, 2008), and many more.

Gestures also benefit interlocutors. For instance, manual expressions that depict something simultaneously mentioned in the speech speed up responses to questions, possibly contributing to the fast turn-taking dynamics (ter Bekke, Drijvers and Holler, 2024). Further, gestures seem to be an important tool for navigating situations where interlocutors struggle with what is being said. When facing interactive breakdown, people commonly engage their hands to support the regaining of shared understanding. They might gesture more extensively or with more precision in the movement (Holler and Wilkin, 2011; Rasenberg *et al.*, 2022).

While perhaps visually most dominant and noticeable, hands are not the only moving part of the upper body. The whole torso is inseparably linked to communication, both in spoken and signed languages. In sign languages, torso movement is partly necessitated by the movement of hands, but it can also mark prosodic or syntactic structure (Tyrone and Mauk, 2016), or be meaningfully employed in embodying a character (Napoli and Sutton-Spence, 2023). In spoken languages, the torso might not be used in such a systematic manner; however, it does interact with speech and discourse in non-trivial ways. For instance, Scheflen (1964) noted that individuals use their posture to mark or punctuate the discourse units in interaction. Shattuck-Hufnagel, Ren and Tauscher (2010) observed that speakers align torso translations with intonational phrases. The evidence is, however, rather anecdotal, and more research is needed to be able to characterize the role of the torso in structuring interaction and its coupling with speech or sign.

Over the last years, scientists have transitioned from assuming that gestures have no effect on speech and serve peripheral functions (Butterworth and Beattie, 1978; Levelt, Richardson and La Heij, 1985; Hadar, 1989) to revisit the way they are integrated with language and speech. By now, gestures figure in several models of speech production that assume a reciprocal link between the vocal and manual modality and the same origin in cognitive processes (Morrel-Samuels and Krauss, 1992; Krauss *et al.*, 1995; de Ruiter, 2000; Kita and Özyürek, 2003; for review, see Wagner, Malisz and Kopp, 2014). While the models differ in specifics, they share a common theoretical perspective: the physical act of gesturing is an output of the underlying cognitive processes.

In the next section, we would like to offer complementary, if not alternative, evidence that supports the idea that gestures are not mere abstract imaginary schemata but actual physical movements performed by real bodies with mass, certain forces, and thus also biomechanical constraints. In this view, the body shall be understood not as an output for internally computed linguistic behavior, but a participant in its creation (Hurley, 2008; Pouw *et al.*, 2014). To appreciate this entanglement, we illustrate this issue by means of recent research in gesture-speech coordination.

2.2.2. How is the upper body linked to physiology?

Dozens of researchers have brought to our attention that gesture and speech co-occur in time (Kendon, 1980; McNeill, Pedelty and Levy, 1990; Renwick, Shattuck-Hufnagel and Yasinnik, 2004; Rochet-Capellan *et al.*, 2008; Chu *et al.*, 2014; Zelic, Kim and Davis, 2015). Prominent moments in gesture, for instance the moments of maximal extension (i.e., apex), have been observed to temporally align with prominent moments in speech, such as pitch accent of a stressed syllable (Leonard and Cummins, 2011). The mainstream argument for such synchrony is that gesture and speech come from a single planning process. As both visual form and spoken utterance conceptually emerge during conceptual formulation and lexical retrieval, they naturally co-occur together (so-called semantic synchrony rule, McNeill, 1992).

However, such accounts do not fully appreciate the vastly different physiologies that constitute articulators of speech versus gesture (Grimme *et al.*, 2011). For instance, the jaw mandible is 21 times lighter in mass than the upper limb (Zhang, Peck and Hannam, 2002; Damavandi, Farahpour and Allard, 2009). As an alternative, we mention one of the recent approaches that argues that such coordination might arise from basic physiological properties on the one hand, and motor control constraints on the other. The model in question is so-called *Gesture-Speech Physics* (Pouw, Harrison and Dixon, 2020).

The Gesture-Speech Physics model builds on the assumption that the upper trunk has a tensegrity architecture (Caldeira, Davids and Araújo, 2021). This means that any forces produced by an effector (i.e., muscle that becomes active) are distributed through the whole medium, requiring specific muscles to stabilize the whole torso posture. Specifically, during extension-flexion movement of the forearm (as in the case of beat gestures), the stabilizing muscles that need to be activated are the same that assist in respiration. Impulsive forces affecting the respiratory system, in turn, increase lung pressure and leave an acoustic “imprint” on the voice. This coupling manifests as a positive correlation between upper limb acceleration and intensity, and secondarily (but not always) *f*0 and has been found in experiments where participants produced simple sustained phonation (Pouw, Harrison and Dixon, 2020; Pouw, Harrison, *et al.*, 2020), fluent speech (Pouw *et al.*, 2021), rhymes (Kadavá *et al.*, 2023) while producing repeated beat or pointing gestures, or in singing performances (Pearson and Pouw, 2022)

Crucially, the acoustic information about the movement is perceivable by others, to the extent for one to synchronize their own movement with the (invisible) producer (Pouw, Paxton, *et al.*, 2020). Similarly, deep neural networks can generate remarkably authentic gestures, based solely on the acoustic speech signal of a speaker (e.g., Yunus, Clavel and Pelachaud, 2021). This is in line with evidence that gestures’ influence is perceived as acoustic prominence (Krahmer and Swerts, 2007), and lexical stress can even shift from one syllable to another depending on where a gesture occurs – a phenomenon similar to the McGurk effect (Bosker and Peeters, 2021).

The Gesture-Speech Physics model is of great importance to our understanding of the physiological basis of multimodal language. It shows that upper limb movement does not participate only in abstract cognitive processes during speech planning, but shapes the utterance – for instance, by changing the way in which we reach prosodic targets. Further, it suggests that gesture-speech timing may not arise purely from a cognitive mechanism, but rather from a biophysical control system.

A similar view on the physiological basis for the cognitive interdependence between gesture and speech has previously been offered by Iverson and Thelen (1999). They pointed out that some language and motor functions share common mechanisms, and there is cross-activation between systems: language tasks recruit motor areas of the brain, and motor tasks engage areas traditionally associated with language. Observing infants in their first year of life, they argued that mouth and hand jointly co-evolve in an interactive fashion, which results in the related systems entraining each other. More recent evidence further corroborates this coupling between early vocal and hand movements (Esteve-Gibert and Prieto, 2014; Borjon *et al.*, 2024). Some even propose that this entanglement goes much deeper into evolutionary history and may be shared with other species (Pouw and Fuchs, 2022).

These findings collectively demonstrate that the relationship between gesture and speech extends beyond a purely cognitive achievement. The acoustic imprint left by upper body movement on vocal production, and the shared neural substrates between motor and language systems, suggest that multimodal language use is rooted in the basic architecture of the human body.

2.3. Phenomena of the lower body

2.3.1. How does the lower body participate in language?

Although often overlooked and less empirically investigated, the lower body, i.e., the hips, legs, and feet, can also contribute to multimodal language. They may be particularly relevant for signalling relational stance, attitude, and emotion.

Studies and popular accounts of nonverbal behaviour often suggest that crossed arms or legs convey closeness, disinterest, or authority. For example, in patient-dentist communication, dentists are advised to avoid crossing legs to foster openness toward patients (Ho *et al.*, 2024). Such claims, however, should be treated with caution, as popular science discussions frequently lack empirical grounding. Some of the interpretations may hold true, but they require rigorous testing and are likely to vary across cultural, hierarchical, gender-related, and situational contexts.

In the scientific literature, O'Reilly (2012) highlighted the underexplored role of the lower body in multimodal communication and proposed what he termed *bipedic gestures*. In a corpus-based video study, previously unfamiliar interlocutors were recorded in spontaneous conversation and later completed questionnaires about their attitudes toward each other. Contrary to common assumptions, crossed legs did not necessarily signal negative attitude; rather, when the leading foot pointed toward the interlocutor, it correlated with positive attitude. A follow-up study using mannequin dolls in various lower body postures confirmed that lower-body positioning alone, even without verbal cues, influenced perceived attitude and emotion.

Another example of the importance of the lower body considers the evolution of conversational flow in a group while walking and talking together through a town in guided visits. "Walking designs a trajectory in which each step projects more steps to come in a way that is both oriented to and anticipated by co-walkers coordinating their walking together… Nonetheless, these 'with' can build different mobile configurations: walking in parallel, in a row, two-by-two, aside or before/behind, etc. These mobile configurations enable different forms of coordinated talk" (Mondada, 2014, p. 397). Using conversation analyses, Mondada described how the interaction of the guide and the visitors unfolds depending on the moving bodies.

Beyond lower body positioning, interacting individuals also spontaneously coordinate their postural sway (i.e., the movement of the body's centre of mass while standing). Research demonstrates that pairs of people engaged in conversation exhibit shared postural configurations that do not arise when co-present individuals are talking but not to each other (Shockley, Santana and Fowler, 2003). Such postural coordination is likely not incidental but reflects functional organization emerging from the interaction, with speech production patterns mediating the coupling between bodies (Shockley *et al.*, 2007)

2.3.2. How is the lower body linked to physiology

The amount of movements available for communicative purposes depends on how much the lower body is involved in maintaining posture. When standing, we balance the entire body on a minimal base of support, i.e., even small movements or perturbations influence stability. In contrast, when sitting, the legs and feet are free to move, as long as balance is maintained.

When the lower limbs face an imbalance like when standing on a wobble board, they may not be able to signal communicatively but first serve to regain equilibrium. Other body parts can also compensate for these disturbances (Taubert, Ziegler and Lehmann, 2024, see especially the supplemental videos). When people learn how to stand on a stabilometer (a kind of wobble board), they initially compensate with the lower trunk and later on learn to anticipate instability and compensate with movements of the upper body, especially the arms. This may suggest that the lower body is in the first place responsible for holding the whole body in place and if this is guaranteed, it may also engage in communicative action.

However, what posture affords extends beyond the lower body itself. For instance, van der Fits *et al.* (1998) reported that patterns in pointing movements vary systematically in standing, sitting, and lying positions. In a recent study with electromyography and ground reaction forces, Pouw *et al.* (2025) showed that voice parameters such as amplitude interacts not only with activity in the muscles of the moving arm as predicted by the Gesture-Speech Physics model, but also with muscles related to postural stabilization. Similarly, in a study by Lagier and colleagues (2010), increased vocal effort was shown to be accompanied by stronger trunk tilt, reflecting larger postural sway. Further, Momsen and Coulson (2025) showed that the vertical dimension of co-speech gestures was more predictive of prosodic features than horizontal movements, suggesting that kinetic properties incorporating gravitational forces are fundamental to the gesture-speech link.

The biomechanical connection between upper body motions, breathing and vocalizations described in *Section 2.3* is also true for the lower body. In a study where participants had to cycle with their legs on a mini-bike, Serré *et al.* (2022) found that the moment of maximal acceleration of the legs co-occurred with intensity peaks of the acoustic speech signal. That means that not only the upper body but also lower body motion can affect the respiratory system and in turn speech acoustics.

Body position is particularly important for language development in infancy. In Ester Thelen's framework, new motor skills emerge from the interaction of the body, environment, and task demands (Thelen, 1979). Motor stereotypies, e.g., repetitive kicking or arm waving, often represent transitional patterns that help infants explore and stabilize new forms of movement. The infant's body posture, whether lying down, sitting, crawling, or walking, crucially shapes the degrees of freedom available to other body parts, influencing what new motor and perceptual skills can emerge next. Eleanor Gibson's work on affordances extends this perspective: each form of locomotion presents different possibilities for action (e.g., Gibson and Schumuckler, 1989). These posture-specific affordances may similarly constrain and enable multimodal communication and interaction with caregivers. An infant who is lying down, sitting, or walking encounters different opportunities for gestural expression, eye contact and other forms of interpersonal coordination.

All in all, the lower body is integral to multimodal communication. While it may not "convey" concepts in a way gestures do, it anchors postural stability that enables upper body movement, shapes interpersonal coordination through positioning and sway, and fundamentally constrains the communicative affordances available across developmental stages and interactive contexts.

3. Summary

The evidence presented in the previous section demonstrates that communication really does engage the whole body, from head to toe. As documented in *Section 2.1*, our laryngeal system is attached to the muscular network that is also active when we move our head. This, in turn, makes head movements systematically affect vocal production. In *Section 2.2* we established how our arms' movement influences the lung volume through the tensegrity architecture of our upper body, and thus the subglottal pressure that constitutes voiced sound. In *Section 2.3,* we showed that the lower body provides the postural foundation for maintaining bipedal balance while simultaneously serving as a communicative resource in interaction. These interdependencies remind us that language is fundamentally physical, shaped and limited by our bodily anatomy. Over evolutionary history, these anatomical constraints have likely influenced how we exploit movement for communication, exemplified by the relationship between bodily movement and prominence.

While these may be the generic biological underpinnings that shape human communicative behavior, we need to acknowledge that 'every culture molds the raw actions according to its own traditions' (Scheflen, 1964, p. 317). In spoken languages, co-speech gestures follow language-specific patterns, differentiating manner and path in motion events according to the grammar of the respective language (Kita and Özyürek, 2003). Yet, the body is not disconnected from the linguistic structure. In a study by Bosker, Hoetjes, Hustin, Pouw and van Maastricht (2025), Dutch learners of Spanish produced beat gestures along with a target word. Even when participants placed acoustic stress correctly, their hand movements emphasized the syllable too early or late, depending on their native language's stress patterns (e.g., Dutch emphasis on the second syllable /proːˈfɛsɔr/ versus Spanish /pɾofeˈsoɾ/ on the third one). Such findings underscore that grammar and body are mutually constitutive: the body serves not as a passive vessel for linguistic output but as an active scaffold that both shapes and is shaped by grammatical structure.

Finally, despite the way we structured this chapter, we must keep in mind that the body is not partitioned into discrete units. The individual parts may have unique structures and unique functions, yet they operate as an interconnected whole. Movement in one region influences and is influenced by movement elsewhere. This interconnectedness points toward the need for a more holistic framework.

4. Beyond individual parts: towards holistic model of multimodal communication

In 1859, the French acrobat Charles Blondin became the first person to cross Niagara Falls on a tightrope. To achieve this, his entire body worked as a coordinated system: his arms held a long balancing pole, his core muscles stabilized his trunk, his head remained level to maintain visual and vestibular equilibrium, and even his breathing had to be controlled to avoid disrupting his center of gravity. If Blondin wanted to speak to the crowd during his crossing – calling out reassurances or acknowledging their cheers – his vocal production had to be integrated with this complex postural dance; even a misplaced breath or a careless resonance could have tipped the equilibrium he so carefully maintained. His communication would not emerge from an isolated speech system, but from his entire body working as a unified, dynamic whole.

This integration of movement, postural control, and communication reflects a principle that traditional linguistic theories have largely overlooked: human language is an integrated system in which articulators from head to toe work in synergy, influencing one another. What our field needs to advance is a holistic model incorporating the above-mentioned phenomena as different dynamic systems into one body. While some models of multimodal communication exist and were mentioned throughout the chapter (e.g., Gesture-Speech Physics, Pouw, Harrison and Dixon, 2020; or SAMP by Momsen, 2024), we envision extensions that reflect the evidence that these interdependencies exist and affect the way we behave. Fortunately, developments in computer vision make methods like motion capture readily available to the broader scientific community (e.g., Pagnon, Domalain and Reveret, 2022), allowing us to ask questions that involve precise quantifications of movement.

Central to this holistic model should be the basic physiological processes such as breathing and postural control. As we have also seen in *Section 2.2*, breathing is supported by the same muscles that work synergistically or antagonistically during gesturing, creating a direct mechanistic link between movement and speech. In *Section 2.3*, we have shown that posture provides the basis for what body parts can move and to what extent, while in *Section 2.1* we have shown that the head, via the vestibular system, may connect multimodal prosody with postural proprioception.

Such model of whole-body involvement in communication must navigate a critical tension: it needs to be broad enough to account for the integration of multiple bodily systems, while remaining constrained enough to generate testable predictions and offer specific mechanistic insights rather than unfalsifiable generalizations. Burning questions that the model could address include: how do we coordinate the whole orchestra of body motions beyond gesture and speech? Are these synergies innate, or learned? How does altered postural stability or lung fitness affect multimodal communication in selected populations? How do individual body properties affect multimodal communication? How is multimodal communication in the absence of gravity, i.e., in space?

The model's implications extend to technological development. Current voice assistants ignore how gestures alter vocal patterns through gesture-speech physics, while speech synthesis systems that coordinate body movements with vocal parameters could produce more natural and convincing artificial speech.

The broader implication is that communication science must become genuinely interdisciplinary, incorporating insights from biomechanics, motor control, and physiological systems alongside traditional linguistic and cognitive approaches. This integration promises a better understanding of how communication works, as well as practical advances in therapy, technology design, and human performance that emerge from recognizing communication as a physical phenomenon.

5. References

Bavelas, J., Gerwing, J., Sutton, C. and Prevost, D. (2008) 'Gesturing on the telephone: Independent effects of dialogue and visibility', *Journal of Memory and Language*, 58(2), pp. 495–520.

ter Bekke, M., Drijvers, L. and Holler, J. (2024) 'Gestures speed up responses to questions', *Language, Cognition and Neuroscience*, 39(4), pp. 423–430.

Blasi, D.E., Moran, S., Moisik, S.R., Widmer, P., Dediu, D. and Bickel, B. (2019) 'Human sound systems are shaped by post-Neolithic changes in bite configuration', *Science*, 363(6432), p. eaav3218.

Borjon, J.I., Abney, D.H., Yu, C. and Smith, L.B. (2024) 'Infant vocal productions coincide with body movements', *Developmental Science*, 27(4), p. e13491.

Borràs-Comes, J., Kaland, C., Prieto, P. and Swerts, M. (2014) ‘Audiovisual Correlates of Interrogativity: A Comparative Analysis of Catalan and Dutch’, *Journal of Nonverbal Behavior*, 38(1), pp. 53–66.

Bosker, H.R., Hoetjes, M., Hustin, D., Pouw, W. and van Maastricht, L. (2025) ‘Foreign Language Learners Show a Kinematic Accent in Their Co-speech Hand Movements’. OSF Preprints.

Bosker, H.R. and Peeters, D. (2021) ‘Beat gestures influence which speech sounds you hear’, *Proceedings of the Royal Society B: Biological Sciences*, 288(1943), p. 20202419.

Brentari, D. (1999) *A Prosodic Model of Sign Language Phonology*. Cambridge, MA, USA: MIT Press (Language, Speech, and Communication).

Bross, F. (2020) ‘Why do we shake our heads?: On the origin of the headshake’, *Gesture*, 19(2–3), pp. 269–298.

Brunner, J., Fuchs, S. and Perrier, P. (2009) ‘On the relationship between palate shape and articulatory behavior’, *The Journal of the Acoustical Society of America*, 125(6), pp. 3936–3949.

Burchardt, L.S., Fuchs, S. and Paschen, L. (2025) ‘Shared Duration of Speech Intervals Across 49 Human Languages’. OSF Preprints.

Butterworth, B. and Beattie, G. (1978) ‘Gesture and Silence as Indicators of Planning in Speech’, in R.N. Campbell and P.T. Smith (eds) *Recent Advances in the Psychology of Language*. Boston, MA: Springer US, pp. 347–360.

Caldeira, P., Davids, K. and Araújo, D. (2021) ‘Neurobiological tensegrity: The basis for understanding inter-individual variations in task performance?’, *Human Movement Science*, 79, p. 102862.

Carignan, C., Esteve-Gibert, N., Lœvenbruck, H., Dohen, M. and D’Imperio, M. (2024) ‘Co-speech head nods are used to enhance prosodic prominence at different levels of narrow focus in French’, *The Journal of the Acoustical Society of America*, 156(3), pp. 1720–1733.

Chen, T.H. and Massaro, D.W. (2008) ‘Seeing pitch: Visual information for lexical tones of Mandarin-Chinese’, *The Journal of the Acoustical Society of America*, 123(4), pp. 2356–2366.

Chu, M. and Kita, S. (2011) ‘The nature of gestures’ beneficial role in spatial problem solving’, *Journal of Experimental Psychology: General*, 140(1), pp. 102–116.

Chu, M., Meyer, A., Foulkes, L. and Kita, S. (2014) ‘Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy’, *Journal of Experimental Psychology. General*, 143(2), pp. 694–709.

Cristia, A. (2013) ‘Input to Language: The Phonetics and Perception of Infant-Directed Speech’, *Language and Linguistics Compass*, 7(3), pp. 157–170.

Ćwiek, A. and Fuchs, S. (2025) ‘The scope and limits of iconic prosody: Head angle predicts f0 changes while object size effects are absent’, *Open Mind*. 9 1959–1981.

Damavandi, M., Farahpour, N. and Allard, P. (2009) ‘Determination of body segment masses and centers of mass using a force plate method in individuals of different morphology’, *Medical Engineering & Physics*, 31(9), pp. 1187–1194.

Darwin, C. (1872) *The expression of the emotions in man and animals*. London, John Murray.

Dediu, D., Janssen, R. and Moisik, S.R. (2019) 'Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels', *Nature Human Behaviour*, 3(10), pp. 1107–1115.

Dediu, D., Moisik, S.R., Baetsen, W.A., Bosman, A.M. and Waters-Rist, A.L. (2021) 'The vocal tract as a time machine: inferences about past speech and language from the anatomy of the speech organs', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1824), p. 20200192.

Erickson, D., Baer, T. and Harris, K.S. (1982) *The role of the strap muscles in pitch lowering*. 70, pp. 275–284.

Esteve-Gibert, N. and Prieto, P. (2014) 'Infants temporally coordinate gesture-speech combinations before they produce their first words', *Speech Communication*, 57, pp. 301–316.

Fagel, S. (2010) 'Effects of Smiling on Articulation: Lips, Larynx and Acoustics', in A. Esposito, N. Campbell, C. Vogel, A. Hussain, and A. Nijholt (eds) *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers*. Berlin, Heidelberg: Springer, pp. 294–303.

van der Fits, I.B., Klip, A.W., van Eykern, L.A. and Hadders-Algra, M. (1998) 'Postural adjustments accompanying fast pointing movements in standing, sitting and lying adults', *Experimental Brain Research*, 120(2), pp. 202–216.

Fuchs, S. and Rochet-Capellan, A. (2021) 'The Respiratory Foundations of Spoken Language', *Annual Review of Linguistics*, 7(1), pp. 13–30.

Gerwing, J. and Bavelas, J. (2004) 'Linguistic influences on gesture's form', *Gesture*, 4(2), pp. 157–195.

Gibson, E.J. and Schumuckler, M.A. (1989) 'Going Somewhere: An Ecological and Experimental Approach to Development of Mobility', *Ecological Psychology*, 1(1), pp. 3–25.

Goldin-Meadow, S., Nusbaum, H., Kelly, S.D. and Wagner, S. (2001) 'Explaining Math: Gesturing Lightens the Load', *Psychological Science*, 12(6), pp. 516–522.

Grimme, B., Fuchs, S., Perrier, P. and Schöner, G. (2011) 'Limb versus Speech Motor Control: A Conceptual Review', *Motor Control*, 15(1), pp. 5–33.

Gullberg, M. (2008) 'Gestures and Second Language Acquisition', in *Handbook of Cognitive Linguistics and Second Language Acquisition*. Routledge.

Hadar, U. (1989) 'Two Types of Gesture and Their Role in Speech Production', *Journal of Language and Social Psychology*, 8(3–4), pp. 221–228.

Ho, J.C.Y., Chai, H.H., Lo, E.C.M., Huang, M.Z. and Chu, C.H. (2024) 'Strategies for Effective Dentist-Patient Communication: A Literature Review', *Patient Preference and Adherence*, 18, pp. 1385–1394.

Holler, J. and Wilkin, K. (2011) 'An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses', *Journal of Pragmatics*, 43(14), pp. 3522–3536.

Honda, K. (2008) 'Physiological Processes of Speech Production', in J. Benesty, M.M. Sondhi, and Y.A. Huang (eds) *Springer Handbook of Speech Processing*. Berlin, Heidelberg: Springer, pp. 7–26.

Hostetter, A.B. and Alibali, M.W. (2008) 'Visible embodiment: Gestures as simulated action', *Psychonomic Bulletin & Review*, 15(3), pp. 495–514.

Hostetter, A.B. and Alibali, M.W. (2019) 'Gesture as simulated action: Revisiting the framework', *Psychonomic Bulletin & Review*, 26(3), pp. 721–752.

Hurley, S. (2008) 'The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading', *Behavioral and Brain Sciences*, 31(1), pp. 1–22.

Hwan Hong, K., Ye, M., Mo Kim, Y., Kevorkian, K.F. and Berke, G.S. (1997) 'The role of strap muscles in phonation—In vivo caninelaryngeal model', *Journal of Voice*, 11(1), pp. 23–32.

Iverson, J.M. and Thelen, E. (1999) 'Hand, mouth and brain. The dynamic emergence of speech and gesture', *Journal of Consciousness Studies*, 6(11–12), pp. 19–40.

Kadavá, Š., Ćwiek, A., Stoltmann, K., Fuchs, S. and Pouw, W. (2023) 'Is gesture-speech physics at work in rhythmic pointing? Evidence from Polish counting-out rhymes'. Open Science Framework.

Kendon, A. (1980) 'Gesticulation and speech: Two aspects of the process of utterance', in M. Key (ed.) *The relationship of verbal and nonverbal communication*. The Hague, the Netherlands: De Gruyter, pp. 207–227.

Kendon, A. (2002) 'Some uses of the head shake', *Gesture*, 2(2), pp. 147–182.

Kendon, A. (2004) *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kettner, V.A. and Carpendale, J.I.M. (2013) 'Developing gestures for no and yes: Head shaking and nodding in infancy', *Gesture*, 13(2), pp. 193–209.

Kirk, E. and Lewis, C. (2017) 'Gesture Facilitates Children's Creative Thinking', *Psychological Science*, 28(2), pp. 225–232.

Kita, S. (2000) 'How representational gestures help speaking', in D. McNeill (ed.) *Language and Gesture*. Cambridge University Press.

Kita, S. and Ide, S. (2007) 'Nodding, *aizuchi*, and final particles in Japanese conversation: How conversation reflects the ideology of communication and social relationships', *Journal of Pragmatics*, 39(7), pp. 1242–1254.

Kita, S. and Özyürek, 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', *Journal of Memory and Language*, 48(1), pp. 16–32.

Krahmer, E. and Swerts, M. (2004) 'More About Brows: A Cross-Linguistic Study via Analysis-by-Synthesis', in Z. Ruttkay and C. Pelachaud (eds) *From Brows to Trust*. Dordrecht: Springer Netherlands (Human-Computer Interaction Series), pp. 191–216.

Krahmer, E. and Swerts, M. (2007) 'The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception', *Journal of Memory and Language*, 57(3), pp. 396–414.

Krauss, R.M., Dushay, R.A., Chen, Y. and Rauscher, F. (1995) 'The Communicative Value of Conversational Hand Gesture', *Journal of Experimental Social Psychology*, 31(6), pp. 533–552.

Lagier, A., Vaugoyeau, M., Ghio, A., Legou, T., Giovanni, A. and Assaiante, C. (2010) 'Coordination between posture and phonation in vocal effort behavior', *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP)*, 62(4), pp. 195–202.

Leonard, T. and Cummins, F. (2011) ‘The temporal relation between beat gestures and speech’, *Language and Cognitive Processes*, 26(10), pp. 1457–1471.

Levelt, W.J.M., Richardson, G. and La Heij, W. (1985) ‘Pointing and voicing in deictic expressions’, *Journal of Memory and Language*, 24(2), pp. 133–164.

Liu, Y., Shamei, A., Chow, U.Y., Soo, R., Pineda Mora, G., de Boer, G. and Gick, B. (2020) ‘F0-related head movement in blind versus sighted speakers’, *The Journal of the Acoustical Society of America*, 148(2), pp. EL190–EL194.

Maynard, S.K. (1986) ‘On back-channel behavior in Japanese and English casual conversation’, 24(6), pp. 1079–1108.

McGurk, H. and McDonald, J. (1976) ‘Hearing lips and seeing voices’, *Nature*, 264(5588), pp. 746–748.

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought*. Chicago, IL, US: University of Chicago Press (Hand and mind: What gestures reveal about thought), pp. xi, 416.

McNeill, D., Pedelty, L.L. and Levy, E.T. (1990) ‘Chapter 7 Speech and Gesture’, in G.R. Hammond (ed.) *Advances in Psychology*. North-Holland (Cerebral Control of Speech and Limb Movements), pp. 203–256.

Miller, N.A., Gregory, J.S., Semple, S.I.K., Aspden, R.M., Stollery, P.J. and Gilbert, F.J. (2012) ‘The Effects of Humming and Pitch on Craniofacial and Craniocervical Morphology Measured Using MRI’, *Journal of Voice*, 26(1), pp. 90–101.

Momsen, J. (2024) *The Conversational Dance: Assessing the Sensorimotor Account of Multimodal Prosody*. UC San Diego.

Momsen, J.P. and Coulson, S. (2025) ‘Decoding Prosodic Information from Motion Capture Data: The Gravity of Co-Speech Gestures’, *Open Mind*, 9, pp. 652–664.

Mondada, L. (2014) ‘Bodies in action: Multimodal analysis of walking and talking’, *Language and Dialogue*, 4(3), pp. 357–403.

Morales, M., Mundy, P. and Rojas, J. (1998) ‘Following the direction of gaze and language development in 6-month-olds’, *Infant Behavior and Development*, 21(2), pp. 373–377.

Morrel-Samuels, P. and Krauss, R.M. (1992) ‘Word familiarity predicts temporal asynchrony of hand gestures and speech’, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), pp. 615–622.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T. and Vatikiotis-Bateson, E. (2004) ‘Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception’, *Psychological Science*, 15(2), pp. 133–137.

Napoli, D.J. and Sutton-Spence, R. (2023) ‘The nature of torso articulation in human expression: Signing and dancing’, *Journal of Cultural Cognitive Science*, 7(3), pp. 269–288.

Nota, N., Trujillo, J.P. and Holler, J. (2021) ‘Facial Signals and Social Actions in Multimodal Face-to-Face Interaction’, *Brain Sciences*, 11(8), p. 1017.

Ohala, J.J. (1978) ‘Production of Tone’, in V.A. Fromkin (ed.) *Tone*. Academic Press, pp. 5–39.

Ohala, J.J. (2005) ‘The acoustic origin of the smile’, *The Journal of the Acoustical Society of America*, 68(S1), p. S33.

Ohala, J.J. and Solé, M.-J. (2008) ‘Turbulence & Phonology’, *UC Berkeley PhonLab Annual Report*, 4(4).

O’Reilly, P. (2012) ‘Bipedics: Towards a new category of kinesics. - An empirical investigation of the expression of attitude, and emotion, through simple leg and foot gesture.’

Pagel, L., Sóskuthy, M., Roessig, S. and Mücke, D. (2023) ‘A kinematic analysis of visual prosody: Head movements in habitual and loud speech’.

Pagnon, D., Domalain, M. and Reveret, L. (2022) ‘Pose2Sim: An open-source Python package for multiview markerless kinematics’, *Journal of Open Source Software*, 7(77), p. 4362.

Paris, T., Kim, J. and Davis, C. (2013) ‘Visual speech form influences the speed of auditory speech processing’, *Brain and Language*, 126(3), pp. 350–356.

Pearson, L. and Pouw, W. (2022) ‘Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement’, *Annals of the New York Academy of Sciences*, 1515(1), pp. 219–236.

Pouw, W. and Fuchs, S. (2022) ‘Origins of vocal-entangled gesture’, *Neuroscience & Biobehavioral Reviews*, 141, p. 104836.

Pouw, W., Harrison, S.J. and Dixon, J.A. (2020) ‘Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony’, *Journal of Experimental Psychology: General*, 149(2), pp. 391–404.

Pouw, W., Harrison, S.J., Esteve-Gibert, N. and Dixon, J.A. (2020) ‘Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures’, *The Journal of the Acoustical Society of America*, 148(3), pp. 1231–1247.

Pouw, W., de Jonge-Hoekstra, L., Harrison, S.J., Paxton, A. and Dixon, J.A. (2021) ‘Gesture–speech physics in fluent speech and rhythmic upper limb movements’, *Annals of the New York Academy of Sciences*, 1491(1), pp. 89–105.

Pouw, W., Paxton, A., Harrison, S.J. and Dixon, J.A. (2020) ‘Acoustic information about upper limb movement in voicing’, *Proceedings of the National Academy of Sciences*, 117(21), pp. 11364–11367.

Pouw, W.T.J.L., de Nooijer, J.A., van Gog, T., Zwaan, R.A. and Paas, F. (2014) ‘Toward a more embedded/extended perspective on the cognitive function of gestures’, *Frontiers in Psychology*, 5.

Raghavan, R., Raviv, L. and Peeters, D. (2023) ‘What’s your point? Insights from virtual reality on the relation between intention and action in the production of pointing gestures’, *Cognition*, 240, p. 105581.

Rasenberg, M., Pouw, W., Özyürek, A. and Dingemanse, M. (2022) ‘The multimodal nature of communicative efficiency in social interaction’, *Scientific Reports*, 12(1), p. 19111.

Renwick, M., Shattuck-Hufnagel, S. and Yasinnik, Y. (2004) ‘The timing of speech-accompanying gestures with respect to prosody’, *The Journal of the Acoustical Society of America*, 115(5_Supplement), p. 2397.

Rochet-Capellan, A., Laboissière, R., Galván, A. and Schwartz, J.-L. (2008) 'The Speech Focus Position Effect on Jaw–Finger Coordination in a Pointing Task', *Journal of Speech, Language, and Hearing Research*, 51(6), pp. 1507–1521.

Rueckert, L., Church, R.B., Avila, A. and Trejo, T. (2017) 'Gesture enhances learning of a complex statistical concept', *Cognitive Research: Principles and Implications*, 2(1), p. 2.

de Ruiter, J.P. (2000) 'The production of gesture and speech', in D. McNeill (ed.) *Language and Gesture*. Cambridge University Press, pp. 284–311.

Sardhaei, N.M., Żygis, M. and Sharifzadeh, H. (2024) 'Facial expressions in different communication settings: A case of whispering and speaking with a face mask in Farsi', *Language and Cognition*, 16(4), pp. 1639–1673.

Scheflen, A.E. (1964) 'The Significance of Posture in Communication Systems†', *Psychiatry*, 27(4), pp. 316–331.

Serré, H., Dohen, M., Fuchs, S., Gerber, S. and Rochet-Capellan, A. (2022) 'Leg movements affect speech intensity', *Journal of Neurophysiology*, 128(5), pp. 1106–1116.

Shattuck-Hufnagel, S., Ren, P.L. and Tauscher, E. (2010) 'Are torso movements during speech timed with intonational phrases?', in. *Proc. SpeechProsody 2010*, p. paper 974.

Shepard, K.G., Spence, M.J. and Sasson, N.J. (2012) 'Distinct Facial Characteristics Differentiate Communicative Intent of Infant-Directed Speech', *Infant and Child Development*, 21(6), pp. 555–578.

Shockley, K., Baker, A.A., Richardson, M.J. and Fowler, C.A. (2007) 'Articulatory constraints on interpersonal postural coordination', *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), pp. 201–208.

Shockley, K., Santana, M.-V. and Fowler, C.A. (2003) 'Mutual interpersonal postural constraints are involved in cooperative conversation', *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), pp. 326–332.

Stone, M. (1991) 'Toward a model of three-dimensional tongue movement', *Journal of Phonetics*, 19(3), pp. 309–320.

Tartter, V.C. (1980) 'Happy talk: Perceptual and acoustic effects of smiling on speech', *Perception & Psychophysics*, 27(1), pp. 24–27.

Tartter, V.C. and Braun, D. (1994) 'Hearing smiles and frowns in normal and whisper registers', *The Journal of the Acoustical Society of America*, 96(4), pp. 2101–2107.

Taubert, M., Ziegler, G. and Lehmann, N. (2024) 'Higher surface folding of the human premotor cortex is associated with better long-term learning capability', *Communications Biology*, 7(1), p. 635.

Thelen, E. (1979) 'Rhythmical stereotypies in normal human infants', *Animal Behaviour*, 27, pp. 699–715.

Titze, I.R. (1994) *Principles of Voice Production*. Prentice Hall.

Trujillo, J.P., Simanova, I., Bekkering, H. and Özyürek, A. (2018) 'Communicative intent modulates production and comprehension of actions and gestures: A Kinect study', *Cognition*, 180, pp. 38–51.

Tyrone, M.E. and Mauk, C.E. (2016) ‘The Phonetics of Head and Body Movement in the Realization of American Sign Language Signs’, *Phonetica*, 73(2), pp. 120–140.

Wagner, P., Malisz, Z. and Kopp, S. (2014) ‘Gesture and speech in interaction: An overview’, *Speech Communication*, 57, pp. 209–232.

Walker, A. and Shipman, P. (1997) *The Wisdom of the Bones: In Search of Human Origins*. Knopf Doubleday Publishing Group.

Weirich, M. and Fuchs, S. (2013) ‘Palatal Morphology Can Influence Speaker-Specific Realizations of Phonemic Contrasts’, *Journal of Speech, Language, and Hearing Research*, 56(6), pp. 1894–1908.

Yunus, F., Clavel, C. and Pelachaud, C. (2021) ‘Sequence-to-Sequence Predictive Model: From Prosody to Communicative Gestures’, in V.G. Duffy (ed.) *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and Behavior*. Cham: Springer International Publishing, pp. 355–374.

Zelic, G., Kim, J. and Davis, C. (2015) ‘Articulatory constraints on spontaneous entrainment between speech and manual gesture’, *Human Movement Science*, 42, pp. 232–245.

Zhang, F., Peck, C.C. and Hannam, A.G. (2002) ‘Mass properties of the human mandible’, *Journal of Biomechanics*, 35(7), pp. 975–978.

Zhang, Z. (2016) ‘Mechanics of human voice production and control’, *The Journal of the Acoustical Society of America*, 140(4), pp. 2614–2635.

Żygis, M. and Fuchs, S. (2023) ‘Communicative constraints affect oro-facial gestures and acoustics: Whispered vs normal speech’, *The Journal of the Acoustical Society of America*, 153(1), pp. 613–626.