Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures

Wim Pouw, Steven J. Harrison, Núria Esteve-Gibert, and James A. Dixon

Citation: The Journal of the Acoustical Society of America **148**, 1231 (2020); doi: 10.1121/10.0001730
View online: https://doi.org/10.1121/10.0001730
View Table of Contents: https://asa.scitation.org/toc/jas/148/3
Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN

A two-stage deep learning algorithm for talker-independent speaker separation in reverberant conditions
The Journal of the Acoustical Society of America **148**, 1157 (2020); https://doi.org/10.1121/10.0001779

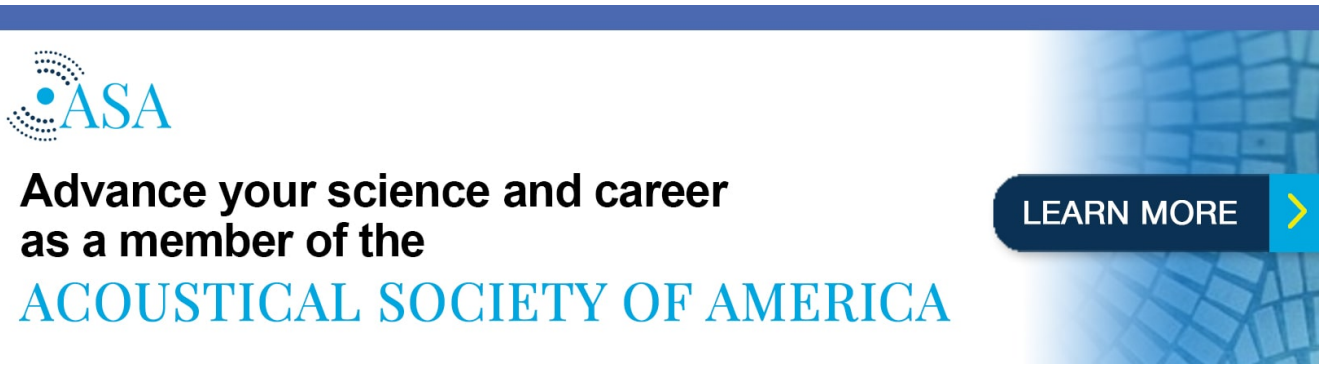
Acoustic time-dependent energy from vibrating surfaces via a generalized radiation impulse response approach
The Journal of the Acoustical Society of America **148**, 1296 (2020); https://doi.org/10.1121/10.0001917

Continuously steerable differential beamformers with null constraints for circular microphone arrays
The Journal of the Acoustical Society of America **148**, 1248 (2020); https://doi.org/10.1121/10.0001770

Near real-time marine mammal monitoring from gliders: Practical challenges, system development, and management implications
The Journal of the Acoustical Society of America **148**, 1215 (2020); https://doi.org/10.1121/10.0001811

A pilot study on the influence of mouth configuration and torso on singing voice directivity
The Journal of the Acoustical Society of America **148**, 1169 (2020); https://doi.org/10.1121/10.0001736

A multimedia speech corpus for audio visual research in virtual reality (L)
The Journal of the Acoustical Society of America **148**, 492 (2020); https://doi.org/10.1121/10.0001670



JASA ARTICLE

Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures

Wim Pouw,[1,a)] Steven J. Harrison,[1,b)] Núria Esteve-Gibert,[2] and James A. Dixon[1]
[1]*Center for the Ecological Study of Perception and Action at the University of Connecticut, 406 Babbidge Road, Storrs, Connecticut 06269, USA*
[2]*Psychology and Education Sciences at the Universitat Oberta de Catalunya, Rambla del Poblenou, 158, 08018, Barcelona, Spain*

ABSTRACT:
Expressive moments in communicative hand gestures often align with emphatic stress in speech. It has recently been found that acoustic markers of emphatic stress arise naturally during steady-state phonation when upper-limb movements impart physical impulses on the body, most likely affecting acoustics via respiratory activity. In this confirmatory study, participants ($N = 29$) repeatedly uttered consonant-vowel (/pa/) mono-syllables while moving in particular phase relations with speech, or not moving the upper limbs. This study shows that respiration-related activity is affected by (especially high-impulse) gesturing when vocalizations occur near peaks in physical impulse. This study further shows that gesture-induced moments of bodily impulses increase the amplitude envelope of speech, while not similarly affecting the Fundamental Frequency ($F0$). Finally, tight relations between respiration-related activity and vocalization were observed, even in the absence of movement, but even more so when upper-limb movement is present. The current findings expand a developing line of research showing that speech is modulated by functional biomechanical linkages between hand gestures and the respiratory system. This identification of gesture-speech biomechanics promises to provide an alternative phylogenetic, ontogenetic, and mechanistic explanatory route of why communicative upper limb movements co-occur with speech in humans.
© *2020 Acoustical Society of America*. https://doi.org/10.1121/10.0001730
(Received 2 December 2019; revised 21 July 2020; accepted 24 July 2020; published online 8 September 2020)
[Editor: James F. Lynch] Pages: 1231–1247

I. INTRODUCTION

Prosody is a concept in speech science that targets the intonational and rhythmic features of utterances, which can serve syntactic, semantic, pragmatic, and affective purposes of communicative exchange. Utterances are often multimodal, such that prosodic aspects of speech unfold through dynamic interplay with prosodic aspects of co-speech gestures (Hübscher and Prieto, 2019; Rusiewicz and Esteve-Gibert, 2018; Shattuck-Hufnagel and Prieto, 2019; Wagner *et al.*, 2014). Prosodically relevant aspects of hand gestures have been roughly defined by salient kinematic moments, such as an expressive stroke or a sudden halt (Loehr, 2012; McClave, 1998; McNeill, 2005), and these moments have been quantified as a gesture's peak in velocity or peak in deceleration (Danner *et al.*, 2018; Leonard and Cummins, 2011; Pouw and Dixon, 2019a,b; Pouw *et al.*, 2020c; Rochet-Capellan *et al.*, 2008). Prosodically relevant acoustic aspects of speech that have been extensively studied in relation to gesture include peaks in the fundamental frequency ($F0$) of contrastively focused speech (e.g., Danner *et al.*, 2018; Esteve-Gibert and Prieto, 2013; Krahmer and Swerts, 2007; Krivokapic *et al.*, 2016; Loehr, 2012; McClave, 1998; Pouw and Dixon, 2019a,b; Pouw *et al.*, 2019a), as well as first and second formants, and intensity and durational modulations (Krahmer and Swerts, 2007; Krivokapic *et al.*, 2016; Pouw *et al.*, 2019a; Krivokapić, 2014). These acoustic markers are together co-constitutive of, for example, emphatic stress on the phrase level (referred to as "pitch accents"; e.g., she sees YOU vs she SEES you). This growing literature is converging on the idea that gesture's kinematic patterns and speech prosodic patterns are tightly but flexibly coupled (Chu and Hagoort, 2014; Parrell *et al.*, 2014; Pouw and Dixon, 2019a; Rusiewicz *et al.*, 2013), raising more fundamental questions about the ontogenetic and cognitive origins of this tightly coupled system.

There is a growing body of evidence that gestures coupling with speech can at times form a single coordinative structure (Chang and Hammond, 1987; Danner *et al.*, 2018; Kelso *et al.*, 1983; Krahmer and Swerts, 2007; Krivokapic *et al.*, 2016; McNeill, 1992; Parrell *et al.*, 2014; Pouw and Dixon, 2019a; Rochet-Capellan and Fuchs, 2014; Rusiewicz *et al.*, 2013; Treffner and Peter, 2002; Zelic *et al.*, 2015). For example, studies on monosyllabic vocalization while manually tapping show that when speech intervals are increasingly shortened there is an attraction towards synchronizing vocalization and taps *in-phase*, rather than an alternating *anti-phase* fashion (Chang and Hammond, 1987; Kelso *et al.*, 1983; Treffner and Peter, 2002). This suggests

[a)]Also at: Donders Institute for Brain, Cognition and Behaviour at the Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands. Electronic mail: w.pouw@psych.ru.nl
[b)]Also at: Department of Kinesiology, University of Connecticut, Storrs, CT 06269, USA.

that stabilities emerge from the interaction between these systems. Further, when vocalizations are emphatically stressed (e.g., co-occurring with an increased mouth aperture), or when finger tapping is given a stress (by increasing the amplitude of the movement), the other modality will tend to also perform a stressed movement or utterance (Parrell *et al.*, 2014; Krahmer and Swerts, 2007). In more natural gesturing contexts, it is found that calling out the name of an object while pointing to the object drives speakers to synchronize the maximum extension of pointing movement with that of the stressed part of the vocalization (Esteve-Gibert and Prieto, 2013; Rochet-Capellan *et al.*, 2008). Even when a gesture-speech system is perturbed by experimenters interfering with the execution of either hand movement or speech production, the other modality flexibly adjusts so as to maintain gesture-speech synchrony (Chu and Hagoort, 2014; McNeill, 1992; Pouw and Dixon, 2019a). In sum, research has demonstrated that synchronization of gesture and speech naturally emerges.

A recently discovered biomechanical constraint for the emergence of gesture-speech synchrony is the role of the physical impulse of gestural upper-limb movement on the respiratory system (Pouw *et al.*, 2019a, Pouw *et al.*, 2020b). When vocalizing a steady-state vowel /ə/ (as in "cinema") while making a higher-impulse arm movement vs lower-impulse wrist movement, especially higher-impulse actions imprint acoustics such that they are distinguishable from vocalizations made without movement. Specifically, positive peaks in *F*0 and the amplitude envelope occur at moments where the upper-limb movement reaches a moment of acceleration or deceleration, i.e., moments of higher physical impulse. Such gestural effects on acoustics were more pronounced when participants were standing vs sitting (Pouw *et al.*, 2019a), indicating an important role for postural stability on the effect of physical impulses on the respiratory-vocal system (Cordo and Nashner, 1982).

In this previous research, we hypothesized that the primary medium for gesture-speech coupling is the physical body. This hypothesis derives from theory (Ingber, 2008; Levin, 2006; Turvey and Fonseca, 2014) and research (Silva *et al.*, 2007) in biomechanics, which hold that the (human) myofascial-skeletal system is a pre-stressed system with tensile (e.g., fascia) and compressive (e.g., bones) elements. This *tensegrity* architecture not only allows for active postural equilibria such as standing upright, or keeping the scapula suspended on the thorax for upper limb movement

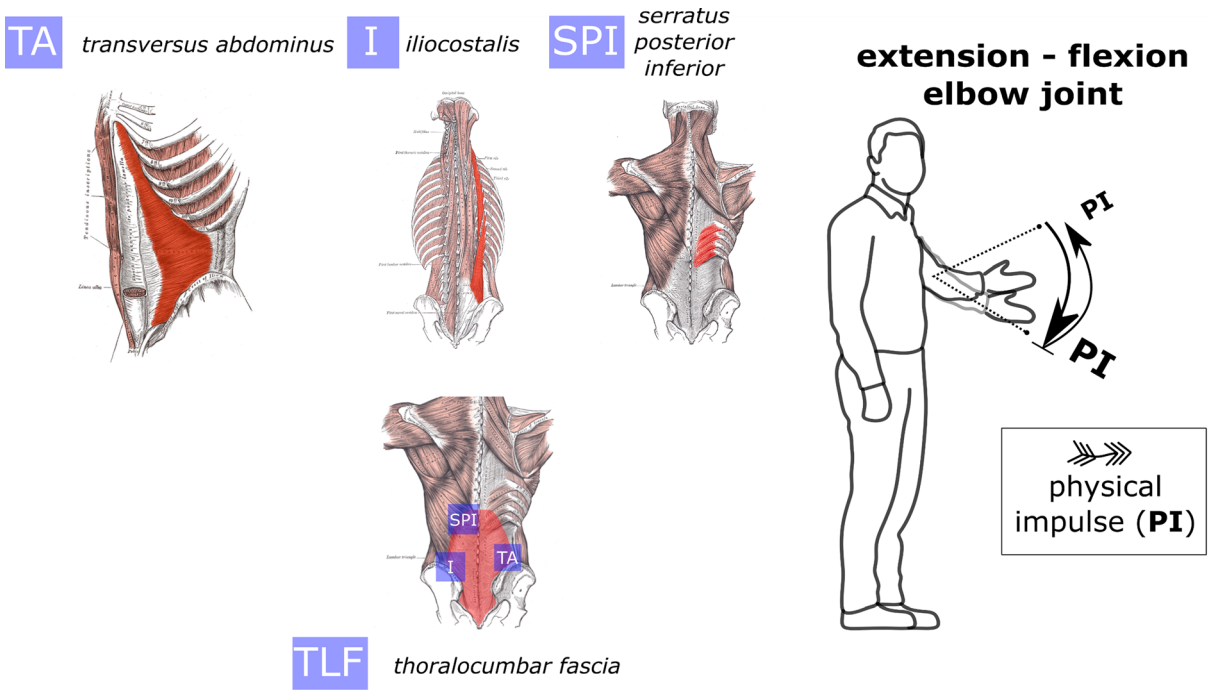


FIG. 1. (Color online) Consider a cyclic extension-flexion movement cycle around the elbow joint such that there is a sudden stop at the moment of maximum extension (a downbeat). The sudden changes in velocity (deceleration/acceleration) at the maximum extension, but also at the maximum flexion, will mechanically transfer forces onto the myo-fascial skeletal system. This force transfer or physical impulse distributes over tensioned muscles which stabilize upper limb action and are associated with respiratory control (e.g., Serratus Anterior). Furthermore, during the maximum extension, anticipatory muscle activations in the back and trunk are recruited to maintain posture (Aruin and Latash, 1995; Cordo and Nashner, 1982; Hodges and Richardson, 1997). Specifically, TA and SPI are implicated in expiratory control, while iliocostalis is implicated in stabilizing inspiratory action by raising the chest wall. Tensioning of these muscles further changes tonus of the heavily innervated thoracolumbar fascia, a prestressed connective tissue that many of the accessory respiratory muscles connect into (Turvey and Fonseca, 2014). For this particular movement, we have found that the physical impulse during the maximum extension increases *F*0 and intensity of steady-state phonation, and thus we must conclude from the current perspective that the net effect of physical impulses leading to an increase in subglottal pressure. Note, however, that the exact muscle contributions in this net positive effect on sub-glottal pressure, affecting vocalization, have not yet been directly studied. Therefore, at the current moment we can only identify *candidate* myofascial areas of interest for particular upper limb movement linkages with the respiratory system. Anatomy pictures are obtained from Wikipedia by Mikael Häggström and the human pose figures were modified from Dimensions.Guide (2019).

https://doi.org/10.1121/10.0001730

(Levin, 2006), but more importantly, it also poises the system to quickly organize multiple degrees of freedom in action and reaction by the fact that forces distribute through a tensioned medium.

Specifically, we suggested that impulsive forces produced by the motions of the upper limb reach the respiratory system, increasing lung pressure and thereby changing phonation acoustics. Figure 1 illustrates a simple cyclic extension-flexion movement of the forearm about the elbow joint. This motion imparts forces on the body. By physical law, the forces transferred from the moving arm to the body depend upon changes in momentum of this effector (momentum of effector = effector mass × effector velocity). Changes in momentum are captured via the physical property of impulse. Given that the mass of the effector is constant, impulse can be equated to acceleration (change in velocity). The sudden arresting of motion at the moment of a downbeat is consequently associated with an extremum in the magnitude of both impulse and acceleration. Changes in impulse will reverberate non-linearly (Levin, 2006) across the tensioned myofascial skeletal system.

Producing a simple cyclic extension-flexion movement of the forearm while also maintaining the posture of the body requires an ensemble of stabilizing muscle activations that are non-coincidentally also accessory respiratory muscles. For example, the scapula must be stabilized during these upper limb motions. This is partially provided by tensioning of the Serratus Anterior, which has been implicated in expiratory control (Smith *et al.*, 2003), though its respiratory function is under debate (Basmajian and De Luca, 1985). Happening more distally from the upper limb movement itself, there are anticipatory muscle activations occurring about 50–70 ms (Aruin and Latash, 1995; Bouisset and Do, 2008) before the peaks in physical impulse that are needed to counteract the destabilizing impulses. Such anticipatory muscles include transvs abdominus, rector abdominus, serratus posterior—all muscles that are also accessory to the control of expiratory flow (e.g., Cordo and Nashner, 1982; Hodges and Richardson, 1997). In sum, for simple extension-flexion upper limb movements there are a number of potential biomechanical linkages with the respiratory system, and for other upper limb movements there are potential other such linkages (e.g., internal rotation of the humerus requires tensioning of the pectoralis major).

In the hierarchy of control for movement construction, as classically proposed by Bernstein (1966), myofascial-skeletal tensioning is serving as the level of "tonus"—an enabling background at which action unfolds by tuning the action's potential (Profeta and Turvey, 2018). Applied to gesture's effect on vocalization, we propose that gesture can be one of several tuning parameters for vocalization. In this way, we emphasize an appreciation of how speech consists of "nested periodicities" (Kelso and Tuller, 1984, p. 933; see also MacNeilage, 1998; Iverson and Thelen, 2005). In the case of prosody, the nested levels of activity can come to resonate (Raja, 2020; Gibson, 1966). A stressed syllable can be realized by organizing a resonating set of components that each contribute to reaching a prosodic target. For example, an increased tension of the vocal folds may resonate with increases in alveolar or lung pressure as supported by mechanical loading of the upper limb movement. Thus, hand gestures can participate as a resonant element in a task-specific device that allows for reaching prosodic targets (Kugler and Turvey, 1987). Of course, gestures are not required to participate, or they may simply fall out of resonance with other levels of activity, as in the obvious case when there is no vocalization during a mechanical loading of gestures on respiration.

If we carve up the human speech system into respiration, vocal cord, and articulatory levels of activity (MacNeilage, 1998), gestures' physical impetus interacts at the level of respiration. During speech production, the main energy source comes from the elastic recoil of the lungs that drives subglottal pressure so as to expire. However, this elastic recoil is modulated by an ensemble of primary (intercostal muscles) as well as accessory muscles. This ensemble of muscles forms a uniquely complex control system that appears in humans to be specifically adapted for speech control (MacLarnon and Hewitt, 1999). In terms of the acoustics, the respiratory system primarily modulates the intensity of phonation (Finnegan *et al.*, 2000). Increased lung pressures also increase the fundamental frequency (perceived as pitch), if the vocal cord tensioning remains unchanged (Baer, 1979; Lieberman, 1996). However, changes in lung pressure need not materialize into changes in $F0$ as the larynx is the primary modulator. With dynamic adjustments of the larynx, the same $F0$ can be maintained under different levels of lung pressures. Nevertheless, the respiration system can come to intentionally participate in modulating $F0$, as it has been observed that short chest-pulses can support the production of emphatically stressed speech accompanied by increases in F0 (Fuchs *et al.*, 2015; Ladefagod, 1968; Ohala, 1990; Petrone *et al.*, 2017). Thus, within the current line of thinking, gesture-induced physical impulses are likely to primarily affect intensity, and potentially F0, as has been observed in previous research with continuous vocalizations (Pouw *et al.*, 2019a; Pouw *et al.*, 2020b).

A. Current study

Although there is evidence suggesting that gestures affect acoustics via the respiratory system and that respiration-related muscle activity is sometimes implicated in emphatically stressed speech, it is yet to be *directly tested* that gesturing and prosodic aspects of speech are linked *via* the respiration system. Additionally, direct evidence for gesture's biomechanical effects on acoustics is still based on a rather rudimentary paradigm (Pouw *et al.*, 2019a; Pouw *et al.*, 2020c) where subjects need to produce steady-state vocalizations. Although there is indirect evidence that gestures might indeed affect spontaneous speech through gesture-speech physics (e.g., Cravotta *et al.*, 2019), direct evidence is needed to show that physical impulse can affect

more speech-like productions through respiratory modulations.

In the current study, participants uttered a consonant-vowel (CV) mono-syllable (/pa/) at 1-second intervals, where participants inhaled once to produce eight CV vocalizations timed by a visual metronome. We studied timing and the magnitude of impulse of an upper limb movement in relation to its potential effect on key acoustic parameters (amplitude envelope and *F*0), as well as its effect on chest-respiratory activity as measured through a respiration belt. Specifically, participants vocalized mono-syllables either without any movement (passive condition) or while moving the wrist or arm vertically at 1-s intervals. Movements consisted of a stress in the downbeat such that faster velocities were produced with more forceful decelerations during the downbeat as compared to the upward (flexion) motion. Crucially, we trained subjects to either time the upper limb downbeat with the vocalization (in-phase condition), or to produce the vocalization when the movement was in a low-impulse flexion movement (90° out-of-phase condition). With this phasing manipulation, we tested whether the moment of physical impulse needs to co-occur with the vocalization (as is the case for the in-phase condition) to maximally impart effects on acoustics, rather than other moments in the same type of upper limb movement (i.e., during the 90° out-of-phase condition).

The current confirmatory study was preceded by preliminary findings (Pouw *et al.*, 2019b) obtained from an exploratory dataset which formed the basis of the pre-registration of the current analysis (https://osf.io/x7zdc/). In "that" exploratory study, we found promising results which we test here in replicatory fashion. We aim to replicate that gesturing in-phase leads to higher maxima in the vocalization parameters (amplitude envelope, *F*0, chest-respiratory activity) as compared to the passive condition. We then assess whether high-impulse arm movements impart more extreme effects on vocalization parameters than lower-impulse wrist movements and whether in-phase movements are reliably different from out-of-phase movement in terms of vocalization maxima. We then quantify how temporal distance from the peak in deceleration relates to acoustic peaks in the vocalization. Finally, we investigate how chest-respiratory activity is related to upper limb movement and vocalization acoustics.

II. METHOD

The data and code supporting this study and the time-stamped version of the pre-registration of the confirmatory analysis can be found on the Open Science Framework (OSF, https://osf.io/e3ukd/). Twenty-nine American-English speaking participants [21 cis-gender females and 8 cis-gender males, Mean *M* (Standard Deviation *SD*) age = 19.35 (1.11), all right-handed] were tested in this experiment. As such, our participants were sampled from a narrow population and we need to be careful in generalizing to other potential populations. The design was fully within-subject, with a two-level movement-type condition (wrist vs arm movement) and a two-level phasing condition (90° out-of-phase vs in-phase). As a control condition, participants also performed a passive condition with no upper-limb movement during vocalizations. Each participant performed four blocked trials containing each crossed condition (in-phase wrist, in-phase arm, 90° out-of-phase wrist, 90° out-of-phase arm), next to a passive condition for each block. Thus, 20 trials in total were performed with eight CV mono-syllabic vocalizations for each trial. A maximum of 4640 vocalizations were expected (29 participants × 20 trials × 8 vocalizations). Given participant errors (e.g., missing a vocalization because of skipping a beat, starting too late within the trial, or stopping too early) and one missed trial due to experimenter error, we obtained 4585 usable vocalizations.

A. Materials, equipment, and measurements

1. Respiration belt

To measure chest-wall kinematic activity, we used a NUL-236 respiration belt (NeuLog, Inc.), sampling at ~30 Hz, 15 ADC resolution, and a range of 0–20 000 arbitrary units. The respiration belt has an air-filled bladder, embedded in the belt, which is fitted around the trunk just below the sternum. The belt was filled with air by the experimenter with a hand-held pump; pressure was increased such that the belt fit comfortably but tightly around the participant's trunk (for video tutorial, see https://neulog.com/respiration-monitor-belt/). This device provided measurements from a pressure sensor within the air-bladder, thereby tracking increasing trunk circumference through higher pressure measurements. The respiration belt captures breathing cycles (with higher pressure indicating inspiration), but also chest-wall kinematics having to do with the tensioning of the muscles around the trunk. We instructed participants to use one breath for each trial, as we wanted to gauge the overall physical activity around the respiratory system within the same breath cycle (henceforth referred to as chest-respiratory activity). Muscle tensions around the trunk will increase chest circumference shown as positive peaks in the respiration-belt readings. The respiration belt measurements were rescaled by maximum and minimum air volume as pretested in a baseline trial. For each trial, we linearly detrended the respiration time series so as to remove the negative trend of chest-circumference due to decreasing air volume during expiration.

2. Audio and hand-motion recording

We used a MicroMic C520 (AKG, Inc.) headset condenser cardioid microphone to obtain audio recordings. A Polhemus Liberty (Polhemus, Inc.) was used for motion tracking, with a single wired light-weight sensor attached to the index finger of the dominant hand. The sampling rate of the motion-tracker was set the same as the respiration belt (~30 Hz), and vertical positions traces (z) and its first (vertical velocity, z') and second derivatives (acceleration, z'') with respect to time were smoothed with a low-pass first-

https://doi.org/10.1121/10.0001730

order 33 Hz Butterworth filter; please see our annotated script for computations https://osf.io/37pzt/.

3. Synchronization of recording

A c++ script was written (see https://osf.io/u2q4f/) to simultaneously record from the respiration belt, microphone, and the motion tracker. The script contains code made available by Richardson (2009) written for the Polhemus motion-tracking system, which was further modified to interface with the respiration belt's application programming interface (https://neulog.com/software/), as well the recording for the audio handled by SFML audio package for c++ (https://www.sfml-dev.org/).

4. Audio processing

We extracted *F*0 and the smoothed amplitude envelope (ENV) from the 44.1 kHz audio. We extracted *F*0 and ENV time series with a 200 Hz sampling rate. We automatically extracted acoustic traces from all audio using custom-written R-scripts. Our *F*0 extraction script (https://osf.io/m43qy/) utilizes R-package wrassp (Bombien *et al.*, 2020) which applies a K. Schaefer-Vincent algorithm for *F*0 detection, with preset ranges for male (50–400 Hz) and female participants (80–640 Hz). For the smoothed amplitude envelope (5 Hz Hanning filter), we rewrote code originally scripted in PRAAT by He and Dellwo (2017) into an R-script (https://osf.io/uvkj6/), which automatically extracted the amplitude envelope of a set of files (Pouw and Trujillo, 2019).

B. Aggregation and post-processing of data

With a custom R-script dedicated for post-processing of the data (https://osf.io/37pzt/), we aligned the acoustic time series data (*F*0 and amplitude envelope) together with the motion tracking and respiration time series data, by up sampling motion and respiration data to 200 Hz. We used ELAN (Wittenburg *et al.*, 2006) to mark the beginning and end of each trial, which was fed as input for the R-script to extract relevant episodes from the data. The individual vocalic events within each trial were automatically identified by assessing the eight longest runs of uninterrupted *F*0 observations within each trial. We observed that some of those runs had intermittent gaps in observations due to some failed tracking of *F*0 and, therefore, we linearly interpolated small gaps in *F*0 observations of up to 25 milliseconds (i.e., maximum five observations) within trials which aided the identification of the eight vocalic events. The script further identified and marked relevant movement phases (e.g., the period of the movement), the timing between kinematic peaks (e.g., peak deceleration), the vowel midpoint, and relevant maxima observed in acoustics and kinematics, which would later be submitted for analysis.

C. Procedure

Participants sat on a chair without arm rests (see Fig. 2).[1] Participants were instructed to sit upright at the front of their seats so as to not touch the back rest of the chair. The respiration belt was then fitted. The motion tracking sensor was attached to the dominant hand's index finger, and the headset microphone was put on by the participant.

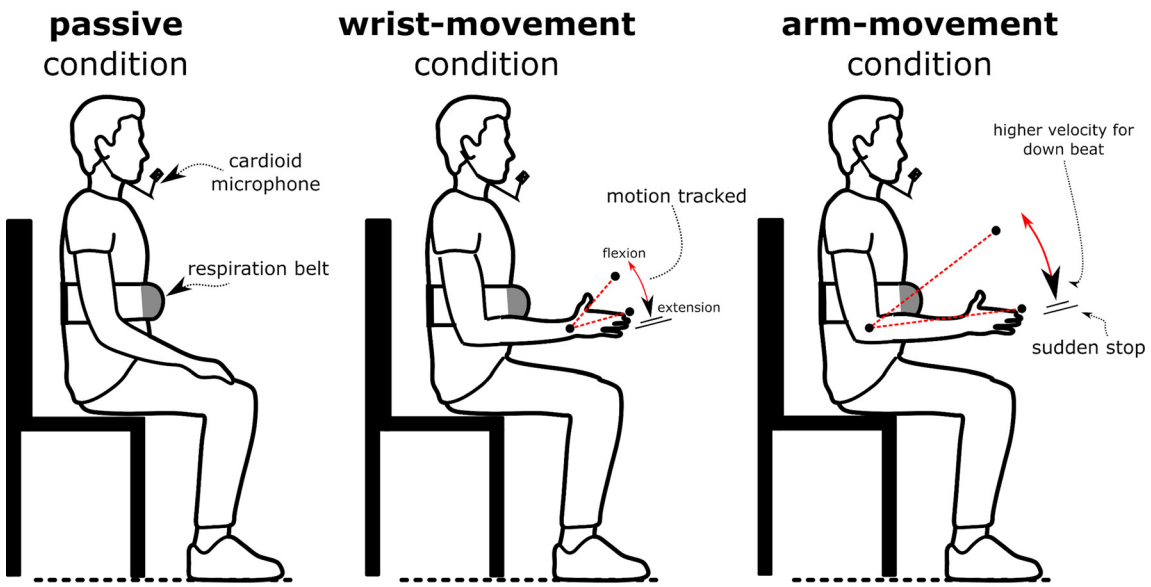


FIG. 2. (Color online) Schematic presentation of the movement conditions and measurements. The upper limb movements were performed in the sagittal plane, moving upwards (flexion phase) and downward (extension phase). There were higher velocities in the extension phase, and there was a sudden stop to be performed as the maximum extension was reached (high deceleration to the point of maximum extension). During these changes in velocity near the maximum extension, there are mechanical loadings of the upper limb onto the body (i.e., physical impulse). The sudden stop is thus controlled by the participant's counteracting flexion movement and not some kind of physical barrier. The human figures have been modified from an open database (Dimension.Guide, 2019).

For a respiration baseline, we measured participants' maximum inspiration and expiration levels. Then participants were shown a visual metronome, showing eight equally spaced tiles. Each tile would light up one-for-one with 1-s intervals (see https://osf.io/qr65j/). The start of the trial was decided by the participant. The start of the first vocalization had to be timed with the illumination of the first tile of the visual metronome. The vocalizations were instructed to be as monotonically performed as possible for the whole experiment ("please keep the same level of pitch and loudness during these vocalizations"). Participants were additionally instructed to vocalize eight times with one expiration. Thus, participants took one inhalation before the start of the trial and were asked to *not* inhale between vocalizations of the ongoing trail. This instruction helps reduce noisy estimates of chest-respiratory activity having to do with the many respiration strategies participants could take (e.g., inhaling every two vocalizations, vs every fourth vocalization). We also asked participants to report if they accidentally inhaled within a trial (or made some other mistake), which only happened seven times. Such trials were immediately redone by the participant. We do have to accept that, in some cases, participants might have subconsciously inhaled within vocalizations which adds noise to our measurements.

Subsequently, participants were familiarized with the movement conditions shown in Fig. 2. For the passive condition, participants held their hands on their legs. For the arm-movement condition, participants were instructed to move their lower arm vertically up and down in the sagittal plane around the elbow joint, with no movement around the wrist or shoulder joint. For the wrist-movement condition, there was only vertical movement around the wrist joint. In the wrist- and arm-movement conditions, participants were instructed to give a beat or contrast in the down-beat extension by more quickly moving downward (as compared to upward) and by quickly halting (decelerating) the movement in the maximum extension. These movements were used to create different levels of physical impulse on the body.

After introduction of the movement conditions, the phasing conditions were introduced (see also Fig. 3). For the in-phase condition, participants produced CV productions together (i.e., *in-phase*) with the beat of the downward (extension) motion. Thus, for the in-phase condition when the downward motion reached zero velocity, then the vocalization needed to be made. This condition allows for the CV vocalization to occur at the moment where there is a physical impulse of upper-limb movement. For the 90° out-of-phase condition, we aimed to have the CV vocalizations occur at the moment of least physical impulse. Participants performed the same movement as before (faster downward

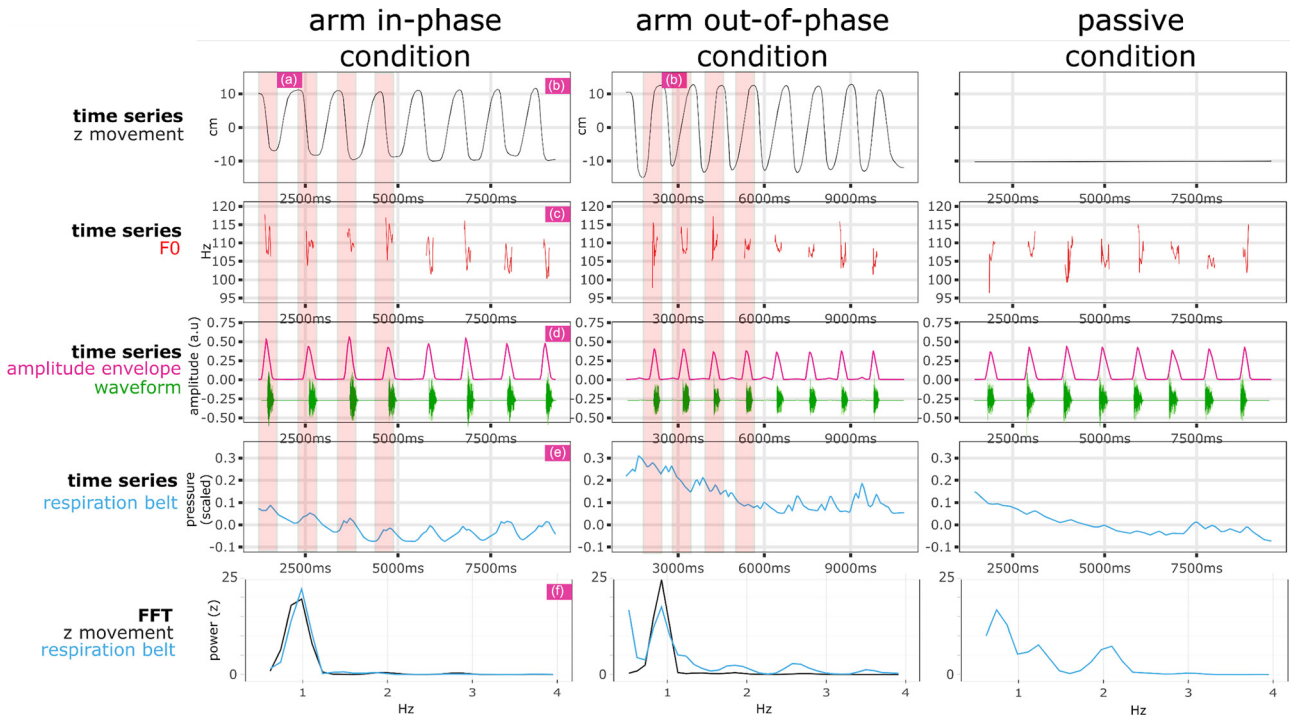


FIG. 3. (Color online) This figure shows example trials performed by a male participant for the passive condition and the two phasing conditions in-phase and the 90° out-of-phase condition. Key measurements are shown: (a) phasing relations of vocalization and movement, where it can be seen that vocalizations are timed around the maximum extension for the in-phase condition (red shaded area indicate extension phases for the first four cycles), while this is not the case for the 90° out-of-phase condition, which has vocalizations in the flexion phases (as indicated by red shaded area). (b) Vertical displacement of the index finger (z movement; shown in black), (c) fundamental frequency ($F0$) of vocalizations shown in red, (d) the smoothed amplitude envelope shown in pink (which roughly traces the outlines of the waveform shown in green), and (e) the concurrent respiration belt measurements shown in blue. (f) To provide in indication how respiration is structured, spectral analysis (Fast Fourier Transform performed on linearly detrended time series) shows that for these trials, there are consistent peaks around the 1 Hz range for the movement conditions, which suggests that the belt is reading movement-related chest-respiratory activity, where more varied periodicities are shown in the 90° out-of-phase and especially the passive condition.

https://doi.org/10.1121/10.0001730

motion, high deceleration in the extension phase), but they timed their CV vocalization at the moment where the wrist or arm was moving upward (flexion phase) after the beat (where velocity is most constant and >0). The participants were explicitly asked to make sure the hand had initiated the upward motion before CV vocalization was made. This is a typical 90° out-of-phase phase relation. This vocalization-movement task was initially difficult for participants, especially as participants needed to flexibly alternate between these coordination regimes. Therefore, participants practiced for about 5 min, with the experimenter providing feedback. Participants were asked to alternate between in-phase movement-vocalization and 90° out-of-phase movement trials (e.g., one out-of-phase with arm, then one in-phase with wrist, and then one out-of-phase with arm and then one in-phase with wrist). If participants were able to successfully complete these alternating trials for wrist and arm movements (so four trials) without clear incorrect timings (as observed by the experimenter) the main experiment would be initiated. If participants made a mistake, they were asked to start the sequence of practice trials over again. This familiarization phase resulted in a satisfactory performance in the final experiment (see manipulation checks in Sec. II D).

D. Manipulation checks

1. Phasing manipulation

Figure 4 shows that, for the in-phase condition, the timing of the vowel midpoint was tightly aligned, but slightly follows, the maximum extension, $M = 84$ ms, $SD = 154$, 95% confidence interval CI [83, 85]. If we look at 90° out-of-phase condition in Fig. 4, it can be seen that vocalizations are further away from the maximum extension at either side. Note that we use the vowel midpoint for timing calculation as this is a signal-based timing anchor for the vocalization, where for each run of $F0$ observations during a vowel the middle observation is chosen as the representative $F0$ (i.e., midpoint $F0$). The bimodal distribution indicates arbitrary leader-follower assignment of movement vs vocalization (depending on which starts first). If we take the absolute value of the timings, we see that these 90° out-of-phase vocalizations occur at half the period of the 1-s movement cycle, $M = 447$ ms, $SD = 133$, 95% CI [446,448]. These timings indicate that our phasing manipulation was successful, with vocalizations in the in-phase condition being closer to the peak in physical impulse as compared to the 90° out-of-phase condition.

2. Movement conditions

From the means and confidence intervals shown in Table I, we can see that moving one's arm vs wrist leads to more extreme positive and negative vertical velocity, as well as higher vertical amplitude of the movement, while having similar movement periods around one second. For the phasing conditions, it can be obtained from the overlapping confidence intervals in Table I that the negative as well as positive velocity, and the movement's vertical amplitude for the flexion and extension phase was comparable for the in-phase and 90° out of phase condition, suggesting that phasing relations did not dramatically affect upper limb kinematics. Note, however, that the 90° out of phase condition did have slightly longer movement periods as compared to in-phase conditions. This is possibly because participants in the 90° out-of-phase condition were extending their movement periods so as to provide a more comfortable amount of time for the vocalization to occur (see, e.g., Stoltmann and Fuchs, 2017), which is corroborated by our finding that flexion as well as extension amplitude is slightly higher during 90° out of phase movement (as compared to in-phase movement).

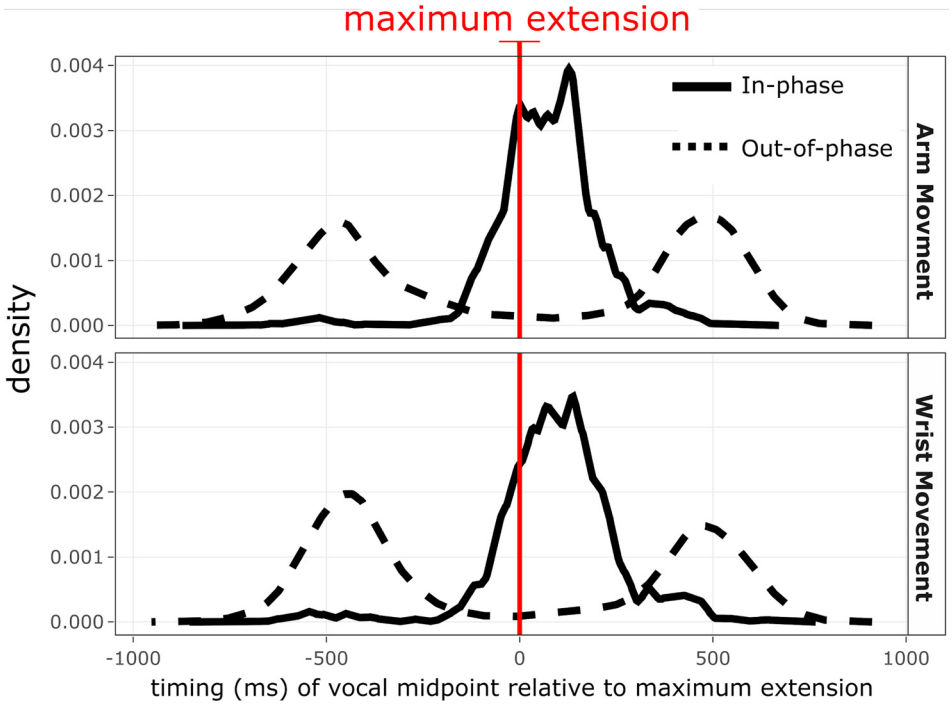


FIG. 4. (Color online) Smoothed density distributions for the in-phase (solid lines) and 90° out of phase condition (dotted line) for the observed timings of the middle of the vowel event relative to the moment of maximum extension of the upper limb movement (indicated by the red vertical bar at timing = 0 ms). The upper panel shows the timing distributions for the arm movement condition and the lower panel shows the wrist movement conditions. Negative timings indicate that the maximum extension followed the vocalization, while positive timings indicate that vocal midpoint followed the maximum extension.

TABLE I. Descriptives of upper limb kinematics per condition.[a]

		Wrist beat Mean (*SD*) [95% *CI*: lower, upper]	One−arm beat *Mean* (*SD*) [95% *CI*: lower, upper]
Max negative velocity *z* (cm/s)	In-phase	−146 (52) [−149, −142]	−269 (192) [−281, −257]
	Out-of-phase	−144 (59) [−147, −140]	−261 (201) [−273, −248]
Max positive velocity *z* (cm/s)	In-phase	66 (34) [64, 68]	146 (160) [137, 156]
	Out-of-phase	78 (37) [76, 80]	165 (171) [154, 175]
Duration (ms) Movement period	In-phase	988 (167) [977, 999]	980 (202) [967, 993]
	Out-of-phase	1010 (156) [1000, 1020]	1007 (193) [995, 1019]
Max Flexion Amplitude (cm)	In-phase	25 (8) [24, 25]	46 (15) [44, 47]
	Out-of-phase	26 (8) [26, 27]	47 (18) [46, 48]
Max Extension Amplitude (cm)	In-phase	25 (8) [25, 26]	47 (22) [45, 49]
	Out-of-phase	27 (9) [26, 28]	49 (24) [47, 51]

[a]Note: The max negative velocity is given in cm/s and indicates the average vertical velocity attained in the extension phases, while the positive velocity indicates the average maximum vertical velocity during of the flexion phases. The duration indicates the time for a movement cycle (period) to complete, which should be around 1 second (the inter-vocalization interval). The flexion and extension amplitude are averaged over all movement periods, except the first and last movement period within a trial so as to exclude possible movements from or to a resting position.

E. Analysis information and alpha restriction

The R analysis script can be found on the OSF (https://osf.io/csk5g/). For all regression analyses, we use mixed regression analysis implemented by R-package nlme (Pinheiro *et al.*, 2019) with maximum likelihood estimation. We also use EMAtools (Kleiman, 2017) to produce Cohen's *d* effect sizes for nlme models. Unless otherwise indicated, we always include a random intercept for participant in our models. As stated in the pre-registration, given that we have four major analyses, we will only treat p's < 0.013 (alpha $= 0.05$ divided by 4) as indication for statistical significance.

III. RESULTS

A. Descriptives

Table II provides a descriptive overview of key vocalization markers. Throughout the analyses that follow, we always use *z*-standardized, acoustic, and respiration measurements. This *z*-transformation is applied within subjects, such that individual variability in these measurements (e.g., differences in *F*0 due to sex) can be ignored. Figure 5 contains an overview of the key results.

B. Movement vs passive control condition

First, we assess whether moving in-phase or 90° out-of-phase leads to heightened peaks in the amplitude envelope, *F*0, and respiration-related movement (Respiration) as compared to the passive control condition. The results of the mixed regression analysis are shown in Table III. The maximum amplitude envelope is higher for the in-phase movement condition and lower for the 90° out-of-phase condition, as compared to the passive control condition. These results are the same for the maximum respiration-related movement. Movement or movement-phasing does not affect *F*0 as compared to the passive condition. Thus, moving the upper limbs so that the physical impulse occurs together with the vocalization increases the amplitude of speech and is associated with more respiration-related activity as compared to not moving during vocalizations.

C. The role of physical impetus and phasing in upper-limb movement vocalizations

Having ascertained that the passive condition is distinguishable from movement conditions, for our next analyses we assess the different roles of movement type (wrist vs

TABLE II. Descriptives: Acoustic measures.[a]

	Passive *Mean* (*SD*)	Wrist in-phase *Mean* (*SD*)	Wrist out-of-phase *Mean* (*SD*)	Arm in-phase *Mean* (*SD*)	Arm out-of-phase *Mean* (*SD*)
Duration phonation (ms)	298 *(107)*	282 *(111)*	278 *(104)*	284 *(118)*	275 *(106)*
Max *F*0 (*z*)	1.28 *(1.01)*	1.28 *(0.98)*	1.21 *(1.17)*	1.26 *(1.19)*	1.23 *(1.34)*
Max *F*0 Male (Hz)	148 *(33)*	150 *(35)*	149 *(34)*	149 *(35)*	150 *(34)*
Max *F*0 Female (Hz)	225 *(30)*	224 *(30)*	223 *(30)*	226 *(30)*	225 *(30)*
Vowel Midpoint *F*0 (*z*)	−0.11 *(0.70)*	−0.07 *(0.69)*	−0.08 *(0.92)*	−0.06 *(0.94)*	−0.07 *(1.00)*
Midpoint *F*0 Male (Hz)	129 *(28)*	132 *(31)*	131 *(29)*	132 *(31)*	132 *(30)*
Midpoint *F*0 Female (Hz)	196 *(30)*	196 *(31)*	197 *(31)*	199 *(30)*	196 *(30)*
Max Amplitude Envelope (*z*)	2.18 *(0.79)*	2.26 *(0.82)*	2.00 *(0.83)*	2.41 *(0.91)*	2.05 *(0.83)*
Max Respiration (*z*)	0.21 *(1.02)*	0.49 *(0.91)*	0.23 *(0.85)*	0.74 *(1.03)*	0.21 *(1.00)*

[a]Note: The average duration of vowel phonation of the CV utterance is given alongside the maximum of the *z*-standardized (*z*) and raw (Hz) measurements of *F*0, Midpoint *F*0, amplitude envelope, and respiration are averaged for each CV utterance.

https://doi.org/10.1121/10.0001730

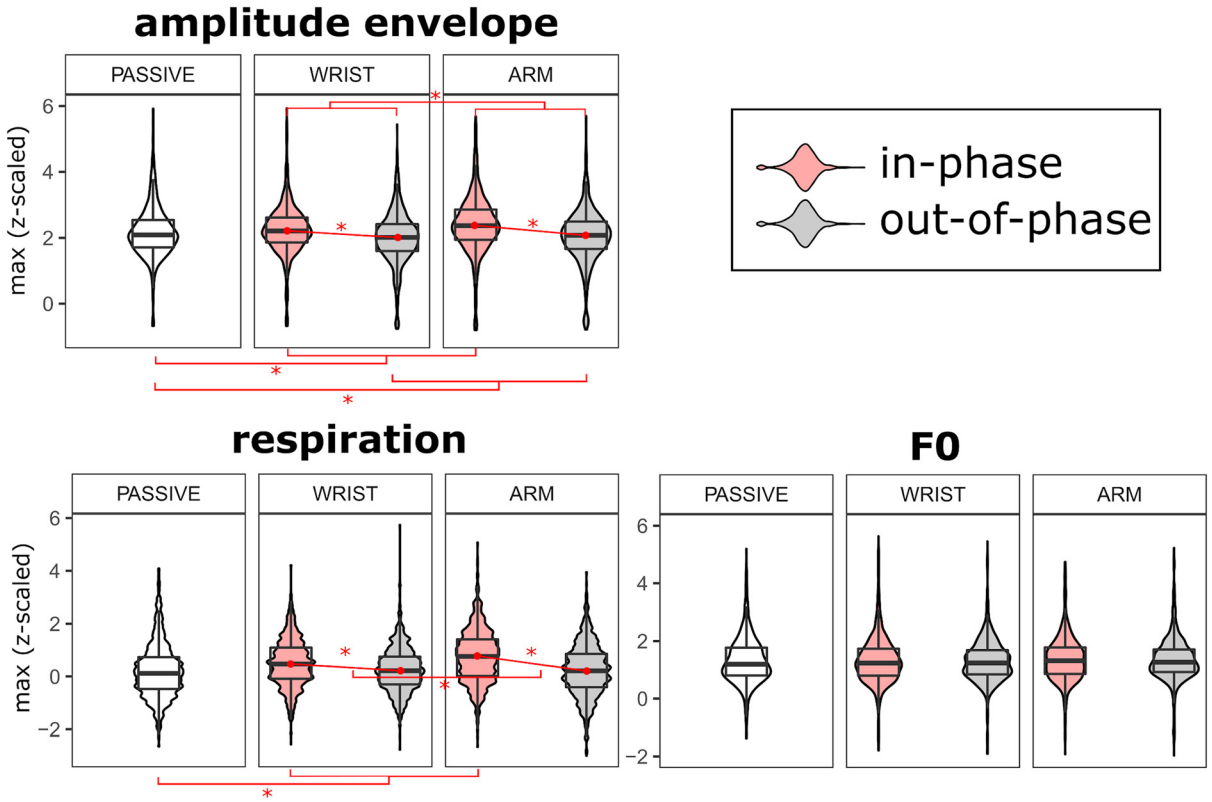


FIG. 5. (Color online) This graph shows the key acoustic measurements of the CV utterances, as well as the statistically reliable contrasts (in red, $*p < 0.013$) we obtained from statistical modeling (reported in Secs. III B and III C). Key measures were the amplitude envelope, *F*0, and respiration-related movement (respiration) maxima per vocalization event, *z*-standardized at the participant level. For each condition, the mean and quartiles are shown as boxplots, and distribution of the observations are shown using violin outlines. We obtain that passive condition had lower respiration and amplitude maxima as compared to in-phase movements, and had higher intensity as compared to out-of-phase movement for intensity. Our follow-up analysis obtained that for respiration and amplitude envelope, there is generally higher respiration and intensity maxima for in-phase vs out-of-phase movement. Finally, arm movements had higher maximum for intensity as compared to wrist movements, and this was the case for respiration as well, but only for in-phase movements. Figure 11 in the Appendix (footnote 2) further shows the overall trajectories of vocalization events as modeled by GAM. Plots of all trajectories for all vocalizations can be found on our supplemental OSF webpage for the amplitude envelope, *F*0, and respiration data.

TABLE III. Mixed regression modeling of passive vs phasing condition.

Max ENV (z-scaled)	***b*** [a]	***t*(4553)** [b]	***p***	**Effect size (*d*)**
Passive (intercept)	2.18	37.17	<0.001	
Passive vs in-phase movement	0.15	4.57	<0.001	0.13
Passive vs out-of-phase movement	−0.16	−4.91	<0.001	−0.14
Max *F*0 (*z*-scaled)	***b***	***t*(4553)**	***p***	**Effect size (*d*)**
Passive (intercept)	1.28	15.88	<0.001	
Passive vs in-phase movement	−0.01	−0.27	0.784	<0.001
Passive vs out-of-phase movement	−0.06	−1.36	0.175	<0.001
Max Respiration (*z*-scaled)	***b***	***t*(4553)**	***p***	**Effect size (*d*)**
Passive (intercept)	0.21	4.11	<0.001	
Passive vs in-phase movement	0.42	10.93	<0.001	0.323
Passive vs out-of-phase movement	0.02	0.53	0.594	0.015

[a]"*b*" provides the slope estimate for the model predictors.
[b]Note that the the high amount of degrees of freedom of the model reflect the high amount of observations (vocalization events).

arm) and phasing for the movement conditions only. Table IV provides the results of the mixed regression analysis, where interactions between movement and phasing are only reported when found reliable as established by a significance test of the change in Chi-squared between model with and without the interaction added.

Again, we find no statistically reliable effects for conditions on maximum F0. However, we do find that the amplitude envelope maximum was higher for arm movements as compared to wrist movements. Additionally, there was a statistically reliable effect of phasing, with in-phase condition leading to higher peaks in the amplitude envelope as compared to the 90° out-phase condition. Similar main effects for movement type and phasing were found for the maximum respiration-related activity (max respiration), next to an interaction effect of movement type and phasing. Further assessing this interaction with *post hoc* analysis (R-package lsmeans; Lenth and Lenth, 2017) with Bonferroni correction, showed that only in the in-phase condition did arm movement lead to higher peaks in respiration-related movement as compared to wrist movement. In the 90° out-of-phase condition, movement type

TABLE IV. Mixed regression passive vs phasing condition.[a]

Max amplitude envelope (*z*-scaled)	***b***	***t*(3628)**	***p***	**Effect size (*d*)**
Intercept	2.38	39.95	<0.001	
Phase (Out-of-phase vs in-phase)	−0.30	−11.48	<0.001	−0.381
Movement (wrist vs arm)	−0.10	−3.75	<0.001	−0.124
Max *F*0 (*z*-scaled)	***b***	***t*(3628)**	***p***	**Effect size (*d*)**
Intercept	1.26	17.47	<0.001	
Phase (Out-of-phase vs in-phase)	−0.05	−1.28	0.200	−0.042
Movement (Wrist vs arm)	0.01	0.15	0.878	0.153
Max respiration (*z*-scaled)	***b***	***t*(3628)**	***p***	**Effect size (*d*)**
Intercept	−0.75	13.61	<0.001	
Phase (In-phase vs out-of-phase)	−0.53	−12.37	<0.001	−0.410
Movement (Wrist vs arm)	−0.25	−5.92	<0.001	−0.196
Phase × Movement	**0.27**	**4.44**	**<0.001**	0.147
Phase × Movement Posthoc	**Estimated difference**	***t*(3628)**	**corrected *p***	
In-phase: Wrist vs Arm	−0.252	−5.92	<0.001	
Out-of-phase: Wrist vs Arm	0.015	0.03	0.719	

[a]Note: The mixed regression coefficients are shown per acoustic parameter. Note that for modeling maximum respiration, we find a statistically reliable interaction between movement type and phasing (indicated in bold and italic), wherein we further probed the relation between movement type for each phasing condition separately. Only in the in-phase condition (but not 90° out-of-phase) we see that wrist movement has lower peaks in respiration as compared to arm movements. Interaction effects of phasing and movement type are not obtained for modeling of amplitude envelope or *F*0 maximas.

was not a statistically reliable predictor. These analyses further confirm that the in-phase condition is associated with higher peaks in amplitude envelope and respiration-related activity, and higher impulse arm movements are associated with higher peaks as compared to lower impulse wrist movements.

Importantly, we also see clearly that the *F*0 is not affected by our experimental manipulations. There is a possibility that maximum *F*0 is a noisy point-estimate not representative of the *F*0 trajectory as it has been found that *F*0 often settles on a stable level at later moments in the vowel occurring after a voiceless plosive consonant, possibly due to vocal fold pre-stiffening (Hanson, 2009; Löfqvist *et al.*, 1989). However, we also explored other markers such as the F0 midpoint to get a measurement of more stable F0 levels reached at later portions in the vowel, and the effects remain the same. Thus, it seems that F0 is largely unaffected by movement type and phasing condition.[2]

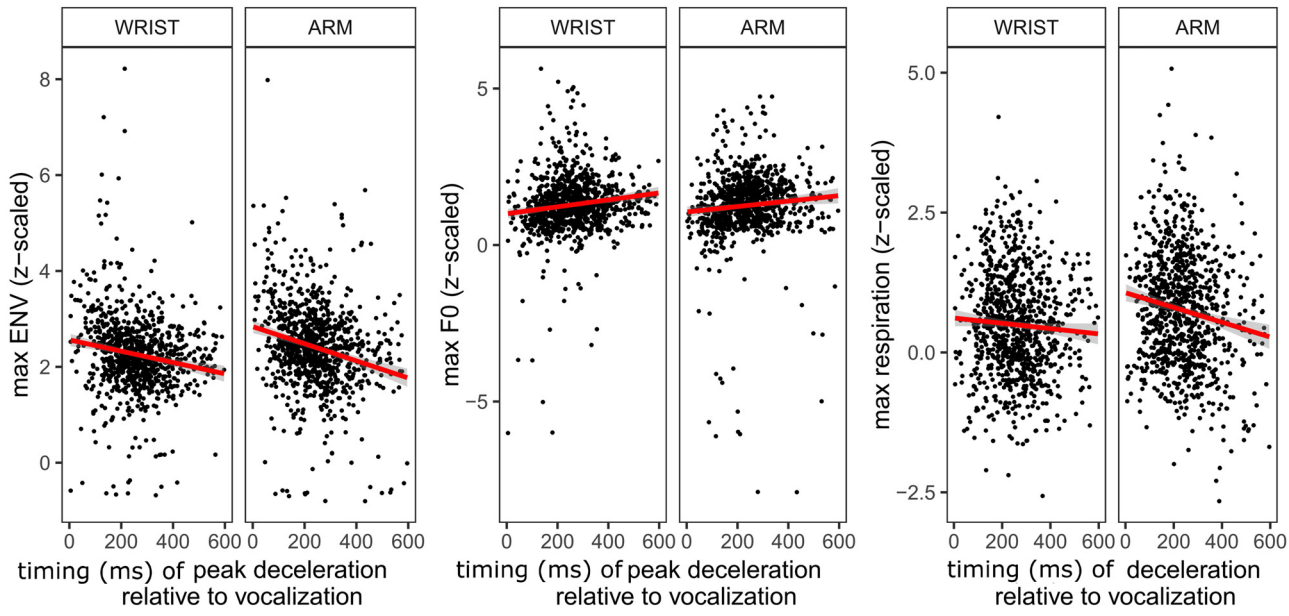


FIG. 6. (Color online) For each acoustic parameter and movement type condition, the *z*-scaled maxima are plotted on the *y* axis and the absolute deviance in milliseconds of the peak in deceleration (occurring at ms = 0) and the midpoint *F*0. It can be seen that, especially for the arm movement condition (but also wrist motion), there is a population level negative trend for amplitude envelope and respiration-related movement, such that the further temporally removed the vocalization was from the peak in deceleration the lower the height of the peak.

https://doi.org/10.1121/10.0001730

TABLE V. Physical impetus timing predicts height of acoustic peaks.

Max ENV (*z*-scaled)	***b***	***t*(1798)**	***p***	**Effect size (*d*)**
Intercept	2.62	33.83	<0.001	
Movement: Wrist vs arm	−0.13	−3.49	<0.001	−0.164
Timing peak dec	−0.0009	−4.33	<0.001	0.203
Max *F*0 (*z*-scaled)	***b***	***t*(1798)**	***p***	**Effect size (*d*)**
Intercept	1.13	11.56	<0.001	
Movement: Wrist vs arm	0.024	5.66	0.620	0.023
Timing peak dec	0.0005	1.83	0.067	0.086
Respiration (*z*-scaled)	***b***	***t*(1798)**	***p***	***Effect size* (*d*)**
Intercept	1.03	11.10	<0.001	
Movement: Wrist vs arm	−0.23	−5.55	<0.001	−0.256
Timing peak dec	−0.0012	−5.08	<0.001	−0.239

D. The role of the temporal distance between the peak of the physical impulse and the vocalization

Having ascertained that the in-phase condition leads to higher peaked amplitude envelope and respiration-related movement, our third investigation will look into this particular phasing condition to see whether we can directly confirm that the temporal co-occurrence of physical impulse with vocalization is crucial for affecting acoustic parameters. We assess for each vocalization how far the midpoint was temporally removed (in milliseconds, and in absolute terms) from the nearest peak in deceleration of the upper limb movement (proxy for peak physical impulse).

Figure 6 shows these correlations, and Table V provides the mixed regression analysis. We find that next to the effect of movement type, the closer the vocalization is to the peak deceleration, the higher the peak in the amplitude envelope and the peak in respiration-related activity. Again, such relations are not found for the *F*0.

E. Relation between respiration and acoustic peaks

It is still possible that the respiration-related movement associations with upper limb movements are an artefact of movement, without playing a role in modulating vocal activity. However, our exploratory analysis shows (see Fig. 7) that even in the passive condition, there is a positive relation between the maximum amplitude envelope and respiration-related movement, $b = 0.20$, b 95% CI [0.15, 0.25], t

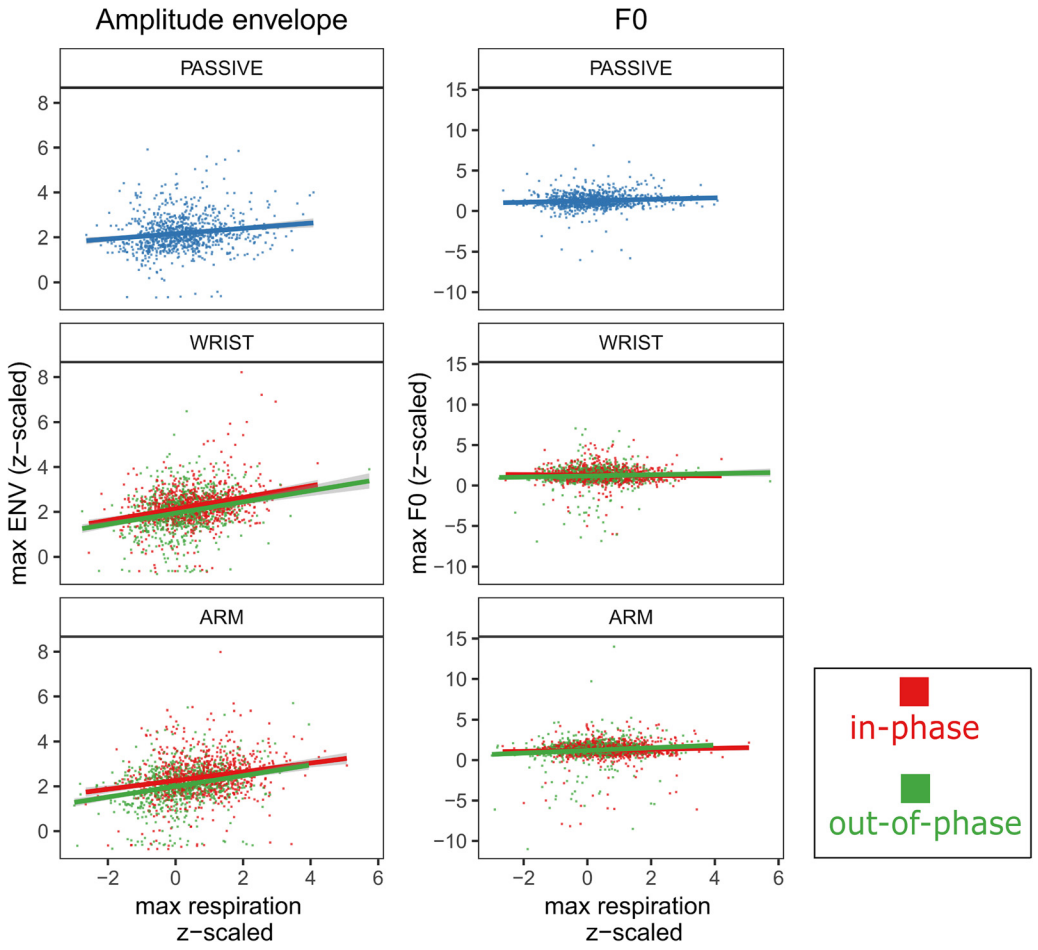


FIG. 7. (Color online) For the acoustic parameters of the vocalization (amplitude envelope and *F*0), we plot on the *y* axis the peak height of each of the vocalizations made against the peak height of the chest-respiratory movement. It can be seen that there are consistent positive relations between respiration-related movement and the amplitude envelope, such that higher amplitude vocalizations occurred more often with higher respiration-related movement. This again is not immediately apparent for *F*0, but participant-level mixed regression analysis did obtain a small effect for a positive respiration-*F*0 scaling for the movement conditions.

(894) = 7.82, $p < 0.001$, Cohen's $d = 0.523$, such that higher vocalization peaks go together with more respiration-related activity. This association of respiration and acoustics for the passive condition was not reliable for *F*0 at our predefined threshold, $b = 0.08$, b 95% CI [0.015, 0.14], t (924) = 2.40, $p = 0.016$, Cohen's $d = 0.160$. The relation between respiration and amplitude envelope was reliably more extreme for the movement conditions, $b = 0.27$, b 95% CI [0.25, 30], t (3630) = 7.82, $p < 0.001$, Cohen's $d = 0.666$, with no moderating role for this respiration-amplitude correlation for movement type or phasing of movement. For the movement conditions, a similar correlation—but of much smaller magnitude—was obtained for *F*0 and respiration, $b = 0.09$, b 95% CI [0.06, 0.13], t (3630) = 4.77, $p < 0.001$, Cohen's $d = 0.155$, suggesting that *F*0 is not entirely immune to movement-related effects via respiration.

For our confirmatory analysis, we also probe the relation between respiration and movement in the in-phase condition alone, so as to predict acoustic peaks based on the degree of frequency coupling between upper-limb movement and respiration-related movement. We do this with cross-wavelet analysis by quantifying the degree to which respiration-related movement periodicities at the 1-s interval range was correlated with the movement's periodicities. The correlation between the periodic structure of two signals (respiration and movement time series) is expressed with a "coherence" coefficient ranging from 0 to 1 (no correlation to perfect correlation). This analysis is reported in the Appendix.[2] Our analyses were fully inconclusive, with all correlations between respiration-movement coherence and acoustic peaks statistically unreliable. We suspect that this lack of a relation was caused by a lack of variability between respiration-movement coupling, which showed on average very high periodicity correlations, *M* coherence = 0.92 (SD = 0.08), with a heavy tailed (non-normal) distribution tending towards perfect correlation of coherence = 1.00 (see Fig. 8). Thus, movement and respiration-related activity was very highly correlated according to these analyses, but these coherence coefficients failed to be predictive for acoustic peaks (*F*0 or Amplitude Envelope), possibly due to said artefacts of non-normality and lack of variability.

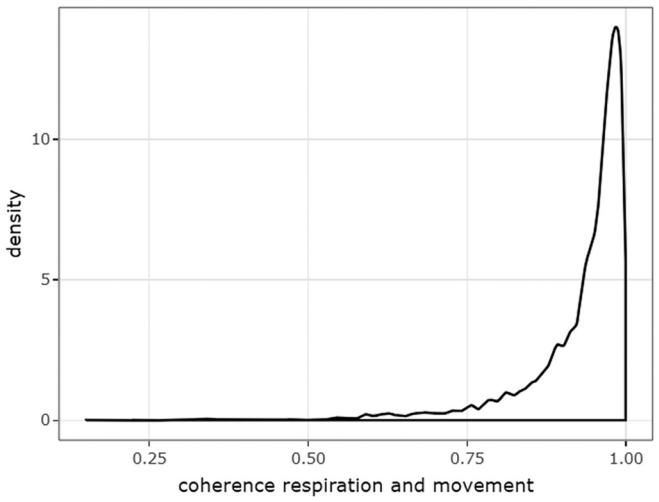


FIG. 8. Density distribution of the coherence estimates for each vocalization. It can be seen that very high coherences are observed for respiration-related movement and upper limb movement, bordering perfect correlations of coherence = 1.

However, note that we find in our previous confirmatory analysis clear effects of movement type and phasing on respiration-related movement, which is now also confirmed by the above coherence analysis. Further, we find in our exploratory analysis that respiration-related movement and the amplitude envelope peaks are strongly correlated.

IV. DISCUSSION

When moving the upper limbs so that the physical impulse of that movement co-occurs near the vocalization (in-phase condition), higher amplitude envelope and more respiration-related activity are found as compared to a passive (no-movement) condition, or as compared 90° out-of-phase movement condition. In all these cases, the heightened acoustic and respiration-related activity effects of the in-phase condition were more pronounced for arm movements as compared to wrist movements, replicating earlier work on more extreme physical impulses on vocalization acoustics when mass of the effector in motion is higher (Pouw *et al.*, 2019a; Pouw *et al.*, 2019c, Pouw *et al.*, 2020b). There is a crucial role of physical impulse in the in-phase condition as vocalizations that occurred closer to the peak in deceleration had higher amplitude and concomitant respiration-related activity as compared to vocalizations that were further temporally removed from the peak physical impulse. Additionally, respiration-related movement is not merely artefactual to upper limb movement but also reliably predicts height of the amplitude envelope peaks when there is no upper limb movement and is more closely related to acoustic peaks when co-occurring with movement, and this was observed for *F*0 as well. However, for all analyses, *F*0 of CV utterances were much less, if at all, affected by physical impulses in this paradigm.

n important reason that might explain why *F*0 is not similarly affected by a gesture-induced physical impetus in the current /pa/ utterances is that the vocal cords can counteract any energy bump that is imparted by an upper limb movement in the maintenance of a steady *F*0 output. Thus, while the amplitude of speech is *necessarily* affected by sudden gesture-induced changes in lung pressure (as this energy needs to be dissipated somehow), changes in *F*0 *need not materialize necessarily* as *F*0 is in most part controlled by the vocal cords.

Another likely important factor for whether upper limb movement affects *F*0 is the prosodic-acoustic target of the vocalization. For example, in a study on pointing, it was found that regardless of whether the vocalization started with a voiceless plosive consonant preceding the vowel of an *accented* syllable, positive *F*0 excursions were found at the moment of the abrupt stop of the pointing movement (i.e., the deceleration of the gesture referred to as the "apex" of the gesture). It is

https://doi.org/10.1121/10.0001730

possible that such excursions are thus either not produced through gesture-speech physics at all, as they are produced via other sensorimotor solutions (Perrier and Fuchs, 2015), or alternatively, gestures' biomechanical effects are allowed to arise as they are congruent with the prosodic-acoustic target of accenting the syllable. Indeed, it is very important to emphasize that in the current task, participants are *deliberately asked to give a stable and monotonic vocal output* regardless of movement and phasing conditions.

The current findings make for a more compelling case that gesture-speech synchrony may be grounded in biomechanical linkages between upper limb movement and the respiratory system. Specifically, in the current case, a clear acoustic marker of speech rhythm (Chandrasekaran *et al.*, 2009; Tilsen and Arvaniti, 2013), the amplitude envelope, is affected by concurrent upper limb movement. As such, gesture-speech entrainment might not be an arbitrary cognitive invention. Rather it is plausible that gesture-speech synchrony occurs because it allows for functional biomechanical linkages that constrain speech production. The non-arbitrariness of vocal productions as it relates to bodily properties dovetails with observations that non-human mammals' vocalizations contain indexical cues of body size (Pisanski *et al.*, 2016). Orangutans even modulate their vocalization acoustics by cupping the hands in front of the mouth, which affects their spectral center of gravity, supposedly so as to be perceived by predators as more threatening in size (Hardus *et al.*, 2009; see also de Boer *et al.*, 2015). As such, much like other speech properties that may have arisen out of embodied constraints (Blasi *et al.*, 2019; Dediu *et al.*, 2019; Ćwiek and Fuchs, 2019; see Fuchs, 2019, for a discussion), gesture-speech synchrony may have arisen phylogenetically as an evolutionary adaptation to respiratory-vocal control.

The current research builds on a host of research showing tight, but flexible, coupling of the gesture and speech system (Chu and Hagoort, 2014; Danner *et al.*, 2018; Esteve-Gibert and Prieto, 2013; Fuchs *et al.*, 2015; Kelso *et al.*, 1983; Krivokapic *et al.*, 2016; Loehr, 2012; McClave, 1998; McNeill, 1992; Parrell *et al.*, 2014; Pouw and Dixon, 2019b,a; Rochet-Capellan *et al.*, 2008; Shattuck-Hufnagel and Prieto, 2019; Treffner and Peter, 2002; Zelic *et al.*, 2015). What the current research adds is that it identifies a biomechanical route for manual gestures' imprint on acoustics. This biomechanical linkage may also offer a more powerful explanation for recent research in machine learning. In this research, it is shown that a person's gestures can be near-perfectly synthesized by a deep neural network (DNN) based on novel input of speech. Here, a DNN is pre-trained on the association between gestural movement and speech acoustics (Alexanderson *et al.*, 2020; Ginosar *et al.*, 2019; Kucherenko *et al.*, 2019). That such DNNs can reconstruct gesture kinematics based on speech acoustics alone suggests that there must be a tight link between acoustics and manual movement that are still to be fully appreciated and understood by gesture-speech researchers. However, we acknowledge that current research provides no evidence that current effects can generalize to fluent speech (see, however, Pouw *et al.*, 2020a).

A. Conclusion and future directions

While it has been known for a long time that gestures beat with the rhythmic aspects of speech (e.g., Efron *et al.*, 1972), an explanation for this phenomenon that works on multiple levels of analysis has been absent (Wagner *et al.*, 2014). If upper limb movements are implicated in respiratory control, we can postulate that gestures might have evolved in our species to gain vocal control. We can further postulate that motor and vocal babbling in infants can invoke a discovery of vocal control through gesture. When upper limb movements can help reach prosodic targets, we can postulate that a purely neural timing mechanism invoked by psychologists to explain gesture-speech timing can be discarded, and rather instantiated in a wider biophysical control system (i.e., a tensegrity system). Thus, a biomechanical linkage between upper limb movement and vocalization holds promise for a unification of phylogenetic, ontogenetic, and cognitive explanations of gesture-speech coordination. The challenge for fulfilling this promise lies in devising experiments that can probe the implications of gesture-speech physics at these different levels of analysis.

Phylogenetically, cross-species research could help in an understanding of how often respiratory-vocal systems form coalitions with accessory control systems that evolved for initially different purposes (e.g., Lancaster *et al.*, 1995). For ontogeny, research could focus on the role of physical impulses during vocal-motor babbling in infants (e.g., Ejiri, 1998). As for an extension of the current line of work, much more is needed to understand how gesture-speech physics plays out in more naturalistic speech (Cravotta *et al.*, 2019; Pouw *et al.*, 2020a). In gesture studies, identification of possible different functions that gestures may take in discourse can be improved by taking into account whether such gestures impart a physical impulse on the body (Cooperrider, 2017). Finally, a complete understanding of gesture-speech physics requires the identification of how different limb and head movements affect expiratory flow during speech by soliciting different gradients of force transmissions in the tensioned body, opening up an avenue of research in biomechanics with said multidisciplinary implications.

ACKNOWLEDGMENTS

This research has been funded by The Netherlands Organisation of Scientific Research (NWO; Rubicon Grant "Acting on Enacted Kinematics," Grant No. 446–16-012; PI W. Pouw).

APPENDIX

1. Results movement-respiratory coherence and acoustic peaks analysis

For our final confirmatory analysis, we wanted to further investigate the relationship between movement and respiration and how this relates to acoustics. Specifically, we hypothesized that a higher coupling between movement of the upper limb

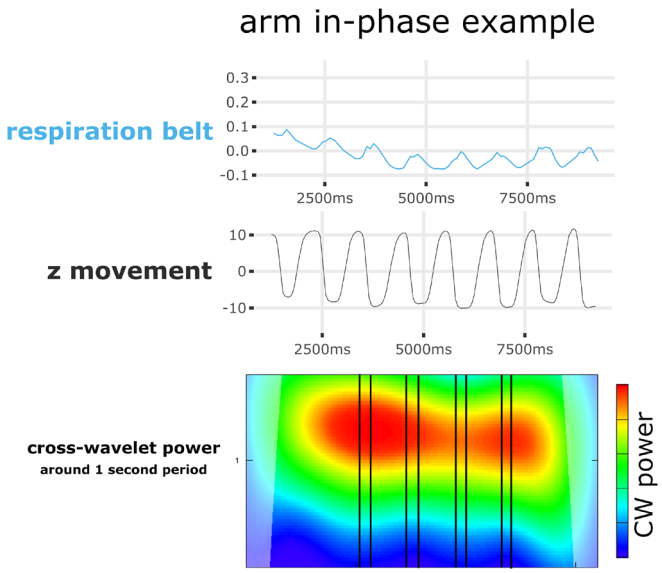


FIG. 9. (Color online) For the cross-wavelet analysis we assessed for each trial the correlation (i.e., coherence) between the periodicity of respiration-related movement (see respiration belt time series in blue) and the periodicity of movement (see vertical z movement time series in black) at the moments that there was vowel vocalization. The shared periodicities are detected by cross-wavelet analysis and a cross-wavelet plot shown indicating a shared periodicity around 1 s. Because coherence estimates are less reliable at the start and tail of the timeseries, we only extracted four maximum coherence estimates per trial for the middle vocalizations (as indicated by the black bars). These coherence estimates could then later be related to acoustic peaks.

with that of respiration-related movement would lead to higher acoustic peaks ($F0$ and amplitude envelope) for the in-phase condition—we ignore the other conditions in this analysis as we have already ascertained that the passive and 90° out-of-phase condition did not lead to higher acoustic peaks. We reasoned that since the movement periods were about 1 s for all participants, we could, therefore, detect in the respiration belt signal similar 1-s periods and see if they were correlated with the movement periods.

To assess the correlation (i.e., coherence) between movement and respiration time series we performed cross-wavelet analysis with R-package WaveletComp (Rosch and Schmidbauer, 2014). Cross-wavelet analysis has the advantage of assessing coherence between two time series in a time-dependent way, such that we can estimate the coherence at a particular point in time, e.g., at the moment that the participants were vocalizing. For the cross-wavelet analysis, we set the period range at 0.8–1.2 s, thereby quantifying the strength of periodicity coupling within these intervals. We entered vertical movement displacement (z) of the upper limb and respiration belt time series into this analysis so as to produce time-dependent coherence estimates. Then, for each vocalization, we determined the maximum observed coherence at the midpoint of the vocalization (see Figs. 9 and 10 for more information).

Subsequently, similar to the previous analysis, we performed mixed linear regressions whereby we assess whether movement-respiratory coherence predicted acoustic peaks

TABLE VI. GAM parametric coefficients per vocalization parameter.

Amplitude envelope	***b***	***t***	***p***
Intercept (passive)	1.37	340.90	<0.001
Wrist in-phase vs passive	0.02	3.87	<0.001
Arm in-phase vs passive	0.10	25.56	<0.001
Wrist out-of-phase vs passive	−0.10	−25.12	<0.001
Arm out-of-phase vs passive	−0.06	−14.83	<0.001
Max *F0* (*z*-scaled)	***b***	***t***	***p***
Intercept (passive)	−0.03	−7.74	<0.001
Wrist in-phase vs passive	0.02	2.56	0.011
Arm in-phase vs passive	0.06	10.80	<0.001
Wrist out-of-phase vs passive	0.03	6.44	<0.001
Arm out-of-phase vs passive	0.06	11.49	<0.001
Respiration	***b***	***t***	***p***
Intercept (passive)	0.004	0.72	0.467
Wrist in-phase vs passive	0.18	32.46	<0.001
Arm in-phase vs passive	0.24	42.48	<0.001
Wrist out-of-phase vs passive	0.01	1.93	0.053
Arm out-of-phase vs passive	−0.08	−14.09	<0.001

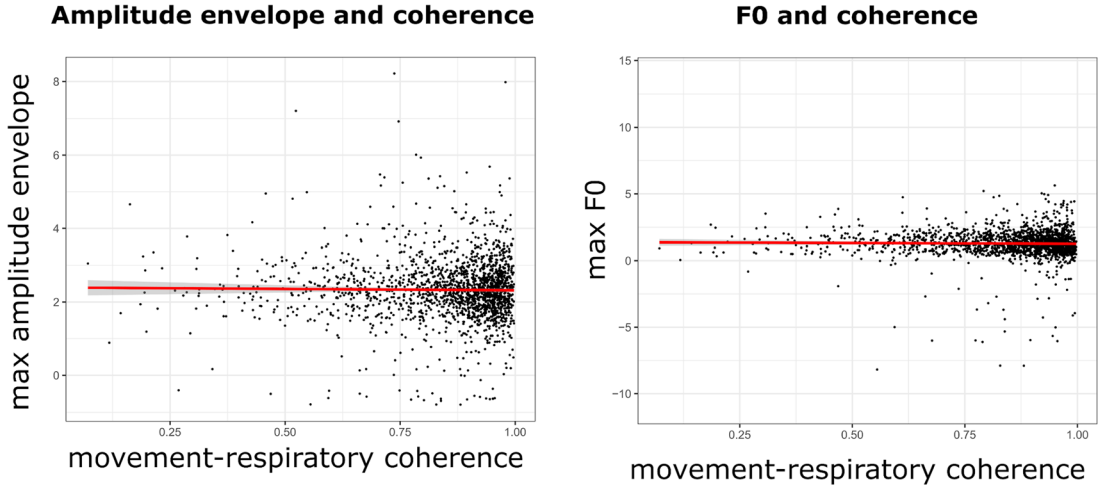


FIG. 10. (Color online) Movement-respiratory coherence estimates and height peaks vocalization.

https://doi.org/10.1121/10.0001730

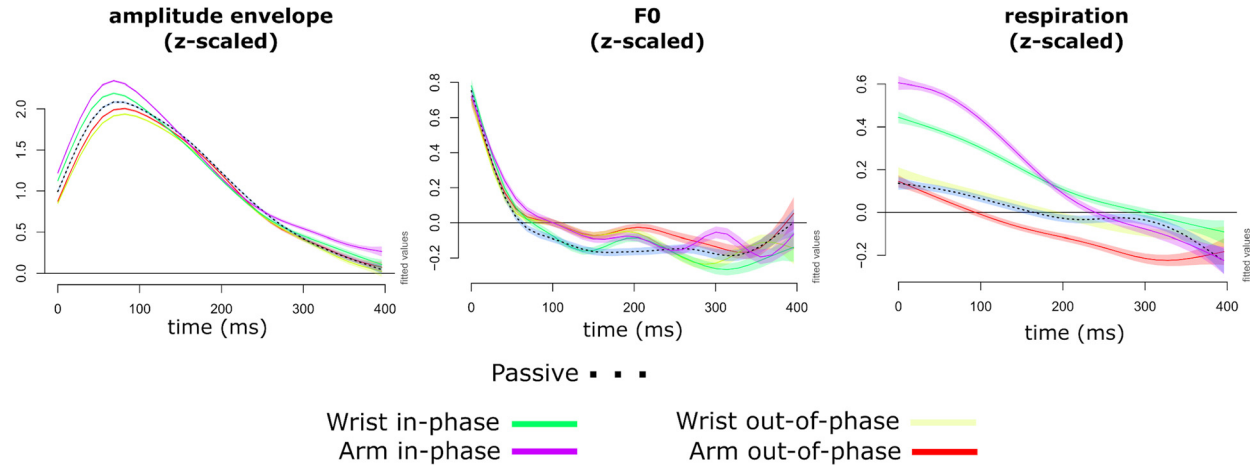


FIG. 11. (Color online) The fitted trajectories are shown for the different vocalization parameters. The shaded areas indicate 95% CIs. It can be seen that relative height of the trajectories replicate our main findings, such that the arm in-phase movement led to highest peaks in amplitude envelope and respiration-related movement, followed by the wrist-in phase, passive, arm 90° out-of-phase, and wrist 90° out-of-phase. For the *F*0 trajectories, we can see that especially around 100 milliseconds into phonation there are clear deviances from the passive control condition with heightened trajectories for high-impetus arm-movements (regardless of phasing).

(see Figure 10 for scatter plots). Coherence was not predictive for maximum amplitude envelope, $b = -0.39$, $p = 0.071$. Movement-respiratory coherence also was not predictive for maximum *F*0, $b = -0.20$, $p = 0.496$.

2. Generalized additive modeling (GAM) of vocalization events

GAM is a method for non-linear multilevel regression of a time series (Arnold *et al.*, 2013; Wieling, 2018; Wood, 2017). Thus, instead of modeling a single observation-point representative of some aspect of the trajectory (e.g., maximum) as we did in the confirmatory analysis, we can also model the non-linear trajectory in its entirety. GAM uses a set of smooth base functions to model non-linear relations between predictors (usually time) and some dependent variable. We performed GAM by adopting a maximum likelihood method, using R packages gam (Hastie and Hastie, 2008) and mgcv (Wood, 2017) for plotting fitted trajectories. Similar to our confirmatory analyses, we use the participant as a random intercept for all the modeling. As factorial, we submitted pooled phase and movement condition (Passive, Wrist-in phase, Arm in-phase, Wrist 90° out-of-phase, Arm 90° out-of-phase) so as to differentiate trajectories against the Passive control condition. We used the start of the vowel vocalization up to 400 ms as the time interval that will be modeled for each vocalization event. This cutoff of 400 was chosen as we know that only a few participants phonated longer than this and therefore modeling estimates become unreliable at the tails of the trajectories. We performed three separate GAMs to assess the different trajectories over time per condition for the amplitude envelope, *F*0, and the respiration belt data.

Table VI provides the GAM results for the parametric coefficients, given for each condition comparison. These coefficients indicate whether the non-linear trajectory is relatively shifted upwards (positive coefficient) or downwards (negative coefficient) relative to the intercept (i.e., passive control condition). Note that in all analysis, the non-linear smooths were significant too, as indeed the trajectories are not linearly related to phonation time (e.g., amplitude envelope rises and falls over time).

The fitted trajectories by GAM are plotted in Fig. 11. The amplitude envelope and respiration results fully confirm our analysis from the main results. Namely, higher amplitude envelope and respiration-related movement are obtained when participants are moving in-phase, and this is more extreme for higher impetus arm movements. For the amplitude envelope, you also see clearly detrimental effects of 90° out-of-phase movement, leading to lower amplitude as compared to the passive condition.

Note that for *F*0, we do now find significant differences of movement vs passive condition, with more extreme effects for when participants were moving with the arm vs wrist. The phasing conditions showed similar coefficients. Thus, we obtain some further exploratory evidence that there are subtle changes in *F*0 trajectories when moving vs not moving with higher physical impulse.

[1]We chose a sitting position, as opposed to standing position, to minimize differences in posture between trials. This does reduce possible exaggerated effects of gesture on speech via anticipatory muscle activations which we have observed in previous research (Pouw, *et al.*, 2019). We also chose for lower-impetus unilateral vs higher-impetus bilateral movements, so as to ensure that we could accurately determine timing relations between vocalization and movement of one hand rather than having to account for two timing relations. We reason that if we still find similar gesture effects for sitting postures for unilateral movements, we have a more solid basis for these effects being generalizable to other contexts.

[2]However, we also have performed more powerful exploratory non-linear regression analyses using GAM, whereby we modeled the entire non-linear trajectory of the acoustic and respiration-related movement. These exploratory results are reported in the Appendix, and while they fully replicate our main current confirmatory conclusions about the amplitude envelope and respiration-related movement, we also obtained that movement (but not phasing) now does lead to higher *F*0 trajectories as opposed to the passive condition, with more pronounced effects of high-impetus

arm vs low impetus- movement. Nevertheless, these effects seem very subtle, and we should be careful with equating such effects with the magnitudes of the effects for the amplitude envelope and the respiration-related movement.

Alexanderson, S., Henter, G. E., Kucherenko, T., and Beskow, J. (**2020**). "Style-controllable speech-driven gesture synthesis using normalising flows," Comput. Graphics Forum **39**(2), 487–496.

Arnold, D., Wagner, P., and Baayen, R. H. (**2013**). "Using generalized additive models and random forests to model prosodic prominence in German," in *Proceedings of Interspeech 2013*, August 25–29, Lyon, France, pp. 272–276.

Aruin, A. S., and Latash, M. L. (**1995**). "Directional specificity of postural muscles in feed-forward postural reactions during fast voluntary arm movements," Exp. Brain Res. **103**, 323–332.

Baer, T. (**1979**). "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes," J. Acoust. Soc. Am. **65**, 1271–1275.

Basmajian, J. V., and De Luca, C. J. (**1985**). *Muscles Alive: Their Functions Revealed By Electromyography*, 5th ed. (Williams and Wilkins, Baltimore, MD).

Bernstein, N. (**1966**). *The Co-Ordination and Regulation of Movements* (Pergamon Press, London, UK).

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), eaav3218.

Bombien, L., Winkelmann, R., and Scheffers, M. (**2020**). "wrassp: An R wrapper to the ASSP Library," R package version 0.1.9, https://cran.r-project.org/web/packages/wrassp/index.html (Last viewed August 18, 2020).

Bouisset, S., and Do, M. C. (**2008**). "Posture, dynamic stability, and voluntary movement," Clin. Neurophysiol. **38**, 345–362.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (**2009**). "The natural statistics of audiovisual speech," PLoS Comput. Biol. **5**(7), e1000436.

Chang, P., and Hammond, G. R. (**1987**). "Mutual interactions between speech and finger movements," J. Motor Behav. **19**(2), 265–274.

Chu, M., and Hagoort, P. (**2014**). "Synchronization of speech and gesture: Evidence for interaction in action," J. Exp. Psychol. General **143**(3), 1726–1741.

Cooperrider, K. (**2017**). "Foreground gesture, background gesture," Gesture **16**(2), 176–202.

Cordo, P. J., and Nashner, L. M. (**1982**). "Properties of postural adjustments associated with rapid arm movements," J. Neurophysiol. **47**(2), 287–302.

Cravotta, A., Busà, M. G., and Prieto, P. (**2019**). "Effects of encouraging the use of gestures on speech," J. Speech Lang. Hear. Res. **62**, 3204–3219.

Ćwiek, A., and Fuchs S. (**2019**). "Iconic prosody is rooted in sensori-motor properties: Fundamental frequency and the vertical space," in *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, July 24–27, Montreal, Canada, pp. 1572–1578.

Danner, S. G., Barbosa, A. V., and Goldstein, L. (**2018**). "Quantitative analysis of multimodal speech data," J. Phon. **71**, 268–283.

Dediu, D., Janssen, R., and Moisik, S. R. (**2019**). "Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels," Nat. Hum. Behav. **3**, 1107.

de Boer, B., Wich, S. A., Hardus, M. E., and Lameira, A. R. (**2015**). "Acoustic models of orangutan hand-assisted alarm calls," J. Exp. Biol. **218**(6), 907–914.

Dimensions.Guide (2019). "Male (side) dimensions & drawings," https://www.dimensions.com/element/sitting-male-side-1 (Last viewed May 1, 2019).

Efron, D., Efron, J. M., and 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, the Hague, the Netherlands).

Ejiri, K. (**1998**). "Relationship between rhythmic behavior and canonical babbling in infant vocal development," Phonetica **55**(4), 226–237.

Esteve-Gibert, N., and Prieto, P. (**2013**). "Prosodic structure shapes the temporal realization of intonation and manual gesture movements," J. Speech Lang. Hear. Res. **56**(3), 850–864.

Finnegan, E. M., Luschei, E. S., and Hoffman, H. T. (**2000**). "Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech," J. Speech Lang. Hear. Res. **43**(4), 934–950.

Fuchs, S. (**2019**). "Vocal tract variations affect vowel sounds," Nat. Hum. Behav. **3**, 1043.

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, W. D., and Koenig, L. L. (**2015**). "Assessing respiratory contributions to F0 declination in German across varying speech tasks and respiratory demands," J. Phon. **52**, 35–45.

Gibson, J. (**1966**). *The Senses Considered as Perceptual Systems* (Houghton–Mifflin, Boston, MA).

Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., and Malik, J. (**2019**). "Learning individual styles of conversational gesture," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 16–20, Long Beach, CA, pp. 3497–3506.

Hanson, H. M. (**2009**). "Effects of obstruent consonants on fundamental frequency at vowel onset in English," J. Acoust. Soc. Am. **125**(1), 425–441.

Hardus, M. E., Lameira, A. R., Schaik, C. S., and 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.

Hastie, T., and Hastie, M. T. (**2018**). "Package 'GAM,'" GAM Package CRAN, https://cran.r-project.org/web/packages/gam/ (Last viewed August 18, 2020).

He, L., and Dellwo, V. (**2017**). "Amplitude envelope kinematics of speech: Parameter extraction and applications," J. Acoust. Soc. Am. **141**(5), 3582.

Hodges, P. W., and 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.

Hübscher, I., and Prieto, P. (**2019**). "Gestural and prosodic development act as sister systems and jointly pave the way for children's sociopragmatic development," Front. Psychol. **10**, 1259.

Ingber, D. W. (**2008**). "Tensegrity and mechanotransduction," J. Bodywork Move. Ther. **12**(3), 198–200.

Iverson, J. M., and Thelen, E. (**2005**). "Hand, mouth and brain: The dynamic emergence of speech and gesture," J. Conscious. Studies **22**, 19–40.

Kelso, S., and Tuller, B. (**1984**). "Converging evidence in support of common dynamical principles for speech and movement coordination," Am. J. Physiol. **246**, 928–935.

Kelso, J. A. S., Tuller, B., and Harris, K. (**1983**). "A 'dynamic Pattern' perspective on the control and coordination of movement," in *The Production of Speech*, edited by P. F. McNeilage (Springer, New York), pp. 137–173.

Kleiman, E. M. (**2017**). "EMAtools: Data management tools for real-time monitoring/ecological momentary assessment data," https://CRAN.R-project.org/package=EMAtools (Last viewed August 18, 2020).

Krahmer, E., and 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.

Krivokapić, J. (**2014**). "Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes," Philos. Trans. R. Soc. B **369**(1658), 20130397.

Krivokapic, J., Tiede, M. K., Tyrone, M. E., and Goldenberg, D. (**2016**). "Speech and manual gesture coordination in a pointing task," in *Proceedings of Speech Prosody 2016*, May 31–June 3, Boston, MA.

Kucherenko, T., Hasegawa, D., Henter, G. E., Kaneko, N., and Kjellström, H. (**2019**). "Analyzing input and output representations for speech-driven gesture generation," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents—IVA '19*, July 2–5, Paris, France, pp. 97–104.

Kugler, P. N., and Turvey, M. T. (**1987**). *Information, Natural Law, and the Self-Assembly of Rhythmic Movement* (L. Erlbaum Associates, Hillsdale, NJ).

Ladefagod, P. (**1968**). "Linguistic aspects of respiratory phenomena," in *Sound Production in Man*, edited by A. Bouhuys (New York Academy of Sciences, New York), pp. 141–151.

Lancaster, W. C., Henson, O. W., and Keating, A. W. (**1995**). "Respiratory muscle activity in relation to vocalization in flying bats," J. Exp. Biol. **198**(1), 175–191.

Lenth, R., and Lenth, M. R. (**2017**). "Package 'lsmeans,'" Am Stat. **34**(4), 216–221.

Leonard, T., and Cummins, F. (**2011**). "The temporal relation between beat gestures and speech," Lang. Cogn. Process. **26**(10), 1457–1471.

Levin, S. M. (**2006**). "Tensegrity: The new biomechanics," in *Textbook of Muscularskeletal Medicine*, edited by M. Hutson and R. Ellis (Oxford University Press, Oxford, UK), pp. 69–80.

Lieberman, P. (**1996**). "Some biological constraints on the analysis of prosody," in *Signal to Syntax*, edited by J. L. Morgan and K. Demuth (Erlbaum, Mahwah, NJ), pp. 67–78.
Loehr, D. P. (**2012**). "Temporal, structural, and pragmatic synchrony between intonation and gesture," Lab. Phonol. **3**(1), 71–89.
Löfqvist, A., Baer, T., McGarr, N. S., and Seider Story, R. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.
MacLarnon, A. M., and 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.
McClave, E. (**1998**). "Pitch and manual gestures," J. Psycholing. Res. **27**(2), 69–89.
McNeill, D. (**1992**). *Hand and Mind: What Gestures Reveal About Thought* (University of Chicago Press, Chicago, IL).
McNeill, D. (**2005**). *Gesture and Thought* (University of Chicago Press, Chicago, IL).
Ohala, J. J. (**1990**). "Respiratory activity in speech," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht, the Netherlands), pp. 22–53.
Parrell, B., Goldstein, L., Lee, S., and Byrd, D. (**2014**). "Spatiotemporal coupling between speech and manual motor actions," J. Phon. **42**, 1–11.
Perrier, P., and Fuchs, S. (**2015**). "Motor equivalence in speech production," in *The Handbook of Speech Production*, edited by M. A. Redford (Wiley, New York), pp. 223–247.
Petrone, C., Fuchs, S., and 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.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Team (**2019**). "nlme: Linear and nonlinear mixed effects models," https://cran.r-project.org/web/packages/nlme/index.html (Last viewed August 18, 2020).
Pisanski, K., Cartei, V., McGettigan, C., Raine, J., and Reby, D. (**2016**). "Voice modulation: Window into the origins of human vocal control?," Trends Cogn. Sci. **20**(4), 304–318.
Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., and Dixon, J. A. (**2020a**). "Gesture-speech physics in fluent speech and rhythmic upper limb movements," Psyarxiv, https://doi.org/10.31234/osf.io/359te.
Pouw, W., and Dixon, J. A. (**2019a**). "Entrainment and modulation of gesture—Speech synchrony under delayed auditory feedback," Cogn. Sci. **43**(3), e12721.
Pouw, W., and Dixon, J. A. (**2019b**). "Quantifying gesture-speech synchrony," in *Proceedings of the 6th Meeting of Gesture and Speech in Interaction*, September 11–13, Paderborn, Germany.
Pouw, W., Harrison, S. H., and Dixon, J. (**2019a**). "Gesture-speech physics: The biomechanical basis of the emergence of gesture-speech synchrony," J. Exp. Psychol. General **149**, 391–404.
Pouw, W., Harrison, S. J., Esteve-Gibert, N., and Dixon, J. A. (**2019b**). "Energy Flows in Gesture-Speech Physics: Exploratory Findings and Pre-Registration of Confirmatory Analysis," *Open Science Framework*, https://doi.org/10.17605/OSF.IO/E3UKD and https://osf.io/x7zdc/ (Last viewed August 18, 2020).
Pouw, W., Paxton, A., Harrison, S. J., and Dixon, J. A. (**2019c**). "Acoustic specification of upper limb movement in voicing," in *Proceedings of the 6th Meeting of Gesture and Speech in Interaction*, September 11–13, Paderborn, Germany, available at https://doi.org/10.17619/UNIPB/1-812.
Pouw, W., Paxton, A., Harrison, S. J., and Dixon, J. A. (**2020b**). "Multimodal origins of the human voice: Acoustic information about upper limb movement in voicing," Proc. Natl. Acad. Sci. **117**(21), 11364–11367.
Pouw, W., and Trujillo, J. P. (**2019**). "Tutorial Gespin2019—Using video-based motion tracking to quantify speech-gesture synchrony," https://doi.org/10.17605/OSF.IO/RXB8J (Last viewed August 18, 2020).
Pouw, W., Trujillo, J. P., and Dixon, J. A. (**2020c**). "The quantification of gesture-speech synchrony: An overview and validation of video-based motion tracking," Behav. Res. Methods **52**, 723–740.
Profeta, V. L., and Turvey, M. T. (**2018**). "Bernstein's levels of movement construction: A contemporary perspective," Hum. Move. Sci. **57**, 111–133.
Raja, V. (**2020**). "Resonance and radical embodiment," Synthese, https://doi.org/10.1007/s11229-020-02610-6.
Richardson, M. (**2009**). "Polhemus applications and example code," http://xkiwilabs.com/software-toolboxes/ (November 1, 2018).
Rochet-Capellan, A., and Fuchs, S. (**2014**). "Take a breath and take the turn: How breathing meets turns in spontaneous dialogue," Philos. Trans. R. Soc. B: Biol. Sci. **369**(1658), 20130399.
Rochet-Capellan, A., Laboissière, R., Galván, A., and 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.
Rosch, A., and Schmidbauer, H. (**2014**). "WaveletComp 1.1: A guided tour through the R package 59," http://www.hs-stat.com/projects/WaveletComp/WaveletComp_guided_tour.pdf (Last viewed August 18, 2020).
Rusiewicz, H. L., and Esteve-Gibert, N. (**2018**). "Set in time: Temporal coordination of prosody and gesture in the development of spoken language production," in *The Development of Prosody in First Language Acquisition*, edited by P. Prieto and N. Esteve-Gibert (John Benjamins Publishing Company, Amsterdam, the Netherlands), pp. 103–124.
Rusiewicz, H. L., Shaiman, S., Iverson, J., and Szuminsky, N. (**2013**). "Effects of prosody and position on the timing of deictic gestures," J. Speech Lang. Hear. Res. **56**(2), 458–470.
Shattuck-Hufnagel, S., and Prieto, P. (**2019**). "Dimensionalizing co-speech gestures," in *Proceedings of the International Congress of Phonetic Sciences 2019*, August 5–9, Melbourne, Australia.
Silva, P., Moreno, M., Mancini, M., Fonseca, S., and 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.
Smith, R., Nyquist-Battie, C., Clark, M., and Rains, J. (**2003**). "Anatomical characteristics of the upper serratus anterior: Cadaver dissection," J. Orthopaed. Sports Phys. Therapy **33**(8), 449–454.
Stoltmann, K., and Fuchs, S. (**2017**). "Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate," J. Multimodal Commun. Stud. **4**, 63–68.
Tilsen, S., and Arvaniti, A. (**2013**). "Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages," J. Acoust. Soc. Am. **134**(1), 628–639.
Treffner, P., and Peter, M. (**2002**). "Intentional and attentional dynamics of speech–hand coordination," Hum. Move. Sci. **21**(5–6), 641–697.
Turvey, M. T., and Fonseca, S. T. (**2014**). "The medium of haptic perception: A tensegrity hypothesis," J. Motor Behav. **46**(3), 143–187.
Wagner, P., Malisz, Z., and Kopp, S. (**2014**). "Gesture and speech in interaction: An overview," Speech Commun. **57**, 209–232.
Wieling, M. (**2018**). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," J. Phon. **70**, 86–116.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (**2006**). "ELAN: A professional framework for multimodality research," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, May 22–28, Genoa, Italy.
Wood, S. N. (**2017**). *Generalized Additive Models: An Introduction with R, Second Edition* (CRC Press, Boca Raton, FL).
Zelic, G., Kim, J., and Davis, C. (**2015**). "Articulatory constraints on spontaneous entrainment between speech and manual gesture," Hum. Move. Sci. **42**, 232–245.