

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/328293969>

Gesture–Speech Physics: The Biomechanical Basis of Gesture–Speech Synchrony

Preprint · October 2018

CITATIONS

0

READS

72

3 authors, including:



Wim Pouw

1. University of Connecticut 2.Erasmus University Rotterdam

23 PUBLICATIONS 233 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



The cognitive function of manual activity in learning and problem solving [View project](#)



The dynamics of gesture-speech synchrony [View project](#)

This is a preprint (version 1), the final version will differ.

Gesture-Speech Physics: The Biomechanical Basis of Gesture-Speech Synchrony

Wim Pouw^{1,2}, Steven J. Harrison¹, & James A. Dixon¹

Center for the Ecological Study of Perception and Action, University of Connecticut¹

Department of Psychology, Education, & Child Studies, Erasmus University Rotterdam²

Word count (in text): 7.671

Author note: Correspondence should be addressed to Wim Pouw (wimpouw@uconn.edu).

Note that parts of the current manuscript may overlap verbatim with the pre-registration.

Open data & Pre-registration: This study has been pre-registered on the Open Science Framework (OSF: <https://osf.io/5aydk/>) and the raw anonymized quantitative data, and analyses scripts supporting this confirmatory study are also available on the OSF.

Funding: This research has been funded by The Netherlands Organisation of Scientific Research (NWO; Rubicon grant “Acting on Enacted Kinematics”, Grant Nr. 446-16-012; PI Wim Pouw).

Abstract (241 words)

Hand gestures and speech move in a common rhythm, as exemplified by the synchrony between prosodic contrasts in gesture movement (e.g., peak velocity; maximum effort) and speech (e.g., peaks in fundamental frequency; F0). This joined rhythmic activity is hypothesized to have a variable set of functions, ranging from self-serving cognitive benefits for the gesturer, to communicational advantages that support listeners' understanding. However, gesture-speech synchrony has been invariably understood as a "neural-cognitive" achievement; i.e., gesture and speech are coupled through neural-cognitive mediation. Yet, it is possible that gesture-speech synchrony emerges out of resonating forces that travel through a common physical medium – the body. In the current pre-registered study, we show that movements with relatively high physical impact affect phonation in a way that accommodates gesture-speech synchrony. Beating with one arm or two arms at a rhythmic pace led to acoustic peaks in the fundamental frequency (F0) and the amplitude envelope of phonation that were entrained with the rhythm of movement. Such effects were not found for upper limb movements with lower physical impetus (wrist movements), nor when participants were phonating without movement. We further provide evidence that postural stability is contributing to the effect of movement on phonation, as entrainment of movement and phonation was more pronounced when participants were standing as opposed to sitting. The current findings suggest that gesture-speech synchrony emerges from biomechanical constraints, potentially obviating the need for a cognitive predictive mechanism that ties gesture and speech in synchrony.

Keywords: Gesture-speech synchrony, Upper limb movement, Phonation, Motion Tracking

Introduction

Hand gesture and speech are closely synchronized (see for an example <https://osf.io/29h8z/>; Chu & Hagoort, 2014; Leonard & Cummins, 2010; Krivokapić, Tiede, Tyrone, & Goldenberg, 2016; Krivokapić, Tiede, Tyrone, 2017; Pouw & Dixon, 2018a, b; Parrel, Goldstein, Lee, & Byrd, 2014; Rochet-Capellan, Shaiman, Iverson, & Szumisky, 2014; Shattuck-Hufnagel & Ren, 2018; Treffner & Peter, 2002; Zelic, Kim, & Davis, 2015).

Specifically, speech' prosodic contrasts, captured by contrasts in the fundamental frequency of speech (F0; perceived as the 'pitch' of speech), structurally aligns with energetic contrasts in gesture (e.g., peak velocity or point of maximum effort; Krivokapić, Tiede, & Tyrone, Goldenberg, 2016; Loehr, 2004; Pouw & Dixon, 2018a, 2018b).

Explanations of why gesture and speech synchronize are varied (Wagner, Malisz, & Kopp, 2014). They include arguments relating to communicative functions, such that the meaning of gesture and speech is less ambiguous (and optimally effective) when performed in synchrony (e.g., Krauss, 2000). Others suggest self-serving (cognitive) functions of gesture for the gesturer, such that gesturing allows for stabilizing the rhythm of speech through entrainment, and vice versa (Pouw & Dixon, 2018b; Rusiewicz, 2011; Rusiewicz & Esteve-Gibert, 2018), or that gestures can stabilize imagination through extra-neural bodily imaginings recruited in talking (Morsella & Krauss, 2004; Pouw & Hostetter, 2016).

Another focus has been put on developmental origins of gesture-speech coupling, such that hand and mouth are solicited to interact from birth on (e.g., bringing food to the mouth and opening the mouth), which readies opportunities for increased entrainment of the manual and speech system during social development (Gentilucci & Corballis, 2006; Iverson & Thelen, 1999; see also Esteve-Gibert & Guellai, 2018). These varied explanations are

united, however, in that gesture-speech synchrony is invariantly understood (or otherwise implied) to be bounded by a strictly cognitive informational linkage. Indeed, as McClave (1997, p. 69) maintains, "coordination of direction of pitch and manual gesture movements is an option available to speakers, but it is not biologically mandated". Thus, gesture researchers have so far disregarded that gesture and speech have a shared physical medium through which synchrony can emerge biomechanically - the body.

Gesture-speech Synchrony and its Medium

The fundamental frequency (F0) of speech is determined by the alveolar/subglottal (lung) air pressure and larynx muscle tonus (Lieberman, 1993). Everything else being equal, increasing the alveolar pressure will produce more acoustic energy in the form of amplitude and will produce an increased fundamental frequency (i.e., perceived as a higher pitch; Lieberman, Knudson, & Mead, 1969). The prime source of acoustic energy that determines speech is the modulation of the expiratory flow. This energy for expiration is primarily delivered by the elastic recoil in the lungs. These elastic forces are so great, that were it not for the adjustive counter forces produced by a set of alveolar muscles that govern expiration (e.g., intercostal muscles, abdominal muscles), the pressures during the inflated phase of the lungs would blow the "vocal tract apart if the speaker attempted to phonate" (p. 61; Lieberman, 1993).

Given the sensitive role of expiration-related muscles and alveolar pressure in the stable production of speech (more specifically phonation), it seems reasonable to suspect that gesture movements could affect the actions of expiration-related muscles and therefore could affect prosodic metrics of speech directly (e.g., contrasts in F0; changes in amplitude). Despite studies that have looked at effects of gross body exercise during

speaking and phonation (Godin & Hansen, 2015; Johaness et al., 2007) and further advanced research on phonation on a plethora of physical constraints on F0 modulation (e.g., breathing cycles, see Bouhuys, 1974; alveolar pressure and volume, Dromey & Ramig, 1998; heart beat cycles, Orlikoff & Baken, 1988), we are unaware of any research in phonetics that has looked at the possible biomechanical effects of upper limb movement and phonation that can be directly informative to gesture-speech dynamics.

There is however one prominent study in gesture research that has found acoustic correlates of body movements that are relevant to the present study (but also see Bernardis & Gentilucci, 2006; Nobe, 1996; McClave, 1998 for comparable observations). Namely, Krahmer and Swerts (2005; experiment 1), assessed whether hand gestures, head nods, or eyebrow raises¹ affected speech. Such movements were produced during either a part of the sentence that was also intended to be produced with a pitch accent, or during a different part of the sentence where there was no pitch accent intended. It was found, that when any type of movement was made, that this increased duration of phonation and also higher frequency for the first formant (higher F1), and this was regardless of whether a pitch accent was actually intended. Yet, these unintended effects of movements on speech were comparable to effects of speech when pitch-accents *were* intended and that were not accompanied by concomitant movement. Namely, higher F1 and increased duration were also observed for pitch accented speech without movements, suggesting that movement versus intended pitch accent affected speech in similar ways on the dimension of duration

¹ Unfortunately, it was not reported what the exact nature of the physical movements were, and what physical momenta they carried.

and F1. These effects arose regardless of movement type, and regardless of whether the movement coincided with the intended pitch accent, which suggest that making any burst-like body movement during speech affects speech acoustics. However, other more prominent characteristics of pitch accent, namely increased amplitude and increased pitch (F0), were not found to be affected by body movement, but were only observed for speech with intended pitch accent. This is surprising as most of what is known about gesture-speech synchrony in natural speech is based on the relation of pitch (F0) and gesture (Wagner et al., 2014). Importantly, although this study is promising for understanding on which dimensions speech and gesture couple, it is left unknown whether the effect of movement on acoustics is related to direct physical impetus of a gesture movement on acoustics. Krahmer & Swerts (2005, pp. 410) do acknowledge that there must be some kind of muscular synergy that gives rise to these effects, wherein “extra effort for one kind of gesture spills over into the other”, but this effect was still conceived of gesture and speech being “handled by the same underlying mechanism” which is still in line with a purely neural-cognitive understanding of gesture-speech synchrony.

There is further mixed evidence for acoustic effects of gesture from more naturalistic contexts, wherein speech is analyzed when participants gesture during a narration task, which is then compared to when participants are not able to gesture (e.g., Cravotta, Busa, & Prietó, 2018; Hoetjes, Krahmer, & Swerts, 2014). Hoetjes and colleagues (2014), showed that gesturing (vs. no gesturing) did not affect minimum and maximum F0/pitch, pitch range, or the mean pitch. In an exploratory study by Cravotta and colleagues (2018), it was reported that when participants are encouraged to gesture (vs. not encouraged) the mean intensity of speech was affected, while F0 metrics (minimum,

maximum, mean) were unaffected. There are two characteristics of this research (see also McClave, 1994) that limit our ability to draw inferences relevant to our question of whether gesture's physical impetus affects acoustics. Firstly, this previous research makes a distinction between gesture vs. no gesture, instead of analyzing how kinematics and physical impetus of gesture might affect speech. Secondly, the acoustic metrics are not analyzed dynamically through time in relation to gesture. This is problematic if we assume that gestures are affecting speech at particular moments in time.

Acoustics and gestures: A possible route

We suggest that there is a viable possibility that upper limb movements in gesturing have direct physical effects on F0 and amplitude of speech. We also think that this physical effect will, perhaps in part, provide a means for gesture-speech synchronization as observed in spontaneous gesture-speech synchrony (e.g., Krivokapić et al., 2017; Wagner et al., 2014). Similar to weakly coupled oscillators that spontaneously synchronize due to vibrations traveling through a shared physical medium (e.g., pressure waves; shared physical platform; see Pikovsky, Rosenblum, Kurths, 2001), the body allows - and is dependent in its functioning on - forces that resonate through its musco-skeletal network (Turvey & Fonseca, 2014). Such forces can *in principle* provide a non-cognitive source for gesture-speech synchrony.

How could such an effect of arm movement on speech F0 and amplitude possibly arise *in practice*? Firstly, when moving the upper limbs various muscles will be recruited in anticipatory fashion as to maintain postural stability within about 100 milliseconds *before* and 50 milliseconds *after* onset of the limb movements (e.g., Aruin & Latash, 1995; Boussiet & Zattara, 1981; Boussiet & Do, 2008; Cordo & Nasher, 1982). In the case of arm

movement, these ‘anticipatory postural adjustments’ (APA) mobilize an interconnected set of muscles including those around the trunk (Hodges & Richardson, 1997a, 1997b).

Specifically, one of the key APA muscles that are recruited for arm movements is the Rectus Abdominus (RA; i.e., “the abs”; Aruin & Latash, 1995; Friedli, Hallet, & Simon, 1984). It turns out that the trunk muscles that are recruited for anticipated postural adjustments (including the Rector Abdominus) are directly involved in the active phase of expiration (Hodges, Gandevia, Richardson, 1997), which is the phase during which we produce speech. These adjusting forces are non-negligible. They produce balancing reactive forces that are counteractive to the forces produced by the kinetic perturbations of moving the arms. Moving the arms faster produces more destabilizing forces and will need to be met with an equally more forceful APA. It is finally important to note that contrary to common wisdom, the forces produced by limb movements themselves (as well as APA’s) are not localized to the limb (Silva, Morena, Mancini, Fonseca, Turvey, 2007; for an overview see Turvey & Fonseca, 2014). Any type of muscle contraction will produce forces that travel throughout tensioned connective network of soft tissues known as fascia and the compressed elements (i.e., bones), and such traveling forces are essential in the effective coordination of movement that involves a synergy of components (i.e., any intentional action; Bernstein, 1966).

Now that we have established a potential route through which gestures can affect speech directly, we might wonder whether gestures really produce non-trivial forces, and whether such forces are a viable source of physical coupling. That gesture related forces are non-trivial is indicated by the experiences of Ian Waterman—a person suffering from almost complete proprioceptive loss—who reported a need to suppress his gestures in

initial stages of his disease because he was afraid of falling from the destabilizing effects of these articulations (Gallagher, 2005; McNeil, 2005).

Furthermore, the forces that gestures produce and their coupling with speech prosody already seem to be entailed by the classic gesture categories that are used in gesture studies (McNeill, 2005). Namely, a common type of gesture that is identified as having the sole function of synchronizing with prosodic contrasts with speech are called 'beat' or 'baton' gestures (McNeill, 2005; Kendon, 2004). Such beat gestures are characterized by burst-like vertical arm movements that "beat" with the rhythm of speech (Leonard & Cummins, 2010). Importantly, beat gestures seem to possess greater physical momentum as compared to other types of gestures as they produce sudden halts ("beats"), and therefore possess greater potential for momentum transfers to the body that further act to destabilize body posture (for real-world examples of beat gestures see <https://osf.io/29h8z/>). Thus, beat gestures might synchronize with speech the way they do, because they are recruited in a way to produce a physical impulse on the body. Beat gestures can be distinguished from the next common overarching type of gestures, which are iconic gestures. Compared to beat gestures, iconic gestures have more complex and often more fluid movement trajectories as they need to iconically present meaning.

Although iconic gestures still have moments of emphasis wherein the perceived moment of maximum effort coincides with prosodic peaks in speech (Wagner et al., 2014; Prieto, Cravotta, Kushch, Rohrer, & Vilà-Giménez, 2018), these gestures are more variably (less tightly) coupled with prosodic contrast in speech (Pouw & Dixon, 2018a, 2018b). This is because iconic gestures have movement trajectories that are not recruited primarily to beat with the rhythm with speech, rather some degree of freedom is reserved for iconic

expression. In sum, we think the role of gesture's physical impetus and the correlation with speech prosody lies dormant in current gesture categorization practices.

Exploratory findings

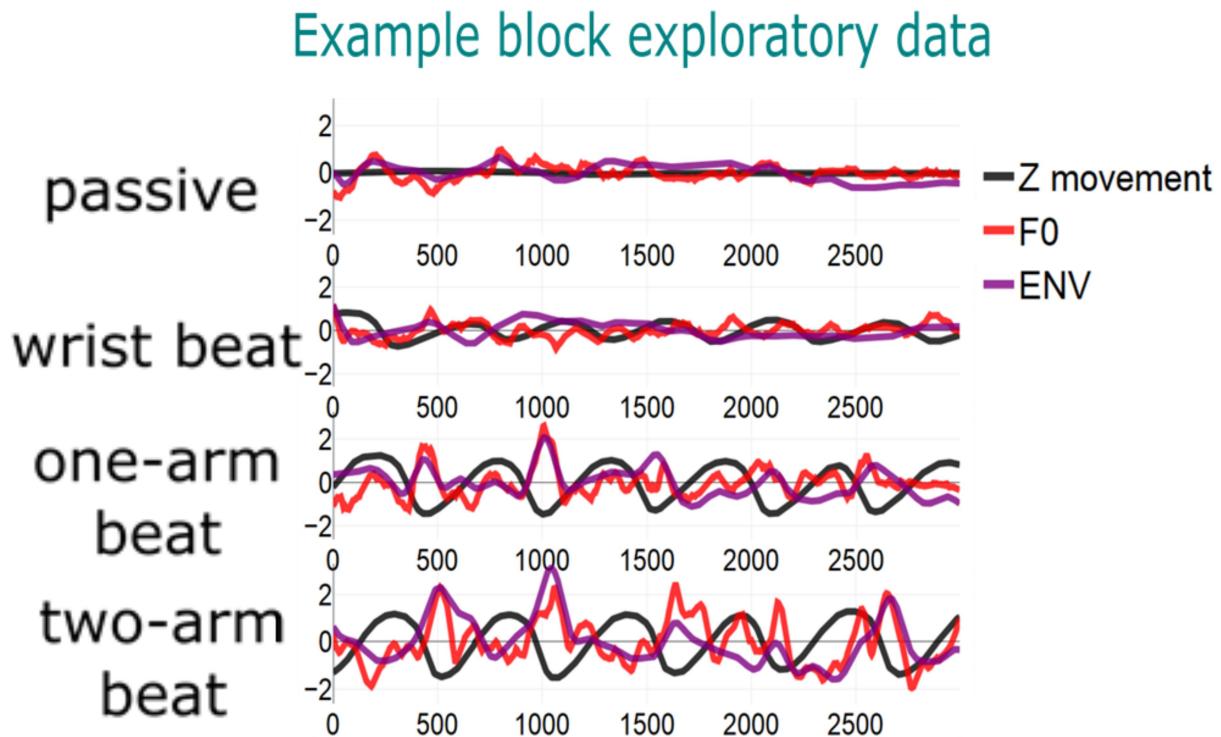
The current pre-registered confirmatory study (see <https://osf.io/5aydk/> for pre-registration documents) was based on a previous exploratory study using motion-tracking (for a full report see: Pouw, Harrison, & Dixon, 2018), where two participants phonated at their own preferred 'pitch' while either moving the wrists, one arm, or both arms, in a beat-like fashion; a vertical movement with a movement contrast (a beat) at the downward (extension) phase. Participants were explicitly instructed to keep their phonating pitch as steady as possible and to resist possible interfering effects when moving the arm(s) or wrist at their own preferred rate. Participants also phonated while not moving the upper limbs (passive condition). In this exploratory study, we found clear individual-level effects of upper limb movements on F0 and the amplitude envelope (see Figure 1). Upper limb movements with high physical impetus (one-arm beat and two-arm beat) were synchronized with peaks in F0 and the amplitude envelope. The reader can listen to audio examples of the trials together with a visual presentation of the amplitude envelope, F0 and vertical hand movement (Z-movement) for the exploratory data here:

<https://osf.io/acmdg/>. These audio examples show clear audible effects of high impact upper limb movements. Further analyses revealed that at about 70 milliseconds before the extension was reached of a one-arm or two-arm beat that there was a peak observed in speech acoustics.

The findings from the exploratory study indicate that when the body is bracing for the impact of the downbeat, involuntary effects on phonation are produced. We speculated

that such effects are related to anticipatory postural adjustments that tension the muscles around the trunk. When the downbeat is less physically destabilizing, as in the case of a wrist movement, no such effects arise. Note that this type of synchronization is 'involuntary' as participants are instructed to keep phonating at a steady pitch level.

Figure 1. Example trials from the exploratory study



Note Figure 1. Example time series (first 3 seconds) from the exploratory study of one block for participant 2 are shown with pitch (F0: in red) amplitude envelope (ENV: in purple) and vertical movement (Z movement). Peaks in F0 and Amplitude Envelope are observed for One-Arm Beat and the Two-Arm Beat condition around the moment that the movement reaches its maximum extension, which is also the moment such movements reach their highest physical impetus on the body.

Confirmatory study

With the current pre-registered study, we aim to replicate our exploratory findings in a larger sample (10 participants; 240 trials) and further assess the possible role of postural stability effects from upper limb movements on phonation. Identical to the exploratory study, participants were asked to phonate the vowel [ə] (as in cinema) during several trials of the passive condition, wrist beat condition, one-arm beat condition, and two-arm beat condition. We also added another within-subjects factor wherein participants performed the upper limb movements while sitting in a chair (sitting condition) or while standing upright (standing condition). We added this condition as it has been shown that anticipatory postural adjustments (APA's) that arise when moving the upper limbs while *standing*, are dramatically diminished when the body is in a more stable *seated* posture (Cordo & Nasher, 1982). If APA's are modulating the current effects, then we would expect to find that the upper limb movements effects on phonation are absent or diminished in the sitting condition relative to upper limb movement effects on phonation in the standing condition.

Method

Design

The current experiment consists of a two-factor within-subject design, with one within-subject factor (movement condition) of 4 levels (passive vs. wrist beat vs. one-arm beat vs. two-arm beat), and another within-subject factor (posture condition) with 2 levels (sitting vs. standing). Ten undergraduate students from the University of Connecticut (5 females and 5 males; 8 right-handed; M age = 19.2, SD age = 1.25) were asked to produce a steady voiced output of the vowel 'a:' (as in 'cinema', [ə]). Participants were asked to stop phonating as soon as they felt that they ran out of air and could not maintain their preferred level of pitch. For each participant we planned to perform 3 blocks of 8 trials (total = 240 trials = 10 participants x 3 blocks x 4 movement condition x 2 posture condition). A total of 239 trials were actually performed².

Procedure

For half of the trials, participants were asked to sit on a chair with their feet firmly on the ground and their backs touching the backrest. The chair did not have armrests. In the other half of the trials, participants were asked to stand upright. For the passive condition, participants were asked to let their hands rest alongside their bodies during phonating when standing, or rest on their lap when sitting. For the one-arm beat condition, participants were asked to continuously move their dominant hand by lifting the hand up and letting it drop with a sudden complete halt (i.e., with energetic contrast, a "beat"). The beat was reached around the point where the elbow flexion angle was 90 degrees. In the

² One trial was not performed because the experimenter accidentally skipped a trial.

“two-arm beat” condition, participants made the same movement in-phase with two arms.

In the “wrist beat” condition, participants were asked to only move in beat-like fashion their dominant hand with only a wrist movement (and no forearm or upper arm movements). Order of condition was randomized for each block of 8 trials.

A crucial change from the exploratory study is that in the current experiment we guided the movement frequency of the participant by a visual presentation. Instead of participants moving at their own preferred frequency, participants were encouraged to move their hands at 80 beats per minute (i.e., 1.3 Hz; oscillation period = 0.77 seconds). This allowed us to analyze the data with a focus on a particular frequency range without having to account for individual differences in preferred moving rate. We programmed in C++ a visual presentation that takes input from the motion tracker so as to visually represent the frequency of the vertical movement to the participant. The visual presentation consisted of a bar that changed size as a function of movement frequency; participants tried to keep the size of the bar within a certain range as specified by two guide bars. The size of the guide bars corresponded to 10% faster or slower than 80 BPM.

Apparatus

Motion and audio recording

We used a Polhemus Liberty to record movement (240Hz), with a sensor attached to the tip of dominant hand's index finger. Since hand movements were primarily in the vertical dimension, we analyzed movement-phonation coordination and computed derivatives (i.e., velocity, acceleration, jerk) only for Z-axis movement. For derivative estimation, we applied a low-pass Butterworth filter of 33 Hz. We recorded audio using a RT20 Audio Technica Cardioid microphone (44.1kHz). We used a modified C++ script made publicly available by Michael Richardson (Richardson, n.d.), so as to simultaneously call and write movement and audio data. We modified this script to enable simultaneous recording of sound from a microphone, using toolbox SFML for C++ (<https://www.sfml-dev.org/>). Using a custom-made script in R (R core Team 2013), the data from PRAAT and the motion tracking data were aggregated (code available on <https://osf.io/5aydk/>).

Phonation Variables

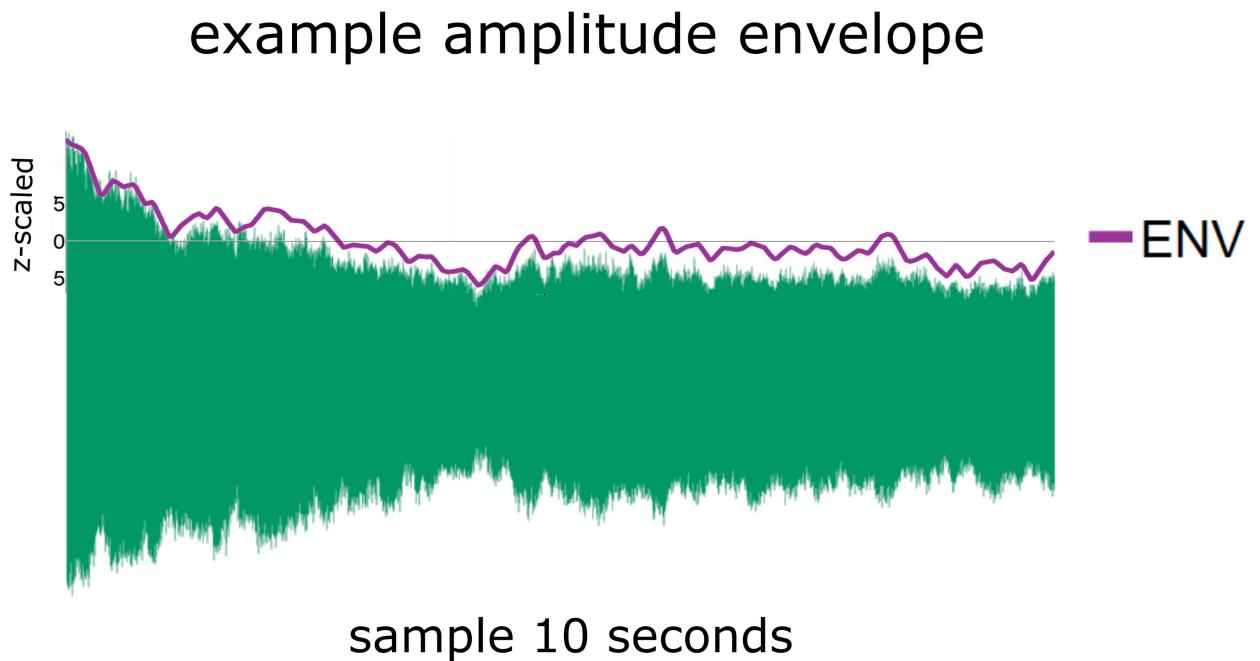
As stated in the pre-registration, our analyses focus on F0 and the amplitude envelope. These acoustic properties are key metrics for prosody in speech. F0 and amplitude time series were sampled at the sampling rate of the motion-tracker (240Hz: 1 sample per 4.16 milliseconds).

Fundamental Frequency (F0; pitch). F0 time series was extracted from the audio using PRAAT (Boersma, 2001) with a range suitable for male (75-500Hz) or female (100-500 Hz) voice range.

Amplitude Envelope (env). A raw speech signal has both fine and gross structural changes, i.e., higher and lower frequency fluctuations. The lower frequency fluctuations are

important for the rhythmic structure of speech (Chandrasekaran et al., 2009; Tilsen & Arvaniti, 2012) and can be captured by the Amplitude Envelope (ENV). ENV can be reconstructed from the raw audio signal using the Hilbert transform (He & Dellwo, 2017). The amplitude envelope (ENV) time series were produced by applying the PRAAT script by He & Dellwo (2017; see also He & Dellwo, 2015). ENV is scaled in Hilbert Units ranging from 0 to 1. Thus each sound recording from a participant is scaled from 0 (minimum amplitude) to 1 (maximum amplitude). See Figure 2 for an example of the amplitude envelope metric.

Figure 2. Example Amplitude Envelope



Note. Standardized amplitude envelope (240 Hz sampling) for a sample of 10 seconds of phonation for the current data (trial = standing one arm beat condition). Essentially the amplitude envelope tracks gross fluctuations of the raw audio waveform.

Results

Descriptives

Descriptives for the fundamental frequency (F0) and the amplitude envelope (ENV) computed for each trial, and averaged across trials per condition, are provided in Table 1. Examination of Table 1 shows that phonation was less stable for the one-arm and two-arm beat trials, as standard deviations are markedly larger (especially for F0). Indeed, when listening to the audio samples it was apparent that phonation was less stable during the arm-movement conditions. The average time for each phonation trial was 8.71 seconds ($SD = 2.69$ seconds), with average duration for passive = 8.48s ($SD = 2.65$), wrist Beat = 8.90s ($SD = 2.93$), one-Arm beat = 8.72s ($SD = 2.85$), two-arm Beat = 8.74s ($SD = 2.62$), and sitting = 8.58s ($SD = 2.62$) and standing = 8.83s ($SD = 2.76$). Trial time was correlated with decreases in F0 ($r = -.199, p < .001$) and ENV ($r = -.134, p < .001$), which indicates that the ability to maintain acoustic energy levels decreased as participants reached the end of their breadths. To prevent spurious effects of time in our time series analyses, we linearly detrended the effect of time for each trial before entering into the analyses. Not surprisingly, ENV and F0 (standardized for each trial) were weakly positively correlated (average $r = .22$, average $p < .017$).

Table 1. Mean and standard deviation of F0 and ENV per condition

		Mean (SD)		
		Passive	Wrist Beat	One-Arm Beat
F0	TOTAL	178.82 (1.81)	178.82 (2.16)	178.13 (2.67)
	Sitting	180.89 (1.63)	177.47 (2.45)	178.35 (2.67)
	Standing	176.47 (1.62)	176.26 (1.88)	179.28 (2.66)
ENV	TOTAL	.257 (.047)	.250 (.047)	.251 (.048)
	Sitting	.212 (.042)	.199 (.035)	.191 (.038)
	Standing	.297 (.051)	.301 (.057)	.309 (.056)

Note. F0 is given in Hertz. Amplitude Envelope (Amp. Env.) is given in Hilbert Units (range = 0-1).

To provide an insight into the kinematics of the current movements, we computed the maximum velocity produced during the extension phase (i.e., negative movement direction) of the movement. This corresponds to the maximum velocity of the downbeat. The average maximum negative velocity is given in Table 2 for each movement condition.

Table 2. Average maximum negative velocity

		Mean (SD)		
		Wrist Beat	One-Arm Beat	Two-Arm Beat
Average Max Negative Velocity				
	Sitting	-75 cm/s (24.59)	-138.88 cm/s (42.71)	-141.97 cm/s (35.70)
	Standing	-98.55 cm/s (23.10)	-172.28 cm/s (31.41)	-166.93 cm/s (22.26)
	Total	-87.21 cm/s (26.28)	-155.87 cm/s (40.71)	-154.24 cm/s (75.86)

Note. Velocity is given in centimeters per second.

Exclusions for analyses

After inspection, we found for several trials that for two participants (participant 7 and 8), PRAAT could not reliably track continuous F0 while participants were in fact phonating continuously, and also showed noisy periodicity estimates of F0 traces. We will

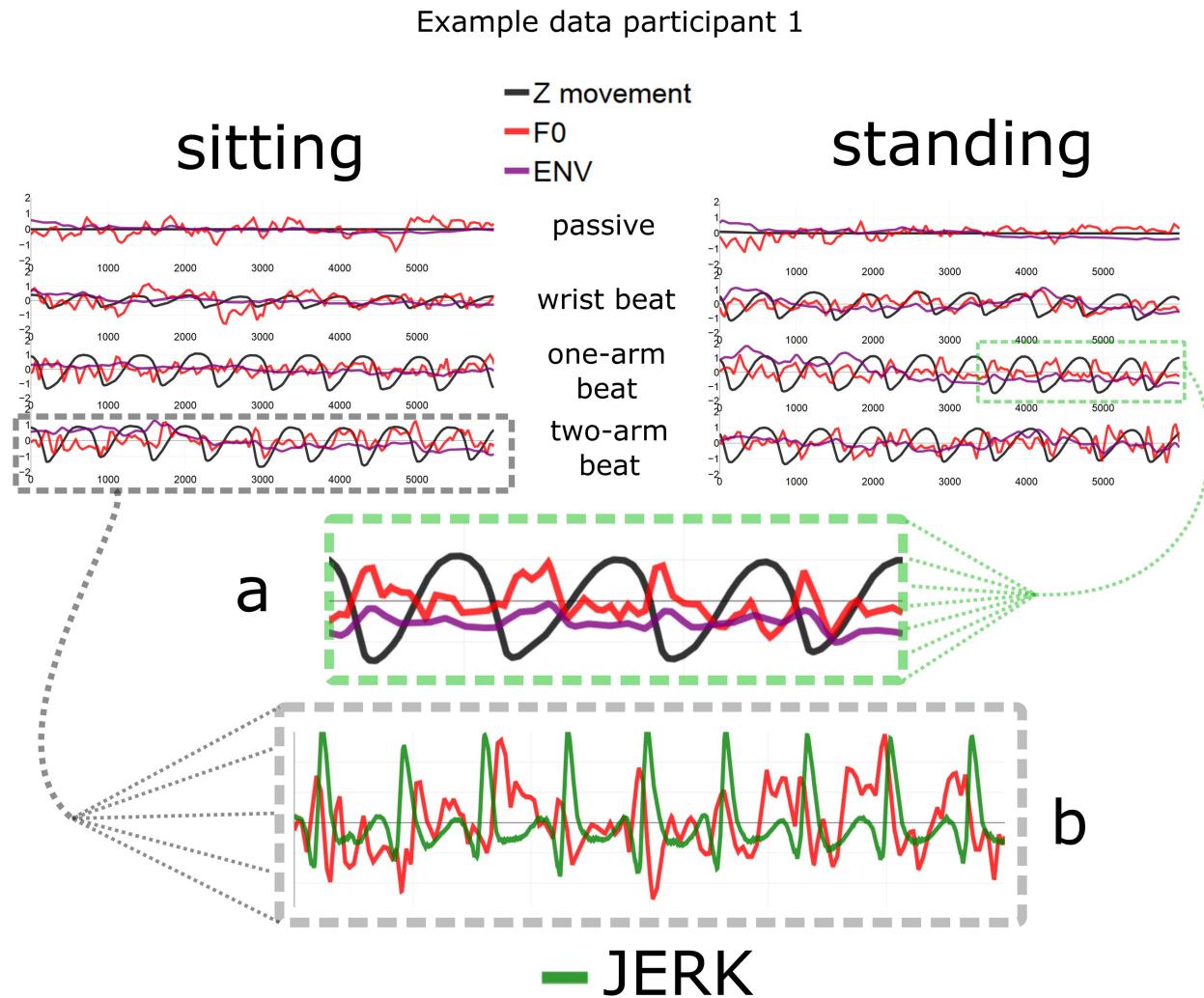
not include these participants in our analyses for fundamental frequency³. The amplitude envelope could be reliably tracked and showed no anomalies.

Time series descriptives

Figure 3. shows an example of the time series for the first participant (female) of the current dataset (for 1 block).

³ The failed F0 tracking was most likely due to too much distance between participants and the microphone (leading to a faint signal). Note that this decision does not affect our conclusions in any way; we have run the analyses with these noisy data included (and this did not affect our main conclusions).

Figure 3. Example phonation and movement time series

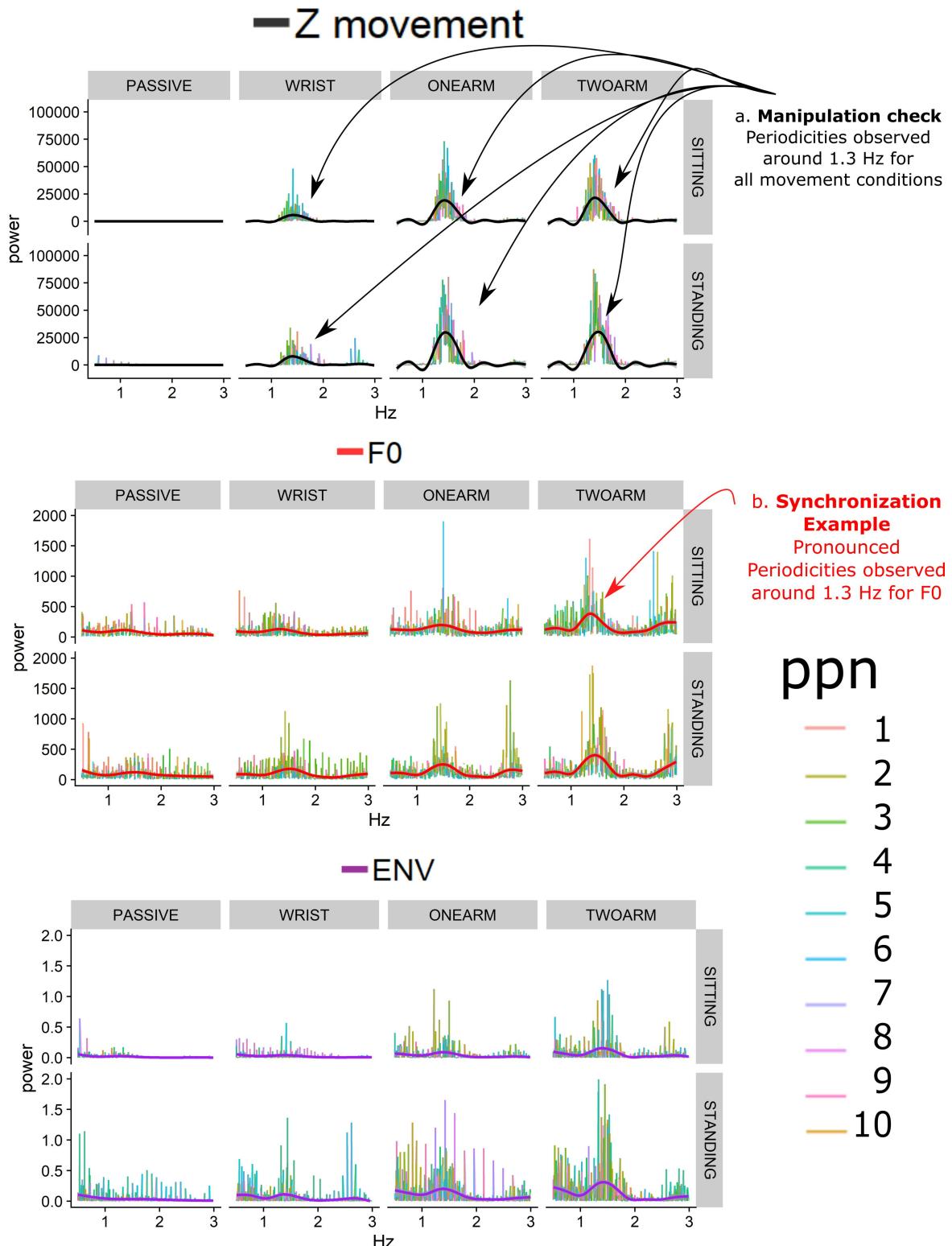


Note Figure 3. Example time series (first 6 seconds) for each movement and posture condition (F0: in red) amplitude envelope (ENV: in purple) and vertical movement (Z movement: in black). All measures shown are z-standardized. Panel (a) shows an enlarged section of the one-arm beat condition, where it is very clear that peaks in F0 and ENV are observed when the vertical movement reaches its maximum extension. Panel (b) shows another representation of the time series for F0 and vertical movement. Jerk is time-

derivative of acceleration, and indicates here that sudden changes in movement acceleration often co-occur with peaks in F0.

To further summarize the (shared) periodic structure of the time series, we performed a spectral decomposition analyses with R package ‘spectral’ (Salmayer, 2016). This analysis used the Fast Fourier Transform (FFT) to assess periodicities in movement and phonation (see Figure 4). It is to be expected that for the movement conditions we find periodicities in vertical movement time series around the target range of 1.3Hz (80 BPM, period = 0.77 seconds). If movements are entraining phonation (as Figure 1 and 3 indicate), we too would expect to observe that periodicities around 1.3Hz are dominant periodicities for the F0 and ENV time series. Visual inspection of Figure 4 indeed indicates that there are likely to be shared periodicities of movement with phonation (F0 and ENV), suggesting that high-impact movements are structurally affecting phonation.

Figure 4. Spectral decomposition of periodicities of movement (black), F0 (red), and amplitude envelope (purple), per condition.

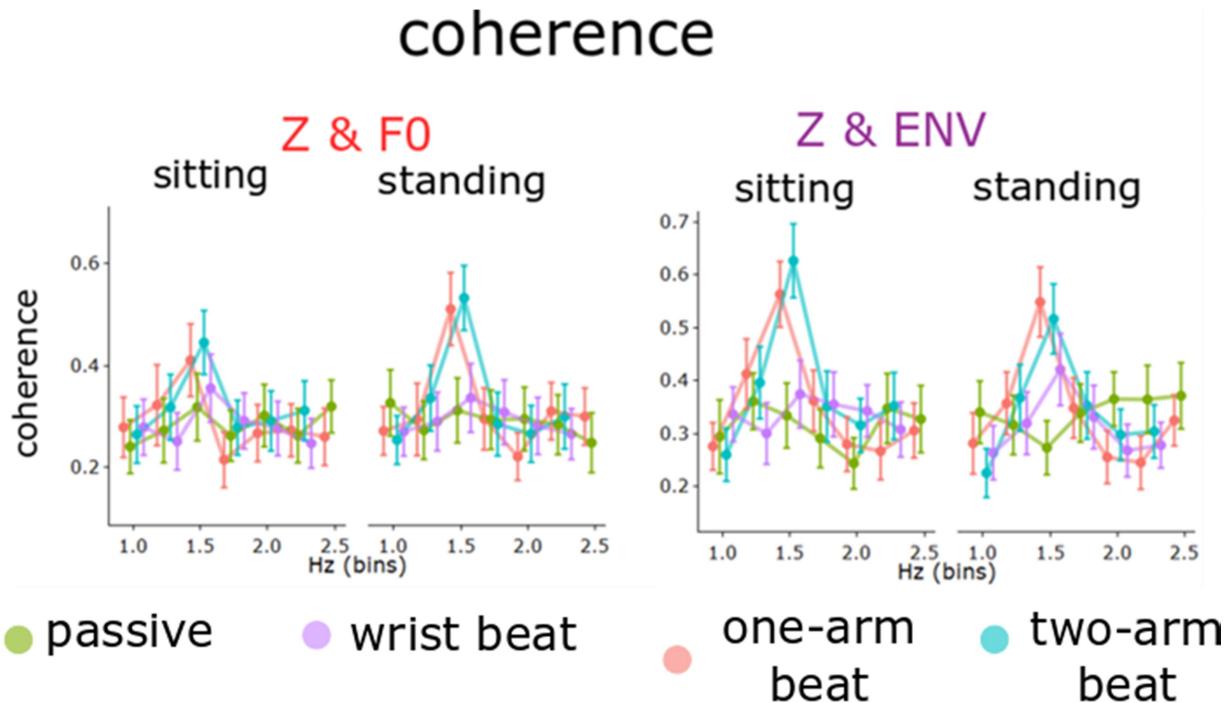


Note Figure 4. Solid lines (black = Z, red = F0, purple = ENV) reflect smoothed mean power densities. Bumps in these solid lines indicate that the time series were defined by periodicities around that frequency. Colors reflect individual participant data. Note, at **a**) that our manipulation to guide participants movement frequency was successful as consistent periodicities are shown. Note, at **b)** that peaks at the movement frequency range are found for several conditions.

Confirmatory Analyses

Coherence

To formally test whether the periodicities of movement and phonating were correlated, we computed coherence between the different spectral density distributions (R package ‘seewave’; Jerome et al., 2018). Coherence is measure that provides a correlation strength of the periodicities, ranging from 0 (no correlation) to 1 (perfect correlation) across a frequency range. Figure 5 provides an overview for the mean coherence per condition between I) movement and ENV, and II) movement and F0. Examination of Figure 5 suggests coherence levels are increased at and around 1.3 Hz for the two-arm beat- and the one-arm beat condition, for both sitting and standing conditions. The effect of movement on F0 in the standing condition seems more pronounced as compared to the sitting condition.

Figure 5. Coherence

Note. This figure shows the coherence levels for each bin (bins have 0.25Hz width) and condition between a) movement (Z) and amplitude envelope (ENV), and b) movement (Z) and fundamental frequency (F0). Data points that are directly adjacent to each other fall within a single bin. Error bars indicate 95% confidence intervals. It can be observed that the passive and wrist condition has generally lower coherence levels, confirming that movement and phonating were not coupled. Consistent with the spectral density results around the 1.3Hz range, there are prominent peaks for the ENV and F0 for the one-arm and two-arm beat conditions, and this seems to be most pronounced for the standing condition. Note that we added the passive condition as a baseline to compare to the other conditions. We should naturally expect that movement and phonation do not have high coherence for when participants are not moving.

To directly test differences in coherence levels we performed mixed regression modeling (R package nlme: participants as random intercept) to predict coherence levels as a function of movement condition, posture condition, and the interaction of movement and posture condition; as stated in our pre-registration. Coherence was assessed in a frequency band around 1.3Hz (0.8-1.8 Hz range).

Coherence between movement and F0 was reliably predicted by movement condition, as compared to a base model which predicted the overall mean ($\chi^2 [6] = 16.88, p < .001$). Adding posture condition to the model, further improved predictions for coherence as compared to the previous model ($\chi^2 [7] = 6.60, p < .010$). Adding an interaction of movement condition x posture condition, did not however lead to further improvement of the model ($\chi^2 [10] = 1.51, p = .679$). The best fitting model with movement and posture condition as predictor showed that the wrist beat condition did not reliably differ from the passive condition ($b = .005, t[1683] = 0.360, p = .718$). The one-arm beat condition ($b = .038, t[1683] = 2.43, p = .015$), and the two-arm beat condition ($b = .055, t[1683] = 3.50, p < .001$), did reliably differ from the passive condition. Furthermore, we found that the standing condition resulted in increased coherence as compared to the sitting condition ($b = .028, t[1683] = 2.57, p = .010$). The one-arm beat ($b = .032, t[1683] = 2.12, p = .034$) and the two-arm beat condition ($b = .049, t[1683] = 3.21, p = .002$) also had reliably higher coherences as compared to the wrist beat condition. These analyses show that fundamental frequency of phonation was entraining to movements with high physical impetus (one-arm beat condition and two-arm beat condition), and that standing resulted in increased coherence suggesting that postural stability is a contributing (but not a necessary) factor.

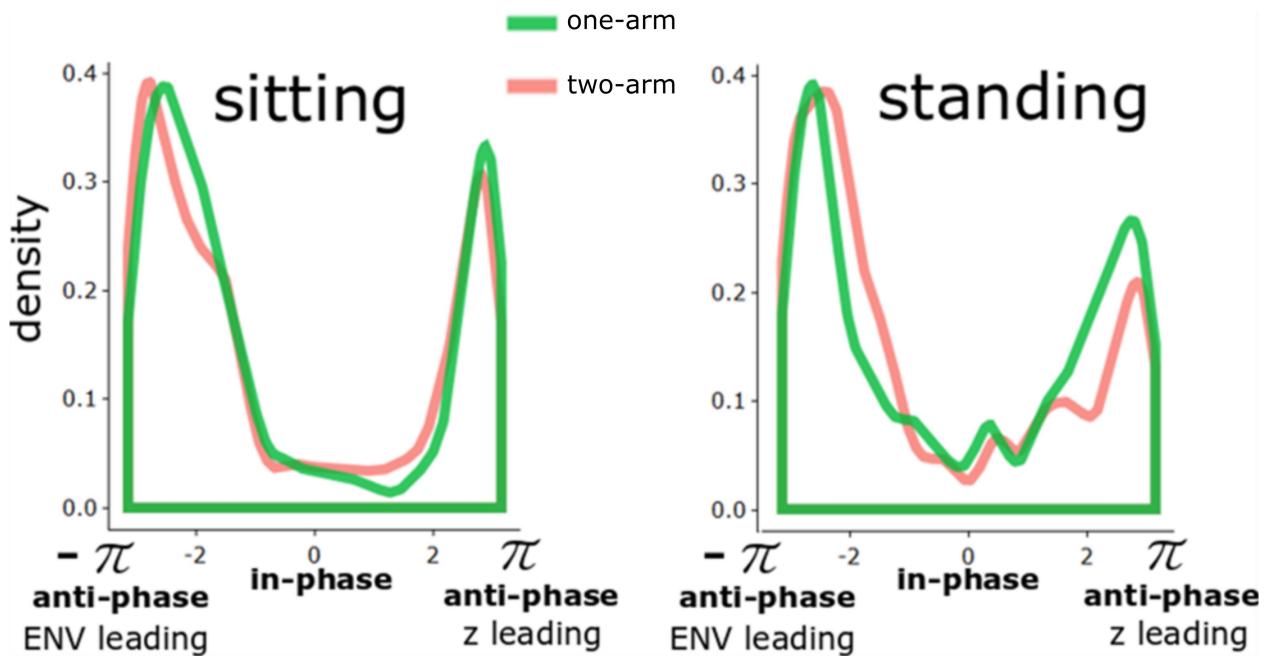
Coherence between movement and ENV was reliably predicted by movement condition, as compared to a base model (change in χ^2 [6] = 36.43, $p < .001$). Adding posture condition to the model did not improve predictions (change in χ^2 [7] = 2.51, $p = 0.112$), and adding the interaction also did not improve (change in χ^2 [10] = 2.64, $p = 0.449$). The best fitting model with movement condition as predictor revealed that the one-arm beat condition ($b = .075$, $t[2027] = 4.76$, $p < .001$) and the two-arm beat condition ($b = .074$, $t[2027] = 4.77$, $p < .001$) had reliably higher coherence levels than the passive condition. The wrist beat condition did not differ in coherence as compared to the passive condition ($b = .018$, $t[2027] = 1.13$, $p = 0.256$). The one-arm beat ($b = .057$, $t[2027] = 3.70$, $p < .001$) and the two-arm beat condition ($b = .057$, $t[2027] = 3.62$, $p < .001$) also had significantly higher coherences as compared to the wrist beat condition. These analyses reveal that arm movement with high physical impetus were also entraining the amplitude envelope of phonation.

Temporal dynamics

From the previous analyses we know that upper limb movements with high physical impetus (one-arm and two-arm beats) are structurally entraining phonation. The next question is how movement and phonation are locking their phases. From the time series example in Figure 3 (and Figure 1), it can be observed that acoustic peaks are observed around the moment where the vertical movement reaches its maximum extension. To assess whether vertical movement is indeed related to phonation in this anti-phase fashion, we computed the relative phases (i.e., phi: Φ) between phonation and the amplitude envelope for the one-arm and the two-arm beat conditions using cross-wavelet analyses (with R package 'WaveletComp'; Rösch & Schmidbauer, 2014; for a helpful tutorial

see Rösch & Schmidbauer, 2016). Cross-wavelet analysis uses a mother wavelet as a basis for decomposing complex time series in dominant periodicities, and further allows for comparison of periodicities between time series (hence *cross*-wavelet). Our analysis used the Morlet wavelet as its daughter wavelet. We entered one-arm and two-arm movement trials into a cross-wavelet analyses (using 50 simulations to compute *p*-values) where we assessed the relative phases of ENV with movement around the frequency 1.3 Hz (i.e., period = 0.77). Figure 6 shows a summary of the results for the cross-wavelet analyses, whereby the observed relative phases at reliability levels $p < .01$ across trial time is plotted. It can be seen that there is a reliable out-of-phase coordination.

Figure 6. Cross-wavelet plots and relative phases of movement (Z) and amplitude Envelope (ENV) for all movement conditions



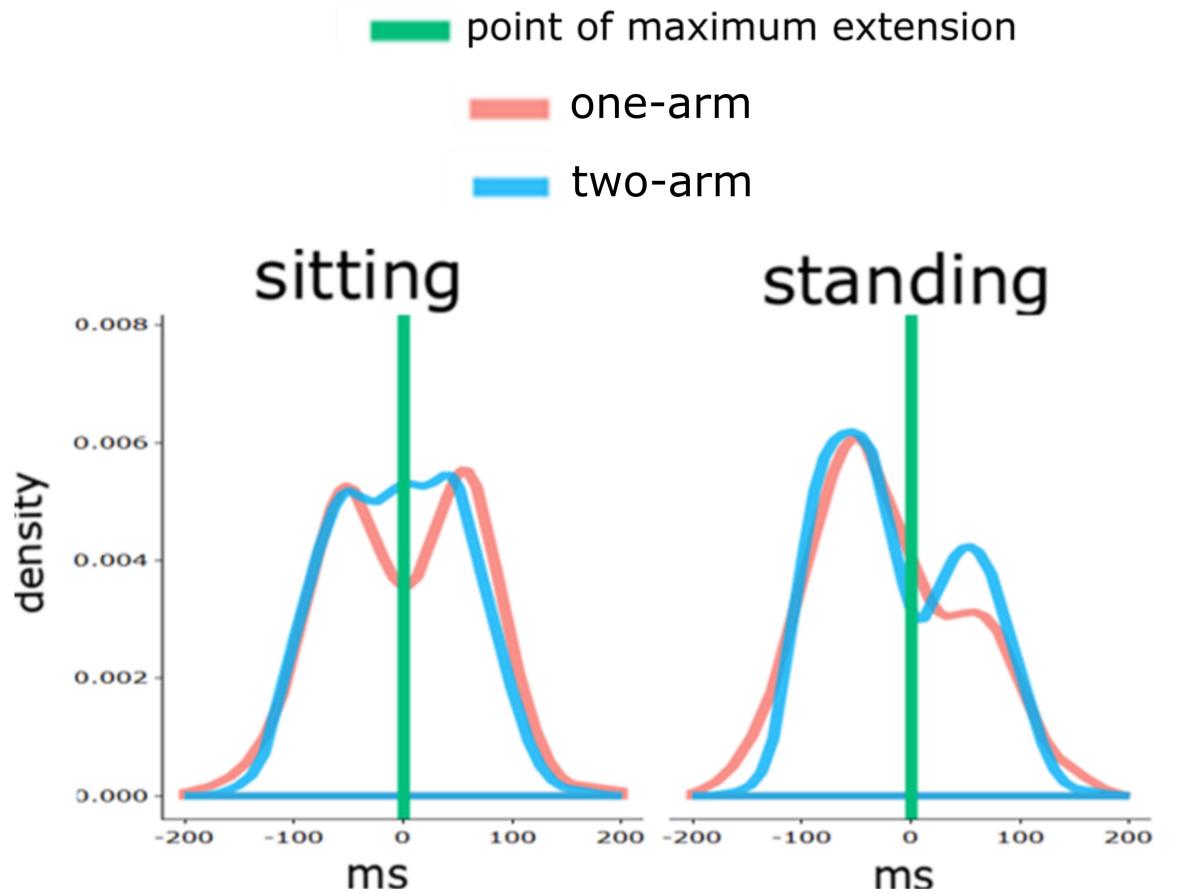
Note. Smoothed density distributions of relative phases (between Z and ENV) observed at $p < .01$ for the one-arm and two-arm condition for the 1.3Hz range. It can be found that reliable relative phases are found at the anti-phase regions ($\Phi > \pi/2$ or $\Phi < -\pi/2$). Furthermore, it seems that phonation predominately leads Z time series (especially in the standing condition), and this is a likely indication that slightly before the maximum extension in the movement, the maximum physical impetus is reached due to the deceleration, which affects phonation.

Given that we now know that there is an anti-phase synchronization between movement and phonation (when movement is in the extension phase, F0 and the amplitude envelope is rising in intensity), we can estimate the time it takes for a gesture movement to

reach phonation. Namely, we can compare when a peak in amplitude envelope is observed relative to the maximum extension of the beating gesture. Thus, we determined the point of maximum extension of the down-beat phase (minimum Z value) and relating this to the nearest peak in the positive rate of change of the amplitude envelope⁴ (in other words, the peak in the positive “acceleration” of the amplitude envelope). Figure 7 shows the main results. It can be obtained that slightly before (mean for negative distribution = -58 ms [$SD = 39$]) and slightly after (mean for positive distribution = 55 ms [$SD = 39$]) the point of maximum extension is reached, there is a positive acceleration in the amplitude envelope. These results suggest that right *before* the moment of maximum extension (right at the moment where *anticipatory* postural adjustments are made to brace for the impact of the beat gesture), and right *after* (i.e., right at the moment where the arm accelerates again for flexion) the peak in phonation is observed. That the *anticipatory* phonation effects are more pronounced for the standing condition is to be expected, as anticipatory postural adjustments are more pronounced when standing.

⁴ We used the amplitude envelope (as opposed to F0) for this estimation as this measure has less fine structural fluctuations as compared to F0 which allows for a reliable estimate of peak positive rate of change (i.e., envelope ‘acceleration’)

Figure 7. Timing of peak change ENV – Point of Maximum extension beat.



Note. Ms (milliseconds) is the temporal distance between the nearest peak of a positive rate of change in the amplitude envelope (peak env “acceleration”) *versus* the maximum extension of the downbeat. If temporal offset in milliseconds is negative, this indicates that peak in change of the amplitude envelope *precedes* the point of maximum extension. For the standing condition as compared to the sitting condition, it seems that the negative distribution becomes more peaked, likely indicating that *anticipatory* postural adjustments that are made *before* the maximum extension (and thus physical impetus) is reached are impacting phonation.

Discussion

In the current pre-registered study, we have replicated our exploratory findings (Pouw, Harrison & Dixon, 2018), showing that repetitive arm movements (moving at 1.3Hz) with relatively high physical impetus (one-arm and two-arm vertical beat movements) structurally entrain phonation. Repetitive arm movements with relatively low physical impetus (wrist beat condition) or making no movements at all (passive condition) did not lead to effects on phonation (as compared to the one-arm and two-arm beat condition). We further found that performing movements when standing, as compared to sitting, increased the degree of entrainment of movement and phonation (but only for F0). This suggests that anticipatory postural adjustments (APA's) are contributing to the movement-phonation synchronization effects. However, since movement effects on phonation still arise when participants are seated (and posture effects were not very pronounced), it is likely that the physical impact of arm movement has direct effects on phonation as opposed to being completely mediated by APA's. The effects of phonation were such that when the arm movement reached its maximum extension, peaks in fundamental frequency and amplitude envelope were found. Such peaks were observed at about 50 ms before, and 50 ms after, the extension maxima (i.e., the beat) was reached, which coincide with the moments where physical impetus is highest (deceleration for stopping extension and acceleration for initiating flexion).

Regardless of the exact mechanisms underlying our effects (see below for theoretical implications), the current data show that merely moving upper limbs affects phonation acoustics. Of note, movement affects acoustics despite the fact that participants were instructed to resist any effects on their phonation. Remarkably, these dramatic effects

of upper limb movements on phonation acoustics can directly accommodate gesture-speech synchrony in naturalistic contexts. Namely, high-impact movement resulted in peaks in F0 and amplitude envelope, and these acoustic peaks are generally equated in speech as prosodic contrasts (i.e., pitch accents). That is, prosody researchers identify stressed pronunciations or a “pitch accent” by searching for prominent peaks in F0 and/or amplitude (e.g., Wagner et al., 2004). Thus, it is one thing that our findings indicate *direct* effects of movement on phonation, because it provides a bodily route for information to travel. It is quite another, that the present physical effects are directly transferable to how prosodic contrasts are made in speech; by sudden increases in pitch and amplitude.

On the basis of the previous literature, the current findings might seem surprising. There have been studies that have assessed direct effects of upper-limb movements and gesture on speech acoustics in natural and more controlled contexts, yielding no (e.g., Hoetjes et al., 2014) or more subtle effects on speech acoustics (e.g., Krahmer & Swerts, 2004). However, it should be noted that previous research has not a) looked at physical impetus as a constraint on gesture-speech synchrony, and b) methods used in these previous research often involves averaging acoustic metrics *over time*, rather than studying movement (with motion tracking) and speech acoustics dynamically *through time* as we did with the current data. Indeed, our results suggest that gesture will affect speech acoustics *at particular moments in time* – when a pronounced physical impetus is reached. Of course, our findings do relate to a range of classic findings on the gesture-speech prosody correlation, which initially revealed a coupling of gesture and pitch (e.g., McClave, 1994). Yet, such effects have never been understood as a physical effect of movement, but rather seen as a cognitive artifact (for an overview see Wagner et al., 2014). In sum, the

current findings clearly indicate that upper limb movements affect acoustics, and in a way that at least partially accommodates findings on gesture-speech synchrony.

Theoretical implications

We think there are a couple of important theoretical implications from the current study. In broad strokes, we think the current data supports a view of gesture-speech synchrony that entails a radically constitutive relation between non-neural bodily dynamics and intentional communicative behavior. This view relates to findings in other domains of perception and action, such as catching a fly ball, whereby hypothesized predictive brain mechanisms *in the agent* are obviated by the discovery of perception-action invariants that are continuously available to the agent in the environment (Michaels & Carello, 1981). In the same vein, given that we find that upper-limb movements entrain acoustics through physical impetus, it seems inadequate to impose a cognitive burden on a supposed neural gesture-speech prediction module that estimates the timing of the movement trajectory of a gesture with the timing of the prosodic contrast in speech (see e.g., de Ruiter, 1998). Prediction is not needed as such, because synchrony emerges from intrinsic bodily dynamics. This view of gesture-speech synchrony further fuels the idea that “biologically mandated” bodily dynamics (cf. McClave, 1994) define the space from which linguistic tools such as prosody can emerge and conventionalize (Fowler, 2010). Prosodic gesture-speech coordination are thus not arbitrary conventions, but *embodied innovations* that are exploited ontogenetically (Iverson & Thelen, 1999; Lee et al., 2009), and possibly serves the biological basis for what seems to be a fundamentally shared cross-cultural gesture-speech prosody correlation (de Ruiter & Wilkins, 1998). This embodied innovation further falls in line with coordination research with humans and other animals, such that

rhythmic movements (e.g., running or wheelchair propulsion) phase-lock to respiration cycles (Amazeen, Amazeen, & Beek, 2001; Bramble & Carrier, 1983) which results in a more stable entrained system, making it more efficient in its energy expenditure as opposed to when such coordination between respiration and locomotion is not achieved (Bernaconi, & Kohl, 1993). Indeed, synchronization phenomenon are pervasive in living and non-living systems (Pikovsky et al., 2001). Even brain oscillations in non-respiratory related brain regions are found to phase-lock with lower frequency respiration rhythms (Rojas-Libano, Wimmer del Solar, Aguilar-Rivera, Montefusco-Siegmund, & Maldanado, 2018). Gesture-speech synchrony might thus be filed under this general synchronization phenomenon, and synchronization may provide a way to stabilize rhythmic activity through generating physical impulses that make the system as a whole more stable in its (linguistic) activity. Indeed, there is evidence that gesture-speech synchrony is more pronounced where speech is interfered due to auditory feedback delay (Pouw & Dixon, 2018a, b), suggesting that gesture-speech entrainment is something than can be modulated to resist instabilities (also see evidence for beneficial effects of beat gestures in second language learning; Gluhareva & Prieto, 2017; Prieto, Llanes-Coromina, & Rohrer, 2018). Thus, in broad strokes, the current data support the idea that gesture-speech synchrony emerges from biomechanical constraints which possibly provide the system with new or improved (linguistic) stabilities.

Questions still need to be addressed about the precise nature of coupling between upper limb movement and phonation and how they relate to gesture. The first question is whether the coupling of upper-limb movement and acoustics entails “simply” a unidirectional effect of force traveling from the arm to lungs leading to changes in acoustic

energy? Such a mechanism is relatable to the now disproven visceral piston theory of locomotor-respiratory coupling, which entailed the that the vertical impulse produced during reactive forces of footfalls when running, provides a rhythmic perturbation to the diaphragm which leads to coupling of limb motions and breathing cycles (Amazeen, Amazeen, & Beek, 2001). Such a theory fell short in the end because it was found that similar rhythmic movements, such as wheel chair propulsion performed with upper limbs, coupled in similar ways with the respiratory system as is the case with running (Amazeen, Amazeen, & Beek, 2001). We too think that the current synchronization mechanism, while emerging from biomechanical constraints of the body as the current data show, involves an interactive synergy between a brain-limb-respiration system, in the case of occurrences of gestures-speech synchrony “in the wild”.

Indeed, the current evidence for physical effects of gesture on acoustics *cannot* explain all occurrences of gesture-speech synchrony. For example, beat gestures can be very small in their movement amplitude and still tightly synchronize with speech (McNeill, 2005), and head movements often serve as prosodic beats while they are not likely to have the physical impact that upper limb movements have (Krahmer & Swerts, 2004). Furthermore, gesture and speech can be affected in their temporal alignment when visual feedback of gesture or speech is perturbed (Chu & Kita, 2014; Rusziewicz et al., 2011; Pouw & Dixon, 2018a, b), and further show large standard deviations in their coupling (e.g., Loehr, 2004; McLave, 1994), suggesting that speech and gesture prosody is not a one-to-one coupling in spontaneous gesture. Finally, speech prosody is not only defined by peaks in F0 and amplitude, but by a myriad of other temporally dynamic features (e.g., modulation of syllable duration) that will likely also constrain gesture coordination. Thus,

bodily resonances cannot fully accommodate for gesture-speech synchrony in all contexts. Yet, it is possible that humans become sensitive to gesture as a reliable “kinesthetic marker” on acoustics, which can then be intentionally, if not ontogenetically exploited for communicative purposes. In other words, the currently discovered internal dynamics of gesture and phonation is something that may be intentionally modulated (at some point in development, and at some point in its emergence in language), as to be competently used in semiotic expression across a wider set of contexts (than the context in which it necessarily emerges). Relatedly, we do not want to deny that neural mechanisms are in play in the establishment and development of gesture-speech synchrony. Our results do however minimally show that the kinetic effects of movement should provide a cheap solution for synchronizing gesture and speech, as opposed to predicting and monitoring where a trajectory of the movement would align with speech’ prosodic contrasts as to time their activity in synchrony, which is maintained by information-processing theories of gesture (e.g., De Ruiter, 2000) or otherwise implied by theories that excluded direct effects of gesture on acoustics (e.g., McClave, 1994). A such, it is likely that neural mechanisms involved in gesture-speech synchrony were driven in their development to exploit *existent bodily dynamics*, rather than unnecessarily *emulate* the activity of gesture and speech as to synchronize their trajectories.

Future research could therefore focus on the precise nature of entrainment between phonation and gestural movements, as well as how current effects might play out in more naturalistic contexts (e.g., Loehr, 2004; Shattuck-Hufnagel & Ren, 2018). Specifically studies are needed that assess the role of physical impetus of naturalistic fluid gestures and its relation to fluid speech. Furthermore, research could focus on the possible structural

effects of upper limb movements on speech. For example, is it the case that providing speech with a consistent physical impulse through gesture, that such speech becomes more rhythmic? Does the degree or rhythmicity of the physical impulse modulate such effects of speech? Furthermore, we think developmental work could focus on how children start to harness internal bodily dynamics for communicative purposes (Iverson & Thelen, 1999; Esteve- Esteve-Gibert & Guellaï, 2018), with a specific focus on physical impetus.

Conclusion

The current results reveal that upper limb motions with high physical impact affect phonation acoustics in a way that accommodates gesture-speech synchrony as observed “in the wild”. The current findings further possibly obviate cognitively costly prediction mechanisms that link the gesture and speech system together in their activity and provides the basis for a cheap radically embodied solution where gesture-speech synchrony emerges out of biomechanical constraints, rather than top-down neural constraints. In any case, the current effects need to be accommodated in our understanding of gesture-speech synchrony, as well as our understanding of the emergence of multimodal language as a whole.

References

- Amazeen, P. G., Amazeen, E. L., & Beek, P. J. (2001). Coupling of breathing and movement during manual wheelchair propulsion. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1243.
- Aruin, A. S. & Latash, M. L. (1995) Directional specificity of postural muscles in feed-forward postural reactions during fast voluntary arm movements. *Experimental Brain Research, 103*, 323-332.
- Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*(2), 178-190. doi: 10.1016/j.neuropsychologia.2005.05.007
- Bernasconi, P., & Kohl, J. (1993). Analysis of co-ordination between breathing and exercise rhythms in man. *The Journal of physiology, 471*(1), 693-706.
- Bernstein, N. (1966). *The co-ordination and regulation of movements*. London: Pergamon Press.
- Bouhuys, A. (1974). *Breathing; physiology, environment and lung disease*. New York: Grune & Stratton.
- Bouisset, S., & Do, M. C. (2008). Posture, dynamic stability, and voluntary movement. *Neurophysiologie Clinique/Clinical Neurophysiology, 38*(6), 345-362. doi: 10.1016/j.neucli.2008.10.001
- Bouisset, S., & Zattara, M. (1981). A sequence of postural movements precedes voluntary movement. *Neuroscience letters, 22*(3), 263-270. doi: 10.1016/0304-3940(81)90117-8

- Bramble, D. M., & Carrier, D. R. (1983). Running and breathing in mammals. *Science*, 219(4582), 251-256.
- Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7), e1000436. doi: 10.1371/journal.pcbi.1000436
- Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, 143(4), 1726-1741. doi: 10.1037/a0036281.
- Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, 143(4), 1726-1741. doi: 10.1037/a0036281.
- Collyer, C. E., Broadbent, H. A., & Church, R. M. (1994). Preferred rates of repetitive tapping and categorical time production. *Perception & Psychophysics*, 55(4), 443-453. doi: 10.3758/BF03205301
- Cordo, P. J., & Nashner, L. M. (1982). Properties of postural adjustments associated with rapid arm movements. *Journal of neurophysiology*, 47(2), 287-302. doi: 10.1152/jn.1982.47.2.287
- Cravotta, A., Busà, M. G., & Prieto, P. (2018). Restraining and encouraging the use of hand gestures: Effects on speech. In *Proceedings 9th International Conference on Speech Prosody 2018*(pp. 206-210).
- Dawson J., & Cole, 2010. <http://www.thearticulatehand.com/ian.html>
- De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture*. New York: Cambridge University Press.

- de Ruiter, J., & Wilkins, D. (1998). The synchronization of Gesture and Speech in Dutch and Arrernte (an Australian Aboriginal language): A Cross-cultural comparison, *Oralité et Gestualité*, 603-607.
- Dromey, C., & Ramig, L. O. (1998). The effect of lung volume on selected phonatory and articulatory variables. *Journal of Speech, Language, and Hearing Research*, 41(3), 491-502.
- Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the Auditory and Visual Domains: A Developmental Perspective. *Frontiers in Psychology*, 9, 338. doi: 10.3389/fpsyg.2018.00338
- Fowler, C. A. (2010). Embodied, embedded language use. *Ecological Psychology*, 22(4), 286-303.
- Friedli, W. G., Hallett, M., & Simon, S. R. (1984). Postural adjustments associated with rapid voluntary arm movements 1. Electromyographic data. *Journal of Neurology, Neurosurgery & Psychiatry*, 47(6), 611-622.
- Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews*, 30(7), 949-960. doi: 10.1016/j.neubiorev.2006.02.004
- Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, 21(5), 609-631.
- Godin, K. W., & Hansen, J. H. (2015). Physical task stress and speaker variability in voice quality. *EURASIP Journal on Audio, Speech, and Music Processing*, 2015(1), 29.

- He, L., & Dellwo V. (2017). Amplitude envelope kinematics of speech signal: parameter extraction and applications. In: Trouvain, Jürgen; Steiner, Ingmar; Möbius, Bernd. Elektronische Sprachsignalverarbeitung 2017. Dresden: TUDpress, 1-8.
- He, L., & Dellwo, V. (2016). A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform. In *Proceedings Interspeech 2016* (pp. 530-534), San Francisco. doi: 10.21437/Interspeech.2016-1447
- Hodges, P. W., & Richardson, C. A. (1997a). Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Experimental Brain Research*, 114(2), 362-370.
- Hodges, P. W., & Richardson, C. A. (1997b). Relationship between limb movement speed and associated contraction of the trunk muscles. *Ergonomics*, 40(11), 1220-1230.
- Hoetjes, M., Krahmer, E., Swerts, M., (2013). Does our speech change when we cannot gesture? *Speech Communication*, 75, 257-267 . doi: <http://dx.doi.org/10.1016/j.specom.2013.06.007>
- Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6(11-12), 19-40.
- Jerome, S., Aubin, T., Simonis, C., Lelouch, L., Brown, E. C., Depraetere, M., Desjonquieres, C., et al. "Package 'seewave'." (2018).
- Johannes, B., Wittels, P., Enne, R., Eisinger, G., Castro, C. A., Thomas, J. L., ... & Gerzer, R. (2007). Non-linear function model of voice pitch dependency on physical and mental load. *European Journal of Applied Physiology*, 101(3), 267-276.

- Krahmer, E., & 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), 396-414. doi: [10.1016/j.jml.2007.06.005](https://doi.org/10.1016/j.jml.2007.06.005)
- Krauss, R. M., Chen, Y., & Gotfexnum, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill. (Ed). *Language and gesture*. New york: Cambridge University Press.
- Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory Phonology*, 8(1), 1-36. doi: 10.5334/labphon.75.
- Lee, D. N., Bootsma, R. J., Land, M., Regan, D., & Gray, R. (2009). Lee's 1976 paper. *Perception*, 38(6), 837-858.
- Leonard, T., Cummins, F. (2010). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457-1471. doi: 10.1080/01690965.2010.500218.
- Lieberman, P. (1996). Some biological constraints on the analysis of prosody. In J. L. Morgan & K. Demuth (Eds). *Signal to Syntax* (pp. 67-78). Mahwah: Lawrence Erlbaum Associates.
- Lieberman, P., Knudson, R., and Mead, \$. (1969). Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *Journal of the Acoustic Society of America*, 45, 1537- 1543.
- Loehr, D. P. (2004). Gesture and intonation (Unpublished doctoral dissertation). Georgetown University, Washington, DC.
- McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, 23(1), 45-66. doi: 10.1007/BF02143175

- McClave, E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research*, 27, 69–89.
- McNeill, D (2005). *Gesture and Thought*. Chicago: University of Chicago press.
- Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American journal of psychology*, 411-424. doi: 10.2307/4149008
- Nobe, S. (1996). Representational gestures, cognitive rhythms, and acoustic aspects of speech.
- Orlikoff, R. F., & Baken, R. J. (1989). Fundamental frequency modulation of the human voice by the heartbeat: preliminary results and possible mechanisms. *The Journal of the Acoustical Society of America*, 85(2), 888-893. doi: 10.1121/1.397560
- Parrell, B., Goldstein, L., Lee, S., & Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of phonetics*, 42, 1-11. doi: 10.1016/j.wocn.2013.11.002
- Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in nonlinear sciences*. Cambridge: Cambridge University Press.
- Pouw, W. T. J. L. & Hostetter, A. (2016). Gesture as predictive action. *Reti, Saperi, Linguaggi: Italian Journal of Cognitive Sciences*, 3, 57-80. doi: 10.12832/83918
- Pouw, W., & Dixon, J. A. (2018a; unpublished preprint). Effects of delayed auditory feedback on gesture-speech synchrony: Pre-registration and exploratory study. doi: 10.17605/OSF.IO/3UHPV
- Pouw, W., & Dixon, J. A. (2018b; unpublished preprint). Entrainment and modulation of gesture-speech synchrony under delayed auditory feedback. doi: 10.17605/OSF.IO/AVJ7M

- Pouw, W., Harrison, S. J., & Dixon, J. A. (2018; unpublished preprint). The physical basis of gesture-speech synchrony. doi: 10.31234/osf.io/9fzsv
- Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: a labelling proposal. In *Proc. 9th International Conference on Speech Prosody 2018* (pp. 201-205).
- Prieto, V., P., Llanes-Coromina, J., & Rohrer, P. L. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. In *Klessa K, Bachan J, Wagner A, Karpiński M, Śledziński D. Proceedings of the 9th International Conference on Speech Prosody; 2018 June 13-16; Poznań, Poland.* [Lous Tourils]: ISCA; 2018. p. 498-502. DOI: 10.21437/SpeechProsody.2018-101. International Speech Communication Association (ISCA).
- Richardson, M. J. (n.d.). Retrieved from <http://xkiwilabs.com/software-toolboxes/>
- Rojas-Líbano, D., Wimmer del Solar, J., Aguilar-Rivera, M., Montefusco-Siegmund, R., & Maldonado, P. E. (2018). Local cortical activity of distant brain areas can phase-lock to olfactory bulb's respiratory rhythm in freely behaving rat. *Journal of neurophysiology*.
- Rösch, A., & Schmidbauer, H. (2014). WaveletComp: Computational wavelet analysis. R package version 1.0. URL <https://cran.r-project.org/package=WaveletComp>.
- Rösch, A., & Schmidbauer, H. (2016). WaveletComp 1.1: A guided tour through the R package. URL: http://www.hs-stat.com/projects/WaveletComp/WaveletComp_guided_tour.pdf

- Rusiewicz, H. L. (2011). Synchronization of speech and gesture: A dynamic systems perspective. In *proceedings 2nd Gesture and Speech in Interaction* (GESPIN), Bielfeld, Germany.
- Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, 57, 283-300. doi: 10.1016/j.specom.2013.06.004.
- Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, 57, 283-300. doi: 10.1016/j.specom.2013.06.004.
- Rusiewicz, H., L., & Esteve-Gibert, N. (2018). Temporal coordination of prosody and gesture in the development of spoken language production. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition*. Amsterdam: John Benjamins.
- Seilmayer, M. (2016). Common Methods of Spectral Data Analysis: Package 'spectral'. Retrieved from <https://cran.r-project.org/web/packages/spectral/spectral.pdf>
- Seilmayer, M. (2016). spectral: Common Methods of Spectral Data Analysis. Retrieved from <https://cran.r-project.org/web/packages/spectral/index.html>
- Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in psychology*, 9. doi: 10.3389/fpsyg.2018.01514
- Silva, P., Moreno, M., Mancini, M., Fonseca, S., & Turvey, M. T. (2007). Steady-state stress at one hand magnifies the amplitude, stiffness, and non-linearity of oscillatory behavior at the other hand. *Neuroscience Letters*, 429(1), 64-68.

speech: A network/threshold model of gesture production. Unpublished dissertation, University of Chicago.

Sueur, J., Aubin, T., Simonis, C., Lelouch, L., Brown, E. C., Depraetere, M., ... & LaZerte, S. (2018). Sound Analysis and Synthesis: Package 'seewave'. Retrieved from <http://cvsup3.pl.freebsd.org/pub/mirrors/CRAN/web/packages/seewave/seewave.pdf>

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, 134(1), 628-639. doi: 10.1121/1.4807565

Treffner, P., & Peter, M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science*, 21(5-6), 641-697. doi: 10.1016/S0167-9457(02)00178-1

Turvey, M. T., & Fonseca, S. T. (2014). The medium of haptic perception: A tensegrity hypothesis. *Journal of Motor Behavior*, 46(3), 143-187. doi: 10.1080/00222895.2013.798252

Wagner, P., Malisz, Z., & Kopp, S (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209-232. doi: 10.1016/j.specom.201

Zelic, G., Kim, J., & Davis, C. (2015). Articulatory constraints on spontaneous entrainment between speech and manual gesture. *Human Movement Science*, 42, 232-245. doi: 10.1016/j.humov.2015.05.009