Gesture-Speech Physics in Fluent Speech and Rhythmic Upper Limb Movements

|  |
| --- |
| Wim Pouw1,2,3, Lisette de Jonge-Hoekstra1,4, Steven J. Harrison1, Alexandra Paxton1,5, & James A. Dixon1 |
| 1 Center for the Ecological Study of Perception and Action, University of Connecticut |
| 2 Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen |
| 3 Institute for Psycholinguistics, Max Planck Nijmegen |
| 4 University of Groningen |
| 5 Department of Psychological Sciences, University of Connecticut |
|  |

# Author note

Correspondence concerning this article should be addressed to Wim Pouw, Donders Institute for Brain, Cognition and Behaviour, Heyendaalseweg 135, 6525 AJ Nijmegen. E-mail: [w.pouw@psych.ru.nl](mailto:w.pouw@psych.ru.nl)

Abstract

Communicative hand gestures are often coordinated with prosodic aspects of speech. Therein, salient moments of gestural movement (e.g., quick changes in speed) often co-occur with salient moments in speech (e.g., near peaks in fundamental frequency and intensity). A common understanding is that such gesture and speech coordination is culturally and cognitively acquired rather than having a biological basis. Recently, however, the physical coupling of arm movements to speech movements via myofascial tissue biomechanics has been investigated as a potentially important factor in understanding the emergence of gesture-speech coordination. Specifically, in the case of steady-state vocalization and mono-syllable utterances, it has been shown that forces produced during gesturing are transfered onto the tensioned body leading to changes in respiratory-related activity and thereby affecting vocalization F0 and intensity during steady-state vocalization and mono-syllable utterances. In the current experiment (N = 34), we show that gesture-speech physics is relevant for fluent speech, too. Compared with non-movement, participants who are producing fluent self-formulated speech while rhythmically moving their limbs demonstrate heightened F0 and amplitude envelope, and such effects are more pronounced for higher-impulse arm versus lower-impulse wrist movement. We replicate that acoustic peaks arise especially during moments of peak-impulse (i.e., the beat) of the movement, namely around deceleration phases of the movement. Finally, higher deceleration rates of higher-mass arm movements were related to higher peaks in acoustics, supporting a role for physical impulse transmissions of gesture onto the tensioned body, affecting speech. These results further support a radically embodied account of the origins of human gesture.

*Keywords:* hand gesture, speech production, speech acoustics, biomechanics, entrainment

Word count: X

Gesture-Speech Physics in Fluent Speech and Rhythmic Upper Limb Movements

Communicative hand gestures are ubiquitous across human cultures. Gestures aid communication by seamlessly interweaving relevant pragmatic, iconic and symbolic expressions of the hands together with speech (Feyereisen, 2017; Holler & Levinson, 2019; Streeck, 2008). For such multi-articulatory utterances to do their communicative work, gesture and speech must be tightly temporally coupled to form a sensible speech-gesture whole. And in fact, gestures’ salient moments are often timed with emphatic stress made in speech, no matter what the hands depict (Shattuck-Hufnagel & Prieto, 2019; Wagner, Malisz, & Kopp, 2014). For such exquisite gesture-speech coordination to get off the ground, the system must functionally constrain its degrees of freedom (Turvey, 1990), and in doing so it will have to utilize (or otherwise account for) intrinsic dynamics arising from the bio-physics of speaking and moving at the same time.  
In this report we provide evidence that movement of the upper limbs constrain fluent self-generated speech acoustics through biomechanics. This replicates and extends results obtained with rudimentary steady-state vocalization and mono-syllabic utterances showing that physical impulses generated by upper limb movements reach the respiratory system, affecting vocalization acoustics (Pouw et al., 2019a, 2019b; Pouw, Paxton, Harrison, & Dixon, 2020). We thereby show that there is a further complexity to the human voice in that it is directly constrained by gestural movement.

## The gesture-speech prosody link

The tight coordination of prosodic aspects of speech with the kinematics of gesture has been long appreciated and is classically referred to as the beat-like quality of co-speech gesture (McNeill, 1992). This phenomenon has been studied in several ways. Human coders have been trained to identify salient moments called the apex in gestures, together with prosodically meaningful moments in fluent speech. That research tradition has found that gesture apices often align with *pitch accents* in speech—accents which are acoustically predominately defined by positive excursions in the fundamental frequency (F0), lowering of the second formant, longer vowel duration, and increased intensity (Loehr, 2012; McClave, 1998; Mendoza-Denton & Jannedy, 2011). Pitch accents can be perceptually differentiated by sudden lowering of F0 as well, but gestures do not seem to align with those events quite as much (Im & Baumann, 2020).  
In more recent motion-tracking studies, these gesture-speech prosody correlations have been obtained as well, showing—for example—that gestures’ peak velocity often co-occurs near peaks in F0, even when such gestures are depicting something (Danner, Barbosa, & Goldstein, 2018; Krivokapić, 2014; Leonard & Cummins, 2011; Pouw & Dixon, 2019a, 2019b). In pointing gestures, stressed syllables align neatly with the maximum extension of the pointing movement, such that the hand movement terminates at the first syllable utterance in strong-weak stressed “PApa” and terminates later during the second syllable utterance in the weak-strong “paPA” (Esteve-Gibert & Prieto, 2013; Rochet-Capellan, Laboissière, Galván, & Jean-Luc, 2008). During finger-tapping and mono-syllabic utterances, when participants are instructed to alternate prominence in their utterances (“pa, PA, pa, PA”), the tapping action spontaneously aligns with the syllable pattern, such that larger movements are made during stressed syllables (Parrell, Goldstein, Lee, & Byrd, 2014). Conversely, if participants are instructed to alternate stress in finger tapping (STRONG-weak-STRONG-weak force production), speech will follow, with larger oral-labial apertures for stressed vs. unstressed tapping movements.  
Even when people do not intend to change the stress patterning of an uttered sentence, gesturing concurrently affects speech acoustics in a way that makes it seem intentionally stressed, inducing an increase in vocalization duration and a lowering of the second formant of co-occurrent speech (Krahmer & Swerts, 2007). Further, gesture and speech cycle rates seem to be attracted towards particular (polyrhythmic) stabilities: In-phase speech-tapping is preferred over anti-phase coordination, and 2:1 speech-to-tapping ratios are preferred over more complex integer ratios such as 5:2 (Kelso & Tuller, 1984; Stoltmann & Fuchs, 2017; Treffner & Peter, 2002; Zelic, Kim, & Davis, 2015). All these previous results indicate that gesture and speech naturally couple their activity, raising questions about origins of its pervasiveness.

## Gesture-speech physics

The current mainstream understanding of the gesture-prosody link is that is not “biologically mandated” (p.69 ; McClave, 1998) but culturally acquired, and requires neural-cognitive timing mechanisms (Ruiter, 2000) that appear only after 16 months of age (Iverson & Thelen, 2005).  
Recently, however, the physical coupling of arm movements with speech via myofascial tissue biomechanics has been investigated, where it was found that hand gesturing physically impacts steady-state vocalizations and mono-syllabic consonant-vowel utterances (Pouw et al., 2019a, 2019b, 2019d, 2020). Specifically, hand and arm movements can transfer a force (a physical impulse) onto the musco-skeletal system, thereby modulating respiration-related muscle activity, leading to changes in vocalization’s intensity. If gesture-induced impulses are not accommodated for by vocal fold adjustments, the Fundamental Frequency (F0) of vocalizations is affected as well. It has been found that higher-impulse arm- versus wrist movements, or dual- versus one handed movements, will induce more pronounced effects on F0 and intensity. This is because the mass of the “object” in motion is higher for arm(s) versus wrist movements thereby changing the momentum of the effector, everything else—such as effector speed—being equal (as momentum effector = effctor mass x effector velocity). The change in momentum is the physical impulse, and the moment where physical impulse is highest, is when the change in velocity (i.e., acceleration) is highest (everything else—such as effector mass—being constant).  
The way in which physical impulses are absorbed by the respiratory system is likely complex and not a simple linear function (Levin, 2006). However, a complete understanding will involve an appreciation of the body as a pre-stressed system (Bernstein, 1967; Profeta & Turvey, 2018), forming an interconnected tensioned network of compressive (e,g., bones) and tensile elements (e.g., fascia, muscles) through which forces may reverberate nonlinearly (Silva, Moreno, Mancini, Fonseca, & Turvey, 2007; Turvey & Fonseca, 2014). Specifically, the upper limb movements are controlled by stabilizing musculo-skeletal actions of the scapula and shoulder joint, which directly implicate accessory expiratory muscles that also stabilize scapula and shoulder joint actions (e.g., the serratus anterior inferior; see Pouw et al., 2019a for an overview).  
Peripheral actions also play a role, as performing an upper limb movement recruits a whole kinetic chain of muscle activity around the trunk (e.g., rector abdominus) to maintain posture (Hodges & Richardson, 1997). Indeed, when people are standing vs. sitting for example, it has been found that the effects of peak physical impulse of gestures onto vocalization acoustics is more pronounced (Pouw et al., 2019b). We reasoned that this is because standing involves more forceful anticipatory postural counter adjustments (Cordo & Nashner, 1982) which reach the respiratory system via accessory expiratory muscles also implicated in keeping postural integrity. Recently, more direct evidence has been found for the gesture-respiratory-speech link: Respiratory-related activity (measured with a respiratory belt) was enhanced during moments of peak-impetus of gesture as opposed to other phases in the gesture movement, and respiratory-related activity itself was related to gesture-related intensity modulations of mono-syllable utterances (Pouw et al., 2019a).  
The evidence reviewed so far has been based on experiments on continuous vocalizations or monosyllabic utterances. Such results cannot directly generalize to fluent, self-generated, full-sentenced speech, although promising indications suggest that gesture-speech physics does generalize to fluent speech. A recent study found that encouraging participants to gesture during cartoon-narration versus giving no instructions lead to 22Hz increase in observation of max F0 and to greater F0 ranges of speech and intensity (Cravotta, Busà, & Prieto, 2019). Furthermore, computational modelers have reported on interesting successes of synthesizing gesture kinematics based on speech acoustics alone (Ginosar et al., 2019; Kucherenko, Hasegawa, Henter, Kaneko, & Kjellström, 2019), indicating that information about body movements inhabits the speech signal. Although promising, such results do not necessitate a role for biomechanics but only suggests a strong connection between gesture and speech.

## Current experiment

The current experiment was conducted as a simple test of the constraints of upper limb movement on fluent speech speech acoustics. Participants were asked to retell a cartoon scene that they had just watched, while either not moving, vertically moving their wrist, or vertically moving their arm at a tempo of 80 beats per minute (1.33Hz). Participants were asked to give a stress or beat in the downward motion with a sudden stop at maximum extension (i.e., sudden deceleration). Participants were asked to not allow movements to affect their speaking performance in any way. Similar to previous experiments (e.g., Pouw et al., 2019c), we assessed the following to conclude that gesture-speech physics is present: \* 1) Does rhythmic co-speech movement change acoustic markers of prosody (i.e., F0 and amplitude envelope)?  
\* 2) At what moments of co-speech movement is change in acoustics observed? \* 3) Does degree of physical impulse (as measured by effector mass or changes in speed) predict acoustic variation?

# Method

## Participants & Design

We tested a total of 37 participants 37 participants (*M* age = 18.76, *SD* age = 0.95, %cis-gender female = 67.57, %cis-gender male = 32.43, %right-handed = 94.59), drawn from an undergraduate participant pool at the University of Connecticut. There were 6 non-native English speaking Asian-undergraduates in the current sample.  
The current design was fully-within subject, with a three-level movement manipulation (passive Vs. wrist-movement Vs. arm-movement condition). Participants performed 419 trials in total lasting about 40 seconds with Movement condition randomly assigned per trial. The study design was approved by the IRB committee of the University of Connecticut (#H18-227).

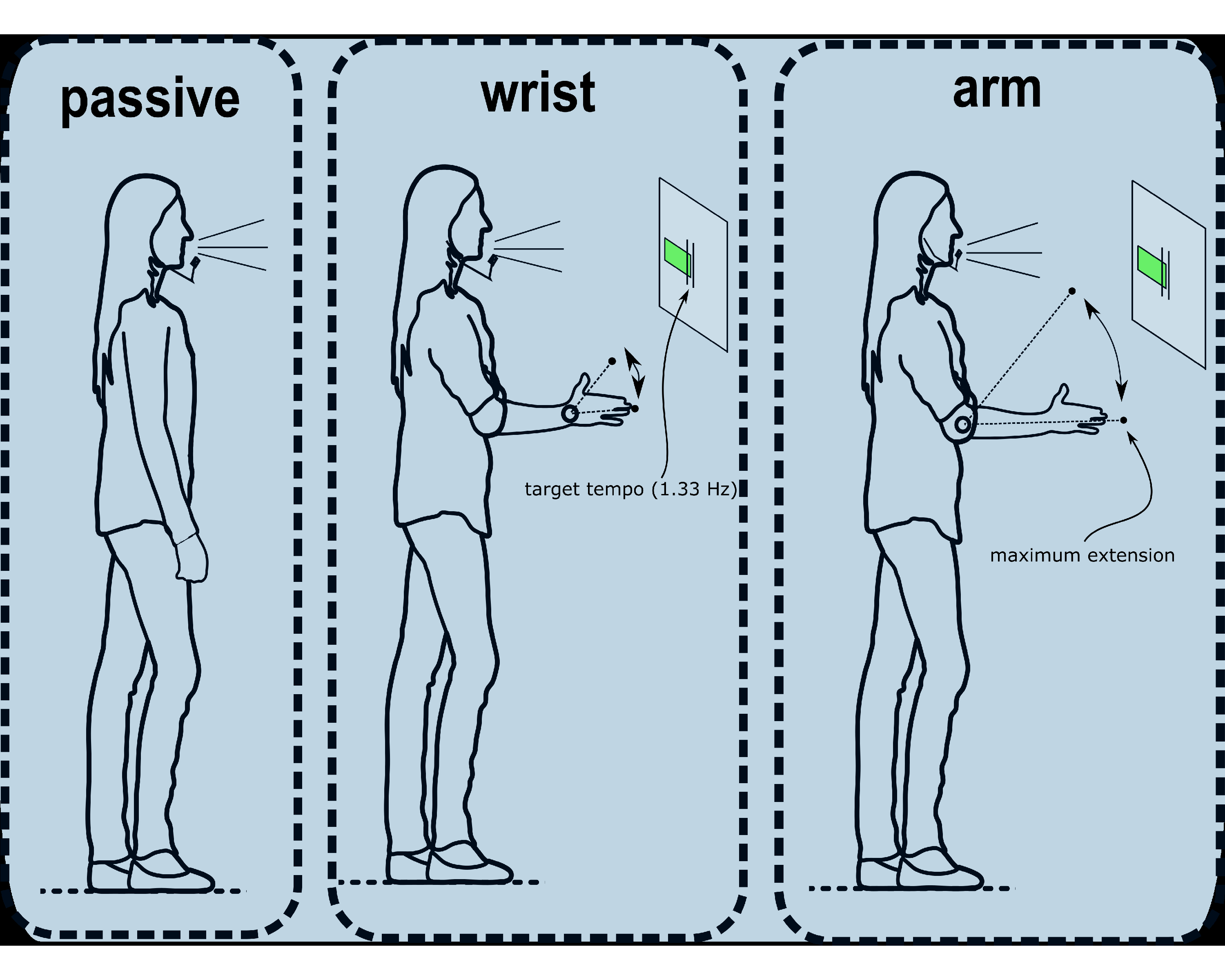
## Material & Equipment

### Cartoon vignettes

Twelve cartoon vignettes were created from the “Canary Row” and “Snow Business” Tweety and Sylvester cartoons, *M* duration vignettes = 59.42seconds (*SD* = 32.11). These cartoons are often used in gesture research (McNeill, 1992). The videos can be accessed here: <https://osf.io/rfj5x/>.

### Audio and Motion Tracking

A MicroMic C520 cardioid condenser microphone headset (AKG, Inc.) was used to record audio. A Polhemus Liberty motion tracking system (Polhemus, Inc.) was used to record the position of the participant’s index finger of the dominant hand, sampling with one 6D sensor at 240 Hz. We applied a first-order Butterworth filter at 30 Hz for the vertical position (z) traces and its derivatives. ## Procedure  
Upon arrival, participants were briefed that this 30-minute experiment entailed retelling cartoon scenes while standing and performing upper-limb movements. A motion sensor was attached to the tip of the index finger of the dominant hand, and a microphone headset was put on. Participants were asked to stand upright and were introduced to three movement conditions (see Figure 1). In the passive condition, participants did not move and kept their arm resting alongside the body. In the wrist-Movement condition, participants were asked to continuously vertically move the hand at the wrist joint while keeping the elbow joint at 90 degrees. In the arm-Movement Condition, participants moved their arm at the elbow joint, without wrist movement. Similar to previous studies (e.g., Pouw et al., 2020)], participants were asked to give emphasis in the downward motion of the movement with a sudden halt—in other words, a beat—at the maximum extension of their movement.  
After introduction of the movements, participants were told that they were to move at a particular tempo, indicated by visual feedback system. The feedback system consisted of a horizontal bar that continually updated to the report on the participant’s movement speed in previous movement cycle. The participant was to keep the horizontal bar between two low and higher boundaries (a 20% region, [72-88] BPM) of the 1.33-Hz target tempo (i.e., 80 BPM). Participants briefly practiced moving at the target rate before starting the experiment. Critically, the participants were not exposed to an external rhythmic signal, like a visual metronome. Subsequently, participants were instructed that they would watch cartoon clips that they would then retell while making one of the instructed movements (or making no movements). Participants were asked to keep their speech as normal as possible while making the movements (or no movement). In the conditions requiring movement, participants were to keep their movement tempo within the target range.  
Subsequently, participants were instructed that they would watch cartoon clips which they would retell after having watched it, while at the same time making one of the instructed movements (or making no movements). Participants were asked to keep their speech as normal as possible while making the movements (or no movement). When moving while speaking, participants were to keep their movement tempo within the target range.  
Twelve cartoon vignettes were readied to be shown before each trial, but if the total experiment time exceeded 30 minutes, the experiment would be terminated without all scenes being retold. To ensure that all Movement conditions would be performed at least once within that time, we set the maximum time per trial at 1 minute. In other words, when participants were still retelling the same scene after 60, seconds the experimenter would terminate the trial and move to the next trial. Mean retelling time was, however, well below 1 minute *M* = 26.00 seconds, *SD* = 7.06.

Figure 1. Graphical overview of movement conditions  *Note*. Movement conditions are shown. Each participant performed all conditions (i.e., within-subjects). To ensure that movement tempo remained relatively constant participants were shown a moving green bar, which indicated whether they moved too fast or too slow relative to a 20% target region of 1.33Hz. Participants were instructed to have an emphasis in the downbeat with an abrupt stop (i.e., beat) at the maximum extension.

## Preprocessing

### Speech acoustics

The fundamental frequency was extracted with sex-appropriate preset ranges (male = 50-400Hz; female = 80-640Hz). We used a previously written (Pouw & Trujillo, 2019) R script (<https://osf.io/m43qy/>) utilizing the R package ‘wrassp’ (Winkelmann, Bombien, & Scheffers, 2018) which applies a K. Schaefer-Vincent algorithm. We also extracted a smoothed (5-Hz Hann window) amplitude envelope using a previously written custom-written R script (<https://osf.io/uvkj6/>, which reimplements in R a procedure from He & Dellwo (2017).

### Data and Exclusions

We collected 189.70 minutes of continuous data (passive condition = 63.45, wrist-movement condition = 63.56, arm-movement = 62.69). However, due to a c++ memory allocation error the precise timing data for the motion-tracking was lost for a smaller portion of the data. Thus the final dataset consists of 124.49 minutes of continuous speech and movement data (passive = 40.08, wrist-movement condition = 42.32, arm-movement condition = 42.10).

## Manipulation Checks and Baseline

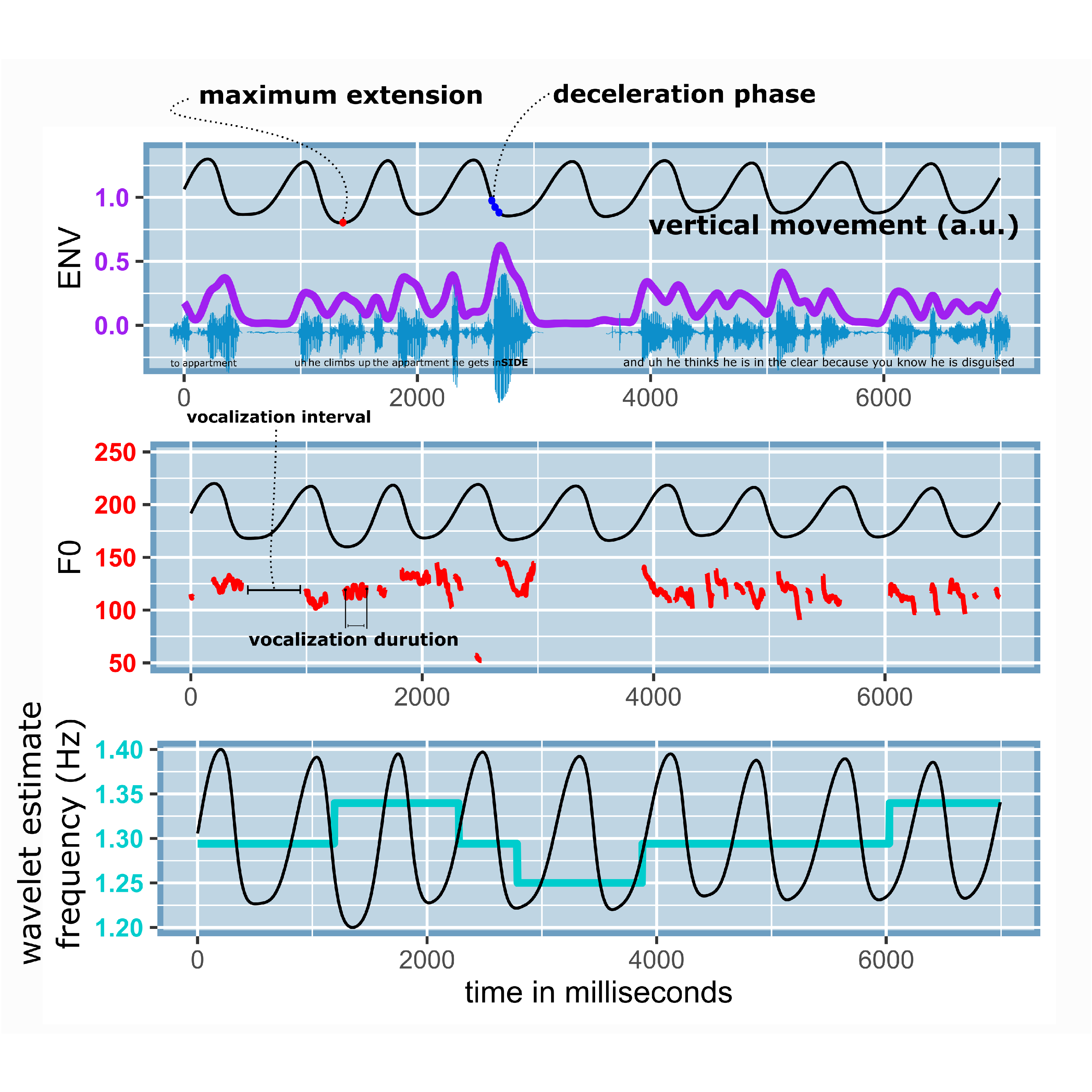
For gesture-speech analysis we also created a surrogate condition using speech of a passive condition trial and randomly pairing this with motion-tracking data from a movement condition for that participant (without scrambling the order). This surrogate randomly paired condition allowed us to exclude the possibility that any effects of movement were due to chance coupling inherent to the structure of speech and movement. We only use this surrogate control condition as a contrast when we are performing analysis on the temporal relation between speech and movement.  
We computed the following measures to check whether our movement manipulation was successful and whether speech rates were comparable across conditions. Figure 2 shows a summary of the results for key manipulation check measures.

### Movement Frequency

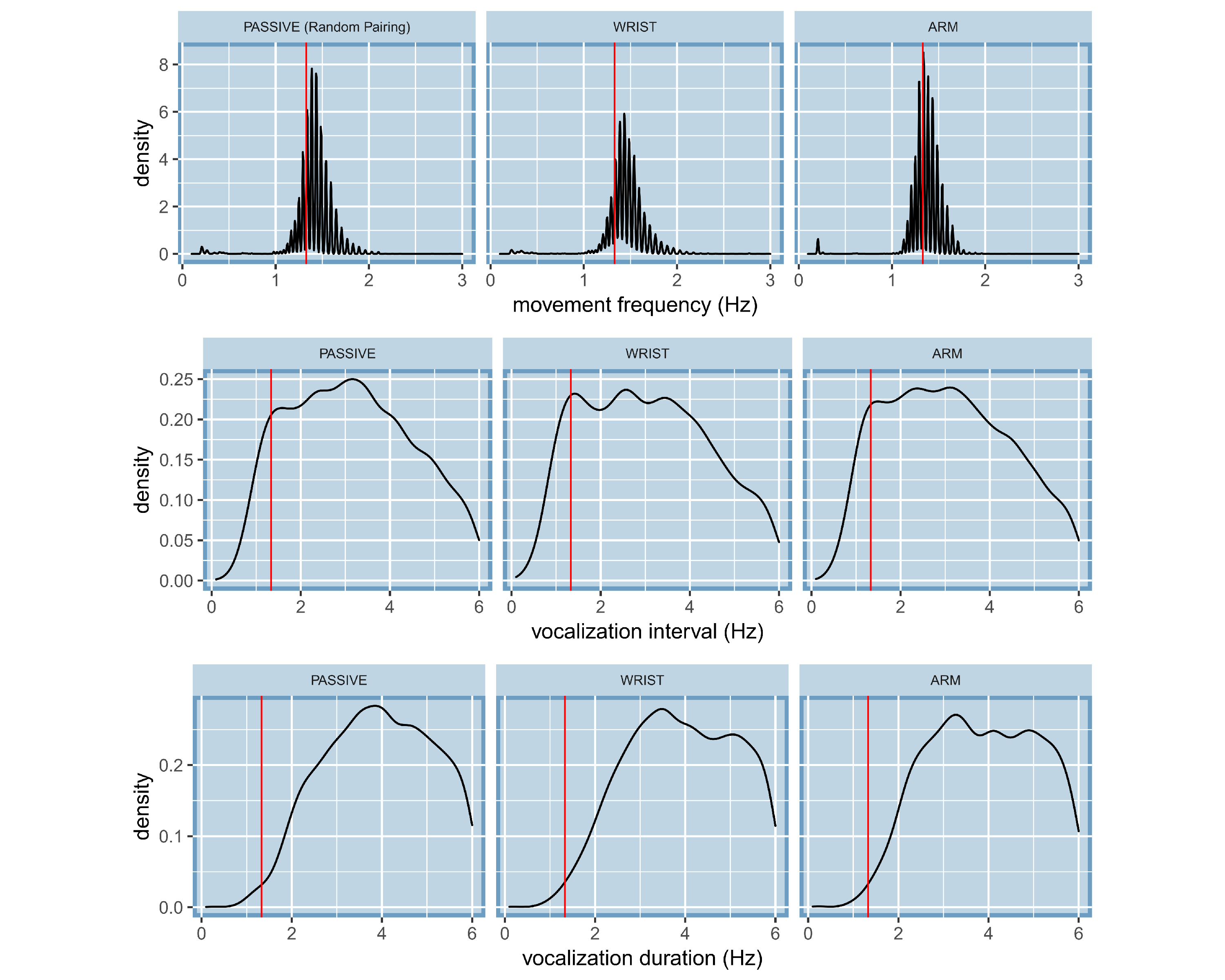
To ascertain whether participants moved their limbs within the target 1.33-Hz range, we performed a wavelet-based analysis (using R-package “WaveletComp”; Rosch & Schmidbauer, 2014). Figure 2 shows that wrist movements were performed at slightly faster rates *M* = 1.44 Hz, *SD* = 0.24, than arm movements, *M* = 1.36 Hz, *SD* = 0.19, but in both cases the movements were distributed over the range 1.33Hz. This confirms that our movement manipulation was successful. For our surrogate control condition, the mean frequency of the artificially paired movement time series fell between both arm- and wrist-movement condition frequency distributions, *M* = 1.41 Hz, *SD* = 0.22.

### Speech Rate

We calculated two measures to provide an indication of speech rate (see Fig. 2 for examples), namely vocalization duration and vocalization interval. Figure 3 shows relatively uniform distributions for speech measures. This shows no clear 1:1 frequency couplings of movement and vocalization duration or vocalization interval nor any other clear signs of poly-rhythmic coupling of movement and speech as has been observed in basic tapping paradigms (Zelic et al., 2015). Thus, we restrict ourselves for the current report to speech vocalization acoustics, instead of focusing on possible temporal changes of speech rate produced under rhythmic movement (e.g., Stoltmann & Fuchs, 2017).  
To compare vocalization rates to movement, we computed the average vocalization duration for each trial by tracking the time of uninterrupted runs of F0 observations and then converting the time in milliseconds to Hz. For the passive condition, the average vocalization duration was *M* = 6.28 Hz, *SD* = 6.03. For the wrist-movement condition the vocalization duration was *M* = 6.24 Hz, *SD* = 5.96, and for the arm-movement condition this was *M* = 6.08 Hz, *SD* = 5.83.  
The average vocalization interval for the passive condition was *M* = 5.17 Hz, *SD* = 6.94. For the wrist-movement condition the vocalization interval was *M* = 5.02 Hz, *SD* = 6.86, and for the arm-movement condition *M* = 4.86 Hz, *SD* = 5.76.

Figure 2. Example movement-, amplitude envelope-, F0- time series, and time-dependent movement frequency estimates 

*Note figure 2.* A sample of about 10 seconds is shown. With the participant’s permission the speech sample is available at <https://osf.io/2qbc6/>. The smoothed amplitude envelope in purple traces the waveform maxima. The F0 traces show the concomitant vocalizations in Hz, with an example of vocalization interval and vocalization duration (which was calculated for all vocalizations). The bottom panel shows the continuously estimated movement frequency in cyan, which hovers around 1.33 Hz. In all these panels, the co-occurring movement is plotted in arbitrary units (a.u.) to show the temporal relation of movement phases and the amplitude envelope, F0, and the movement frequency estimate. In our analysis, we refer to the maximum extension and deceleration phases as relevant moments for speech modulations. In this example, a particularly dramatic acoustic excursion occurs during a moment of deceleration of the arm movement, possibly an example of gesture-speech physics.

Figure 3. Summary of movement-frequency, vocalization duration and vocalization interval  


*Note Figure 3*. Density distributions of movement frequencies, vocalization interval, and vocalization duration are shown. Note, that for the passive condition there was no movement, but we have in this case Random Pairing movement time series for the passive (Random Pairing) condition for which frequency information is shown. The red vertical line indicates the target movement frequency at 1.3Hz.

# Results

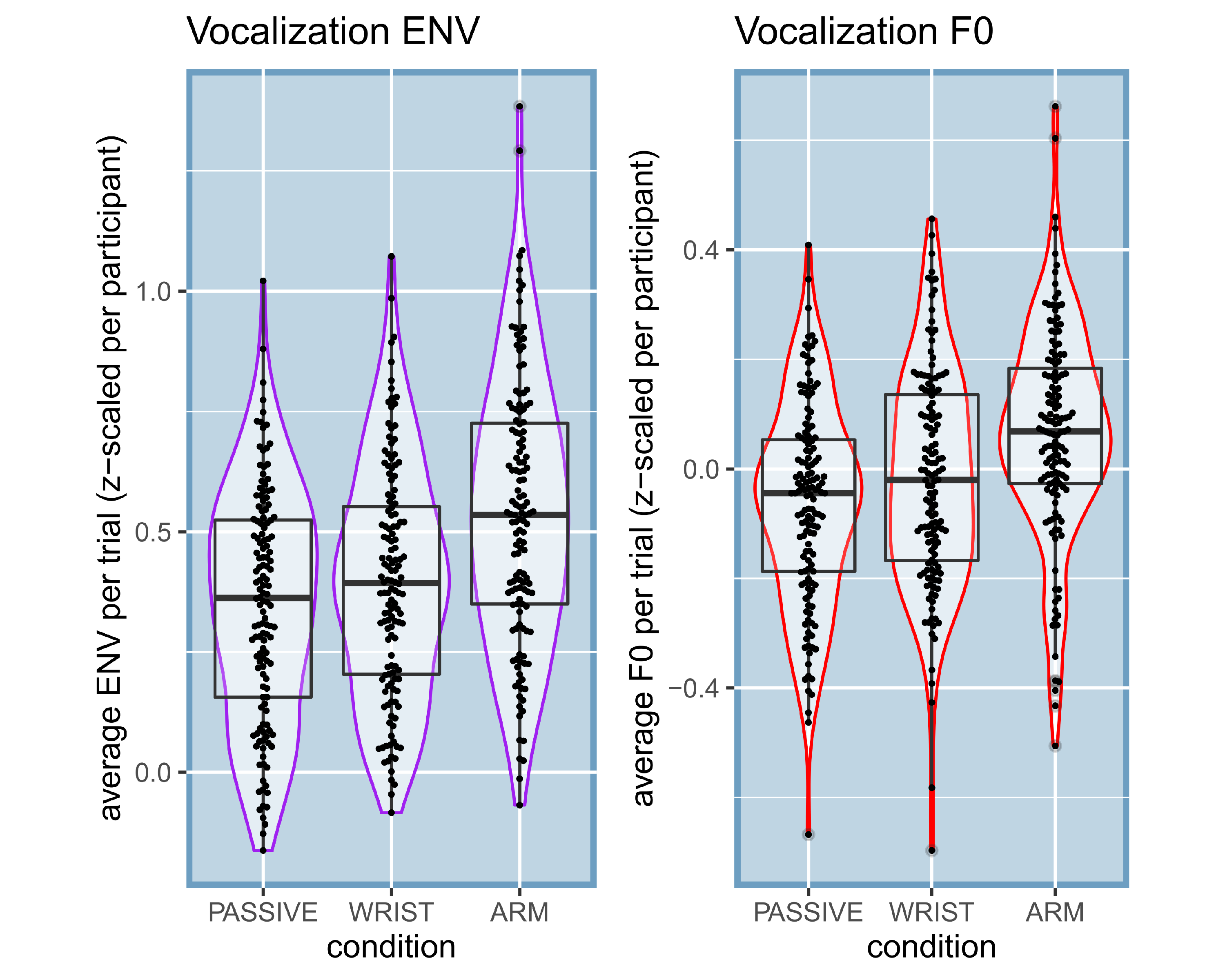
## Overview analysis

We report three main analyses to show that gesture-speech physics is present in fluent speech. Firstly, we assess overall effects of movement condition on vocalization acoustics (F0 and the amplitude envelope); these would support our hypothesis that (especially high impulse) upper limb movement constrains fluent speech acoustics. Secondly, we assess whether vocalization acoustic modulations are observed at particular phases of the movement cycle, which gesture-speech physics holds should occur at moments of peaks in deceleration. Thirdly, we assess whether a continuous estimate of upper limb physical impulse through deceleration rate predicts vocalization acoustic peaks, which would support the gesture-speech physics hypothesis that physical impulses are transferred onto the vocalization system. The following generally applies to all analyses. For hypothesis testing, we performed mixed linear regression models (using R-package “nlme”; Pinheiro, Bates, DebRoy, Sarkar, & R Team, 2019), and non-linear generalized additive modeling or GAM (using R-package “gam”; Hastie, 2019) with random intercept for participants by default.

## Acousic correlates of movement condition

Figure 4 shows the average F0 and amplitude envelope (z-scaled for participants) per trial per condition. The passive condition had generally lower levels of F0 and Amplitude Envelope as compared to the arm- and wrist-Movement conditions. Furthermore, the higher-impulse arm-movement condition generally had higher levels of F0 and Amplitude envelope as compared to lower-impulse wrist-movement condition.  
Table 1 shows the results of mixed linear regression analysis. For the amplitude envelope, the passive condition had a lower average amplitude envelope as compared to the the wrist-movement condition (*p* < .005), as well as the arm-movement condition (*p* < .0001). We further obtain that after accounting for differences in F0 for gender (males had generally 73Hz lower F0), wrist-movement condition has about 1.6 Hz increase in average as compared to the passive condition, but this was not statistically reliable (*p* = .059). Further, the arm-movement condition increased in F0 by 3.5 Hz over the passive Condition (*p* < .001).

Figure 4. Average F0 and Amplitude Envelope (ENV) per trial per condition



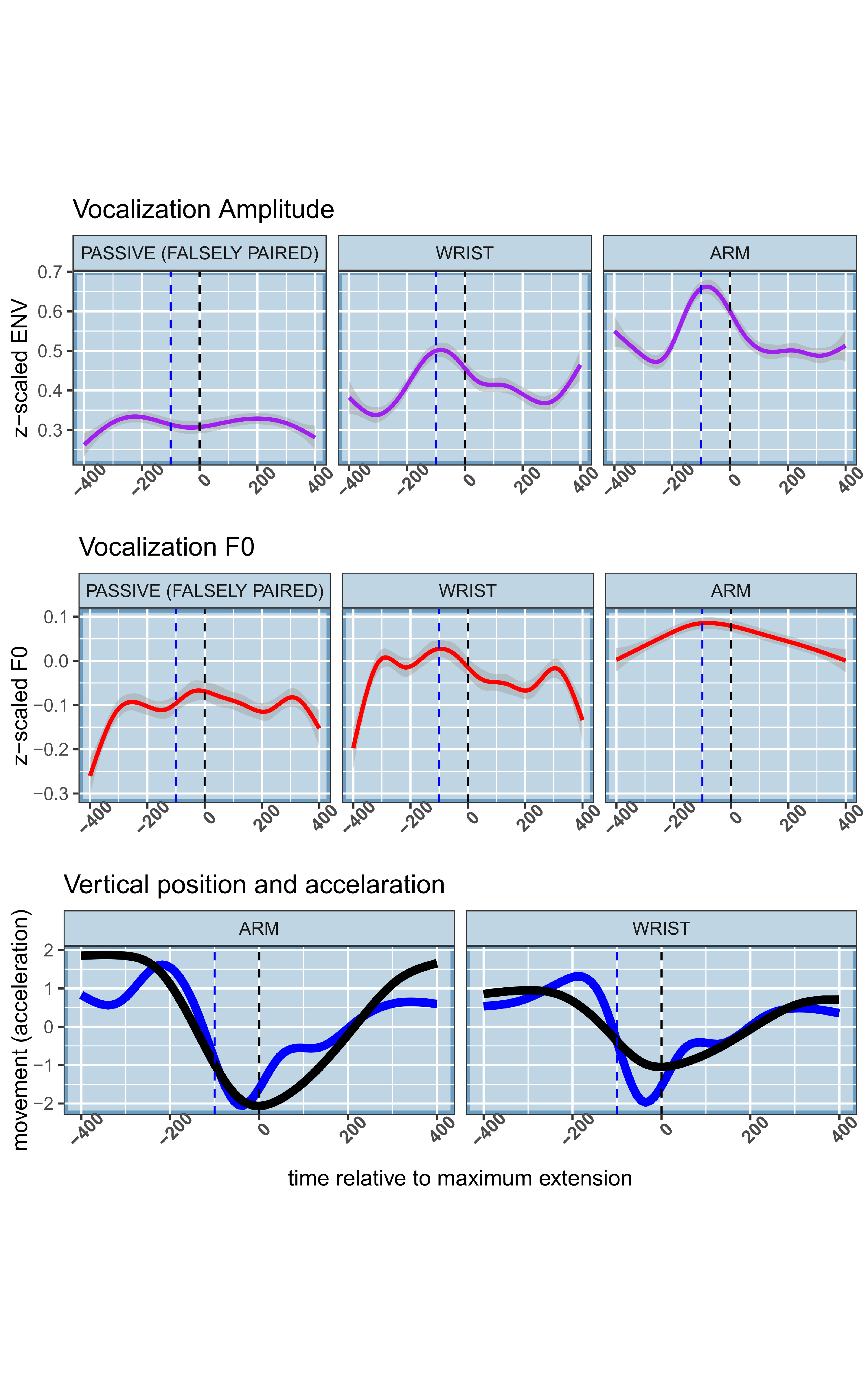
*Note Figure 4*. Violin and box plots are shown for average F0 (Hz) and amplitude envelope (z-scaled) per trial (jitters show observation).

Table 1. Linear mixed effects for effects of condition on F0 and amplitude envelope (ENV)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | contrast | b | SE | df | p |
| ENV (z-scaled) | intercept | 0.347 | 0.033 | 382 | < .0001 |
|  | Wrist vs. Passive | 0.048 | 0.022 | 382 | 0.0287 |
|  | Arm vs. Passive | 0.189 | 0.022 | 382 | < .0001 |

F0 (Hz) intercept 186.916 3.25 382 < .0001 Male vs. Female -73.049 5.511 33 < .0001 Wrist vs. Passive 1.419 0.673 382 0.0357  
Arm vs. Passive 3.208 0.674 382 < .0001

## Coupling of vocalization duration and movement

Having ascertained in the previous analysis that there is some kind of acoustic modulation for movement versus no movement, we further need to confirm that such modulations occur at particular moments in the movement cycle. Figure 5 show the main results for all data, where we model over time the acoustic patterning in vocalizations around the maximum extension of the movement cycle, for all movement cycles that occurred. If there are particular moments in the movement cycle where vocalization is affected, for example the moment when the hand starts decelerating (estimated from the data as shown in Figure 5), we would expect acoustic modulations (peaks) at such moments of the movement cycle.  
Just before the moment of maximum extension, the observed amplitude envelope shows a clear peak, most dramatically for the arm-movement condition, but also present for the wrist-movement condition. For randomly paired movement and passive condition speech, this was not the case; this provides strong evidence that the results observed in the arm- and wrist-movement conditions are not due to mere chance. For F0, the pattern is somewhat less clear, but positive peaks still occur just before the maximum extension. These findings replicate our earlier work on steady-state vocalization and mono-syllabic utterances, showing that moments of peak deceleration also show peaks in acoustics (Pouw et al., 2019b, 2019c).  
To test whether trajectories are indeed non-linear and are reliably different from the passive condition, we performed GAM, a type of non-linear mixed effects procedure. GAM is a popular time-series analysis in phonetics, and allows the automatic modeling of more (and less) complex non-linear patterns by combining a set of smooth basis functions. Furthermore, GAM allows for testing whether those non-linear trajectories are modulated depending on some grouping of the data (see e.g., Wieling, 2018). We assessed with GAM the trajectory of acoustics around 800 milliseconds of the maximum extension of the movement. We chose 800 milliseconds (-400, 400) as this is about the time for a 1.33Hz cycle (1000/1.33Hz = 752 ms) with an added margin of error of about 50ms. The model results with random slopes and intercept for participant are shown in table 2.  
Firstly, for all models tests for non-linearity of the trajectories were statistically reliable (*p*’s < .0001), meaning that there are peaks or valleys in acoustics over the movement cycle rather than a flat linear trend (Figure 6). As shown in Table 2, our results replicate the general finding that the Wrist Movements condition led to reliably different non-linear peaks in acoustics as compared to the passive condition (*p* < .001). Moreover, this effect—relative to the passive condition—is even more extreme for the Arm Movement condition (*p* < .001). Figure 6 provides the fitted trajectories for the GAM models.  
For readers interested in individual differences in trajectories, we have created interactive graphs for each participant’s average amplitude envelope trajectories (<https://osf.io/a423h/>) and F0 trajectories (<https://osf.io/fdzwj/>).  
 Figure 5. Average observed vocalization acoustics relative to the moment of maximum extension  


*Note Figure 5*. For the upper two panels the average acoustic trajectory is shown around the moment of maximum extension (t = 0, dashed black line). In the lower panel, we have plotted the z-scaled average vertical displacement of the hand and the z-scaled acceleration trace. The blue dashed vertical line marks the moment where the deceleration phase starts, which aligns with peaks in acoustics.

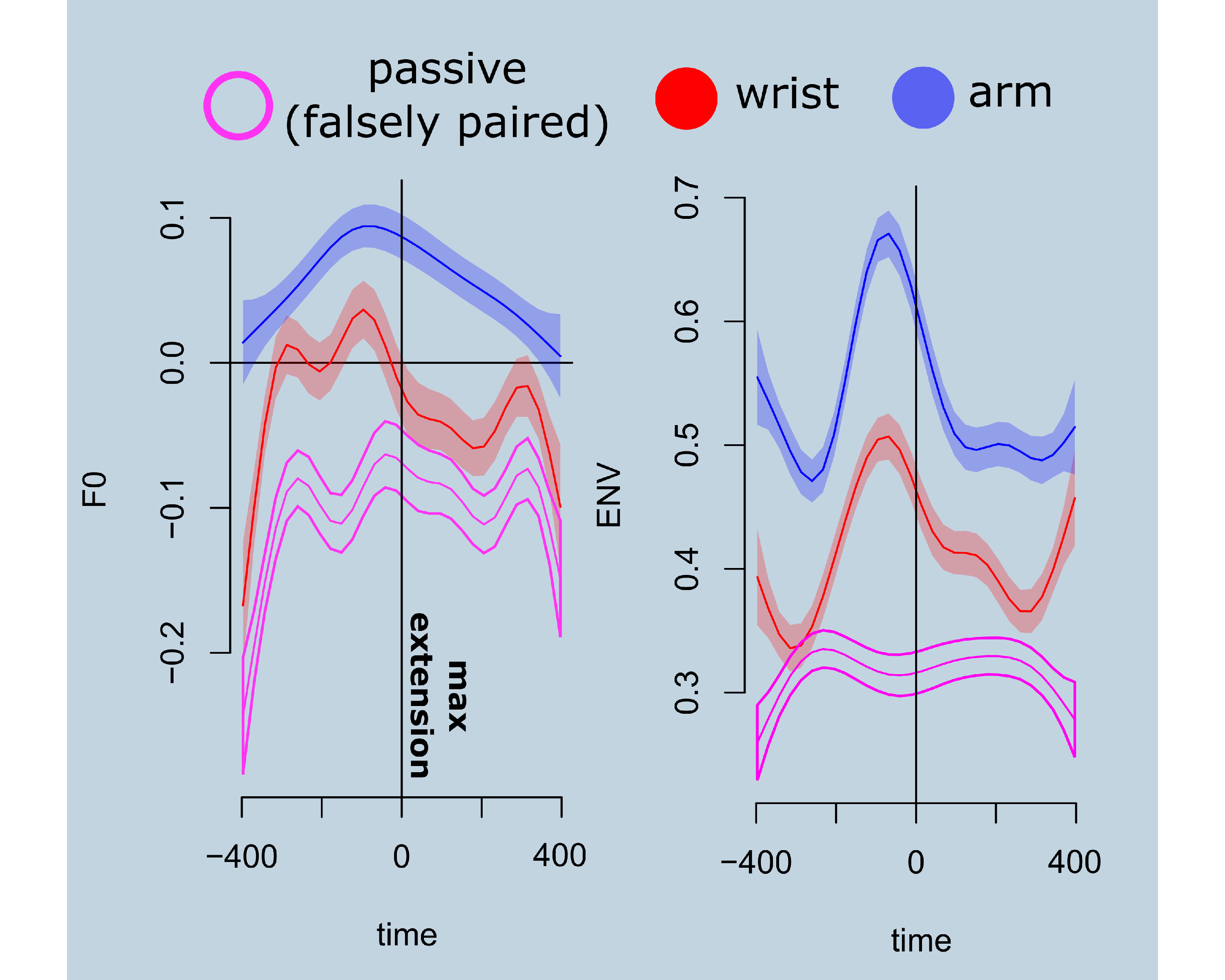
Figure 6. Fitted trajectories GAM 

Table 2. Model results for GAM analysis

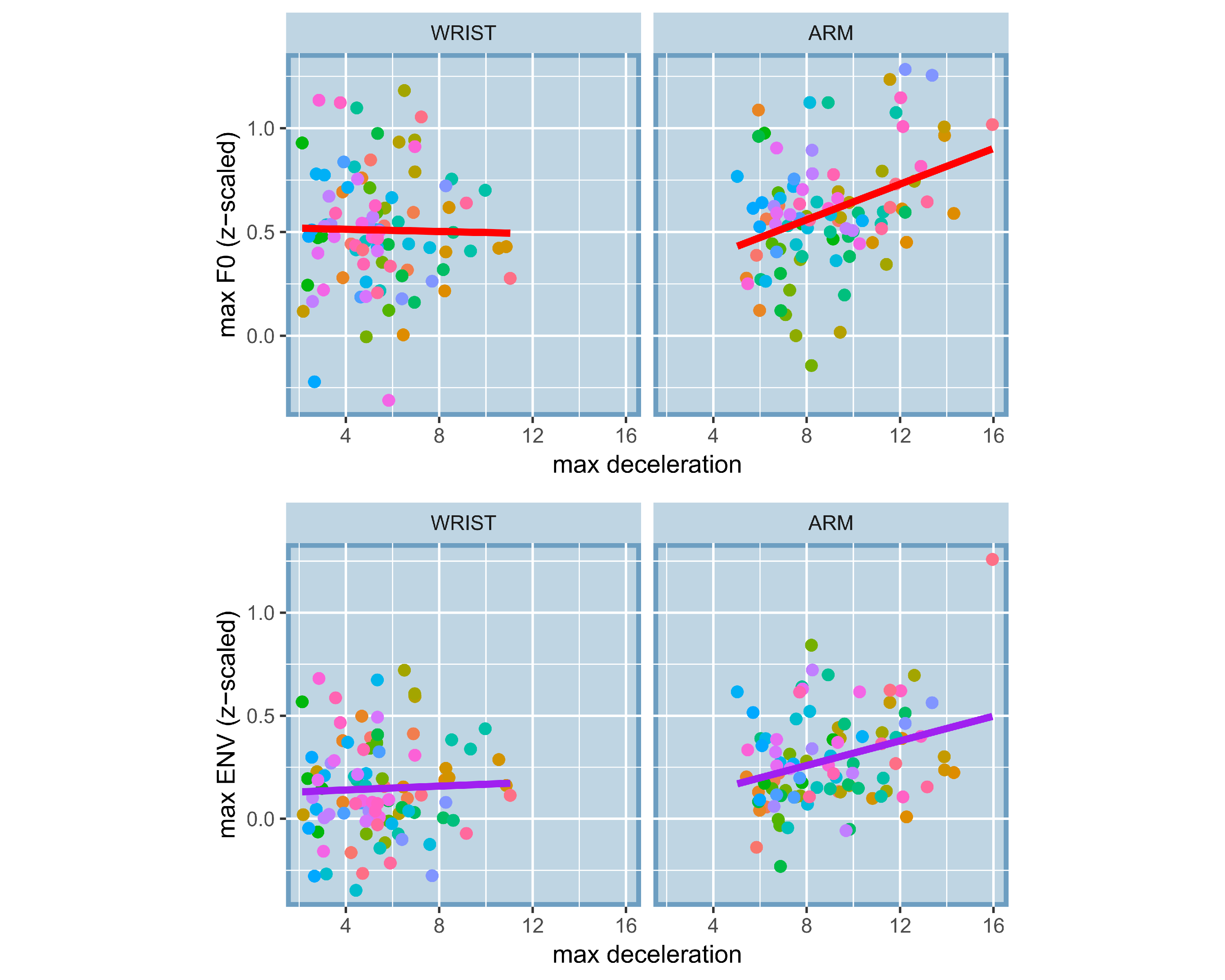
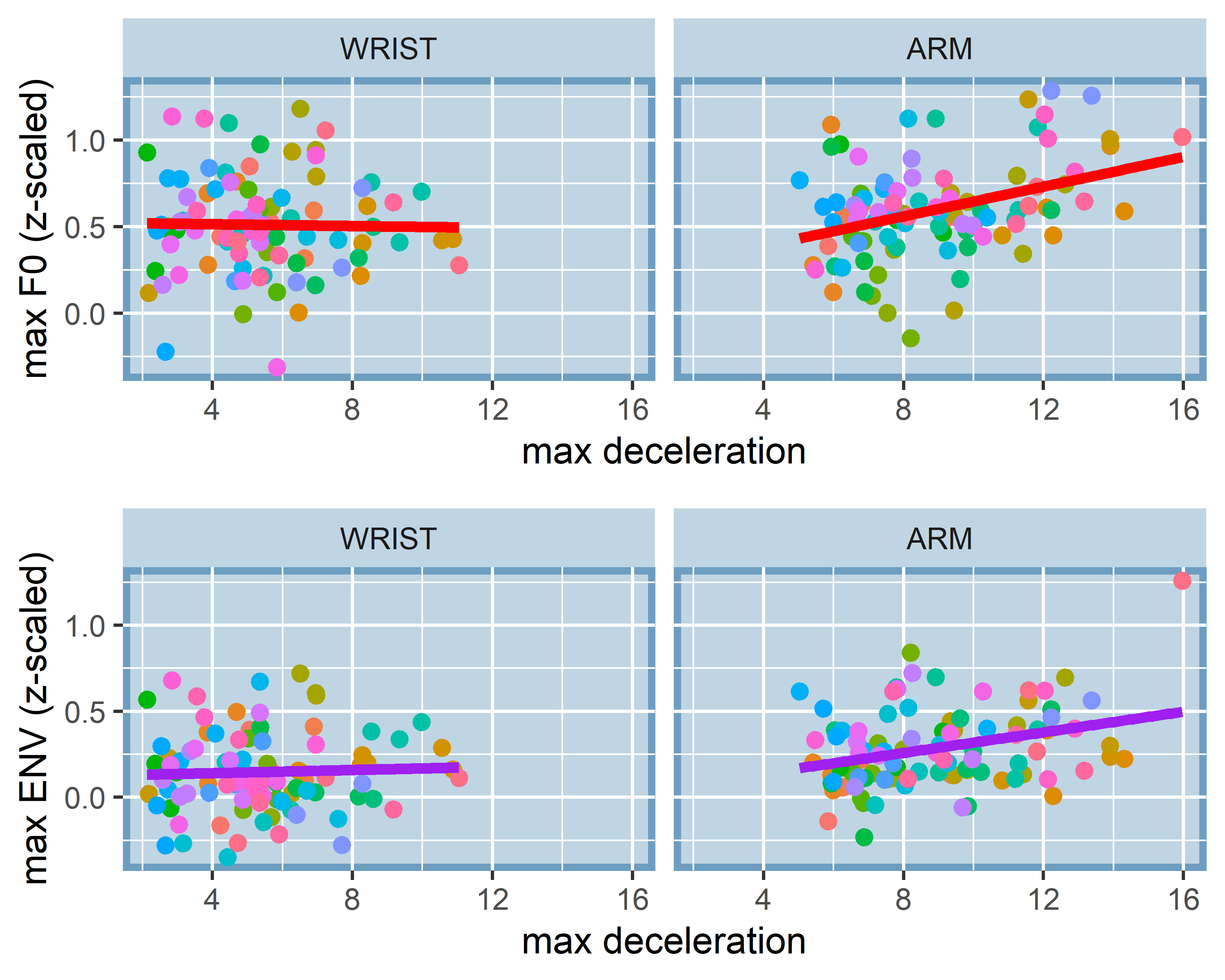
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | contrast | b | SE | df | p |
| ENV (z-scaled) | intercept | 0.237 | 0.006 | 36.923 | < .0001 |
|  | Wrist vs. Passive | 0.096 | 0.009 | 10.579 | < .0001 |
|  | Arm vs. Passive | 0.152 | 0.009 | 16.862 | < .0001 |

F0 (Hz) intercept -0.061 0.006 -8.351 < .0001 Male vs. Female -0.019 0.009 -4.29 < .0001 Wrist vs. Passive 0.101 0.009 10.222 < .0001 Arm vs. Passive 0.094 0.103 9.546 < .0001

*Note*. Model results are shown for the amplitude envelope (ENV; z-scaled) and F0 (Hz). For F0, we accounted for sex differences when estimating independent effects of condition.

## Degree of physical impetus and acoustic peaks

We have confirmed that speech acoustics are modulated around moments of the deceleration phase, about 0-200 ms before the maximum extension. The effect of gesture-speech physics can be further examined by looking at how the forces produced by the upper limb movement predict acoustic peaks. Therefore, for all vocalizations that occurred between 200–0ms before the maximum extension, we assessed whether acoustic peak (i.e., maximum F0 or maximum amplitude envelope) was predicted by the maximum deceleration value (i.e., minimum acceleration observation) observed in that 200 ms window. In previous research, we found that higher deceleration was related to higher amplitude envelope observations but not F0 (Pouw et al., 2019a).  
Figure 7 shows the general pattern of the results for the wrist- and arm-movement condition. Here we averaged per trial the maximum deceleration values of max F0 and max ENV for each vocalization event. Table 3 shows the model results of linear mixed effects model with random intercept and slopes for participants, in which we regressed the trial-averaged max observed deceleration against the co-occurring trial-averaged vocalization acoustic peaks for amplitude envelope and F0 (separately). Higher deceleration indeed predicted higher amplitude envelope and F0 for arm movements, as indicated by a statistically reliable interaction between condition x max deceleration effect (*p*’s < .001). This confirms that both acceleration and the mass of the effector are involved in producing a physical impulse.

Figure 7. Relation max deceleration and height acoustic peak 

*Note Figure 7*. On the x-axis the average maximum deceleration is shown per trial is shown (absolutized negative acceleration value), where 0 indicates no deceleration (absolutized) and positive values indicate higher deceleration rates in cm/seconds squared. It can be seen that deceleration rates are more extreme for the Arm versus the Wrist condition. On the y-axis we have the average maximum observed amplitude envelope (lower panel) and F0 (upper panel) for those moments of deceleration. Higher decelerations co-occur with higher peaks in acoustics for arm movements (but not wrist movements).

Table 4. Linear mixed effects of deceleration and acoustic peaks

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | contrast | b | SE | df | p |
| ENV (z-scaled) | Intercept | 0.007 | 0.134 | 339717 | 0.9568 |
|  | Arm vs. Wrist | -0.043 | 0.003 | 339717 | < .0001 |
|  | Max Deceleration | 0.025 | 0.02 | 339717 | 0.2013 |
|  | Arm x Max Deceleration | 0.009 | 0 | 339717 | < .0001 |

F0 (z-scaled) intercept 0.665 0.128 339717 < .0001 Arm vs. Wrist -0.198 0.004 339717 < .0001 Max Deceleration -0.024 0.019 339717 0.1955  
Arm x Max Deceleration 0.034 0.001 339717 < .0001

# Discussion

In the current study, we demonstrated biomechanical effects of flexion-extension upper limb movements on speech by replicating effects obtained in steady-state vocalization and mono-syllabic utterances in fluent speech. We showed that rhythmically moving the wrist or arm affects vocalization acoustics by heightening F0 and amplitude envelope of speech vocalizations, as compared to a passive control and statistical control conditions. We further show that acoustic modulations of speech vocalizations are especially found around moments of the high-impetus beat (i.e., where the movement abruptly decelerates), thereby producing a physical impetus on the body. We finally show that higher deceleration rates observed within 200 milliseconds before the moment of the maximum extension of the arm movement materializes into more extreme acoustic peaks, demonstrating a role for acceleration and effector mass for gesture’s effect onto speech, i.e., an effect of physical impulse. Indeed, in all analyses, we observe that higher-mass arm versus wrist movements affect speech more clearly.  
Thus it is now possible that stabilities in speaking can arise out of gesture-speech biomechanics, also in fluent speech. This does not mean that speech prosody necessarily requires gesture for reaching prosodic targets. Indeed, other sensorimotor solutions are available for modulating F0 and intensity, for example in the form of vocal-fold tensioning or respiratory actions (Perrier & Fuchs, 2015). However, we do argue that the biomehechanical coupling of gesture and speech provides a ‘smart’ mechanism for ‘timing’ acoustic and movement expressions.

## Wider implications

Gesture-speech physics has promise for a revision of our understanding for the emergence of communicative gesture in anatomically modern humans, both ontogenetically and phylogenetically.  
It is well known that infants produce concurrent vocal-motor babblings. Furtuermore, increased rhythmicity or frequency of motor babbling predicts speech-like maturation of vocalization (Ejiri, 1998; Ejiri & Masataka, 2001). Rather than a primarily neural development that instantiates gesture-speech synchrony (Iverson & Thelen, 2005), we suggest that during such vocal-motor babblings gesture-speech physics is discovered, which provides the basis for infants to develop novel stable sensorimotor solutions for communication, such as a synchronized pointing gesture with a vocalization. Such sensorimotor solutions are, of course, likely solicited and practiced through support of caretakers, yet without the biomorphological scaffolding, gesture-speech synchrony would not get off the ground ontogenetically.  
Phylogenetic accounts have forefronted the depiction and referential function of gesture as the driver for its modern day instantiation (Fröhlich, Sievers, Townsend, Gruber, & Schaik, 2019; Kendon, 2017; Tomasello, 2008). However, it should be considered that peripheral body movements may have served as a control parameter of an under-evolved vocal system. It has already been proposed that vocal system may have been evolutionarily exapted from rhythmic abilities in the loco-motor domain (Larsson, Richter, & Ravignani, 2019; Ravignani et al., 2019), and upper limb movements as having constrained the vocal systems evolution fits neatly in such views. Specifically, when our species became bipedal it has been argued that the respiratory system was thereby liberated from upper-limb locomotary perturbations. Namely, we know that breathing (and vocalization) cycles often rigidly couple 1:1 with locomotion cycles in quadrupeds (Carrier et al., 1984), rigidly limiting what can be done (communicated) in one breath. For example, vocalization acoustics of flying bats are synchronized with their wing beats (Lancaster, Henson, & Keating, 1995). Bipedalism, however, did only free respiration from locomotion, it freed the upper limbs too, allowing these highly skilled articulators to modulate a possibly less killed respiratory-vocal system. Gestures may then have played a role in the complexification of the respiratory system in our species which has been attributed to have occurred to serve speech evolution (MacLarnon & Hewitt, 1999).  
Thus gesture-speech physics is not in principle culturally specific, as animals such as bats can do it too. It can further be related to other species, such as Orangutangs who deepen their vocalizations by cupping their hands in front of their mouth (Hardus, Lameira, Schaik, & Wich, 2009). Other animals have been found to be sensitive to body-related information in sound in that body size and strength can be detected from vocalizations alone (Ghazanfar et al., 2007; Pisanski et al., 2016a), and humans are able to do this with some accuracy as well (Pisanski, Fraccaro, Tigue, O’Connor, & Feinberg, 2014), even when they are blind from birth (Pisanski et al., 2016b). In a recent experiment we have found that listeners are exquisitely sensitive to gesture-modulated acoustics as listeners can synchronize their own upper limb movement by listening to a vocalizer producing a steady-state vocalization while rhythmically moving their wrist or arm (Pouw et al., 2019d, 2020). Thus, bodily dynamics can imprint the (human) voice and this can be informative for listeners.  
To conclude, gesture-speech physics opens up the possibility that gesture may have evolved as a control parameter on vocal actions. This radically embodied revision of gesture-speech coupling provides a solid phylogenetic basis for an evolution of multimodal behavior, whereby peripheral bodily tensioning naturally formed coalitions with sound-producing organs that were still very much under development.

# References

Bernstein, N. (1967). *The Co-ordination and Regulations of Movements* ([1st English ed.] edition.). Pergamon Press.

Carrier, D. R., Kapoor, A. K., Kimura, T., Nickels, M. K., Satwanti, Scott, E. C., … Trinkaus, E. (1984). The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply]. *Current Anthropology*, *25*(4), 483–495. Retrieved from <https://www.jstor.org/stable/2742907>

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](https://doi.org/10.1152/jn.1982.47.2.287)

Cravotta, A., Busà, M. G., & Prieto, P. (2019). Effects of Encouraging the Use of Gestures on Speech. *Journal of Speech, Language, and Hearing Research*. doi:[10.21437/SpeechProsody.2018-42](https://doi.org/10.21437/SpeechProsody.2018-42)

Danner, S. G., Barbosa, A. V., & Goldstein, L. (2018). Quantitative analysis of multimodal speech data. *Journal of Phonetics*, *71*, 268–283. doi:[10.1016/j.wocn.2018.09.007](https://doi.org/10.1016/j.wocn.2018.09.007)

Ejiri, K. (1998). Relationship between Rhythmic Behavior and Canonical Babbling in Infant Vocal Development. *Phonetica*, *55*(4), 226–237. doi:[10.1159/000028434](https://doi.org/10.1159/000028434)

Ejiri, K., & Masataka, N. (2001). Co-occurences of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, *4*(1), 40–48. doi:[10.1111/1467-7687.00147](https://doi.org/10.1111/1467-7687.00147)

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal Realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864. doi:[10.1044/1092-4388(2012/12-0049)](https://doi.org/10.1044/1092-4388(2012/12-0049))

Feyereisen, P. (2017). *The Cognitive Psychology of Speech-Related Gesture*. New York: Routledge. Retrieved from <http://books.google.com?id=nJguDwAAQBAJ>

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & Schaik, C. P. van. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. doi:[10.1111/brv.12535](https://doi.org/10.1111/brv.12535)

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, *17*(5-2), 425–430. doi:[10.1016/j.cub.2007.01.029](https://doi.org/10.1016/j.cub.2007.01.029)

Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., & Malik, J. (2019). Learning individual styles of conversational gesture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3497–3506). Retrieved from <https://arxiv.org/abs/1906.04160>

Hardus, M. E., Lameira, A. R., Schaik, C. S., & Wich, S. A. (2009). Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3689–3694. doi:[10.1098/rspb.2009.1027](https://doi.org/10.1098/rspb.2009.1027)

Hastie, T. (2019). Gam: Generalized Additive Models (Version 1.16.1). Retrieved from <https://CRAN.R-project.org/package=gam>

He, L., & Dellwo, V. (2017). Amplitude envelope kinematics of speech: Parameter extraction and applications. *The Journal of the Acoustical Society of America*, *141*(5), 3582–3582. doi:[10.1121/1.4987638](https://doi.org/10.1121/1.4987638)

Hodges, P. W., & Richardson, C. A. (1997). Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Experimental Brain Research*, *114*(2), 362–370. doi:[10.1007/pl00005644](https://doi.org/10.1007/pl00005644)

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. doi:[10.1016/j.tics.2019.05.006](https://doi.org/10.1016/j.tics.2019.05.006)

Im, S., & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America*, *5*(1, 1), 685–697. doi:[10.3765/plsa.v5i1.4755](https://doi.org/10.3765/plsa.v5i1.4755)

Iverson, J. M., & Thelen, E. (2005). Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 22.

Kelso, J. A., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. *The American Journal of Physiology*, *246*(6 Pt 2), R928–935. doi:[10.1152/ajpregu.1984.246.6.R928](https://doi.org/10.1152/ajpregu.1984.246.6.R928)

Kendon, A. (2017). Reflections on the “gesture-first” hypothesis of language origins. *Psychonomic Bulletin & Review*, *24*(1), 163–170. doi:[10.3758/s13423-016-1117-3](https://doi.org/10.3758/s13423-016-1117-3)

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)

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658). doi:[10.1098/rstb.2013.0397](https://doi.org/10.1098/rstb.2013.0397)

Kucherenko, T., Hasegawa, D., Henter, G. E., Kaneko, N., & 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* (pp. 97–104). Paris, France: ACM Press. doi:[10.1145/3308532.3329472](https://doi.org/10.1145/3308532.3329472)

Lancaster, W. C., Henson, O. W., & Keating, A. W. (1995). Respiratory muscle activity in relation to vocalization in flying bats. *Journal of Experimental Biology*, *198*(1), 175–191. Retrieved from <https://jeb.biologists.org/content/198/1/175>

Larsson, M., Richter, J., & Ravignani, A. (2019). Bipedal Steps in the Development of Rhythmic Behavior in Humans. *Music & Science*, *2*, 2059204319892617. doi:[10.1177/2059204319892617](https://doi.org/10.1177/2059204319892617)

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471. doi:[10.1080/01690965.2010.500218](https://doi.org/10.1080/01690965.2010.500218)

Levin, S. M. (2006). Tensegrity: The new biomechanics. In M. Hutson & R. Ellis (Eds.), *Textbook of muscularskeletal medicine* (pp. 69–80). Oxford, England: Oxford University Press.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1), 71–89. doi:[10.1515/lp-2012-0006](https://doi.org/10.1515/lp-2012-0006)

MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology*, *109*(3), 341–363. doi:[10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-8644(199907)109:3%3C341::AID-AJPA5%3E3.0.CO;2-2)

McClave, E. (1998). Pitch and Manual Gestures. *Journal of Psycholinguistic Research*, *27*(2), 69–89. doi:[https://doi.org/10.1023/A:1023274823974](https://doi.org/https://doi.org/10.1023/A:1023274823974)

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University Of Chicago Press.

Mendoza-Denton, N., & Jannedy, S. (2011). Semiotic Layering through Gesture and Intonation: A Case Study of Complementary and Supplementary Multimodality in Political Speech. *Journal of English Linguistics*, *39*(3), 265–299. doi:[10.1177/0075424211405941](https://doi.org/10.1177/0075424211405941)

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](https://doi.org/10.1016/j.wocn.2013.11.002)

Perrier, P., & Fuchs, S. (2015). Motor Equivalence in Speech Production. In *The Handbook of Speech Production* (pp. 223–247). John Wiley & Sons, Ltd. doi:[10.1002/9781118584156.ch11](https://doi.org/10.1002/9781118584156.ch11)

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Team, R. C. (2019). *Nlme: Linear and nonlinear mixed effects models*.

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016a). Voice modulation: A window into the origins of guman vocal control? *Trends in Cognitive Sciences*, *20*(4), 304–318. doi:[10.1016/j.tics.2016.01.002](https://doi.org/10.1016/j.tics.2016.01.002)

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O’Connor, J. J. M., & Feinberg, D. R. (2014). Return to Oz: Voice pitch facilitates assessments of men’s body size. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1316–1331. doi:[10.1037/a0036956](https://doi.org/10.1037/a0036956)

Pisanski, K., Oleszkiewicz, A., & Sorokowska, A. (2016b). Can blind persons accurately assess body size from the voice? *Biology Letters*, *12*(4), 20160063. doi:[10.1098/rsbl.2016.0063](https://doi.org/10.1098/rsbl.2016.0063)

Pouw, W., & Dixon, J. A. (2019a). Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cognitive Science*, *43*(3), e12721. doi:[10.1111/cogs.12721](https://doi.org/10.1111/cogs.12721)

Pouw, W., & Dixon, J. A. (2019b). Quantifying gesture-speech synchrony. In *Proceedings of the 6th meeting of Gesture and Speech in Interaction* (pp. 68–74). Paderborn: Universitaetsbibliothek Paderborn. doi:[10.17619/UNIPB/1-812](https://doi.org/10.17619/UNIPB/1-812)

Pouw, W., Esteve-Gibert, N., Harrison, S. H., & Dixon, J. A. (2019a). *Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures*. Retrieved from <https://psyarxiv.com/rnpav>

Pouw, W., Harrison, S. H., & Dixon, J. A. (2019b). Gesture-speech physics: The biomechanical basis of the emergence of gesture-speech synchrony. *Journal of Experimental Psychology: General*. doi:[10.1037/xge0000646](https://doi.org/10.1037/xge0000646)

Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A. (2019c). Energy flows in gesture-speech physics: Exploratory findings and pre-registration of confirmatory analysis. doi:[10.31234/osf.io/c7456](https://doi.org/10.31234/osf.io/c7456)

Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2019d). Acoustic specification of upper limb movement in voicing. In *Proceedings of the 6th meeting of Gesture and Speech in Interaction* (pp. 75–80). Paderborn, Germany: Universitaetsbibliothek Paderborn. doi:[10.17619/UNIPB/1-812](https://doi.org/10.17619/UNIPB/1-812)

Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*. doi:[10.1073/pnas.2004163117](https://doi.org/10.1073/pnas.2004163117)

Pouw, W., & Trujillo, J. P. (2019). *Materials Tutorial Gespin2019 - Using video-based motion tracking to quantify speech-gesture synchrony*. Retrieved from <10.17605/OSF.IO/RXB8J>

Profeta, V. L. S., & Turvey, M. T. (2018). Bernstein’s levels of movement construction: A contemporary perspective. *Human Movement Science*, *57*, 111–133. doi:[10.1016/j.humov.2017.11.013](https://doi.org/10.1016/j.humov.2017.11.013)

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., & Kotz, S. (2019). Rhythm in speech and animal vocalizations: A cross‐species perspective. *Annals of the New York Academy of Sciences*. doi:[10.1111/nyas.14166](https://doi.org/10.1111/nyas.14166)

Rochet-Capellan, A., Laboissière, R., Galván, A., & Jean-Luc, S. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1507–1521. doi:[10.1044/1092-4388(2008/07-0173)](https://doi.org/10.1044/1092-4388(2008/07-0173))

Rosch, A., & Schmidbauer, H. (2014). WaveletComp 1.1: A guided tour through the R package, 59.

Ruiter, J. P. de. (2000, August). *The production of gesture and speech*. *Language and Gesture*. doi:[10.1017/CBO9780511620850.018](https://doi.org/10.1017/CBO9780511620850.018)

Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. In *Proceedings of the International Congress of Phonetic Sciences 2019* (p. 5). Melbourne, Australia.

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. doi:[10.1016/j.neulet.2007.09.066](https://doi.org/10.1016/j.neulet.2007.09.066)

Stoltmann, K., & Fuchs, S. (2017). Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate. *Journal of Multimodal Communication Studies*, *4*(1-2), 63–68.

Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*(3), 285–301. doi:[10.1075/gest.8.3.02str](https://doi.org/10.1075/gest.8.3.02str)

Tomasello, M. (2008). *The origins of human communication*. Cambdride, MA: MIT press.

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](https://doi.org/10.1016/S0167-9457(02)00178-1)

Turvey, M. T. (1990). Coordination. *American Psychologist*, *45*(8), 938–953. doi:[10.1037/0003-066X.45.8.938](https://doi.org/10.1037/0003-066X.45.8.938)

Turvey, M. T., & Fonseca, S. T. (2014). The Medium of Haptic Perception: A Tensegrity Hypothesis. *Journal of Motor Behavior*, *46*(3), 143–187. doi:[10.1080/00222895.2013.798252](https://doi.org/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.2013.09.008](https://doi.org/10.1016/j.specom.2013.09.008)

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. *Journal of Phonetics*, *70*, 86–116. doi:[10.1016/j.wocn.2018.03.002](https://doi.org/10.1016/j.wocn.2018.03.002)

Winkelmann, R., Bombien, L., & Scheffers, M. (2018). Wrassp: Interface to the ’ASSP’ Library (Version 0.1.8). Retrieved from <https://CRAN.R-project.org/package=wrassp>

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](https://doi.org/10.1016/j.humov.2015.05.009)

Bernstein, N. (1967). *The Co-ordination and Regulations of Movements* ([1st English ed.] edition.). Pergamon Press.

Carrier, D. R., Kapoor, A. K., Kimura, T., Nickels, M. K., Satwanti, Scott, E. C., … Trinkaus, E. (1984). The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply]. *Current Anthropology*, *25*(4), 483–495. Retrieved from <https://www.jstor.org/stable/2742907>

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](https://doi.org/10.1152/jn.1982.47.2.287)

Cravotta, A., Busà, M. G., & Prieto, P. (2019). Effects of Encouraging the Use of Gestures on Speech. *Journal of Speech, Language, and Hearing Research*. doi:[10.21437/SpeechProsody.2018-42](https://doi.org/10.21437/SpeechProsody.2018-42)

Danner, S. G., Barbosa, A. V., & Goldstein, L. (2018). Quantitative analysis of multimodal speech data. *Journal of Phonetics*, *71*, 268–283. doi:[10.1016/j.wocn.2018.09.007](https://doi.org/10.1016/j.wocn.2018.09.007)

Ejiri, K. (1998). Relationship between Rhythmic Behavior and Canonical Babbling in Infant Vocal Development. *Phonetica*, *55*(4), 226–237. doi:[10.1159/000028434](https://doi.org/10.1159/000028434)

Ejiri, K., & Masataka, N. (2001). Co-occurences of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, *4*(1), 40–48. doi:[10.1111/1467-7687.00147](https://doi.org/10.1111/1467-7687.00147)

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal Realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864. doi:[10.1044/1092-4388(2012/12-0049)](https://doi.org/10.1044/1092-4388(2012/12-0049))

Feyereisen, P. (2017). *The Cognitive Psychology of Speech-Related Gesture*. New York: Routledge. Retrieved from <http://books.google.com?id=nJguDwAAQBAJ>

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & Schaik, C. P. van. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. doi:[10.1111/brv.12535](https://doi.org/10.1111/brv.12535)

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, *17*(5-2), 425–430. doi:[10.1016/j.cub.2007.01.029](https://doi.org/10.1016/j.cub.2007.01.029)

Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., & Malik, J. (2019). Learning individual styles of conversational gesture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3497–3506). Retrieved from <https://arxiv.org/abs/1906.04160>

Hardus, M. E., Lameira, A. R., Schaik, C. S., & Wich, S. A. (2009). Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3689–3694. doi:[10.1098/rspb.2009.1027](https://doi.org/10.1098/rspb.2009.1027)

Hastie, T. (2019). Gam: Generalized Additive Models (Version 1.16.1). Retrieved from <https://CRAN.R-project.org/package=gam>

He, L., & Dellwo, V. (2017). Amplitude envelope kinematics of speech: Parameter extraction and applications. *The Journal of the Acoustical Society of America*, *141*(5), 3582–3582. doi:[10.1121/1.4987638](https://doi.org/10.1121/1.4987638)

Hodges, P. W., & Richardson, C. A. (1997). Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Experimental Brain Research*, *114*(2), 362–370. doi:[10.1007/pl00005644](https://doi.org/10.1007/pl00005644)

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. doi:[10.1016/j.tics.2019.05.006](https://doi.org/10.1016/j.tics.2019.05.006)

Im, S., & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America*, *5*(1, 1), 685–697. doi:[10.3765/plsa.v5i1.4755](https://doi.org/10.3765/plsa.v5i1.4755)

Iverson, J. M., & Thelen, E. (2005). Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 22.

Kelso, J. A., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. *The American Journal of Physiology*, *246*(6 Pt 2), R928–935. doi:[10.1152/ajpregu.1984.246.6.R928](https://doi.org/10.1152/ajpregu.1984.246.6.R928)

Kendon, A. (2017). Reflections on the “gesture-first” hypothesis of language origins. *Psychonomic Bulletin & Review*, *24*(1), 163–170. doi:[10.3758/s13423-016-1117-3](https://doi.org/10.3758/s13423-016-1117-3)

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)

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658). doi:[10.1098/rstb.2013.0397](https://doi.org/10.1098/rstb.2013.0397)

Kucherenko, T., Hasegawa, D., Henter, G. E., Kaneko, N., & 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* (pp. 97–104). Paris, France: ACM Press. doi:[10.1145/3308532.3329472](https://doi.org/10.1145/3308532.3329472)

Lancaster, W. C., Henson, O. W., & Keating, A. W. (1995). Respiratory muscle activity in relation to vocalization in flying bats. *Journal of Experimental Biology*, *198*(1), 175–191. Retrieved from <https://jeb.biologists.org/content/198/1/175>

Larsson, M., Richter, J., & Ravignani, A. (2019). Bipedal Steps in the Development of Rhythmic Behavior in Humans. *Music & Science*, *2*, 2059204319892617. doi:[10.1177/2059204319892617](https://doi.org/10.1177/2059204319892617)

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471. doi:[10.1080/01690965.2010.500218](https://doi.org/10.1080/01690965.2010.500218)

Levin, S. M. (2006). Tensegrity: The new biomechanics. In M. Hutson & R. Ellis (Eds.), *Textbook of muscularskeletal medicine* (pp. 69–80). Oxford, England: Oxford University Press.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1), 71–89. doi:[10.1515/lp-2012-0006](https://doi.org/10.1515/lp-2012-0006)

MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology*, *109*(3), 341–363. doi:[10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-8644(199907)109:3%3C341::AID-AJPA5%3E3.0.CO;2-2)

McClave, E. (1998). Pitch and Manual Gestures. *Journal of Psycholinguistic Research*, *27*(2), 69–89. doi:[https://doi.org/10.1023/A:1023274823974](https://doi.org/https://doi.org/10.1023/A:1023274823974)

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University Of Chicago Press.

Mendoza-Denton, N., & Jannedy, S. (2011). Semiotic Layering through Gesture and Intonation: A Case Study of Complementary and Supplementary Multimodality in Political Speech. *Journal of English Linguistics*, *39*(3), 265–299. doi:[10.1177/0075424211405941](https://doi.org/10.1177/0075424211405941)

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](https://doi.org/10.1016/j.wocn.2013.11.002)

Perrier, P., & Fuchs, S. (2015). Motor Equivalence in Speech Production. In *The Handbook of Speech Production* (pp. 223–247). John Wiley & Sons, Ltd. doi:[10.1002/9781118584156.ch11](https://doi.org/10.1002/9781118584156.ch11)

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Team, R. C. (2019). *Nlme: Linear and nonlinear mixed effects models*.

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016a). Voice modulation: A window into the origins of guman vocal control? *Trends in Cognitive Sciences*, *20*(4), 304–318. doi:[10.1016/j.tics.2016.01.002](https://doi.org/10.1016/j.tics.2016.01.002)

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O’Connor, J. J. M., & Feinberg, D. R. (2014). Return to Oz: Voice pitch facilitates assessments of men’s body size. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1316–1331. doi:[10.1037/a0036956](https://doi.org/10.1037/a0036956)

Pisanski, K., Oleszkiewicz, A., & Sorokowska, A. (2016b). Can blind persons accurately assess body size from the voice? *Biology Letters*, *12*(4), 20160063. doi:[10.1098/rsbl.2016.0063](https://doi.org/10.1098/rsbl.2016.0063)

Pouw, W., & Dixon, J. A. (2019a). Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cognitive Science*, *43*(3), e12721. doi:[10.1111/cogs.12721](https://doi.org/10.1111/cogs.12721)

Pouw, W., & Dixon, J. A. (2019b). Quantifying gesture-speech synchrony. In *Proceedings of the 6th meeting of Gesture and Speech in Interaction* (pp. 68–74). Paderborn: Universitaetsbibliothek Paderborn. doi:[10.17619/UNIPB/1-812](https://doi.org/10.17619/UNIPB/1-812)

Pouw, W., Esteve-Gibert, N., Harrison, S. H., & Dixon, J. A. (2019a). *Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures*. Retrieved from <https://psyarxiv.com/rnpav>

Pouw, W., Harrison, S. H., & Dixon, J. A. (2019b). Gesture-speech physics: The biomechanical basis of the emergence of gesture-speech synchrony. *Journal of Experimental Psychology: General*. doi:[10.1037/xge0000646](https://doi.org/10.1037/xge0000646)

Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A. (2019c). Energy flows in gesture-speech physics: Exploratory findings and pre-registration of confirmatory analysis. doi:[10.31234/osf.io/c7456](https://doi.org/10.31234/osf.io/c7456)

Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2019d). Acoustic specification of upper limb movement in voicing. In *Proceedings of the 6th meeting of Gesture and Speech in Interaction* (pp. 75–80). Paderborn, Germany: Universitaetsbibliothek Paderborn. doi:[10.17619/UNIPB/1-812](https://doi.org/10.17619/UNIPB/1-812)

Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*. doi:[10.1073/pnas.2004163117](https://doi.org/10.1073/pnas.2004163117)

Pouw, W., & Trujillo, J. P. (2019). *Materials Tutorial Gespin2019 - Using video-based motion tracking to quantify speech-gesture synchrony*. Retrieved from <10.17605/OSF.IO/RXB8J>

Profeta, V. L. S., & Turvey, M. T. (2018). Bernstein’s levels of movement construction: A contemporary perspective. *Human Movement Science*, *57*, 111–133. doi:[10.1016/j.humov.2017.11.013](https://doi.org/10.1016/j.humov.2017.11.013)

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., & Kotz, S. (2019). Rhythm in speech and animal vocalizations: A cross‐species perspective. *Annals of the New York Academy of Sciences*. doi:[10.1111/nyas.14166](https://doi.org/10.1111/nyas.14166)

Rochet-Capellan, A., Laboissière, R., Galván, A., & Jean-Luc, S. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1507–1521. doi:[10.1044/1092-4388(2008/07-0173)](https://doi.org/10.1044/1092-4388(2008/07-0173))

Rosch, A., & Schmidbauer, H. (2014). WaveletComp 1.1: A guided tour through the R package, 59.

Ruiter, J. P. de. (2000, August). *The production of gesture and speech*. *Language and Gesture*. doi:[10.1017/CBO9780511620850.018](https://doi.org/10.1017/CBO9780511620850.018)

Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. In *Proceedings of the International Congress of Phonetic Sciences 2019* (p. 5). Melbourne, Australia.

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. doi:[10.1016/j.neulet.2007.09.066](https://doi.org/10.1016/j.neulet.2007.09.066)

Stoltmann, K., & Fuchs, S. (2017). Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate. *Journal of Multimodal Communication Studies*, *4*(1-2), 63–68.

Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*(3), 285–301. doi:[10.1075/gest.8.3.02str](https://doi.org/10.1075/gest.8.3.02str)

Tomasello, M. (2008). *The origins of human communication*. Cambdride, MA: MIT press.

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](https://doi.org/10.1016/S0167-9457(02)00178-1)

Turvey, M. T. (1990). Coordination. *American Psychologist*, *45*(8), 938–953. doi:[10.1037/0003-066X.45.8.938](https://doi.org/10.1037/0003-066X.45.8.938)

Turvey, M. T., & Fonseca, S. T. (2014). The Medium of Haptic Perception: A Tensegrity Hypothesis. *Journal of Motor Behavior*, *46*(3), 143–187. doi:[10.1080/00222895.2013.798252](https://doi.org/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.2013.09.008](https://doi.org/10.1016/j.specom.2013.09.008)

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. *Journal of Phonetics*, *70*, 86–116. doi:[10.1016/j.wocn.2018.03.002](https://doi.org/10.1016/j.wocn.2018.03.002)

Winkelmann, R., Bombien, L., & Scheffers, M. (2018). Wrassp: Interface to the ’ASSP’ Library (Version 0.1.8). Retrieved from <https://CRAN.R-project.org/package=wrassp>

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](https://doi.org/10.1016/j.humov.2015.05.009)