Version 1: This is an unpublished preprint of a preliminary data report and pre-registration (OSF time stamped on 20-8-2018)

The physical basis of gesture-speech synchrony: Exploratory study and pre-registration

Wim Pouw[1,2], Steven J. Harrison[1], & James A. Dixon[1]

Center for the Ecological Study of Perception and Action, University of Connecticut[1]

Department of Psychology, Education, & Child Studies, Erasmus University Rotterdam[2]

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

Open data & Pre-registration: The raw anonymized quantitative data, and analyses scripts supporting this exploratory study are available at https://osf.io/5aydk/. The final data supporting the confirmatory study will also be made available here.

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

Hand gestures during speech move in a common rhythm, as exemplified by the synchrony between prosodic contrasts in gesture movement (e.g., peak velocity) and speech (e.g., peaks in Fundamental Frequency). 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. The current paper presents an exploratory study together with a pre-registration of a larger scale confirmatory study. We provide preliminary evidence that upper limb motions with greater momentum (i.e., physical impetus) affects Fundamental Frequency and the Amplitude Envelope of phonation in a way that accommodates research on gesture-speech synchrony. We speculate that anticipatory postural adjustments and related physical effects of upper limb movements on the musculoskeletal connective system could lead to changes in alveolar (lung) pressure that can impart (prosodic) contrasts in phonation. Here we pre-registered a confirmatory study to more comprehensively address this hypothesis.

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; 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; Esteve-Gibert & Guellaï, 2018). 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, 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 the possibility that gesture and speech have a shared physical medium through which synchrony can emerge biomechanically - the body.

Gesture-speech Synchrony and its Medium

As Lieberman (1993) summarizes, the fundamental frequency (F0) of speech is determined by the alveolar/subglottal (lung) air pressure and larynx muscle tonus. 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 essential and sensitive role of expiration-related muscles and alveolar pressure in the stable production of speech (more specifically phonation), we must wonder whether arm movements in the form of gestures 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 raises[1] 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 movement was made, that this increased duration of phonation and also higher frequency for the first formant (higher F1) regardless of whether a pitch accent was actually intended. These effects of movements on speech were related to how pitch-accents are made without 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 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

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

movement during speech affects speech acoustics. However, other more prominent characteristics of pitch accent, namely increased amplitude and increased pitch (F0 fundamental frequency), 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 gesture-speech synchrony.

Despite having been largely disregarded by phoneticians and gesture researchers, we think that there is a viable possibility that upper limb movements in gesturing have direct physical effects on F0 and amplitude which, 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 anticipation postural adjustment (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 a reactive force that is counteractive, and thus equal in magnitude (if not to fall over) 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 kept locally contained (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).

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. 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). Beat gestures possess greater physical momentum as compared to other types of gestures, and therefore possess greater potential for momentum transfers to the body that might act to destabilize body posture (for real-world examples of beat gestures see https://osf.io/29h8z/). That such forces are non-trivial is indicated by Ian Waterman, a person suffering from lack of proprioception, who reported that he suppressed his gestures in initial stages of his disease because he was afraid of falling over by the destabilizing effects of his gestures (McNeil, 2005; Gallagher, 2005). In contrast to beat gestures, the next common overarching type of gestures, are iconic gestures. These gestures have more complex and often more fluid movement trajectories as they need to iconically present meaning. These gestures are often also coupled to speech prosody in a similar way as beat gestures (Wagner et al., 2014; Prieto, P., Cravotta, Kushch, Rohrer, & Vilà-Giménez, 2018), yet these gestures do seem to be more variably (less tightly) coupled with prosodic contrast in speech (Pouw & Dixon, 2018a, 2018b). This is possibly because iconic gestures have movement trajectories that are not recruited primarily to impart physical impetus on the body, rather some degree of freedom is reserved for iconic expression. Similarly, beat gestures might synchronize with speech

the way they do, because of they are recruited in a way to produce physical impulse on the body.

Current Exploratory Study & Aim current paper

In the current exploratory study 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 down-ward phase. They 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. Under these conditions we aimed to disentangle possible synchrony between phonation and moving the body, and the possible effects of the physical impetus of those movements on synchrony. The exploratory study forms the basis for our larger-scale confirmatory study, which we discuss in the pre-registration in the final section of this paper. In this proposed confirmatory study we will attempt to replicate the current results and expand the design to more directly test a viable hypothesis that postural stability is a key aspect in the sensorimotor synchronization of gesture and speech.

Exploratory Study

Design, method, procedure

The current experiment consists of a within-subject design with one factor (condition) of 4 levels (passive, wrist beat, one-arm beat, two-arm beat).Two right-handed participants (one female and one male) were asked to stand upright and produce a steady voiced output of the vowel 'a:' (as in 'cinem**a**'). 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 the passive condition, participants had their hands resting alongside their bodies during phonating. For the "one-arm beat" condition participants were asked to continuously move their dominant hand at their own preferred rate by lifting the hand up and letting it drop with a sudden complete halt (i.e., with energetic contrast, a "beat"). In the "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. For each condition we performed 4 blocks of 4 trials (total = 32 trials = 2 participants x 4 blocks x 4 condition). Order of condition was randomized for each block.

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 in reference to the Z-axis movements. 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.), as to simultaneously call and write movement and audio data. We modified this script as to enable 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) PRAAT and motion tracking data were aggregated (code available on https://osf.io/5aydk/).

Phonation Variables

A raw speech signal has both fine and gross structure 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 0:1, and is thus scaled for individual differences between participants in amplitude.

Fundamental Frequency (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. We matched the sampling rate of pitch with that of the motion tracker (240Hz: 1 sample per 4.16 milliseconds).

Results Exploratory Study

Descriptives

Table 1 provides an overview of the average F0, Amplitude Envelope (ENV) for each participant and condition. To listen to audio examples of the trials with the envelope, F0 and Z-movement data go to https://osf.io/acmdg/; Effects of hand movement are readily apparent when listening to the audio samples. Indeed, we find higher standard deviations for ENV and F0 for the one-arm and two-arm conditions for both participants, confirming that in these conditions phonating became more unstable. Average phonating duration was 4.19 seconds (*SD* = 2.65 seconds), with average duration for Passive = 4.23s, Wrist Beat = 4.66s, One-Arm beat = 4.00s, Two-arm Beat = 3.75s). Within a trial, amplitude (r =- .50) and F0 decreased (r = -.17) as a function of time (p's < . 0001), indicating that participants energy for phonating diminished as they were reaching the ends of their breadths. To exclude possible artifacts of time in our estimation of the effect of condition, all further analyses have been performed on F0 and ENV time series that were linearly detrended for the effect of trial time for each separate trial. As to be expected, ENV and F0 were positively correlated (mean correlation per trial r = .503, p's < .001).

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

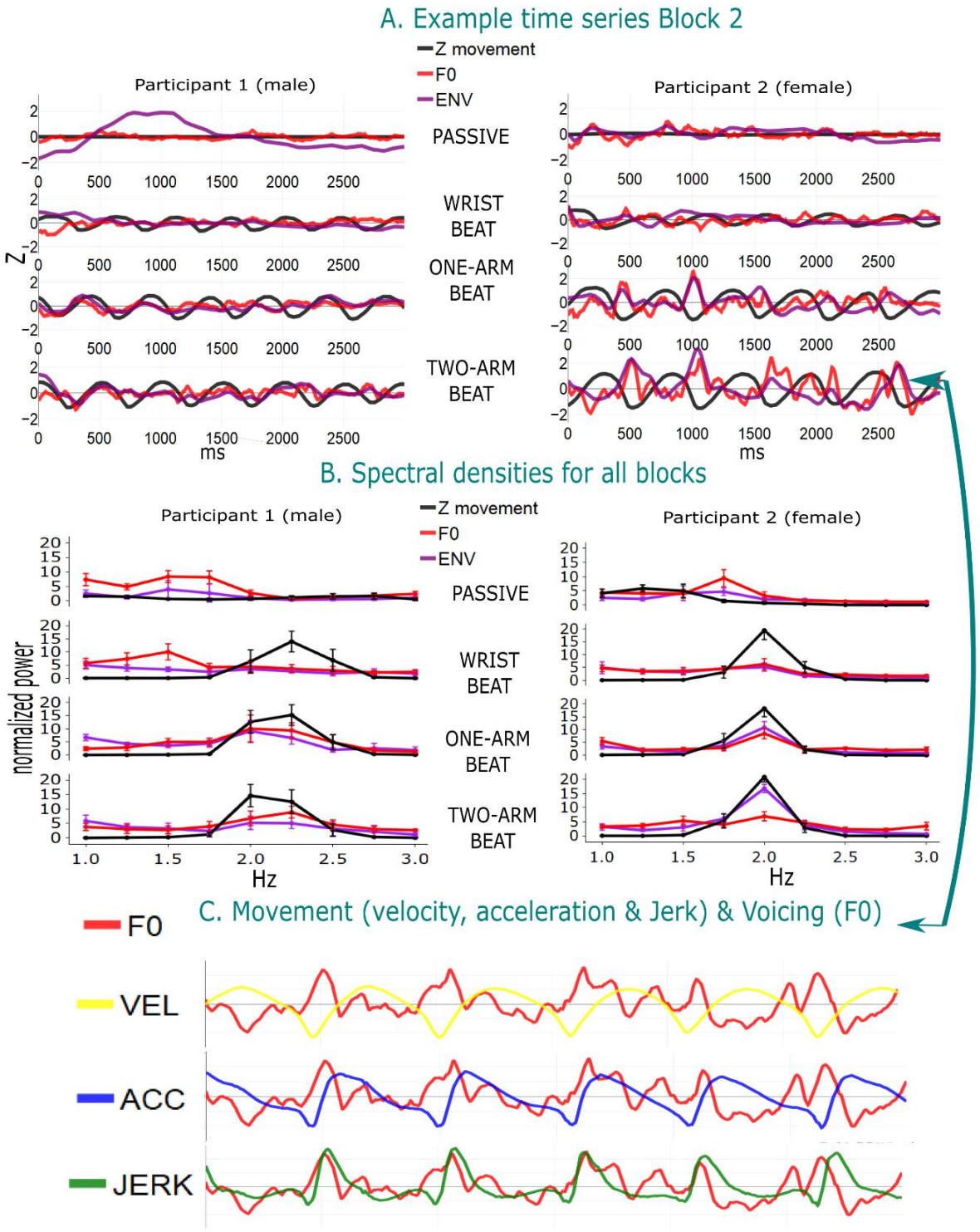
		Mean ***(SD)***			
		Passive	**Wrist Beat**	**One-Arm Beat**	**Two-Arm Beat**
F0	**TOTAL**	160.94 ***(1.18)***	*161.64* ***(1.46)***	*164.08* ***(2.71)***	*162.45* ***(4.42)***
	ppn1	*109.15* ***(0.95)***	*110.37* ***(1.17)***	*113.35* ***(1.94)***	*112.39* ***(2.17)***
	ppn2	*212.73* ***(1.41)***	*212.92* ***(1.74)***	*214.81* ***(3.48)***	*212.51* ***(6.27)***
ENV	**TOTAL**	.169 ***(.033)***	*.156* ***(.033)***	.184 ***(.040)***	.171 ***(.046)***
	ppn1	.158 ***(.020)***	.142 ***(.028)***	.177 ***(.026)***	.188 ***(.037)***
	ppn2	.181 ***(.045)***	.169 ***(.038)***	.191 ***(.053)***	*.155* ***(.055)***

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

Spectral density and coherence

When participants moved their hands, they moved them vertically (i.e., movement on Z-axis) with a preferred frequency of about 2 cycles per second (i.e., 2Hz), a rate commonly observed in humans for spontaneous movements (e.g., Collyer, Broadbent, Church, 1994). If movement structurally affects phonating, rather than simply inducing instabilities in phonating, we would expect perturbations to occur in phonating with about the same frequency. We firstly performed a spectral decomposition analyses with R package 'spectral' (Salmayer, 2016) applying the Fast Fourier Transform (FFT) as to assess periodicities in movement and phonation (see Figure 1, panel B).

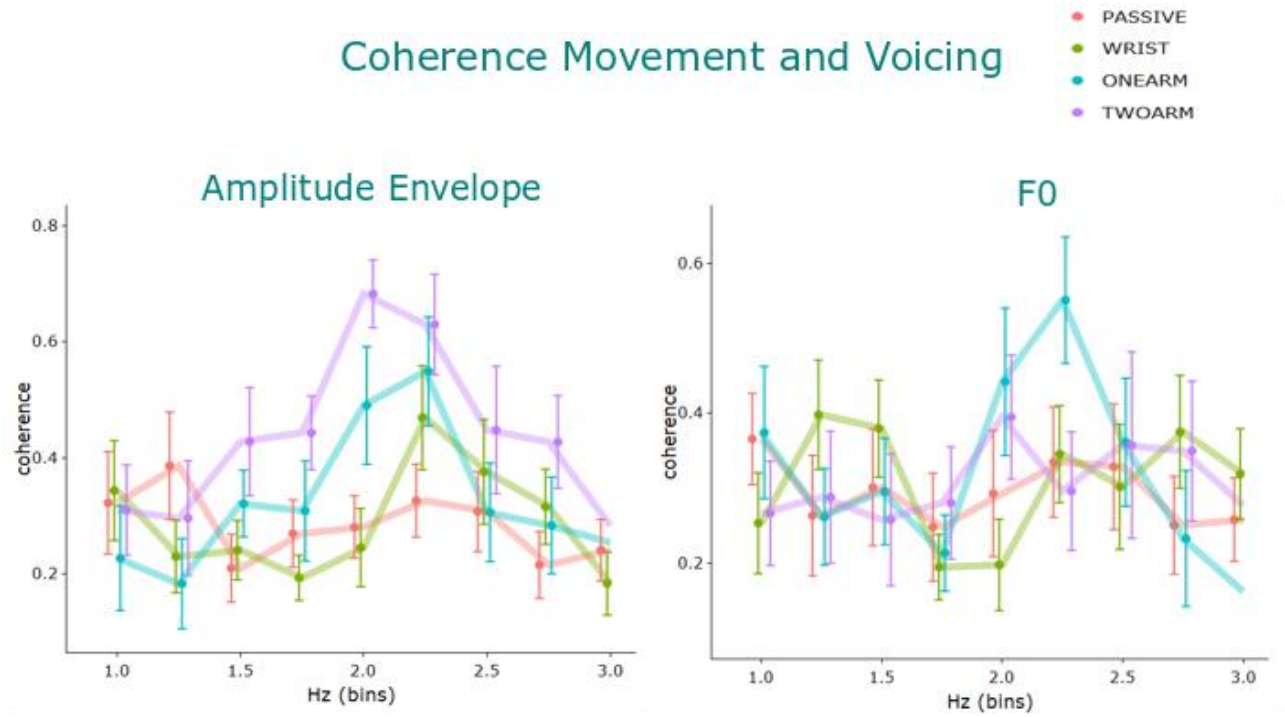
Figure 1. Example F0 speech and movement



Note Figure 1 (continued). Panel A: Example time series (first 3 seconds) of one block for each participant are shown with pitch (F0: in red) amplitude envelope (ENV: in purple) and vertical movement (Z movement). Panel B: Are the normalized spectral density estimates for the 1-3 frequency range. Panel C: Example time series for derivatives and pitch to evaluate relation of kinematics and peaks F0.

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 2 provides an overview for the mean coherence per condition between I) movement and ENV, and II) movement and F0. It can be observed that indeed around the 2Hz frequency range there seems an increased coherence levels for the Two-Arm Beat- and the One-Arm Beat condition.

Figure 2. Coherence



***Note*.** This figure shows the coherence levels for each condition between a) movement (Z) and Amplitude Envelope (ENV), and b) movement (Z) and F0. Error bars indicate 95% confidence intervals. It can be observed that the passive condition has generally lower coherence levels, suggesting that movement and phonating were not coupled. Consistent with the spectral density results around the 2Hz range there are prominent peaks for the ENV and F0 for the one-arm and two-arm beat conditions. Note that we also added in the vertical movement (Z) time series during the passive condition and phonation as a baseline to compare to the other conditions, and to assess whether bodily sway may synchronize with changes is phonation.

We statistically tested the effects of condition on coherence with a mixed regression model (R package nlme: participants as random intercept). We assessed coherence within a

range of 1-3 Hz, with bins 0.5 Hz width. Firstly we entered condition as main predictor for coherence of Amplitude Envelope and movement, which added predictive value as compared to a model that predicts the overall mean (change in $\chi 2$ [6] = 83.44, $p < .001$, AIC = -162.92). This model (which does not take into account frequency range), already shows increased coherence for the one-arm-, $b = .041$, t (1121), 2.148, $p = 0.032$, and the two-arm beat condition, $b = .159$, t (1121), 8.254, $p < 0.001$, as compared to the passive condition. The Wrist Beat condition did not differ in coherence from the passive condition, $b = -0.008$, $t(1121) = 0.462$, $p = .644$.

We expanded the model by entering frequency as a factor (in bin sizes of 0.5, range = 1- 3 Hz) and its interaction with condition as a predictor for coherence. This further improved the model as compared to a model containing only condition as predictor (change in $\chi 2$ [22] = 214.312, $p < .001$, AIC = -345.23). Statistically reliable interactions (ps < .05) were found on the set of bins ranging from 1.5-2.5Hz for one-arm and two-arm beats, such that higher coherence was found for these frequencies as compared to the passive condition. As is also visible in figure 2, the highest rise in coherence as compared to the passive condition was found at the 2Hz-2.5 Hz range, one-arm beat, $b = .358$, $t(1105) = 6.893$, $p < .001$, and two-arm beat, $b = .398$, $t(1105) = 7.492$, $p < .001$. Interestingly, we also found increased coherence for the Wrist Beat condition for the 2Hz range, $b = .127$, $t(1105) = 2.53$, $p = .011$. It should be noted, however, that this effect is less reliable than the effects of one-arm beat and two-arm beat condition.

We repeated these analyses for the effects of condition of coherence between F0 and movement. We obtained that condition was not a significant contributor to the model as compared to a model predicting the overall mean (change in $\chi 2$ [22] = 170.63, AIC = -

225.67, p = .459). However, adding frequency range and its interaction with condition to the model did increase predictive value as compared to model containing condition only (Change in $\chi 2$ [22] = 77.43, p < .001, AIC = -271.10). Similarly to previous analyses a reliable interaction was found such that there was an increased coherence on the 2-2.5 frequency range for the one-arm beat condition, b = .166, t(1105) = 3.088, p = .002, as compared to the passive condition. The two-arm condition showed a similar effect, but on a higher frequency (2.5-3 Hz), b = .044, t(1105) = 0.070, p = .070, but this effect was not statistically reliable in this sample.

All in all we have promising results (from a limited sample) that movement with relatively high physical impact (one-arm and two-arm beat conditions) as compared to movement with relatively low physical impact (wrist beat condition) and no movement (passive condition), are structrually affecting phonating.

Exploration of kinetic aspects

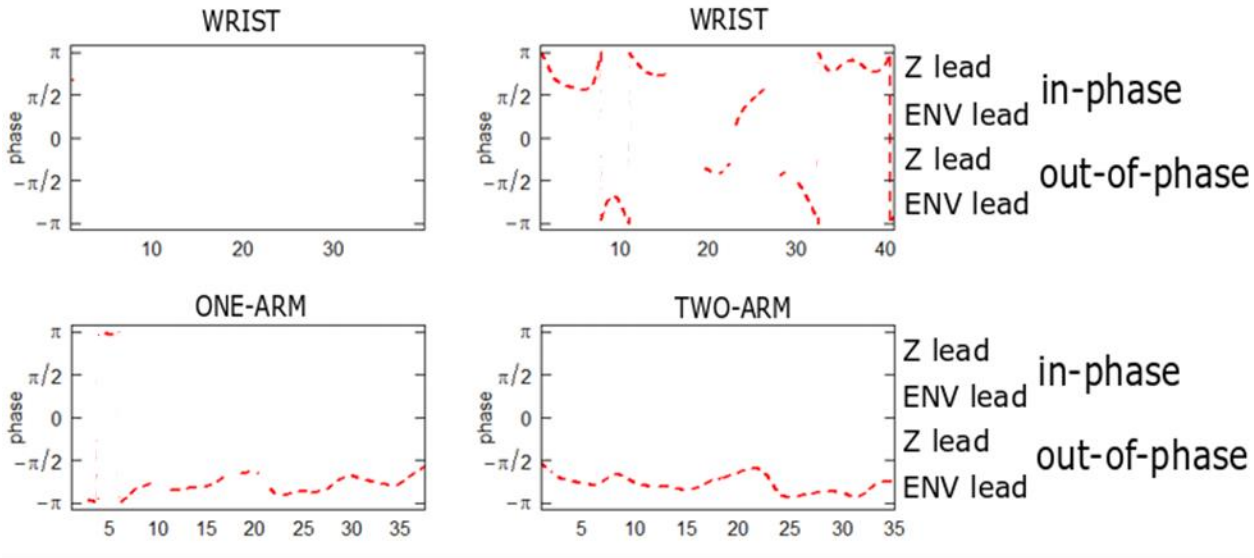
The previous analyses suggest that movement structurally affects phonation. The important question that follows is how arm movements affect phonation. The relation between kinematic properties and changes in phonation could provide insight in the underlying physical dynamics of how movement feeds into phonation. As can be observed from Figure 1 panel A, the effects of arm movement on speech is sometimes visually very apparent in that the downbeat to upward movement phase of the beat seems to coincide with a peak in ENV and F0. This could be because in this phase of the movement the highest physical momentum is generated, as a beat is made and an upward movement subsequently follows to proceed to the upward phase of the movement. Figure 1 panel C

further shows that changes in the rate of acceleration (jerk) are almost perfectly in-phase with changes in F0 (for that particular trial).

To assess when a change in movement imparts a change in phonation we first assessed the relative phase (i.e., phi: Φ) of movement with phonating using cross-wavelet analyses (Grinsted, Moore, & Jevrejeva, 2004)) using R package 'WaveletComp' (Rösch & Schmidbauer, 2014; for a helpful tutorial see Rösch & Schmidbauer, 2016). Cross-wavelet analyses utilizes a Morlet wavelet transform that allows you to decompose complex time series in dominant periodicities, and further allows you to compare periodicities between time series (hence *cross*-wavelet). We used this particular analysis to isolate at a particular relevant frequency range the relative timing of changes (relative phases) between movement and phonation through time. We performed this relative phase analysis only for participant 2 as she was most consistent in her frequency of movement at the 2 Hz range across conditions (see Figure 1). We concatenated for each condition the time series, which we entered into a cross-wavelet analyses (using 50 simulations to compute *p*-values) where we assessed the relative phases of ENV with movement for the frequency range of 2 Hz (i.e., period = 0.5), which was a dominant shared frequency of movement and phonating for participant 2 as obtained in our previous analyses. Figure 3 shows a summary of the results of these analyses. It can be observed that there are reliable relative phases for the one arm and two arm condition at $p < .01$ given the continuous presence of lines. Furthermore, it can be seen that Z movement is throughout primarily out-of phase with amplitude envelope (i.e., **Φ** is negative), which was most pronounced for the two-arm beat condition. Presumably movement and amplitude were out of phase because movement's

physical impact was highest when the movement reached its "beat" (i.e., maximum extension in the downward phase).

Figure 3. Relative phases of movement (Z) and amplitude Envelope (ENV) for participant 1 at the 2Hz range

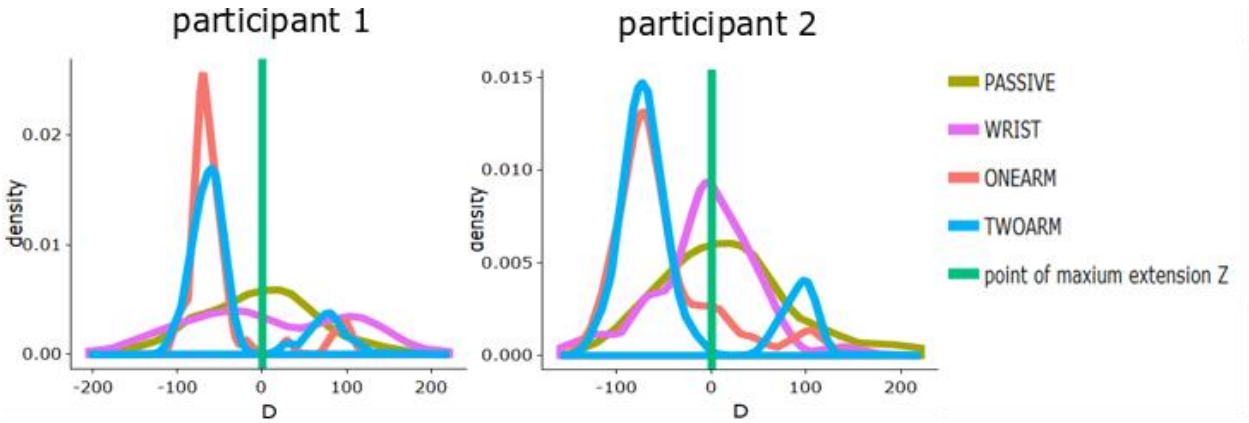


Note. Relative phase estimates per condition for the 0.5 (i.e., 2 Hz) period range. Note that a line segment indicates a statistically reliable phase relation at $p < .01$. It can be seen that for ONE-ARM and TWO-ARM conditions that a positive in change in phonation (ENV) is related to a negative change in movement (Z) as the relative phase in the out-of-phase range.

For the concluding analyses we assess how fast the forces of the arms reach the phonation system. We did this by determining the point of maximum extension of the down-beat phase, and relating this to the nearest peak in the positive rate of change (i.e., second derivative of ENV; i.e., "acceleration") of the amplitude envelope. We used the amplitude envelope 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 4 shows the main results, which provide a very clear picture such that the one-arm and two-arm beat condition lead to very reliably changes in the amplitude envelope, such that 70 ms before the point of maximum extension of the movement a positive change in the amplitude envelope was observed. However, the distributions were bimodal (especially for the two-arm beat condition) as about 70ms *after* maximum extension, changes in ENV were also observed (although to a lesser degree than changes before maximum extension is reached). To provide a formal estimate of the temporal dynamics, we assessed the mean and standard deviation of the second distribution for ONEARM and TWOARM condition combined values where $D > 0$, which showed that an effect in amplitude envelope arose at 68 ms ($SD = 21$). For the first most prominent distribution ($D < 0$), this was -73ms ($SD = 33$). These results seems to suggest that right before the moment of maximum extension - i.e., right at the moment where anticipatory postural adjustments are made to brace for the impact of the beat gesture – a change in phonation is observed.

Figure 4. Timing of peak change ENV – Point of Maximum extension beat.



Note. D 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 D is negative this indicates that peak in change of the amplitude envelope precedes the point of maximum extension.

Brief Discussion[2]

The following hierarchy of results is obtained from the exploratory study. We find that moving the arms with relatively high physical impetus (vertically moving with one arm or two arms with a contrast or beat in the down phase) makes phonation less stable as indicated by increased standard deviations around the mean as compared to beating with wrist movement only, or a passive condition. Spectral density and coherence analyses show that this variability is structural and not random, as we obtain that phonation assimilates to the rhythm of arm movements at around 2 Hz. Relative phase analyses suggested that common oscillation was structured such that during the downbeat phase of the upper limb movement changes in phonation are observed. Further, we estimated the precise timing of the relative phases, where we obtained that when the downbeat reaches its maximum extension 70 ms before and 70 ms after peaks of changes in phonation (amplitude envelope) are observed.

The findings seem to indicate that when the body is bracing for the impact of the downbeat, involuntary effects on phonation are produced. We speculate 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. As such, the current findings might signal that gesture-speech synchrony has its roots in biomechanics. To confirm this

[2] Note that we will not discuss the implications of the results in any further detail as this goes beyond the nature of this pre-registration report.

promising preliminary result, we propose a confirmatory study in the next section which is an extended version of the current preliminary study.

An important caveat however is that it is likely that gestures are not merely synchronized with speech because of biomechanical effects of gesture, if only because beat gestures can be very small in their movement amplitude and still tightly synchronize with speech (McNeill, 2005). Furthermore, gesture and speech can loosen 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. Yet even when bodily resonances cannot fully accommodate for gesture-speech synchrony, it is possible that humans become sensitive to this reliable "kinesthetic marker" which can then be intentionally, if not ontogenetically exploited, in achieving gesture-speech synchrony that can come to function in semiotic expression. In other words, sensing kinetic effects of movement could provide a cheap solution to for synchronizing and gesture and speech, as opposed to predicting and monitoring where a trajectory of the movement would align with speech' prosodic contrasts, which is maintained by information-processing theories of gesture (De Ruiter, 2000). Perhaps such a physical explanation could provide a kinesthetic anchor point for dynamic-systems ontogenetic accounts of how gesture and speech become prosodically entrained in the first place (Iverson & Thelen, 1999; Rusiewicz & Esteve-Gibert, 2018). Such a story would entail that infants and children become sensitive to the effects that arm movements have on their phonation (e.g., Lee, Bootsma, Land, Regan, & Gray, 2009), and become sensitive to how these structural effects can be exploited for communication.

Pre-registration Confirmatory Study

We will directly replicate the current findings and the analyses with a larger sample. Since the effects seem to be very pronounced and observable on the individual level we will recruit 10 participants (5 males and 5 females). Two notable changes are made to the experiment design. Firstly, we will add another within-subjects factor wherein participants perform the same movement while sitting in a chair (sitting condition). We add 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 sitting position (Cordo & Nasher, 1982). This could also explain that Krahmer & Swerts (2006) did not find effects of body movements on F0 and intensity measures as their participants were sitting throughout the experiments (also see Hoetjes, Krahmer, & Swerts, 2013). Thus if APA's are driving the current effects on phonation then we would find that the effects of upper limb movements on phonation are absent or diminished in the sitting condition relative to upper limb movement effects on phonation in the standing condition. Thus, our final design is as follow: 2-factor within-subject design with one factor (posture condition) of two levels (sitting vs. standing) and one factor (movement condition) with 4 levels (passive, wrist beat, one-arm beat, two-arm beat). Again, for each condition we will perform 4 blocks of 4 trials (total = 320 trials = 10 participants x 4 blocks x 4 movement condition x 2 posture condition). Order of conditions will be randomized for each block.

A second crucial change from the exploratory study is that we will guide the movement frequency of the participant by a visual presentation. Instead of participants moving at their own preferred frequency, participants will be encouraged to move their

hands at 80 Beats per minute (i.e., 1.3 Hz). This will allow us to analyze the data with a focus in a particular frequency range without having to account for individual differences in preferred moving rate. We have programmed in c++ a visual presentation that takes input from the motion tracker as to visually represent the frequency of the vertical movement to the participant. The visual presentation consists of a bar that changes size as a function of movement frequency and which participants keep between a certain threshold as indicated by two blue threshold bars. If movement frequencies are observed that are 10% faster or slower than the 90 BPM than the participant can read this off from the visual presentation and can adjust their movement accordingly (e.g., slow down). Note that we explicitly refrain from using a metronome as this can provide unwanted rhythmic signal that participants might entrain to, which would jeopardize answering our research question of whether physical synchronization happens between upper limb movements and phonation.

Data exclusion

We anticipate no data exclusions. Any exclusion that is made will be reported in the final paper and its effects on the results will be quantified through exploratory analyses.

Analyses

For the confirmatory analyses, we will reproduce the exact same analyses as in the exploratory study (analysis scripts that are provided at https://osf.io/5aydk/). However, we will add posture and its interaction with movement condition as additional predictors to the mixed regression models, and we will focus on a particular frequency range as to assess common periodicities between movement and phonation. Thus, similar to the current analyses with the exploratory data, we will do a mixed regression (with participant

as random intercept) predicting coherence (F0 & Z movement; ENV & Z movement) on the 1.3 Hz (a single bin of 1 Hz width; range 0.8-1.8Hz), where we will test in the final model the effect on coherence (in the prescribed frequency range) of posture (sitting vs. standing), movement condition (passive, wrist, one-arm, two-arm) as well as posture x movement condition. Since we are looking at coherence for a particular frequency range (since we are guiding movement frequency of the participants), we do need to assess differences in coherence across a broad spectrum range as in the current analyses (in other words we do not need to enter frequency range as a predictor). Furthermore, we will perform the relative phase analyses using cross-wavelet methods for *all* participants this time, as we can now ensure that participants move on a single 1.5 Hz frequency. All additional exploratory analyses will be reported as such in the final research report. For example, we might also look at other acoustic parameters such as the formants F1 and F2.

Open data and analyses

All anonymized data collected for the confirmatory study will be made publicly available on the Open Science Framework (https://osf.io/5aydk/). All final analyses R scripts will also be made available online.

References

Aruin, A. S. & Latash, M. L. (1995) Directional specificity of postural muscles in feed-forward postural reactions during fast voluntary arm movements. *Experimental Brain Researc*h, *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

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

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

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.

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

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

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., Lellouch, L., Brown, E. C., Depraetere, M., Desjonqueres, 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

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

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. (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. (2018b; unpublished preprint). Entrainment and modulation of gesture-speech synchrony under delayed auditory feedback. doi: 10.17605/OSF.IO/AVJ7M

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).

Richardson, M. J. (n.d.). Retrieved from http://xkiwilabs.com/software-toolboxes/)

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

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., Lellouch, 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