AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 5, 1469–1483
https://doi.org/10.1037/xge0001346

Gesture–Speech Coupling in Persons With Aphasia: A Kinematic-Acoustic Analysis

Theodore Jenkins[1] and Wim Pouw[2]
[1] Communication Department, Rhode Island College
[2] Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen

Aphasia is a profound language pathology hampering speech production and/or comprehension. People With Aphasia (PWA) use more manual gestures than Non-Brain Injured (NBI) individuals. This intuitively invokes the idea that gesture is compensatory in some way, but there is variable evidence of a gesture-boosting effect on speech processes. The status quo in gesture research with PWA is an emphasis on categorical analysis of gesture types, focusing on how often they are recruited, and whether more or less gesturing aids communication or speaking. However, there are increasingly louder calls for the investigation of gesture and speech as continuous entangled modes of expression. In NBI adults, expressive moments of gesture and speech are synchronized on the prosodic level. It has been neglected how this *multimodal prosody* is instantiated in PWA. In the current study, we perform the first acoustic-kinematic gesture–speech analysis in Persons With Aphasia (i.e., Wernicke's, Broca's, Anomic) relative to age-matched controls, where we apply several multimodal signal analysis methods. Specifically, we related the speech peaks (smoothed amplitude envelope change) with that of the nearest peaks in the gesture acceleration profile. We obtained that the magnitude of gesture versus speech peaks are positively related across the groups, though more variably for PWA, and such coupling was related to less severe Aphasia-related symptoms. No differences were found between controls and PWA in terms of temporal ordering of speech envelope versus acceleration peaks. Finally, we show that both gesture and speech have slower quasi-rhythmic structure, indicating that next to speech, gesture is slowed down too. The current results indicate that there is a basic gesture–speech coupling mechanism that is not fully reliant on core linguistic competences, as it is found relatively intact in PWA. This resonates with a recent biomechanical theory of gesture, which renders gesture-vocal coupling as fundamental and a priori to the (evolutionary) development of core linguistic competences.

Keywords: aphasia, multimodal prosody, gesture, speech, gesture–speech synchrony, AphasiaBank

Human spoken language is characteristically multimodal. We gesture and speak at the same time. Manual gestures and speech show entanglement at multiple levels of description (Feyereisen, 2017). They are not only semantically co-expressive, but also temporally coherent on the level of prosody (Loehr, 2012; Rochet-Capellan et al., 2008; Wagner et al., 2014). Indeed, a gesture's kinematic profile tends to couple with a prosodic profile of concomitant speech (Loehr, 2012; Wagner et al., 2014), even when the kinematic profile is *also* designed to represent through depiction or enaction (Pouw & Dixon, 2019, 2020; Shattuck-Hufnagel & Prieto, 2019). Recent machine learning research attest to this, showing not only that the presence or absence of gesture can be predicted from speech alone (Yunus et al., 2020), but also that gesture kinematics can be convincingly simulated from novel speech acoustics by deep neural networks trained on gesture kinematic and speech acoustic associations (Alexanderson et al., 2020; Ferstl et al., 2020; Ginosar et al., 2019). These machine gestures are not yet semantically laden but do look believably human (see video from Ginosar et al., 2019). This suggests that beat-like gestures in synchrony with prosodic aspects of speech is an important co-regulatory process. Indeed, there is now ample evidence that gesture and speech are flexibly coordinating so as to maintain a certain temporal interrelationship (Chu & Hagoort, 2014; Pouw & Dixon, 2020), though such (or related) research is almost exclusively performed with NBIs (but see de Marchena & Eigsti, 2010; Lozano-Goupil et al., 2022; Pouw et al., 2022), and it has simply not yet been pursued for research on speech pathologies.

A broader inquiry into multimodal communication will inform a range of theories about multimodal communication. In De Ruiter's extended Levelt (1993) model (de Ruiter, 2000), it is assumed that at later formulation stages, gesture and speech become "executed separately by the two underlying processing modules, *which are*

This article was published Online First April 13, 2023.
Theodore Jenkins https://orcid.org/0000-0001-5542-9638
Wim Pouw https://orcid.org/0000-0003-2729-6502
Theodore Jenkins and Wim Pouw share first authorship.
This work has been partially supported by a VENI grant (VI.Veni.201G.047) awarded by the Dutch Research Council (NWO) to Wim Pouw (PI).
The data are available at https://osf.io/bwfmp/ (Pouw & Jenkins, 2022).
Theodore Jenkins performed the original study and produced new ELAN annotations needed to perform this work. Wim Pouw performed the analysis. Wim Pouw and Theodore Jenkins co-wrote the manuscript.
Correspondence concerning this article should be addressed to Wim Pouw, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, Thomas van Aquinostraat 4, 6525 GD Nijmegen, The Netherlands. Email: wim.pouw@donders.ru.nl

temporally coordinated by a signaling mechanism" (p. 3; emphasis added; de Beer et al., 2020). Indeed, when assuming some kind of *independence* while accounting for the fact that gesture and speech are tightly temporally *dependent* in NBIs, one needs to invoke some "signaling" mechanism that keeps two putatively independent processes in sync. An oft-cited alternative to the assertion that gesture and speech are formulated separately, is the work by David McNeill (McNeill, 2008). There it is emphasized that gesture and speech are formulated by a single linguistic formulation process (i.e., the growth point) sometimes ascribed to "a specific thought-language-hand link in the brain" (p. 236). Although the putative independence or dependence of gesture and speech can be considered theoretical assertions, we find it difficult to ascertain what testable mechanisms are posited that are responsible for said *in*dependence or *de*pendence. An alternative recent view about the coupling of gesture and speech on the prosodic level is based on biomechanics (Pouw & Fuchs, 2022). Here it is assumed that the coupling of vocalization and upper limb movement is already present pre-linguistically, in the sense that it is in rudimentary form already present in infant vocal-motor babbling and its core mechanism is shared with other non-human animals. The core mechanism is based on cross-species evidence that upper/pectoral limb movements can affect rib-cage movements and thereby respiratory functioning, which will affect vocal degrees of freedom (Cooper & Goller, 2004; Lancaster et al., 1995; Pouw, de Jonge-Hoekstra, et al., 2020). The biomechanical theory is to be differentiated from the aforementioned views sketched, in that it suggests that gesture and speech are coupled in a rudimentary way that is not completely governed by a dedicated linguistic system. However, while gesture–speech coupling may emerge from biomechanics, it is of course regulated by the prosodic regimes of the particular language that is spoken, or the vocal music that is sung (Pearson & Pouw, 2022).

The current article contributes to the effort of understanding gesture–speech coupling on the prosodic level by assessing whether Persons With Aphasia (PWA) versus NBI persons differ on several characterizations of such gesture–speech coupling. If gesture in PWA has similar temporal characteristics in kinematics as compared to gestures from age-matched controls, it would mean that some of the known formulation processes that are compromised in Aphasia are independent of gesture formulation. From a McNeillian perspective, we might hold that gesture kinematics cannot be coordinated with speech as the mechanism that orchestrates them relies on core-linguistic competences that are severely impaired in persons with Aphasia (McNeill, 2008; McNeill et al., 2010). From this perspective, we expect that gesture and speech are temporally different from age-matched controls and there is no reliable coordination between modalities. Finally, if we find statistically reliable gesture–speech coordination, this would signal a rudimentary pre-linguistic gesture-vocal coupling that is asserted by the aforementioned biomechanical theory (Pouw & Fuchs, 2022). Next to these theoretical implications, this research further contributes to the identification of possible multimodal markers of speech pathology, which could enrich diagnostic tools that focus on one modality at a time.

Aphasia and Multimodal Characteristics

In the literature, three common types of Aphasia are discerned. *Wernicke's Aphasia* is often referred to as a "fluent" aphasia that is typified by largely preserved fluent language *production*, but impaired *comprehension* (Robson et al., 2012). A typical profile/speech sample leads to syntactically correct speech, but inappropriate semantic content. *Anomic* Aphasia is a milder and also a fluent aphasia, where language comprehension and production are relatively intact, but is often presented with issues of lexical retrieval (Goodglass & Wingfield, 1997). *Broca's Aphasia* is a non-fluent aphasia that is typified by preserved language *comprehension,* but impaired *production* (Dronkers et al., 2010). Typically, a Broca's Aphasia speech sample will have appropriate word selection, but impoverished syntactic morphology (e.g., omission of English suffixes like *-ed*, *-ing*, *-s*).

PWA exhibit a significantly increased gesture rate compared to NBIs (e.g., Sekine & Rose, 2013). Some theorize that in PWA gesture compensates for their linguistic impairments (Rose, 2006). As both aging NBIs (>60 years of age) and aphasic populations are affected in linguistic performance tasks of fluency, word selection, and comprehension, and given that said linguistic performance relates to gesture increases (Kong, Law, Kwan, et al., 2015; Kong, Law, Wat, & Lai, 2015). It may be concluded that there is compensatory support by gesture in communication (de Beer et al., 2020; de Ruiter, 2000). Research on the nature of the compensation suggests that it might lie in the gestural clarification or disambiguation of difficult-to-interpret speech, as well as resolving word retrieval issues (Akhavan et al., 2018; de Beer et al., 2020; Kistner et al., 2019; Rose et al., 2017; van Nispen et al., 2017).

We keep our overview of mainstream gesture research and the implications for possible "box-and-arrow" theories of gesture production in Aphasia somewhat brief (see however de Beer et al., 2020). This is because they simply are not directly engaging with the phenomenon of *multimodal prosody*—that is, the way that gesture modulations (in kinematics) relate to (acoustic markers of) speech prosody (Wagner et al., 2014). That mainstream theories on gesture as related to Aphasia have overlooked this phenomenon is a pity: The prosodic modulations together with gestures may be an important semiotic means to communicate in Aphasia despite impaired speech processes on other levels (e.g., syntactic and lexical). For example, rhythmic beat-like gestures seem to be a common (if not the dominant) mode of gesturing in PWA (Ferré, 2021), though there seem to be marked differences between aphasia typology (e.g., people with Anomic Aphasia seem to use more beat gestures as compared to other Aphasia types; Ferré, 2022).

So why would rhythmic beat gestures be common in PWA? Consider that it has been demonstrated in a case study by Charles Goodwin that prosody *together* with meaningful gestures can carry an important communicative load in Aphasia (Goodwin, 2010). As Goodwin remarks on a person named "Chil" who has a severe form of Aphasia:

> "[the] rich prosody is a most important component of his combinatorial resources, and contributes to his ability to build action by combining different kinds of signs into meaningful wholes. Through prosody, Chil's limited vocabulary becomes capable of participating in the construction of richly varied utterances and actions. (p. 381)

Though Goodwin's analyses were not about pulsing gestures per se, it highlights the importance of context in *languaging* with Aphasia. Namely, when more free-standing decontextualized construction of meaning in the forms of lexically and syntactically well-formed sentences is limited, the use of other contextual resources becomes extra important (Doedens & Meteyard, 2022). Meaningful gestures can, for example, indicate complicated agent–patient relations (e.g., "I hammered the glass"). Speech prosody can help indicate the informative elements in speech and gesture (i.e., "the glass"). But it is the synchronization of gesture and speech prosody that allows for processing of a contextual rendering of limited speech content. Thus synchronization can serve the marking of important moments in speech and gesture that are otherwise difficult to identify.

In a recent study with four PWA by Ferré (2021), the open question of multimodal prosody and temporal aspects of gesture in aphasia was directly addressed, and the study, therefore, serves as an important initial understanding of multimodal prosody in PWA. Firstly, it was obtained that gesture phases (initiation, pre-stroke hold, stroke, post-stroke hold) are not markedly different on average for PWA versus NBI, except maybe for higher duration in pre-stroke hold in PWA versus NBI—suggestive perhaps for gestures waiting for speech to initiate. Though on average gesture temporal aspects may not show clear differences, much higher variability was observed in the gesture phase timing distributions in PWA as opposed to NBI. Acoustic markers of prosody were reported to be different, especially in the use of (rise-)falling F0 contours that were more prevalent in PWA as compared to NBI. These differences in acoustic markers do not mean "impaired" prosodic modulations though, but rather differently "situated" (Doedens & Meteyard, 2022) ways of using prosody. Interestingly, Ferré (2021) further reports that the beat gestures were very prevalent (even more so than NBI) and also suggests that these beat gestures seemed to serve as anchors for the "pace of speech and [thereby creating] a rhythmic effect of entrainment…" (p. 13; which is also reported by Orgassa, 2005). Though in some cases when retrieval issues were severe the emphatic function of beat gestures is not clear and may function to prolong the holding of a turn in a conversation.

So to provide a preliminary conclusion, it is certainly not the case that (multimodal) prosody is never affected on some levels in some forms of Aphasia (Danly & Shapiro, 1982; Ferré, 2021; Gandour et al., 1988). If anything there seems to be more variability in temporal aspects of gesture and thus potentially gesture–speech dynamics (Ferré, 2021). Nevertheless, in PWA, gestures do still seem to function so as to entrain with speech. Taking all this into account, we deem it still very much an open question whether statistical trends of multimodal prosody can be reliably detected across different forms of Aphasia. Before outlining the current study, we will overview relevant research that has quantified multimodal prosody and revealed some possible mechanisms between modalities.

Gesture–Speech Coupling Mechanisms

What mechanisms account for gesture–speech temporal coordination? The available research on this suggests that gesture and speech is not simply an internal neural timing system that keeps gesture and speech formulated plans temporally in check. Rather it is something that also incorporates ongoing (re)afferent feedback of the body as well as biomechanical factors (Pouw et al., 2022; Pouw & Fuchs, 2022).

Gesture and speech naturally align their activity. Stressing a syllable while pointing will attract the apex of the pointing gesture stroke to occur during the stressed syllable (Esteve-Gibert & Prieto, 2013; Rochet-Capellan et al., 2008). Even when one does not intend to change the hand motion during contrasts in speech prosody, gesture and speech will adjust accordingly if gesture kinematics is contrasted (Krahmer & Swerts, 2007; Parrell et al., 2014). Firstly, there is now ample research showing that gesture–speech coordination dynamically adjusts relative to feedback about gesture or speech, specifically visual, auditory, and proprioceptive feedback play a role in maintaining gesture–speech synchrony (Chu & Hagoort, 2014; Dohen & Roustan, 2017; Kelso & Tuller, 1984; Pouw & Dixon, 2020; Pouw et al., 2022). For example, when the hand is midway of performing a pointing gesture and the visual feedback of the hand is disturbed through virtual reality goggles, this in turn delays the gesture movement, speech adjusts, and waits for the gesture to reach the apex (Chu & Hagoort, 2014). In a single case study (McNeill, 1992) with someone who has lost the sense of proprioception, it has been found that visual feedback is important to maintain gesture–speech synchrony (Pouw et al., 2022). Further, it has been observed that when speech is slurred due to delayed auditory feedback, gesture adjusts and slows down too, so as to maintain gesture–speech synchrony (McNeill, 1992; Pouw & Dixon, 2020; Rusiewicz et al., 2013). Finally, in children 4–5 years old, it has been observed that head gestures entrain to, and seem to help realize contrastive focus in speech (Esteve-Gibert et al., 2021).

It has recently been found that the coupling of gesture and speech also involves biomechanical loops. That is, acoustic–prosodic modulations in speech and vocalizations "timed" with gesture kinematic contrasts (e.g., "beating" moments in gestures) are partially instantiated by interactions between mechanical loading of the upper limbs onto the respiratory–vocal system. Especially *kinematically contrastive upper limb movements*, such as gestures with a beat quality, can produce forces onto the musculoskeletal system that can affect the Fundamental Frequency (F0). This also affects the amplitude of the voice through interactions with respiratory–vocal system (Pouw, de Wit, et al., 2021; Pouw, Dingemanse, et al., 2021; Pouw & Dixon, 2020) and F0 and the amplitude are key markers of prosodic contrasts. These experiments and several related findings (Cravotta et al., 2019;) provide support for the idea that there is a feedback of the gesture and speech trajectory that instantiate the gesture and speech formulation process partly in biomechanical processes. It is currently unclear how such processes play out in people with more or less severe neuro-atypicalities, but in principle, such biomechanical dispositions should be able to be exploited even when core linguistic processes are affected. Evidence that speaks to a robust gesture–speech coupling, was obtained with a person who has lost the sense of proprioception but seems able to exploit the beat-quality of gesture when vision is not available. It was suggested that by recruiting more forceful gestures the biomechanical process was exploited, which synchronizes speech and gesture through interactions with the respiratory system (Pouw et al., 2022). As such, gesture–speech synchrony on the prosodic level can be said to emerge through dynamic causal loops that are in part instantiated by the body rather than the brain (Hurley, 1998).

From the biomechanical view reviewed above, we predict that we should expect some statistically reliable gesture–speech coupling in people with aphasia, as we know that they move their hands vigorously during the production of speech sounds. This alone should lead to emergent synchrony driven by physical interactions between respiratory-vocal aspects of speech and gesture kinematics, even for semantically laden iconic gestures.

The Current Article

In this study, we performed the first kinematic-acoustic analysis of gesture–speech synchrony on the prosodic level in PWA using audiovisual data obtained from AphasiaBank (MacWhinney et al., 2011), which were previously analyzed for gesture occurrences (Jenkins et al., 2017). We further applied video-based motion tracking to extract information about manual gesture kinematics, which we related to the smoothed amplitude envelope in speech. With this analysis approach, we directly answer recent calls in this field to apply more quantitative and continuous analysis of gesture and speech in PWA (Stark et al., 2021a). Namely, we perform several analyses that provide fully reproducible procedures to probe whether speech and gesture activity are co-modulated, whether they synchronize in time, and how gesture and speech rhythms are related. This approach is comparable to recent methods that have been reported in several other studies (Lozano-Goupil et al., 2022; Pouw, de Jonge-Hoekstra, et al., 2020; Pouw & Dixon, 2019, 2020; Pouw et al., 2022; Pouw, Trujillo, & Dixon, 2020). The following research questions will be answered:

1. Is there evidence for gesture–speech coupling in people with aphasia versus controls?
   a. Can we find magnitude coupling, such that peaks in changes in the amplitude envelope are related to the nearest peaks in the acceleration profile of gesture? Note that, coupling betweenacceleration and amplitude envelope coupling are likely quantitative markers of gesture–speech biomechanics (Pouw, de Jonge-Hoekstra, et al., 2020).
   b. Is there temporal coupling, such that peaks in changes in amplitude envelope consistently time with nearest peaks in the acceleration profile of gesture?
2. If we treat gesture kinematics and speech envelopes as continuous oscillating events, what quasi-rhythmic structure appears for the different groups? Does gesture follow the rhythms of speech?

Method

Participants

There were a total of 42 participants (left-handed = 2, right-handed = 39, ambidextrous = 1, unknown = 1) included for this analysis using video samples of persons with aphasia (PWA) and non-brain injured (NBI). Participant data were drawn from an online database (AphasiaBank; MacWhinney et al., 2011). This database includes video and audio samples, as well as results from a test of aphasia severity (Western Aphasia Battery Aphasia Quotient, WAB-AQ; Kertesz, 1982) and naming ability (Boston Naming Test, BNT; Mack et al., 1992). Furthermore, the PWA group was further categorized into the following three subtypes based on scores on the WAB (Kertesz, 1982): Anomic, Broca's, and Wernicke's Aphasia. For further information see Table 1.

Please see our online supplemental materials (https://osf.io/3qpg2) for detailed anonymized information (i.e., participants' age, available information about lesions, aphasia type, as well as the original studies for which these participants were recruited).

The current PWA sample includes individuals who are all monolingual speakers of English, at least 40 years of age, and have an aphasia severity of mild-moderate as indicated by a WAB-AQ of at least 45. This aphasia score ensures that all of the speakers are able to produce at least some discernible connected language. Additionally, aphasia is often co-morbid with issues of hemiparesis which may affect gesture production. The AphasiaBank protocol does not have reliably reported information on the presence or extent of such co-morbidities.

Note that, originally, this study selected at random 60 narratives from Aphasiabank of PWA (i.e., total $n = 45$; 15 Anomic, 15 Broca's, 15 Wernicke's) and NBI ($n = 15$). However, the current sample is smaller because some of this data was of poor video and/or audio quality, or because gestures were partly out of view of the camera frame. The original sample size for each group was based on the availability and usability of language and video data for the fewest number of participants in a single group (i.e., $n = 15$ for Wernicke's). The other groups were assembled through random-matching of age, gender (6 females per group), and the presence of some spoken language to meet the group size of Wernicke's group. Large group sizes are rare in studies of PWA, and these numbers were chosen to maximize the power of findings (Wang & Krishnan, 2014).

Procedure and Tasks

The following tasks were performed by the participants of the different studies for which they were originally recruited (MacWhinney et al., 2011).

Narrative Task

As a part of the Aphasiabank protocol, each participant was asked to do a narrative retelling of the Cinderella story. Participants (i.e., PWA and NBI) were shown a 36-panel wordless picture book outlining the story. After participants examined the book at their own pace, it was taken away. They were then each asked to retell the story in their own words. The language samples had been collected as part of a battery of tests according to an IRB-approved protocol at several medical and academic settings to sample such abilities as working

Table 1
Information About the Current Sample

Group	*n*	WAB AQ	BNT	Age	Gender
Anomic	11	Range: 68.5–93.4 mean: 80.9	Range: 1.00–15.0 mean: 9.91	Range: 41.4–83.2 mean: 65.1	6 male; 5 female
Broca's	10	Range: 45.5–71.1 mean: 57.7	Range: 0.00–12.0 mean: 3.80	Range: 42.2–78.3 mean: 62.4	5 male; 5 female
Wernicke's	12	Range: 53.0–74.4 mean: 63.9	Range: 1.00–14.0 mean: 6.33	Range: 42.6–91.7 mean: 67.8	6 male; 5 female
NBI	9	N/A[a]	N/A[a]	Range: 42.2–85.1 mean: 68.6	7 male; 2 female

Note. WAB-AQ = Western Aphasia Battery Aphasia Quotient; BNT = Boston Naming Test; NBI = non-brain injured.
[a]NBI were not tested on the WAB AQ and BNT.

memory, aphasia severity, and narrative production. Trained research associates at each institution tested all participants included in the database.

Aphasia Severity (i.e., WAB-AQ)

Information on aphasia severity was collected through the administration of the Western Aphasia Battery (WAB-AQ; Kertesz, 1982). This is a standardized battery of language tasks, including lexical retrieval, linguistic comprehension, and fluency. The metric produces a score from 1 (*severely impaired*) to 100 (*not impaired*). Additionally, based on the scoring on certain subtests on the WAB, it produces classifications for certain subtypes of aphasia (e.g., Anomic, Broca's, Wernicke's). Conventionally studies involved with PWA have relied on the WAB as a diagnostic metric to classify individuals to a certain sub-type of aphasia (e.g., Anomic, Broca's, Wernicke's). The AphasiaBank database relies upon this for participant classification for its group analysis. It should be noted that there are studies showing that participant diagnoses primarily depend on verbal output during WAB administration (e.g., Ellis et al., 2021; Rao et al., 2022).

Boston Naming Test (i.e., BNT)

The BNT is a confrontational naming test consisting of 60 line drawings of different objects of increasing difficulty (Kaplan et al., 1983). The resulting score will produce a score of 0 (poor/low performance) to 60 (perfect/high performance).

Mildness: Aggregate Measure of BNT and AQWAB-AQ

We will further refer to an aggregate measure that provides information about the *mildness* (or severity) of speech pathology in general. In such instances, we submit for analysis a particular gesture–speech measure (e.g., gesture–speech mean synchrony) that will be related to normalized scores of WAB-AQ and the BNT. We then set Test Type (WAB-AQ vs. BNT) as a random intercept.

Post-Processing and Analyses

Gesture Annotations

Gestures were originally coded and assessed for inter-rater reliability in an earlier study about gesture prevalence in PWA (Jenkins et al., 2017). For the current study, these codings were recoded into ELAN (Wittenburg et al., 2006) with additional information about gesture visibility. The coding procedure for gesture segmentation, type, and event timing was adopted from the following.

Gesture Segmentation

For each narrative, manual movements were identified and segmented for gestures. For any such movement to be considered a gesture, the following three phases were required: (a) preparation, where the hand(s) is moved from a place of rest/inactivity; (b) stroke, in which the hand(s) is moved in front of the body; and (c) retraction, in which the hand(s) returns to a position of rest (McNeill, 1992). Other studies of language and gesture have shown that participants do not always follow this exact path (e.g., hands pause in mid-space and change in obvious trajectory and hand shape; Kong, Law, Wat, & Lai, 2015). In coding such events, this was considered to be a boundary between separate gestures. Additionally, hand movements that were thought to be non-linguistic in nature (e.g., scratching one's nose, running hands through the hair) were not included in gestural coding and analysis (i.e., self-adapters; Jacobs & Garnham, 2007).

Representational Versus Non-Representational Gesture Coding

McNeill (1992) identifies a general difference between *representational* (i.e., iconic, metaphoric, and deictic) and beat gestures, which include some general communicative information. This distinction has been elaborated on to over 20 specific categories in recent years (Kong, Law, Wat, & Lai, 2015; Sekine & Rose, 2013). In the interest of reliability gesture classification and identification, this study uses McNeill's original distinction of representational versus beat. This can be quantified through the application of a beat filter (McNeill, 1992). Briefly, the beat filter is a series of questions that examine each gesture and in turn, gives a point value based on those movements as to how "beat-like" the gesture is. For example, two of the beat filter questions ask whether a particular gesture has only two movements (i.e., up and down) and is made in the central part of the body (i.e., near the abdomen). The beat filter assumes that beat movements have minimal movements (e.g., two; up and then down), and are made in minimal space (McNeill, 1992). Based on the beat filter questions, if the gesture was rated as being very beat like (i.e., beat filter rating of 1), it was classified as beat, otherwise, it was considered to be representational (e.g., beat filter rating of 5). The coding scripts in ELAN were diametrically categorized as either being representational or non-representational.

To reduce artifacts of dexterity in our analysis, only gestures that included movement of the dominant hand were considered. In the case of two-handed gestures, the kinematics (e.g., speed) was averaged over the two effectors. If the hands were partly out of the frame, we annotated these instances in ELAN, and then excluded these gestures for kinematic-acoustic analysis.

Acoustic and Motion Tracking

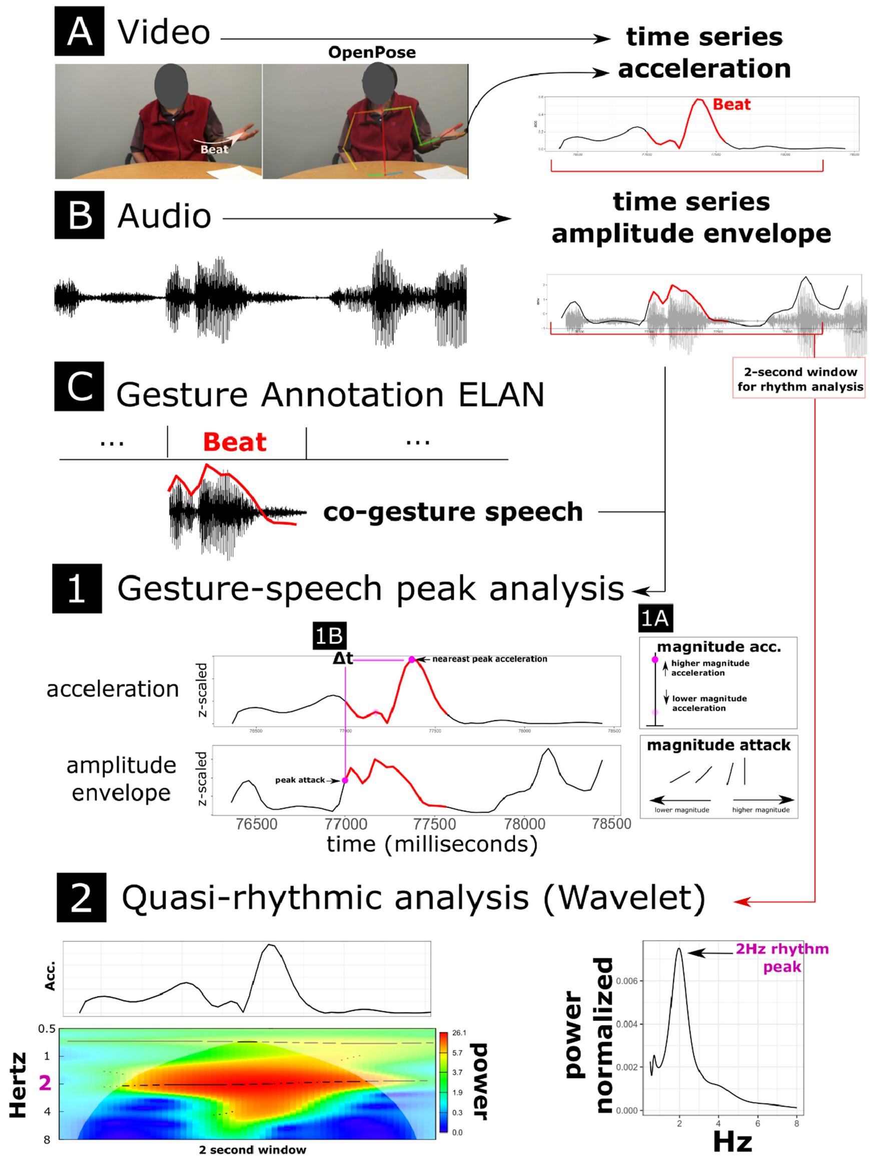
Speech: Amplitude Envelope

A smoothed amplitude envelope was extracted from the audio signal (He & Dellwo, 2017). This was done by applying the Hilbert transform to the raw waveform, and taking the complex modulus from the complex analytical signal. We then applied an 8 Hz hanning window for smoothing, and resampled the signal to 30 Hz. All this yields a 1D time series that roughly traces the amplitude of the speech waveform signal, that is, the smoothed amplitude envelope of speech (see Figure 1C). For the R code to extract the amplitude envelope from speech please see Pouw and Trujillo (2019). For our analysis, we also take the first derivative of the amplitude envelope (the change in amplitude envelope) to measure the peak attack phase of the signal (i.e., sudden rises of the amplitude envelope; see e.g., MacIntyre, Cai, & Scott, 2022).

Gesture Kinematics: OpenPose

The video frames (30 frames per second) were tracked with OpenPose, yielding human pose information at 30 Hz temporal resolution. To filter out tracking inaccuracies that lead to jitter in the

Figure 1
Overview Methods and Analysis



Note. This figure overviews the current study procedure. The video data (panel A) is processed for hand motions, extracting the absolutized acceleration profile of the dominant hand. The audio (panel B) is processed by extracting the smoothed amplitude envelope. Gesture annotation (panel C) is performed in ELAN based on the video, allowing us to select relevant gesture and co-gesture speech events. Once we are able to query gesture events, we perform analysis (1) where we assess whether the magnitude of the attack in the amplitude envelope (a more extreme burst of intensity change leads to a higher magnitude attack) scales with the magnitude of the nearest peak in acceleration. In (1B), the timing between the peak in positive change in amplitude envelope (attack) with the nearest peak in the gesture acceleration (i.e., yielding gesture–speech timing). Finally, in analysis (2), we extract 2 second windows around the gesture event, and we yield the average power over the 0.5–8 Hz range through a wavelet analysis of the amplitude envelope (not shown here) as well as the gesture acceleration (shown here). In this example, this beat gesture event had a dominant quasi-rhythm at 2 Hz. See the online article for the color version of this figure.

position traces, we applied a third order Kolmogorov–Zurbenko filter with a span of five frames. We apply this filter again for derivative measures, such as speed and acceleration. The main body key points of interest were the index finger of the right and left hand. From these keypoint position x and y traces, we calculated the 2D speed. For the analysis, we use the first derivative of speed (i.e., acceleration) as there is reason to believe that this kinematic feature is especially informative about gesture–speech dynamics (e.g., Ferstl et al., 2020; Pouw, de Jonge-Hoekstra, et al., 2020) and has an important status in the biomechanical theory of gesture–speech coupling as it relates to the amount of forces produced by the gestures (Pouw & Fuchs, 2022). Acceleration was absolutized (so higher values can mean higher acceleration or deceleration).

Key Analyses

Figure 1 provides a general overview of the analysis procedure of this report. As referred to in the introduction we perform: (1A) gesture–speech magnitude coupling, (1B) gesture–speech temporal coupling, and (2) quasi-rhythmic gesture–speech profiling.

Gesture–Speech Magnitude and Temporal Coupling Analysis

Firstly, for each gesture event, we assess whether the positive peak in the amplitude envelope change (informative about attack phase of speech) is related to the magnitude of the nearest observed peak in the acceleration profile of the gesture. We choose the amplitude envelope change because this marker is important for perceiving beats in speech (Rathcke et al., 2021). To do this, for each gesture event, we simply take the global maximum for the amplitude envelope change. If there was no speech occurring during the gesture, we skip the procedure and go to the next gesture event. Then we apply a peak finding algorithm on the gesture acceleration profile that identifies the local maxima's peak, but only if they exceed 25% of the standard deviation of the acceleration signal (this thus excludes minor magnitude peaks). We are then left for each gesture event, a value that indicates the magnitude of the attack phase of speech and the magnitude of the nearest peak in the acceleration of the gesture. If these values correlate, it would mean that more effortful/forceful gestures are related to modulated speech, that is, that speech and gesture activity couple in their magnitude, which in previous research has been related to biomechanical coupling (Pouw, de Jonge-Hoekstra, et al., 2020).

We also relate gesture–speech magnitude coupling to the Mildness score of PWA (see results). The same analysis is performed, but this time we are interested in the moments when these peaks in speech and acceleration happen. We thus simply calculate the temporal difference in speech and acceleration peaks as defined above, and assess the gesture–speech asynchrony distribution (e.g., mean, kurtosis), which will also be related to mildness scores.

Gesture-Speech Continuous Analysis

Finally, in an exploratory analysis, we also provide information about more continuous properties of the speech and gesture signal. Namely, what are the dominant periodicities in the signal? For example, if speech is slurred we would predict that the speech envelope is defined by slower periodicities (i.e., higher powers at slower rhythms), and we also assess whether this has similar repercussions for gesture. We applied wavelet analysis using R package WaveletComp (Rosch & Schmidbauer, 2014). For each gesture event, we selected a time window of at least 1 s around the gesture event, so that we have comparable lengths of the gesture and co-gesture speech time series. Then we submit for each event, the amplitude envelope and the acceleration profile of the gesture to a wavelet analysis where we compute the average (normalized) wavelet powers at the 2–8 Hz range (this range aligns with what has been found as relevant timescales for both speech and manual communication; Brookshire et al., 2017; Pouw, Jaramillo, et al., 2020).

Open Data and Scripts

Since we cannot share the raw video and audio data due to privacy concerns, we have only made available the processed data (https://osf.io/xjkh2/). However, our processing script that aggregates all raw data streams into a processed time series file can be found here: https://osf.io/tcwvs/. Our main analysis script that takes the processed data input can be found here: https://osf.io/4fs29/. Note further that most of the basic multimodal signal processing steps (motion tracking, acoustic signal processing, merging ELAN annotations) are described and explained in this methods paper: Pouw, Trujillo, and Dixon (2020), and this online tutorial: Pouw and Trujillo (2019).

Alpha Correction

Given that we ask three exploratory questions (magnitude coupling, temporal coupling, and rhythmic coupling), we perform our study with a relatively small sample size due to the rarity of the population, and given that PWA will likely have much more variable underlying distributions of behavior, it would be good to reduce risks of overinterpreting results. We have therefore reduced our alpha level to a Bonferroni thresholded level, where we treat a result as statistically reliable when the *p*-value is equal or lower than .016.

Results

Descriptives

The total length of the recordings were 165.90 min (Control = 28.42 m, Wernicke = 51.55 m, Broca = 31.60 m, Anomic = 54.33 m). We annotated 944 gestures that included the dominant hand(s) (Control = 60, $M = 22.82$, $SD = 17.95$, min = 0, max = 47; Wernicke = 377, $M = 31.42$, $SD = 23.13$, min = 6, max = 84; Broca = 258, $M = 25.80$, $SD = 11.61$, min = 8, max = 47; Anomic = 249, $M = 22.64$, $SD = 17.24$, min = 0, max = 47). There were three control participants and two anomic participants who did not gesture, or for whom we did not have information about the dominant hand, or whose gestures were not in full view and thus not analyzable with our methods. These participants are therefore excluded for any analysis that involved gesture kinematic analysis. In total, there were 49.89% beat gestures (control $N = 65\%$, Wernicke $N = 49.07\%$, Broca $N = 39.53\%$, Anomic = 58.23%) and 50.11% representational gestures (control $N = 35\%$, Wernicke $N = 50.93\%$, Broca $N = 60.47\%$, Anomic = 41.78%) annotated.

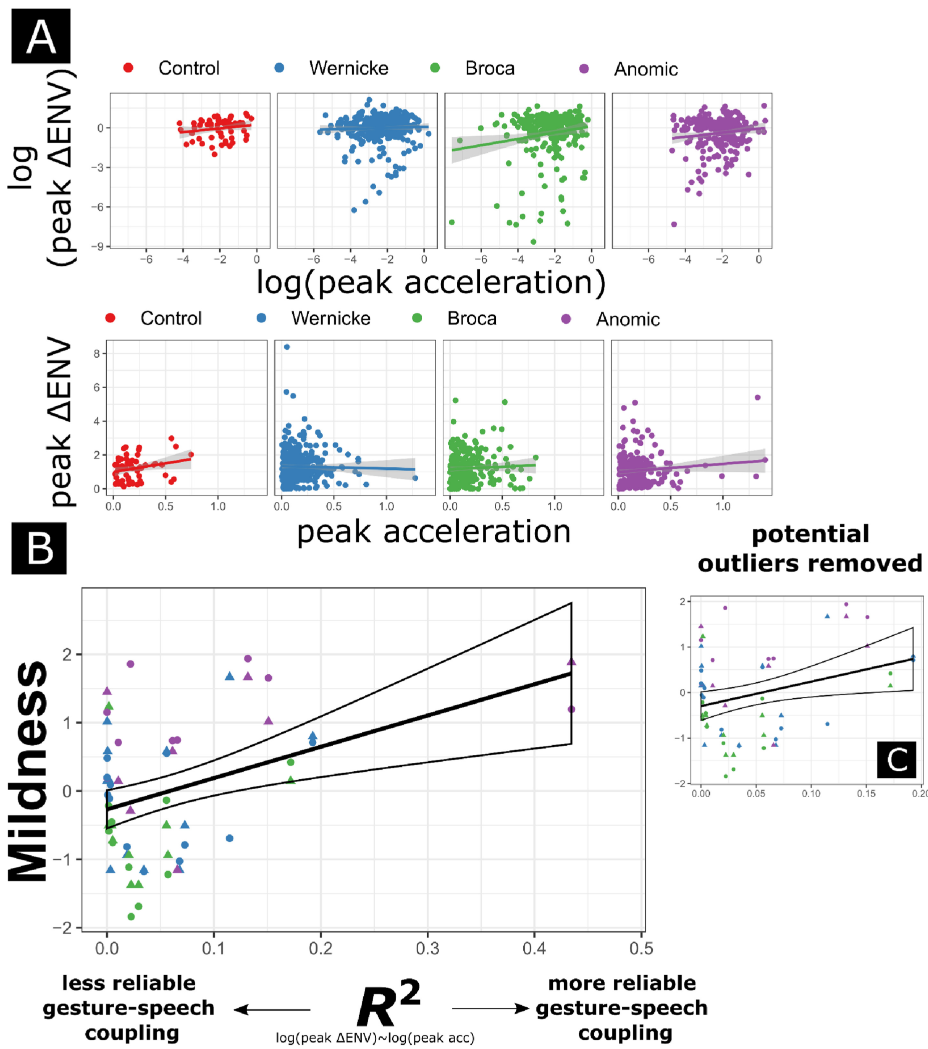
Gesture–Speech Magnitude Coupling

Our first analysis concerns whether the effort in one modality spills over to another. We call this gesture–speech magnitude coupling analysis, and it assesses whether the highest peak in the change in the amplitude envelope (peak ΔENV) of co-gesture speech event is predictive of the magnitude of the nearest peak in the absolutized acceleration profile of the co-speech gesture.

As can be seen from the graphical results reported in Figure 2, untransformed values for peak ΔENV did not relate linearly with peak acceleration, and we also did not find statistically reliable results for analyses with these variables. However, similar to previous research (Pearson & Pouw, 2022), these variables showed a power relationship, as indicated by linear trends emerging after log–log plotting of the data. We performed a linear mixed regression with the log-transformed peak ΔENV (log peak ΔENV) as a dependent variable and participants as a random intercept. A model that included log transformed peak acceleration (log peak acc.) was reliably related to log peak ΔENV when compared to a base model predicting the overall mean change in $\chi^2(1) = 12.98$, $p = .003$. Adding group, or Group × Log peak acc., to the model did not further improve predictability of log peak ΔENV ($p > .134$). The most reliable model with the main effect of log peak acc. showed that the higher the acceleration peak the higher the peak ΔENV, $b = .142$, $t(899) = 3.612$, $p = .003$.

Figure 2

Gesture–Speech Magnitude Coupling



Note. Panel A shows for each gesture event the peak in change in the amplitude envelope (y-axis) versus the nearest absolutized acceleration peak of that gesture. The lower row shows the raw values, and the upper row shows the log transformed values. It can be seen that higher peak accelerations are generally associated with higher changes in amplitude envelope, and given that this relationship is apparent for the log-log plots, we can conclude that the scaling relationship is not simply linear, but roughly follows a power relation. Note further, that there is a very low variability in the control group, possibly suggesting that gesture–speech magnitude coupling is more stable. To capture something about stability, we can compute the R^2 of the correlation between speech and movement peaks per participant. Panel B shows the severity results (y-axis) with information about the test (AQ and BNT) in our sample relative to the R^2 of that person with Aphasia. It can be seen that a more stable gesture–speech magnitude coupling is related to lower scores on tests that gauge Aphasia severity. Panel C shows the same trend when removing the two extreme data points representing individuals with very high gesture–speech coupling. AQ = Aphasia Quotient; BNT = Boston Naming Test. See the online article for the color version of this figure.

Since we log-transformed the variables, the beta indicates that for every 100% increase in the acceleration, there is a 14.2% increase to be expected in the peak ΔENV (i.e., a positive power-law approximation of about 1/7). The lack of interaction with the group suggests that there were no reliable differences between the groups in terms of the gesture–speech magnitude coupling.

However, from the upper panels in Figure 2, we see that there was much more variability of gesture–speech magnitude coupling for PWA as compared to the control participants. We further assess whether the degree of stable gesture–speech magnitude coupling is predictive of the mildness in Aphasia symptoms. We capture the variability in the gesture–speech magnitude coupling by computing for each participant a single R^2 of the Pearson correlation between log peak ΔENV∼log peak acc. We then performed a mixed linear regression with Aphasia Mildness as a dependent variable and the test type as random intercept. A base model predicting the overall mean Aphasia Mildness was outperformed by a model containing the gesture–speech magnitude coupling R^2 with a random slope for test change in $\chi^2(1) = 11.160$, $p < .011$. We obtain that the more stable gesture–speech magnitude coupling was milder the Aphasia symptoms were, $b = 4.616$, $t(59) = 3.44$, $p < .002$. However, it is clear from Figure 2B that there might be two extreme outliers that dramatically inflate our correlation estimate. Removing these two data points that are extreme in their gesture–speech magnitude coupling leads to a similar conclusion (see Figure 2C), but caution is warranted because the p-value slightly exceeds our Bonferroni corrected alpha of .016. With these potential outliers removed, we obtain a more stable gesture–speech magnitude coupling related to milder Aphasia symptoms $b = 2.49$, $t(57) = 2.49$, $p = .018$.

To conclude, PWA couple, the magnitude of their speech fluctuations with the movement effort produced in their gestures. Furthermore, we observed that the less stable this gesture–speech magnitude coupling is for a particular person with Aphasia, the higher the severity of Aphasia-related symptoms.

Gesture–Speech Temporal Coupling

Now that we know that physical effort of a gesture (acceleration peaks) seems to scale with acoustic effort (peak change in the amplitude envelope) we may wonder whether these peaks are synchronized in time in a similar way across groups. Figure 3 shows the main results of this comparison. We show timings per gesture type given that representational gestures tend to have a longer duration which could affect the results. But it is clear that differences in gesture–speech synchrony are minimal, though beat gestures seem to have a more stable gesture–speech coupling for control participants as compared to the other groups (as indicated by a more peaked timing distribution for control participants).

To assess differences in groups in terms of gesture–speech synchrony, we performed two analyses. Firstly, we assess whether there are differences in gesture–speech asynchrony *offset*, such that gesture may lead or follow speech depending on the group. We performed a linear mixed regression with offset per gesture as a dependent variable and participant as random intercept. Adding group (and random slopes per participant) was not a reliable model for predicting gesture–speech asynchrony offset as compared to a model predicting the overall mean, change in $\chi^2(1) = 3.47$, $p = .99$). Adding gesture type, leads to a statistically reliable improvement relative to a model with only group as predictor, change in $\chi^2(1) = 6.40$, $p = .011$. Adding an interaction between gesture type and group, did not further improve the model, change in $\chi^2(1) = 3.18$, $p = .365$. The second most reliable model showed no reliable differences per group, p's $> .126$, but did show that representational gestures were less in the lead in terms acceleration peaks relative to speech as compared to beat gestures, $b = 30.67$, $t(905) = 2.56$, $p = .011$. This analysis indicates that the trend that we see in Figure 3, where PWA tends to have gesture peaks before speech peaks, is not statistically reliable.

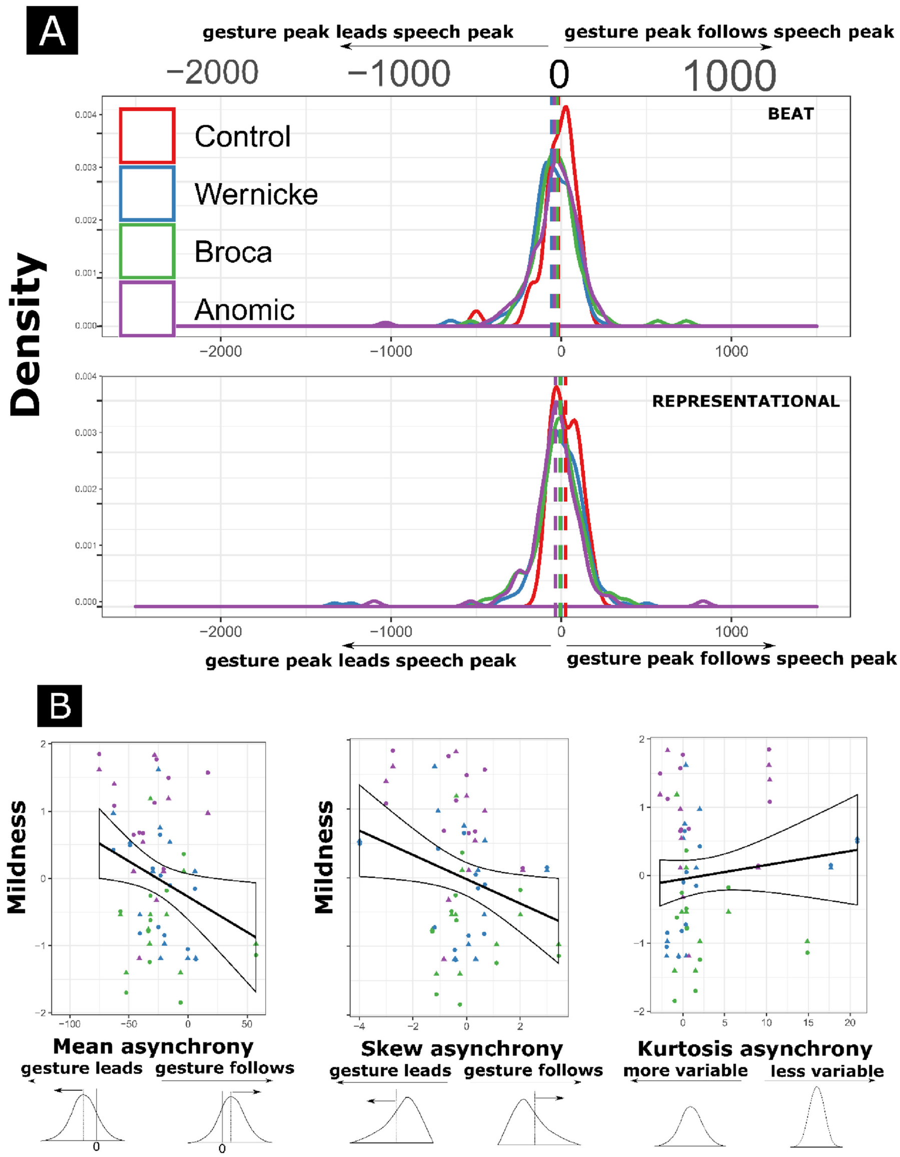
To assess whether the magnitude of the gesture–speech asynchrony differs per group, we computed the absolutized gesture–speech asynchrony per gesture relative to the mean of that participant. This indicates whether there is a lot of variability in gesture–speech timing, thereby serving as a proxy for gesture–speech stability (higher values more instable, lower values more stable). Adding group (and random slopes per participant) was not a reliable model for predicting absolute asynchrony as compared to a model predicting the overall mean, change in $\chi^2(1) = 5.05$, $p = .956$. Adding gesture type did not reliably improve the model further, change in $\chi^2(1) = 3.39$, $p = .066$, nor did add an interaction between gesture type and group, change in $\chi^2(1) = 1.94$, $p = .586$.

We further assessed within the PWA group whether mildness/severity of symptoms can be predicted from features of the gesture–speech asynchrony distributions. To do this, we assessed three properties of each person's gesture–speech synchrony distribution (see Figure 3 for visual information about distribution features). Firstly, we looked at the mean offset (does gesture lead or follow speech peaks?). Relatedly, we can assess the skewness of the distribution indicating whether gesture tends to lead or follow in relatively more extreme amounts. Finally, we can compute for each person the degree of coupling stability by looking at the "peakedness" of the distribution by computing a kurtosis value for each person's gesture–speech asynchrony distribution.

The results were quite consistent but not very reliable, showing that when gesture acceleration peaks tend to be timed later relative to speech peaks, then the severity of the aphasia symptoms is more pronounced, but not at the corrected alpha level of .016. Our analyses showed that relative to a base model predicting the overall mean of Aphasia Mildness (with test type as random intercept), a model with mean gesture–speech asynchrony offset was not reliably more predictive, change in $\chi^2(1) = 5.07$, $p = .024$. When gesture peaks are timed earlier relative speech peaks, then there was an increase in Aphasia Mildness, $b = -0.011$, $t(61) = -2.26$, $p = .027$, but again, not reliably so. Similarly, a model with gesture–speech asynchrony skewness was marginally more reliable than a base model, change in $\chi^2(1) = 4.84$, $p = .028$. These results were comparable to the mean offset, such that distributions that skewed toward gestures being timed earlier relative to speech peaks were related to higher Aphasia Mildness, $b = -0.177$, $t\,(61) = -2.21$, $p = .031$. The gesture–speech asynchrony stability was however not more reliable than a base model predicting overall mean of Mildness, change in $\chi^2(1) = 0.96$, $p = .327$.

In conclusion, gesture–speech temporal coupling, as determined by the nearest peak analysis, does not show statistically reliable differences in gesture–speech timing off-sets. Though we find similar trends that indices of gesture-leading speech relates to less severe markers of Aphasia, these analyses are not as clear-cut as our other results, given that they do not reach statistical reliability for the more conservative alpha level.

Figure 3
Gesture–Speech Temporal Coupling



Note. Gesture–speech timing distributions are shown in panel A, for beat gestures (upper panel) and representational gesture (lower panel) separately. It can be seen that gesture–speech timing tends to center on ideal synchrony of a value of 0. Though statistical analysis could not confirm these trends, there is a tendency for control participants to time their gesture a little later relative to speech as compared to the Aphasia groups. Further, especially for beats, the control participants seemed to have a more stable timing, as indicated by the pronounced peak in the gesture–speech asynchrony distribution. Panel B shows the Aphasia Mildness score relative to different features of the gesture–speech asynchrony distribution per person. When distributions were such that the mean and skewness shows a preference for gesture peaks to lead speech peaks relatively more, then there were more Mild Aphasia symptoms. The peakedness as a proxy of gesture–speech synchrony stability does not relate clearly to aphasia Mildness. See the online article for the color version of this figure.

Exploratory Analysis: Gesture–Speech Rhythms (Continuous Analysis)

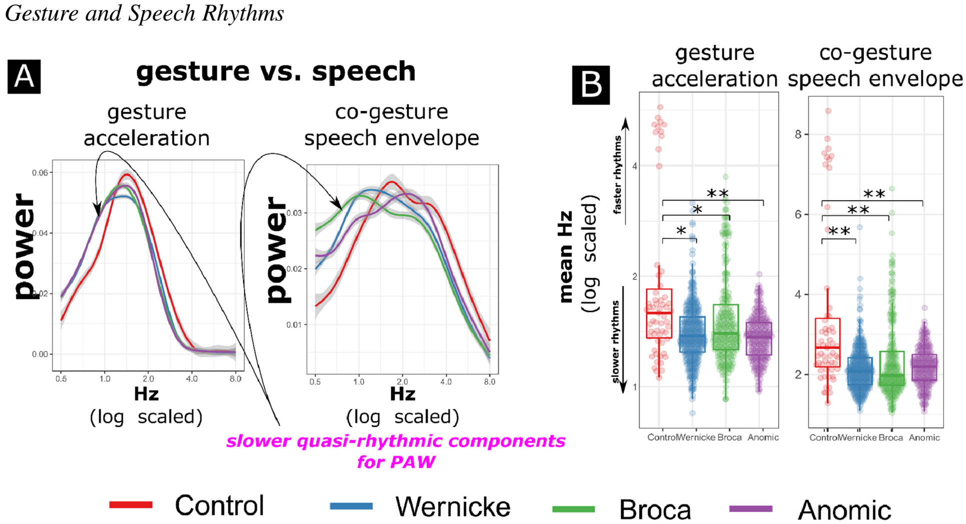
So far we have extracted point estimates of what are continuous gesture–speech events. Here we take a different, more continuous, approach by quantifying on the one hand the quasi-rhythmic structure of the speech amplitude envelope, and on the other hand the quasi-rhythmic structure of the gesture acceleration profile. We thereby simply quantify the rhythms that tend to be present in these *continuous* signals. We do this by selecting a gesture window of at least a second around a gesture event, and then performing a wavelet transform on the gesture acceleration and co-gesture speech amplitude envelope, so as to extract the average power over the 0.5–8 Hz range, which is normalized by the total observed power. The selected time scale range is important for both speech (~2–8 Hz) and gesture (~0.5–6 Hz), where speech is generally a faster oscillating system than the manual system (Brookshire et al., 2017; Pouw & Dixon, 2020). To compare the differences in rhythms we take the mean of the power distributions for each gesture and co-gesture speech event, yielding an average rhythm in Hertz.

Figure 4 provides the main results for this analysis. We obtain that, in general, gestures and co-gesture speech produced by control participants are more defined by faster rhythms, while PWA tends to have slower rhythms in gesture as well as speech (Figure 4A). For speech, this is of course to be predicted given the limitations in speech fluency. However, it seems that gesture is slurred too in PWA, thereby following speech. When comparing a base model (with participant as random intercept) predicting the overall mean rhythm of co-gesture speech events, to a model with group comparisons (control vs. PWA), we find that group comparisons indeed increased model performance, change in $\chi^2(1) = 13.13$, $p = .004$. A model with group comparisons also increased model performance relative to a base model for gesture mean rhythm, change in $\chi^2(1) = 8.06$, $p = .044$. Table 2 provides the parameter estimates for these models. They confirm that comparisons between control versus PWA (i.e., Wernicke, Broca, Anomic) are all statistically different for speech rhythms ($p < .002$), but also for gesture rhythms ($p < .018$, though bordering our Bonferroni restricted alpha of .016), such that slower rhythms are observed for especially speech, but also gesture, in PWA as compared to Control participants. Speech that is affected in its rhythmic composition is not surprising, but that gesture follows, is an important finding that warrants more investigation in the future.

Discussion

Little is known about intermodal timing in multimodal language, especially in neurally atypical populations such as PWA. We, firstly, find that gesture kinematic and speech acoustic coupling is present in PWA, as well as NBI participants. Co-gesture speech acoustic modulations (degree of attack in the amplitude envelope) are positively related to the magnitude of the nearest peak in acceleration or deceleration (a measure of physical impact of a gesture). This relation is obtained across the sample that includes different aphasia types and NBI. However, the variability in the gesture–speech magnitude coupling is much higher than controls, and this seems to align with the recent gesture–speech analysis by Ferré (2021). We further find that the more stable the gesture–speech magnitude coupling is for a person with aphasia, the less severe the symptoms of Aphasia are. This first analysis already shows that gesture–speech coupling might be an interesting avenue for research and diagnostic evaluation with PWA. We do not find dramatic differences between groups in terms of gesture–speech *temporal* timing differences as operationalized by a nearest-peak analysis. Finally, in an exploratory analysis we applied a fully continuous approach, analyzing the quasi-rhythmic structure of co-gesture speech envelope and gesture acceleration. As to be expected, we find that the speech signal has slower rhythmic components, fully in line with speech disfluencies in PWA. However, we also find statistically reliable differences in gesture, where we also observe slower rhythms in gestures.

Figure 4
Gesture and Speech Rhythms



Note. Panel A shows the normalized power over the log scaled temporal range of 0.5–8 Hz. Higher power at a particular frequency, means that the signal was more defined by that time scale. It can be seen that especially for speech, but also for gesture, there is a higher power observed for lower frequency ranges around 0.5–1 Hz for PWA as compared to controls. This means that speech and gesture, generally had slower quasi-rhythmic components, in line with speech disfluencies generally in PWA and a novel finding with respect to gesture. Panel B shows the mean rhythm per gesture and co-gesture event per group, indicating that indeed control participants generally had faster mean rhythms as compared to PWA (for statistical result see Table 1). PWA= people with aphasia. Panel B: $^{**}p < .016$, $^{*}p < .05$. See the online article for the color version of this figure.

Table 2
Group Comparisons for Mean Rhythm for Gesture and Speech

	B	*t*	*p*
Gesture			
Intercept (mean control)	2.31	8.72	<.001
Control versus Wernicke	−0.79	−2.52	.017
Control versus Broca	−0.81	−5.49	.018
Control versus Anomic	−0.92	−2.85	.008
Co-gesture speech			
Intercept (mean control)	3.78	9.99	<.001
Control versus Wernicke	−1.50	−3.35	.002
Control versus Broca	−1.59	−3.46	<.002
Control versus Anomic	−1.67	−3.64	<.001

Note. Linear mixed regression models predicting mean rhythms for gesture and speech (random intercept for participant).

In sum, using point-wise and continuous kinematic-acoustic analysis, we show that gesture–speech coupling is variable but apparent in PWA as compared to controls, and both the gesture–speech coupling strength and the temporal phasing were related to Aphasia severity. This study thus underlines the importance of what we might call intermodal as opposed to multimodal aspects of aphasia, where gesture and speech patterning are analyzed *in a continuous relation to each other*. Further, since we find statistical signatures of intermodal coordination, we think it supports earlier research suggesting that prosody might have important semiotic roles to play in the communicative repertoire of PWA (Goodwin, 2010). Finally, we can conclude from this analysis that gestures still seem to entrain (Ferré, 2021; Pouw & Dixon, 2020) to a speech system with an impaired overall activity; gesture effort couples to speech modulation; gesture and speech peaks are timed similarly as NBIs; and gesture rhythms are slower, similar to slower rhythms in speech. Future more controlled research should focus on whether these aphasia-related kinematic-acoustic features in gesture–speech behavior are also differentiable within the different aphasia types. Though not the main focus of the study, we have not found striking differences between different Aphasia types. Though persons with Broca's aphasia seem to have a more variable gesture–speech magnitude coupling relation and slower rhythmic components in their speech envelope (see Figures 2 and 4A). Of course, the current study focuses on a narrow set of performance variables that does not exhaust the differences in speech or gesture production between Aphasia types. Thus the current study also generates important questions for future avenues in this line of inquiry.

There are clear theoretical conclusions to be drawn for the current study. Firstly, it seems that positing two independent formulation processes for gesture and speech is untenable. After all, gesture and vocal aspects of speech coadapt in Aphasia as we show. Additionally, positing that gesture and speech are coordinated from the start of the formulation process because they are regulated by a special linguistic system is also not tenable (McNeill, 2008). After all, core linguistic processes are disturbed in Aphasia, yet on some level of multimodal communication there is still coupling between sub-systems. This coupling on the prosody level can however be explained by the biomechanical theory of gesture-vocal synchrony, as it is assumed the temporal coordination of upper limb movement and speech-constituting vocalizations can arise in a rudimentary form out of pre-linguistic processes (Pouw & Fuchs, 2022). This coupling putatively arises out of acceleration of the upper limbs, which affects rib-cage movements, which constrains the voice. Of course, it does not mean that all there is to gesture–speech coupling is biomechanics. It merely suggests that coupling arises out of pre-linguistic processes, which can then be aligned, counteracted, or adapted relative to linguistic processes (for a nuanced discussion on this see Pouw & Fuchs, 2022).

There are several shortcomings to the current study. Firstly, our sample sizes are small, especially for the number of NBI control subjects. The results are bolstered however by our analysis showing relations between gesture–speech coupling parameters and aphasia-related symptoms within the PWA group. Another shortcoming is that the original AphasiaBank data (MacWhinney et al., 2011) was not collected with the acoustic-kinematic analysis in mind. Thus video angles were not consistent, speech acoustic quality varied, and therefore our analysis is inevitably injected with such measurement-related noise. It is important to emphasize though that our analysis was specifically designed to be more dimensionless. For example, we do not compare the absolute speed between gestures, as indeed the approximations of this would be tainted by camera positions. Instead, we normalized all the signals, and characterized the signals in a way that is less sensitive to acoustic or video differences between participants; namely by looking at gesture–speech timing and magnitude within participants and by analyzing gesture and speech rhythms. Finally, it has been shown that video-based kinematic analysis as performed here can be reliably used for gesture–speech synchrony analysis as compared to a gold-standard device-based motion tracking (Pouw, Trujillo, & Dixon, 2020). All in all, we believe the current study should be interpreted as a very promising *proof-of-concept* for gesture–speech kinematic-acoustic analysis of intermodal processes in aphasia, which can be fruitfully employed in unison with the categorically and conversation analytic-oriented study of gesture and speech (for a good example see: Ferré, 2021).

Additionally, there has been some debate pertaining to the reliability of aphasia subtype diagnosis based on the Western Aphasia Battery. Some recent studies have provided evidence that this metric is primarily based on verbal output (e.g., Ellis et al., 2021; Rao et al., 2022). Although this metric remains the primary method of subtype identification for aphasia research, we acknowledge that there may be some bias in classification of Anomic, Broca's, and Wernicke's Aphasia.

From the current study, potential multimodal markers of aphasia severity are observed that can and must be further subjected for confirmatory research for developing a multimodal diagnostic approach to aphasia. Firstly, if kinematic-acoustic magnitude coupling is low or absent, this relates to more severe aphasia symptoms. Additionally, the quasi-rhythmic structure of gesture tends to be slowed down in people with aphasia. This is not necessarily a negative symptom. Slowing down gestures can be an adaptation. After all, by slowing down gestures similar to speech, gesture–speech synchrony can be maintained.

We believe, together with other researchers (Stark et al., 2021b), that the promises of the current line of research will be fully fulfilled when more controlled confirmatory research is conducted where researchers have these particular multimodal signal processing analysis in mind. This requires high-quality video and audio data that are homogeneously parametrized across participants; if possible high-performance 3D-motion tracking is used. Under such more or less improved measurement conditions a whole host of avenues for further research opens up. This would not only allow for a more finegrained analysis of gesture–speech coupling (Alviar et al., 2020; Krivokapic et al., 2017; Pedersen et al., 2022; Wagner et al., 2014). It could also expand on research by relating to studies showing that gesture kinematics change depending on the communicative context (Trujillo et al., 2018) and psychological background (Trujillo, Özyürek, et al., 2021), and that such modulation of kinematics can affect smooth conversational turn-taking (Trujillo, Levinson, & Holler, 2021). It would further be possible to analyze the iconic and symbolizing forms that are produced in PWA versus control subjects, where analysis of gesture inter-relationships can be utilized to show how gesture constellates into more or less coherent and systematic ways (Pouw, Dingemanse, et al., 2021; Pouw & Dixon, 2020; Pouw, de Wit, et al., 2021). Finally, it is important to triangulate more continuous kinematic-acoustic analysis, with that of the categorical approach that has so far forwarded our knowledge on gesture research in aphasia and psycholinguistics in general (Stark et al., 2021a).

References

Akhavan, N., Göksun, T., & Nozari, N. (2018). Integrity and function of gestures in aphasia. *Aphasiology*, *32*(11), 1310–1335. https://doi.org/10.1080/02687038.2017.1396573

Alexanderson, S., Henter, G. E., Kucherenko, T., & Beskow, J. (2020). Style-controllable speech-driven gesture synthesis using normalising flows. *Computer Graphics Forum*, *39*(2), 487–496. https://doi.org/10.1111/cgf.13946

Alviar, C., Dale, R., Dewitt, A., & Kello, C. (2020). Multimodal coordination of sound and movement in music and speech. *Discourse Processes*, *57*(8), 682–702. https://doi.org/10.1080/0163853X.2020.1768500

Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences*, *114*(24), 6352–6357. https://doi.org/10.1073/pnas.1620350114

Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, *143*(4), 1726–1741. https://doi.org/10.1037/a0036281

Cooper, B. G., & Goller, F. (2004). Multimodal signals: enhancement and constraint of song motor patterns by visual display. *Science*, *303*(5657), 544–546.

Cravotta, A., Busà, M. G., & Prieto, P. (2019). Effects of encouraging the use of gestures on speech. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3204–3219. https://doi.org/10.1044/2019_JSLHR-S-18-0493

Danly, M., & Shapiro, B. (1982). Speech prosody in Broca's aphasia. *Brain and Language*, *16*(2), 171–190. https://doi.org/10.1016/0093-934X(82)90082-7

de Beer, C., Hogrefe, K., Hielscher-Fastabend, M., & de Ruiter, J. P. (2020). Evaluating models of gesture and speech production for people with aphasia. *Cognitive Science*, *44*(9), Article e12890. https://doi.org/10.1111/cogs.12890

de Marchena, A., & Eigsti, I.-M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research*, *3*(6), 311–322. https://doi.org/10.1002/aur.159

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.018

Doedens, W. J., & Meteyard, L. (2022). What is functional communication? A theoretical framework for real-world communication applied to aphasia rehabilitation. *Neuropsychology Review*, *32*(4), 937–973. https://doi.org/10.1007/s11065-021-09531-2

Dohen, M., & Roustan, B. (2017). Co-production of speech and pointing gestures in clear and perturbed interactive tasks: Multimodal designation strategies. *Interspeech 2017 – 18th Annual Conference of the International Speech Communication Association*. https://hal.archives-ouvertes.fr/hal-02367749

Dronkers, N., & Baldo, J. V. (2010). Language: Aphasia. *Encyclopedia of neuroscience*, 343–348.

Ellis, C., Peach, R. K., & Rothermich, K. (2021). Relative weight analysis of the Western Aphasia Battery. *Aphasiology*, *35*(10), 1281–1292. https://doi.org/10.1080/02687038.2020.1787947

Esteve-Gibert, N., Lœvenbruck, H., Dohen, M., & D'Imperio, M. (2021). Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech. *Developmental Science*, *25*(3), Article e13154. https://doi.org/10.1111/desc.13154

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. https://doi.org/10.1044/1092-4388(2012/12-0049)

Ferré, G. (2021). Gesture, prosody and verbal content in non-fluent aphasic speech. *Multimodal Communication*, *10*(1), 73–91. https://doi.org/10.1515/mc-2020-0016

Ferré, G. (2022). Gesture profile of fluent and non-fluent people with aphasia. *Corela. Cognition, Représentation, Langage*, *20*(1), 1–30. https://doi.org/10.4000/corela.14742

Ferstl, Y., Neff, M., & McDonnell, R. (2020). Understanding the predictability of gesture parameters from speech and their perceptual importance. *Proceedings of the 20th ACM international conference on intelligent virtual agents* (pp. 1–8). https://doi.org/10.1145/3383652.3423882

Feyereisen, P. (2017). *The cognitive psychology of speech-related gesture*. Routledge.

Gandour, J., Petty, S. H., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language*, *35*(2), 201–240. https://doi.org/10.1016/0093-934X(88)90109-5

Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., & Malik, J. (2019). Learning individual styles of conversational gesture. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3497–3506). https://arxiv.org/abs/1906.04160

Goodglass, H., & Wingfield, A. (Eds.). (1997). *Anomia: Neuroanatomical and cognitive correlates*. Academic Press.

Goodwin, C. (2010). Constructing meaning through prosody in aphasia. In D. Barth-Weingarten & E. Reber (Eds.), *Prosody in interaction* (pp. 373–394). John Benjamins.

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. https://doi.org/10.1121/1.4987638

Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291–303. https://doi.org/10.1016/j.jml.2006.07.011

Jenkins, T., Coppola, M., & Coelho, C. (2017). Effects of gesture restriction on quality of narrative production. *Gesture*, *16*(3), 416–431. https://doi.org/10.1075/gest.00003.jen

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test* (2nd ed.). Lea & Febiger.

Kelso, J. A. S., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. *The American Journal of Physiology*, *246*(6), R928–R935. https://doi.org/10.1152/ajpregu.1984.246.6.R928

Kertesz, A. (1982). *Western Aphasia battery*. Grune & Stratton.

Kistner, J., Dipper, L. T., & Marshall, J. (2019). The use and function of gestures in word-finding difficulties in aphasia. *Aphasiology*, *33*(11), 1372–1392. https://doi.org/10.1080/02687038.2018.1541343

Kong, A. P.-H., Law, S.-P., Kwan, C. C.-Y., Lai, C., & Lam, V. (2015). A coding system with independent annotations of gesture forms and functions during verbal communication: Development of a Database of Speech and GEsture (DoSaGE). *Journal of Nonverbal Behavior*, *39*(1), 93–111. https://doi.org/10.1007/s10919-014-0200-6

Kong, A. P.-H., Law, S.-P., Wat, W. K.-C., & Lai, C. (2015). Co-verbal gestures among speakers with aphasia: Influence of aphasia severity, linguistic and semantic skills, and hemiplegia on gesture employment in oral discourse. *Journal of Communication Disorders*, *56*(6), 88–102. https://doi.org/10.1016/j.jcomdis.2015.06.007

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. https://doi.org/10.1016/j.jml.2007.06.005

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), Article 3. https://doi.org/10.5334/labphon.75

Lancaster, W. C., Henson, O. W., & Keating, A. W. (1995). Respiratory muscle activity in relation to vocalization in flying bats. *The Journal of Experimental Biology*, *198*(1), 175–191.

Levelt, W. J. M. (1993). Lexical access in speech production. In E. Reuland & W. Abraham (Eds.), *Knowledge and language: Volume I from Orwell's problem to Plato's problem* (pp. 241–251). Springer Netherlands. https://doi.org/10.1007/978-94-011-1840-8_11

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

Lozano-Goupil, J., Raffard, S., Capdevielle, D., Aigoin, E., & Marin, L. (2022). Gesture–speech synchrony in schizophrenia: A pilot study using a kinematic-acoustic analysis. *Neuropsychologia*, *174*, Article 108347. https://doi.org/10.1016/j.neuropsychologia.2022.108347

MacIntyre, A. D., Cai, C. Q., & Scott, S. K. (2022). Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques. *The Journal of the Acoustical Society of America*, *151*(3), 2002–2026. https://doi.org/10.1121/10.0009844

Mack, W. J., Freed, D. M., Williams, B. W., & Henderson, V. W. (1992). Boston naming test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology*, *47*(3), P154–P158.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). Aphasiabank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University Of Chicago Press.

Mcneill, D. (2008). *Gesture and thought*. University of Chicago press.

Mcneill, D., Quaeghebeur, L., & Duncan, S. (2010). IW- "The man who lost his body". In *Handbook of phenomenology and cognitive science* (pp. 519–543). Springer.

Orgassa, A. (2005). Co-speech gesture in anomic aphasia. *Toegepaste Taalwetenschap in Artikelen*, *73*(1), 85–97. https://doi.org/10.1075/ttwia.73.09org

Parrell, B., Goldstein, L., Lee, S., & Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of Phonetics*, *42*, 1–11. https://doi.org/10.1016/j.wocn.2013.11.002

Pearson, L., & Pouw, W. (2022). Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement. *Annals of the New York Academy of Sciences*, *1515*(1), 219–236. https://doi.org/10.1111/nyas.14806

Pedersen, N. F., Dau, T., Hansen, L. K., & Hjortkjær, J. (2022). Modulation transfer functions for audiovisual speech. *PLOS Computational Biology*, *18*(7), e1010273. https://doi.org/10.1371/journal.pcbi.1010273

Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2020). Gesture–speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, *1491*(1), 89–105. https://doi.org/10.1111/nyas.14532

Pouw, W., Dingemanse, M., Motamedi, Y., & Özyürek, A. (2021). A systematic investigation of gesture kinematics in evolving manual languages in the lab. *Cognitive Science*, *45*(7), e13014. https://doi.org/10.31219/osf.io/heu24

Pouw, W., & Dixon, J. A. (2019). *Quantifying gesture–speech synchrony*. Proceedings of the 6th meeting of gesture and speech in interaction (pp. 68–74). https://doi.org/10.17619/UNIPB/1-812

Pouw, W., & Dixon, J. A. (2020). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discourse Processes*, *57*(4), 301–319. https://doi.org/10.1080/0163853X.2019.1678967

Pouw, W., & Fuchs, S. (2022). Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews*, *141*, Article 104836. https://doi.org/10.1016/j.neubiorev.2022.104836

Pouw, W., Harrison, S. J., & Dixon, J. A. (2022). The importance of visual control and biomechanics in the regulation of gesture–speech synchrony for an individual deprived of proprioceptive feedback of body position. *Scientific Reports*, *12*(1), Article 14775. https://doi.org/10.1038/s41598-022-18300-x

Pouw, W., Jaramillo, J. J., Ozyurek, A., & Dixon, J. A. (2020, September 7–9). Quasi-rhythmic features of hand gestures show unique modulations within languages: Evidence from bilingual speakers. *Proceedings of the 7th meeting of gesture and speech in interaction. GESPIN2020*, Stockholm, Sweden.

Pouw, W., & Jenkins, T. (2022). *Gesture–speech coupling in persons with aphasia: A kinematic-acoustic analysis*. https://osf.io/qex93/

Pouw, W., & Trujillo, J. P. (2019). *Materials Tutorial Gespin2019—Using video-based motion tracking to quantify speech-gesture synchrony*. https://doi.org/10.17605/OSF.IO/RXB8J

Pouw, W., Trujillo, J. P., & Dixon, J. A. (2020). The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, *52*(2), 723–740. https://doi.org/10.3758/s13428-019-01271-9

Pouw, W., de Wit, J., Bögels, S., Rasenberg, M., Milivojevic, B., & Ozyurek, A. (2021, July 24–29). *Semantically related gestures move alike: Towards a distributional semantics of gesture kinematics*. In Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and Behavior: 12th International Conference, DHM 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, Proceedings, Part I (pp. 269–287). Springer International Publishing.

Rao, L. A., Roberts, A. C., Schafer, R., Rademaker, A., Blaze, E., Esparza, M., Salley, E., Coventry, C., Weintraub, S., Mesulam, M. M., & Rogalski, E. (2022). The reliability of telepractice administration of the Western aphasia battery—Revised in persons with primary progressive aphasia. *American Journal of Speech-Language Pathology*, *31*(2), 881–895. https://doi.org/10.1044/2021_AJSLP-21-00150

Rathcke, T., Lin, C.-Y., Falk, S., & Bella, S. D. (2021). Tapping into linguistic rhythm. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *12*(1), 1–32. https://doi.org/10.5334/labphon.248

Robson, H., Sage, K., & Ralph, M. A. L. (2012). Wernicke's aphasia reflects a combination of acoustic-phonological and semantic control deficits: a case-series comparison of Wernicke's aphasia, semantic dementia and semantic aphasia. *Neuropsychologia*, *50*(2), 266–275.

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. (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. 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* (p. 59).

Rose, M. L. (2006). The utility of arm and hand gestures in the treatment of aphasia. *Advances in Speech Language Pathology*, *8*(2), 92–109. https://doi.org/10.1080/14417040600657948

Rose, M. L., Mok, Z., & Sekine, K. (2017). Communicative effectiveness of pantomime gesture in people with aphasia. *International Journal of Language & Communication Disorders*, *52*(2), 227–237. https://doi.org/10.1111/1460-6984.12268

Rusiewicz, H. L., Susan, S., Iverson, J., & Szuminsky, N. (2013). Effects of prosody and position on the timing of deictic gestures. *Journal of Speech, Language, and Hearing Research*, *56*(2), 458–470. https://doi.org/10.1044/1092-4388(2012/11-0283)

Sekine, K., & Rose, M. L. (2013). The relationship of aphasia type and gesture production in people with aphasia. *American Journal of Speech-Language Pathology*, *22*(4), 662–672. https://doi.org/10.1044/1058-0360(2013/12-0030)

Shattuck-Hufnagel, S., & Prieto, P. (2019). *Dimensionalizing co-speech gestures*. Proceedings of the international Congress of phonetic sciences 2019 (pp. 1490–1494).

Stark, B. C., Clough, S., & Duff, M. (2021a). Suggestions for improving the investigation of gesture in aphasia. *Journal of Speech, Language, and Hearing Research*, *64*(10), 4004–4013. https://doi.org/10.1044/2021_JSLHR-21-00125

Stark, B. C., Clough, S., & Duff, M. (2021b). *Suggestions for improving the investigation of gesture in aphasia*. PsyArXiv. 10.31234/osf.io/4hbzr

Trujillo, J. P., Levinson, S. C., & Holler, J. (2021, July 24–29). *Visual information in computer-mediated interaction matters: Investigating the association between the availability of gesture and turn transition timing in conversation*. In Human-Computer Interaction. Design and User Experience Case Studies: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, Proceedings, Part III 23 (pp. 643–657). Springer International Publishing.

Trujillo, J. P., Özyürek, A., Kan, C. C., Sheftel-Simanova, I., & Bekkering, H. (2021). Differences in the production and perception of communicative kinematics in autism. *Autism Research*, *14*(12), 2640–2653. https://doi.org/10.1002/aur.2611

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition*, *180*, 38–51. https://doi.org/10.1016/j.cognition.2018.04.003

van Nispen, K., van de Sandt-Koenderman, M., Sekine, K., Krahmer, E., & Rose, M. L. (2017). Part of the message comes in gesture: How people with aphasia convey information in different gesture types as compared with information in their speech. *Aphasiology*, *31*(9), 1078–1103. https://doi.org/10.1080/02687038.2017.1301368

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232. https://doi.org/10.1016/j.specom.2013.09.008

Wang, W., & Krishnan, E. (2014). Big data and clinicians: a review on the state of the science. *JMIR Medical Informatics*, *2*(1), Article e2913.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research* (p. 4).

Yunus, F., Clavel, C., & Pelachaud, C. (2020). *Sequence-to-sequence predictive model: From prosody to communicative gestures*. ArXiv:2008.07643 [Cs, Eess]. http://arxiv.org/abs/2008.07643

Received November 5, 2021
Revision received October 1, 2022
Accepted November 10, 2022 ■