

Research report

N400 amplitude, latency, and variability reflect temporal integration of beat gesture and pitch accent during language processing

Laura M. Morett^{a,*}, Nicole Landi^b, Julia Irwin^c, James C. McPartland^d

2012

^a University of Alabama, United States^b Haskins Laboratories, University of Connecticut, United States^c Haskins Laboratories, Southern Connecticut State University, United States^d Yale University, United States

HIGHLIGHTS

- This work examined how the N400 ERP reflects temporal asynchrony of beat gesture relative to pitch accent.
- Average N400 amplitude and latency reflect temporal integration of beat gesture with pitch accent.
- Higher N400 trial-by-trial variability reflects temporal asynchrony between beat gesture and pitch accent.

ARTICLE INFO

Keywords:

N400

Evoked response variability

Beat gesture

Pitch accent

Temporal integration

ABSTRACT

This study examines how across-trial (average) and trial-by-trial (variability in) amplitude and latency of the N400 event-related potential (ERP) reflect temporal integration of pitch accent and beat gesture. Thirty native English speakers viewed videos of a talker producing sentences with beat gesture co-occurring with a pitch accented focus word (synchronous), beat gesture co-occurring with the onset of a subsequent non-focused word (asynchronous), or the absence of beat gesture (no beat). Across trials, increased amplitude and earlier latency were observed when beat gesture was temporally asynchronous with pitch accenting than when it was temporally synchronous with pitch accenting or absent. Moreover, temporal asynchrony of beat gesture relative to pitch accent increased trial-by-trial variability of N400 amplitude and latency and influenced the relationship between across-trial and trial-by-trial N400 latency. These results indicate that across-trial and trial-by-trial amplitude and latency of the N400 ERP reflect temporal integration of beat gesture and pitch accent during language comprehension, supporting extension of the integrated systems hypothesis of gesture-speech processing and neural noise theories to focus processing in typical adult populations.

1. Introduction

Successful language comprehension requires listeners to identify and attend to the most important information in discourse. To do so, they make use of a variety of cues that emphasize focus (informational importance), including discursive context (Cutler & Fodor, 1979), syntactic constructions (Birch et al., 2000), pitch accent (Birch et al., 2000; Nooteboom & Krut, 1987), and beat gesture (Biau et al., 2018; Holle et al., 2012; McNeill, 2005). In natural conversation, beat gesture and pitch accent often occur conjointly, and behavioral and neural evidence indicates that they exert both independent and interactive effects on language processing (Dimitrova et al., 2016; Holle et al., 2012; Krahmer & Swerts, 2007; Kushch et al., 2018; Kushch & Prieto,

2016; Llanes-Coromina et al., 2018; Morett et al., 2020a,b; Morett & Fraundorf, 2019; Wang & Chu, 2013). Although behavioral evidence indicates that the close temporal relationship between beat gesture and pitch accent influences language processing (Esteve-Gibert & Prieto, 2013; Leonard & Cummins, 2011; Roustan & Dohen, 2010a,b; Rusiewicz et al., 2013, 2014), it is unclear how this temporal relationship is represented via the neurobiological signatures of language processing. The current study addresses this issue by examining how the N400 event-related potential (ERP), which has been argued to reflect predictive processing and semantic integration effort in language processing generally (Borovsky et al., 2012; Lau et al., 2013; Van Berkum et al., 1999) and gesture-speech integration specifically (He et al., 2020; Holle & Gunter, 2007; Kelly et al., 2004; Zhang et al.,

* Corresponding author.

E-mail address: lmoret@ua.edu (L.M. Morett).<https://doi.org/10.1016/j.brainres.2020.147059>

Received 25 March 2020; Received in revised form 3 August 2020; Accepted 12 August 2020

Available online 17 August 2020

0006-8993/© 2020 Elsevier B.V. All rights reserved.

2020), differs as a function of the temporal synchrony of beat gesture relative to pitch accenting. In doing so, it provides insight into how the temporal relationship between speech and gesture is represented—and processed—in the brain, and how it affects language processing.

1.1. Beat gesture and pitch accent as cues to focus

In spoken English, one of the most common cues to focus is pitch accent, which is used to direct listeners' attention to new or contrastive information (Wang et al., 2014). Acoustically, pitch accent consists of changes in the fundamental frequency (f_0), duration, and intensity of speech (Ladd, 2008). Inappropriate use of pitch accent in relation to focus, such that non-focused information is pitch accented or focused information is unaccented, hinders online and offline language processing (Bock & Mazzella, 1983; Dahan et al., 2002; Morett et al., 2020a,b; Nooteboom & Kruyt, 1987; Zellin et al., 2011). Sensitivity to the focusing function of pitch accent is also evident in ERPs, which show a larger posterior negativity (N400) reflecting increased difficulty of semantic integration when non-focused information is pitch accented (Dimitrova et al., 2012) and when focused information is de-accented (Schumacher & Baumann, 2010; Wang & Chu, 2013) relative to when focused information is pitch accented. These N400 effects are similar to those observed for focusing discourse structures and syntactic constructions such as *it*-clefts (Birch & Rayner, 1997; Bredart & Modolo, 1988; Cowles et al., 2007) and word order in German (Bornkessel et al., 2003), indicating that they reflect semantic processing of focus cues.

Nonverbally, beat gesture conveys focus analogously to pitch accent by highlighting new and contrastive information in co-occurring speech, serving as a “yellow gestural highlighter” (McNeill, 2006). In natural discourse, beat gesture most often consists of non-referential, rhythmic downward hand flicks (McNeill, 1992, 2005), but it can be produced using other parts of the body (e.g., finger movements, head nods, foot taps), in other orientations (e.g., horizontal, oblique, curved), and with multiple components (Shattuck-Hufnagel & Ren, 2018). Recent evidence indicates that beat gesture enhances memory for focused information in spoken discourse (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018; Morett & Fraundorf, 2019) and facilitates online processing of focused information in spoken discourse, particularly when it occurs in combination with pitch accent (Morett et al., 2020a). The results of ERP research are consistent with these findings. Specifically, they demonstrate that the absence of beat gesture where expected based on focus elicits an N400 larger in amplitude than the presence of beat gesture in conjunction with focus (Wang & Chu, 2013). Moreover, the presence of beat gesture relative to non-focused, pitch accented words elicits an increased late positivity relative to the absence of beat gesture relative to non-focused, pitch accented words (Dimitrova et al., 2016). These findings are consistent with work demonstrating that beat gesture facilitates disambiguation of syntactically-ambiguous spoken sentences when it occurs in conjunction with the subject of a non-preferred but correct reading, as evidenced by a reduction in the P600 ERP (Biau et al., 2018; Holle et al., 2012). Taken together, these findings indicate that beat gesture is processed similarly to other focus cues and is integrated with them during online language comprehension, resulting in an interpretation of focus that takes all relevant cues into account.

Findings concerning beat gesture's focusing function and its co-occurrence with pitch accent in natural speech are consistent with frameworks of language processing positing that gesture and speech are integrated during language comprehension, each contributing to multimodal representations of language (Holler & Levinson, 2019; Kelly et al., 2008, 2010; McNeill, 1992). One such framework, the integrated systems hypothesis (Kelly et al., 2010), is based on the finding that representational gestures and co-occurring speech conveying the same meaning facilitate language processing (e.g., Silverman, Bennetto, Campana, & Tanenhaus, 2010), whereas representational gestures and co-occurring speech conveying different meanings disrupt language

processing (e.g., S.D. Kelly et al., 2010). Much of the research supporting this hypothesis has shown that the N400 reflects semantic influence of representational gesture and co-occurring speech on one another (Holle & Gunter, 2007; Kelly et al., 2004, 2009; Özyürek et al., 2007; Wu & Coulson, 2005, 2007). Additionally, some research has shown that representational gesture's semantic influence on co-occurring speech is reflected in activity in posterior superior temporal sulcus, which subserves audiovisual integration (Dick et al., 2009; Green et al., 2009; Holle et al., 2008; Willems et al., 2009) and Broca's area, which subserves language processing (Skipper et al., 2007; Willems et al., 2007). Applied to focus, the integrated systems hypothesis posits that beat gesture and pitch accent should influence one another, disrupting processing of focused information when one cue occurs in the absence of the other. Thus, temporal asynchrony of beat gesture and pitch accent should disrupt processing of focused information, and the temporal relationship between these cues should be reflected in neural signatures of language processing, such as the N400.

1.2. Temporal alignment of beat gesture and pitch accent

In natural discourse, co-speech gesture is closely temporally aligned with lexical affiliates (i.e., words or phrases conveying the same emphasis or meaning). This temporal alignment has been interpreted as evidence that co-speech gesture and lexical affiliates are products of a shared computational stage (McNeill, 1985). Indeed, in production, greater temporal synchrony has been observed between representational gestures and highly-familiar lexical affiliates relative to less-familiar lexical affiliates (Morrel-Samuels & Krauss, 1992). This finding suggests that temporal gesture-speech synchrony and semantic gesture-speech integration are related. However, more recent work examining production suggests that the amount of time by which representational gestures precede their lexical affiliates in questions is not associated with the latency of responses to these questions (Ter Bekke et al., 2020). This finding suggests that temporal synchrony between representational gesture and speech production may not affect prediction, particularly across speakers. During comprehension, N400 amplitude differs based on semantic congruency of representational gesture and lexical affiliates within a time window in which representational gesture's stroke precedes its lexical affiliate by no more than 200 ms and follows its lexical affiliate by no more than 120 ms. (Habets et al., 2011; Obermeier & Gunter, 2014). Moreover, viewing representational gesture affects N400 amplitude to lexical affiliates heard beyond this time window when semantic gesture-speech congruency is judged (Obermeier et al., 2011). Even during a passive viewing task, however, source localization of ERPs following lexical affiliate onsets by ~150 – 180 ms revealed less activity in posterior superior temporal sulcus and posterior ventral frontal region when related representational gestures preceded lexical affiliates than when representational gestures were absent (Skipper, 2014). Together, these findings demonstrate that temporal gesture-speech synchrony affects semantic gesture-speech integration (albeit perhaps not predictive processing). Moreover, they provide evidence that the N400 ERP reflects the relationship between temporal gesture-speech synchrony and semantic gesture-speech integration.

Among all types of co-speech gesture, the temporal synchrony between beat gesture and its lexical affiliates is the closest and most stable. Indeed, it has recently been proposed that temporal synchrony with speech prosody, rather than type of handshape and motion trajectory, identifies beat gesture (Prieto et al., 2018; Shattuck-Hufnagel et al., 2016). Accordingly, close temporal alignment has been observed between points of maximum extension (*apices*) of beat gesture and F_0 peaks of the stressed syllable of pitch accented words, the former of which typically precedes the latter by only about 200 ms during production (Leonard & Cummins, 2011). Critically, this close temporal alignment is present not only for beat gestures produced spontaneously in conjunction with spoken discourse, but also for elicited body

movements temporally synchronous with focused information (Esteve-Gibert & Prieto, 2013; Roustan & Dohen, 2010a,b; Rusiewicz et al., 2013). Comprehenders are highly sensitive to the close temporal relationship between beat gesture and pitch accent and can reliably detect asynchronies 600 ms and greater when beat gesture precedes the stressed syllable of pitch accented words and 200 ms and greater when beat gesture follows the stressed syllable of pitch accented words (Leonard & Cummins, 2011). Indeed, even 9-month-old infants, who are not yet able to produce gesture in conjunction with speech, are sensitive to temporal synchrony between beat gesture and syllabic stress (Esteve-Gibert et al., 2015). Thus, sensitivity to the temporal relationship between beat gesture and pitch accent in comprehension likely precedes production of these cues.

To date, no research has examined the neural substrates of the temporal relationship between beat gesture and pitch accent. However, similar to viewing the downward strokes of prototypical beat gestures, viewing horizontal hand movements co-occurring with pitch accented lexical affiliates elicits a smaller N400 than the absence of hand movement relative to these words (Wang & Chu, 2013). This finding suggests that the presence of hand movements co-occurring with pitch accenting facilitates processing of focused information in spoken discourse. Because the timing of beat gesture and pitch accent relative to one another was not manipulated in this work, however, it is unclear whether processing of focused information was enhanced merely by the presence of these cues—or, rather, by their timing—relative to one another. Related work has shown that, during language comprehension, the presence of beat gesture relative to spoken lexical affiliates elicits a larger early positivity prior to 100 ms as well as a larger auditory P2 component around 200 ms relative to the absence of beat gesture (Biau & Soto-Faraco, 2013). With respect to loci of processing, fMRI research has shown that activity in planum temporale, which subserves secondary auditory processing, while viewing beat gesture accompanied by speech exceeds that while viewing nonsense hand movements accompanied by speech. Moreover, activity in posterior superior temporal sulcus, which subserves multimodal integration, while viewing beat gesture accompanied by speech exceeds that while viewing beat gesture unaccompanied by speech (Hubbard et al., 2009). Together, these findings provide insight into the neural signatures of enhanced semantic processing resulting from the presence of beat gesture relative to focused information.

Some fMRI research has probed the neural loci of the temporal relationship between beat gesture and lexical affiliates. This work has shown that functional activity in inferior frontal gyrus, which subserves semantic gesture-speech integration, as well as middle temporal gyrus, which subserves audiovisual speech perception, is reduced when lexical affiliates precede beat gesture by 800 ms relative to when beat gesture and lexical affiliates are temporally synchronous during language comprehension (Biau et al., 2016). Given that functional activity in these same brain regions is affected by the semantic relationship between representational gesture and lexical affiliates during language comprehension (Dick et al., 2009; Green et al., 2009; Holle et al., 2008; Willems et al., 2009), this finding is consistent with extension of the integrated systems hypothesis to account for temporal synchrony of beat gesture with co-occurring speech. Because ERPs have higher temporal resolution than the BOLD signal, however, modulation of the N400 ERP by temporal synchrony of beat gesture with co-occurring speech would provide even more compelling evidence for extending the integrated systems hypothesis to encompass the influence of temporal gesture-speech synchrony on semantic gesture-speech processing. Moreover, differences in trial-by-trial N400 ERP variation, which has even higher temporal resolution than across-trial¹ N400 ERP averages,

would provide insight into how the N400 reflects temporal synchrony's effect on semantic integration of beat gesture and pitch accent during focus processing. We now turn to discussing the significance of trial-by-trial ERP variation in-depth.

1.3. Trial-by-trial evoked response variability and beat gesture-pitch accent integration

To date, most published research investigating the neural signatures of beat gesture-speech integration during language comprehension has examined across-trial averages. This work has been pivotal in identifying key ERP components and brain regions involved in integrating beat gesture and co-occurring speech. However, growing evidence suggests that trial-by-trial evoked response variability reflects sensitivity to the temporal synchrony of multimodal stimuli, which is crucial to multisensory integration. Trial-by-trial evoked response variability is abnormally high in disorders characterized by deficits in temporal multisensory integration, such as autism spectrum disorder (Dinstein et al., 2012; Haigh et al., 2015; Milne, 2011; Simon & Wallace, 2016) and developmental dyslexia (Hornickel & Kraus, 2013; Power et al., 2013). These observations have led to the development of “neural noise” theories of multisensory processing deficits (Dinstein et al., 2015; Hancock et al., 2017), which posit that neural excitability and neural noise are critical to the precise timing mechanisms necessary for integration of multisensory stimuli. At present, however, it is unclear whether, in typical populations, neural variability increases or decreases during processing of temporally synchronous stimuli in healthy adult populations. On one hand, low trial-by-trial evoked response variability is associated with more effective speech perception in noise (Anderson et al., 2010) and object recognition (Schurger et al., 2015; Xue et al., 2010); on the other, high trial-by-trial evoked response variability is associated with superior visual perceptual matching, attentional cuing, and delayed match-to-sample performance (Garrett et al., 2011) as well as reading skill (Malins et al., 2018). Moreover, it is currently unclear whether differences in trial-by-trial evoked response variability result from spontaneous fluctuations in the power of high-frequency electrical activity intrinsically generated in the brain that are indirectly associated with multimodal integration, or whether they are directly attributable to evoked neural responses to temporally (a)synchronous stimuli during multisensory processing. Indeed, another possibility is that atypical or mistimed stimuli may result in a variety of reactions (surprise, puzzlement, suppression, etc.), which may increase trial-by-trial variability of evoked potentials. For these reasons, it is important to consider how trial-by-trial variability is related to across-trial average evoked responses to multimodal stimuli varying in temporal synchrony.

We are not aware of any research to date that has examined how trial-by-trial evoked response variability differs as a function of the timing of related stimuli such as beat gesture and pitch accent relative to one another. However, trial-by-trial variability of the N400 ERP in particular inherently reflects both temporal and semantic processing on a fine-grained level. Thus, trial-by-trial variability is well-positioned to provide insight into how the temporal relationship between beat gesture and pitch accent is represented in the brain, in turn affecting processing of focused information. Unlike across-trial ERP amplitude and latency, which represent condition-level averages, trial-by-trial ERP variability reveals fine-grained temporal asynchrony of post-synaptic potentials measured at the scalp during a given time window (Duann et al., 2002; Nunez & Srinivasan, 2006).

Previous work has shown that trial-by-trial variability reflects entrainment to sensory stimuli, which may result in downstream effects on oscillatory activity in various frequency ranges (Giraud & Poeppel,

¹ Here and elsewhere in the paper, the term “across-trial” refers to neural signals that are averaged across trials for each condition—i.e., standard ERP analysis. By contrast, the term “trial-by-trial” refers to analyses of neural signal

(footnote continued)
variability.

2012). Indeed, oscillatory activity during integration of beat gesture with pitch accent shows increases in phase-locking in the 5–6 Hz theta range and the 8–10 Hz alpha range (Biau et al., 2015). With respect to spoken language comprehension, evidence examining the relationship between oscillatory activity and semantic processing is mixed. On one hand, some work shows a decrease in power in the 8–10 Hz alpha range and 12.5–30 Hz beta range during the N400 time window during processing of semantically-incongruent language (Bastiaansen et al., 2009; Luo et al., 2010; Wang et al., 2012). On the other hand, other work indicates that such power decreases are distinct from semantic processing (Davidson & Indefrey, 2007; Kiellar et al., 2014). Additionally, other work demonstrates that these power decreases may not be specific to the N400 time window (Drijvers et al., 2018; He et al., 2015, 2018). These findings suggest that direct measures of trial-by-trial ERP variability such as median absolute deviation (MAD) may provide a more informative measure of temporal integration of beat gesture with pitch accent than oscillatory activity. Although MAD has been used to examine trial-by-trial variation in the P1 ERP (Milne, 2011), to our knowledge, it hasn't been used to examine trial-by-trial variation in the N400 ERP. Moreover, to our knowledge, MAD as a measure of trial-by-trial variation hasn't been correlated with across-trial N400 ERPs. Therefore, although we believe that MAD has the potential be particularly informative as a measure of trial-by-trial evoked response variability, our examination of it should be considered exploratory and supplemental to across-trial N400 amplitude and latency in the current study.

1.4. The current study

The current study is the first to manipulate the timing of the two most closely temporally-linked focus cues in English—beat gesture and pitch accent—to examine the neural signatures of the temporal relationship between them. To do so, it uses ERPs, which have a much higher temporal resolution than fMRI and should therefore reflect sensitivity to manipulations of temporal beat gesture-pitch accent synchrony with high fidelity. In particular, the current study focuses on the N400 ERP, thereby providing insight into how the temporal relationship between beat gesture and pitch accent affects processing of focused information. Notably, it examines *whether* the N400 differs on the basis of temporal synchrony in terms of both its amplitude, which was previously examined in Wang & Chu (2013), as well as its latency, which has not previously been examined during beat gesture-pitch accent integration. Moreover, it explores *how* the N400 differs on the basis of temporal synchrony in terms of its trial-by-trial variability in amplitude and latency using MAD, which reflects variability in ERP expression across trials. Finally, it examines the relationship between across-trial (average) and trial-by-trial (variability in) measures of N400 amplitude and latency based on temporal synchrony. In doing so, it provides insight into how trial-by-trial variability in N400 expression relates to across-trial differences in N400 amplitude and latency during temporal and semantic integration of beat gesture and pitch accent.

In the current study, all experimental stimuli took the form of videos of a talker producing unrelated sentences consisting of a subject, a verb, a direct object, and an indirect object, as in Wang and Chu (2013) (e.g., Yesterday, Anne brought colorful *lilies* to the room). In these sentences, beat gesture was manipulated in relation to the direct object (the critical word; CW, italicized), which was focused via pitch accent. Thus, the timing of beat gesture relative to pitch accent was highly predictable on the basis of information structure. To examine how the timing of beat gesture relative to pitch accenting affects processing of focused information, the timing of beat gesture relative to pitch accented CWs was manipulated. This manipulation of the timing of beat gesture relative to pitch accenting differed from Wang and Chu (2013), in which the presence and form—but not the timing—of beat gesture relative to pitch accent was manipulated. To confirm how the timing of beat gesture relative to pitch accenting influenced ERPs, two

complementary analyses time-locked to different events were conducted. The primary analysis examined ERPs time-locked to CW onset, and a secondary analysis examined ERPs time-locked to the onset of the indirect object (IO), which co-occurred with temporally asynchronous beat gesture.

Based on the findings of previous research, several predictions were advanced concerning across-trial and trial-by-trial N400 amplitude and latency. First, we predicted that temporal asynchrony of beat gesture in relation to pitch accent would elicit larger and earlier N400 ERPs relative to temporally synchronous beat gesture or the absence of beat gesture. This finding would demonstrate that temporal asynchrony of beat gesture relative to pitch accent disrupts processing of focused information to a greater degree than the presence or—notably—even the absence of beat gesture relative to pitch accent. Second, in our exploratory analyses of MAD, we anticipated that temporal asynchrony of beat gesture relative to pitch accent would (1) increase trial-by-trial variability during the N400 window and (2) elicit a positive correlation between trial-by-trial variability and across-trial averages. These findings would provide evidence that the temporal relationship between beat gesture and pitch accent is reflected in trial-by-trial variability of the N400, an ERP that reflects semantic and temporal processing. Together, these results would provide convergent evidence that temporal synchrony of beat gesture relative to pitch accent affects processing of focused information during language comprehension.

2. Results

2.1. Critical word onset analyses

We first examined across-trial and trial-by-trial amplitude and latency, as well as the respective relationships between them, for ERPs in the 200–800 ms time window time-locked to onsets of pitch accented CWs. This constituted our primary analysis.

2.1.1. Across-trial amplitude and latency (standard ERP analysis).

Our main goal was to determine whether ERPs during the N400 time window differed in amplitude, latency, and topology as a function of the temporal relationship of beat gesture relative to pitch accenting. To achieve this goal, we conducted a standard ERP analysis. In this analysis, we compared the mean amplitudes and peak latencies of grand average ERPs in the left and right posterior “regions of interest” (ROIs) elicited by temporally synchronous beat gesture, temporally asynchronous beat gesture, and no beat gesture using linear mixed effect models. Grand average ERPs averaged across channels comprising each ROI as well as scalp voltage topographies are presented by condition in Fig. 1. Parameter estimates for the models for mean ERP amplitude and peak ERP latency are displayed in Tables 1 and 2, respectively.

For mean amplitudes in the 200–800 ms time window, a main effect of condition was observed for both comparisons (see Table 1). Tukey HSD-adjusted planned comparisons indicated that mean amplitudes in the no beat and asynchronous beat conditions were more negative (larger) than those in the synchronous beat condition ($B = -1.04$, $SE = 0.23$, $t = -4.45$, $p < .001$; $B = -2.61$, $SE = 0.23$, $t = -11.13$, $p < .001$). Furthermore, mean amplitudes in the asynchronous beat condition were more negative than those in the no beat condition ($B = -1.57$, $SE = 0.23$, $t = -6.68$, $p < .001$). Moreover, a main effect of ROI was observed, indicating that mean amplitudes were more negative in the right than the left posterior ROI (see Table 1). Crucially, a two-way interaction between the asynchronous vs. synchronous and no beat conditions and ROI was observed (see Table 1). Tukey HSD-adjusted planned comparisons indicated that mean amplitudes in the asynchronous beat condition were more negative than mean amplitudes in the synchronous beat condition in the right posterior ROI ($B = 2.04$, $SE = 0.33$, $t = 6.15$, $p < .001$), but not the left posterior ROI ($B = -0.28$, $SE = 0.33$, $t = -0.84$, $p = .96$). These results indicate that, relative to synchronous beat gesture, asynchronous beat gesture

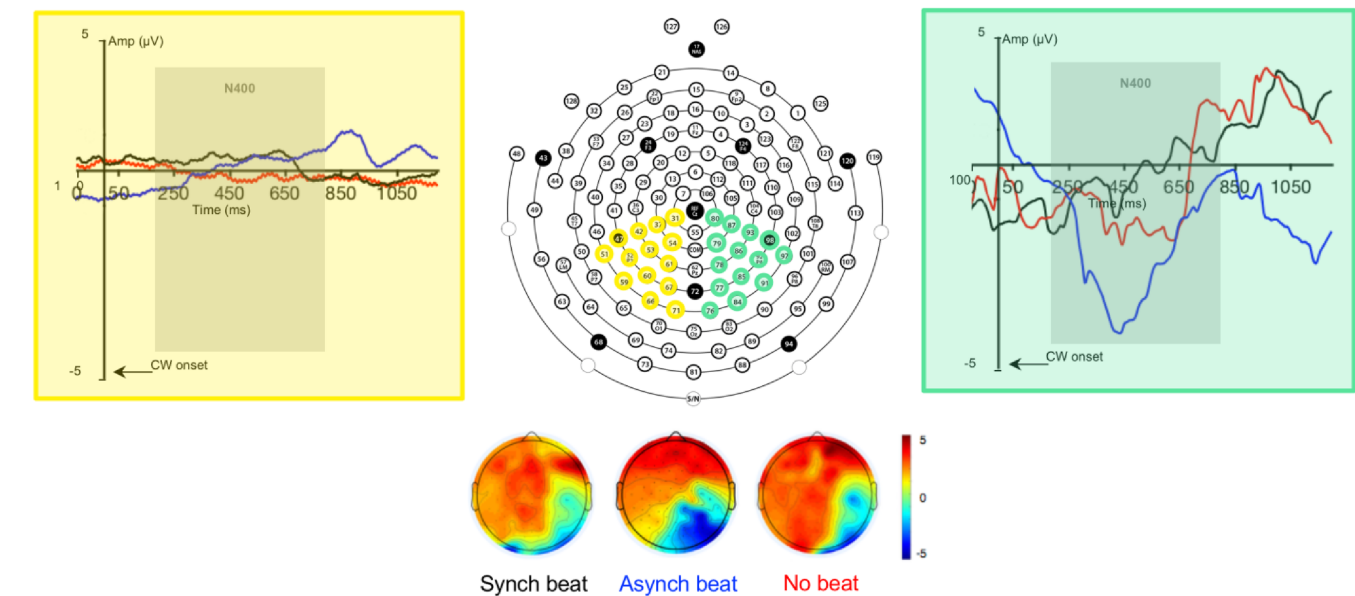


Fig. 1. 128-Channel Montage Used for EEG Recording in This Study with Channels in Left Posterior (Yellow) and Right Posterior (Green) ROIs Highlighted with Grand Average ERPs Evoked by Temporal Synchrony of Beat Gesture Relative to Pitch Accented CW Onset Averaged Across Channels and Scalp Voltage Topographies During the 200–800 Time Window.

Table 1
Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of Mean Amplitudes of Event-Related Potentials by Condition and Region of Interest for 200–800 ms Time Window Relative to CW Onset (Observations = 2100).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	−0.09	0.51	−0.19	0.85
Condition 1 (synch beat vs. no beat)	1.04	0.23	4.45	< 0.001
Condition 2 (asynch beat vs. synch + no beat)	−2.09	0.2	−10.29	< 0.001***
ROI (left vs. right)	−0.56	0.19	−2.91	0.004**
Condition 1 × ROI	−0.16	0.41	−0.41	0.69
Condition 2 × ROI	1.56	0.47	3.34	< 0.001***
Random effect				s ²
Participant				2.50
Channel				0.01

Table 2
Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of Peak Latencies of Event-Related Potentials by Condition and Region of Interest for 200–800 ms Time Window Relative to CW Onset (Observations = 2100).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	531.91	17.76	29.96	< 0.001***
Condition 1 (synch beat vs. no beat)	−9.04	9.87	−0.92	0.36
Condition 2 (asynch beat vs. synch + no beat)	3.19	8.55	0.37	0.71
ROI (left vs. right)	29.8	8.06	3.7	< 0.001***
Condition 1 × ROI	−1.29	17.09	−0.08	0.94
Condition 2 × ROI	40.87	19.74	2.07	0.04*
Random effect				s ²
Participant				86.46
Channel				0.01

elicited a right-lateralized negative ERP during the N400 time window, whereas the absence of gesture did not.

For peak latencies in the 200–800 ms time window, no main effect

of condition was observed (see Table 2). A main effect of ROI was observed, however, indicating that peak amplitudes across conditions occurred earlier in the right than the left posterior ROI (see Table 2). Moreover, a two-way interaction between the asynchronous vs. synchronous and no beat conditions and ROI was observed (see Table 2). This interaction indicates that this effect was driven by earlier peak amplitudes in the asynchronous beat than the no beat and synchronous beat conditions in the right posterior ROI. Tukey HSD-adjusted planned comparisons confirmed that peak amplitudes in the asynchronous beat condition occurred earlier in the right posterior ROI than they did in the left posterior ROI ($B = -50.66$, $SE = 14.0$, $t = -3.63$, $p = .004$). Furthermore, peak amplitudes in the asynchronous beat condition occurred earlier than they did in the no beat condition in the right posterior ROI ($B = -47.51$, $SE = 14.0$, $t = -3.40$, $p = .009$). No other planned comparisons reached significance. These results indicate that the right-lateralized negative ERP elicited by the asynchronous beat gesture condition during the N400 time window peaked earlier than the right-lateralized ERPs elicited by the synchronous and no gesture conditions.

2.1.2. Trial-by-trial variability

Our secondary goal was to determine whether trial-by-trial variability of the N400 ERP differed between the synchronous beat, asynchronous beat, and no beat conditions. To achieve this goal, we compared MAD (median absolute deviation) of peak amplitude by trial (Fig. 2A) and by sample (Fig. 2B) and MAD of peak latency by trial (Fig. 2C) within the 200–800 ms time window across these conditions using linear mixed effect models. Parameter estimates of these models for MAD of peak amplitude by trial and sample and MAD of peak latency by trial are displayed in Tables 3–5, respectively.

For MAD of peak amplitude by trial, a main effect of condition was observed for both comparisons (see Table 3 and Fig. 2A). Tukey HSD-adjusted paired comparisons indicated that MAD of peak amplitude by trial was greater in the asynchronous beat and no beat conditions than in the synchronous beat condition ($B = 0.02$, $SE = 0.01$, $t = 2.77$, $p = .01$; $B = 0.05$, $SE = 0.01$, $t = 5.48$, $p < .001$). Furthermore, MAD of peak amplitude by trial was greater in the asynchronous beat condition than in the no beat condition ($B = 0.02$, $SE = 0.01$, $t = 2.96$, $p = .009$). For MAD of peak amplitude by sample, a main effect of condition was also observed for both comparisons (see Table 4 and

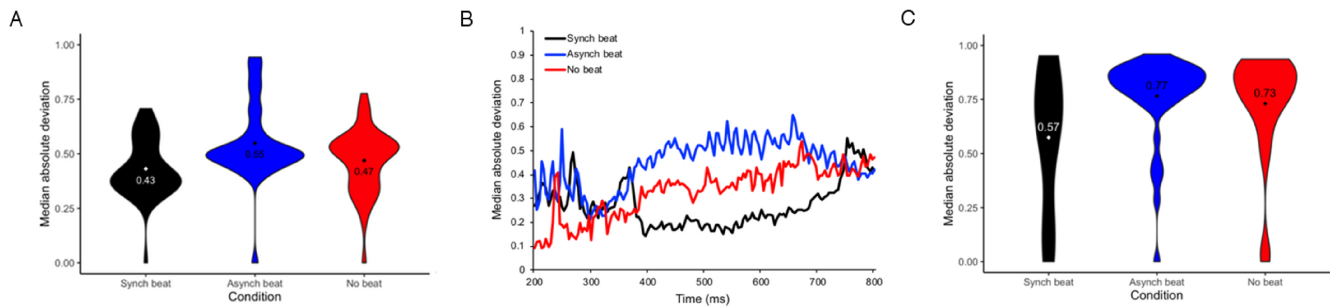


Fig. 2. Estimates of Trial-By-Trial ERP Variability at 200–800 ms for Beat Gesture Relative to Pitch Accented CWs: (A) Normalized MAD of Peak Amplitude by Trial; (B) Normalized MAD of Peak Amplitude by Sample; (C) Normalized MAD of Peak Latency by Trial.

Table 3

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of MAD of Peak Amplitudes of Event-Related Potentials in the Right ROI by Condition for 200–800 ms Time Window Relative to CW Onset by Trial (Observations = 3782).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.48	0.02	21.85	< 0.001***
Condition 1 (synch beat vs. no beat)	−0.02	0.01	−2.77	0.006**
Condition 2 (asynch beat vs. synch + no beat)	−0.04	0.01	−4.88	< 0.001***
Random effect				
s^2				
Participant				0.10
Trial				0.08

Table 4

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of MAD of Peak Amplitudes of Event-Related Potentials in the Right ROI by Condition for 200–800 ms Time Window Relative to CW Onset by Sample (Observations = 11325).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.50	0.02	23.76	< 0.001***
Condition 1 (synch beat vs. no beat)	−0.04	0.01	−15.36	< 0.001***
Condition 2 (asynch beat vs. synch + no beat)	0.02	0.01	7.65	< 0.001***
Random effect				
s^2				
Participant				0.11
Sample				0.01

Table 5

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of MAD of Peak Latencies of Event-Related Potentials in the Right ROI by Condition for 200–800 ms Time Window Relative to CW Onset by Trial (Observations = 3782).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.76	0.02	40.05	< 0.001***
Condition 1 (synch beat vs. no beat)	0.02	0.01	3.48	< 0.001***
Condition 2 (asynch beat vs. synch + no beat)	0.01	0.01	0.97	0.33
Random effect				
s^2				
Participant				0.04
Trial				0.15

Fig. 2B). Tukey HSD-adjusted paired comparisons indicated that MAD of peak amplitude by sample was greater in the asynchronous and no beat conditions than the synchronous beat condition ($B = 0.04$, $SE = 0.01$, $t = 15.36$, $p < .001$; $B = 0.04$, $SE = 0.01$, $t = 14.30$,

$p < .001$). No difference in MAD of peak amplitude by sample was found between the asynchronous beat and no beat conditions, however ($B = 0.01$, $SE = 0.01$, $t = 1.06$, $p = .54$). Finally, for MAD of peak amplitude by trial, a main effect of condition was also observed for the synchronous beat vs. no beat comparison (see Table 5 and Fig. 2C). Tukey HSD-adjusted paired comparisons indicated that that MAD of peak latency by trial was greater in the asynchronous beat and no beat conditions than it was in the synchronous beat condition ($B = 0.02$, $SE = 0.01$, $t = 3.48$, $p = .001$; $B = 0.01$, $SE = 0.01$, $t = 2.52$, $p = .03$). MAD of peak latency by trial did not differ between the asynchronous beat and no beat conditions, however ($B = 0.01$, $SE = 0.01$, $t = 0.90$, $p = .64$). Considered as a whole, these results provide evidence that temporal asynchrony and absence of beat gesture relative to pitch accent are associated with higher variability of amplitude and latency during the N400 epoch than temporal synchrony of beat gesture relative to pitch accent.

2.1.3. Across trial – trial-by-trial ERP correlations.

Finally, we examined whether trial-by-trial variability of amplitude and latency during the N400 time window correlates with ERP amplitude and latency, as well as whether this relationship differs based on temporal synchrony of beat gesture with pitch accent. To do so, we regressed MAD of N400 peak amplitude and latency onto mean N400 amplitude and peak N400 latency both across and within conditions. Relationships between these factors are presented in Fig. 3. No overall relationship was observed between MAD of peak amplitude and N400 mean amplitude, $B = 0.01$, $SE = 0.01$, $t = 0.13$, $p = 0.93$, nor did this relationship differ by condition (see Fig. 3A). However, a positive overall relationship was observed between MAD of peak latency and peak N400 latency, $B = 0.01$, $SE = 0.01$, $t = 2.04$, $p = 0.04$. This relationship differed between the asynchronous and synchronous beat conditions, $B = 0.12$, $SE = 0.06$, $t = 2.09$, $p = 0.04$, as well as the no beat and synchronous beat conditions, $B = 0.12$, $SE = 0.07$, $t = 2.12$, $p = 0.04$, such that MAD of peak latency and peak N400 latency positively correlated in the asynchronous and no beat conditions but not the synchronous beat condition (see Fig. 3B). These results provide evidence that MAD of peak latency and peak N400 latency during the 200–800 ms epoch are positively related and that temporally synchronous beat gesture relative to pitch accent reverses this relationship.

2.2. Indirect object onset analyses

We next probed the generalizability of the results of the critical word onset analyses. To do so, we analyzed across-trial averages and trial-by-trial variability, as well as the relationship between them, for ERPs in the 200–800 ms time window time-locked to onsets of the IO of each sentence. Although the talker was not explicitly instructed to pitch accent IOs, their position proximal to a phrasal boundary resulted in implicit pitch accenting (see Table 5 for descriptive statistics of acoustic characteristics). This constituted our secondary analysis.

To facilitate comparison of results between the primary (CW) and

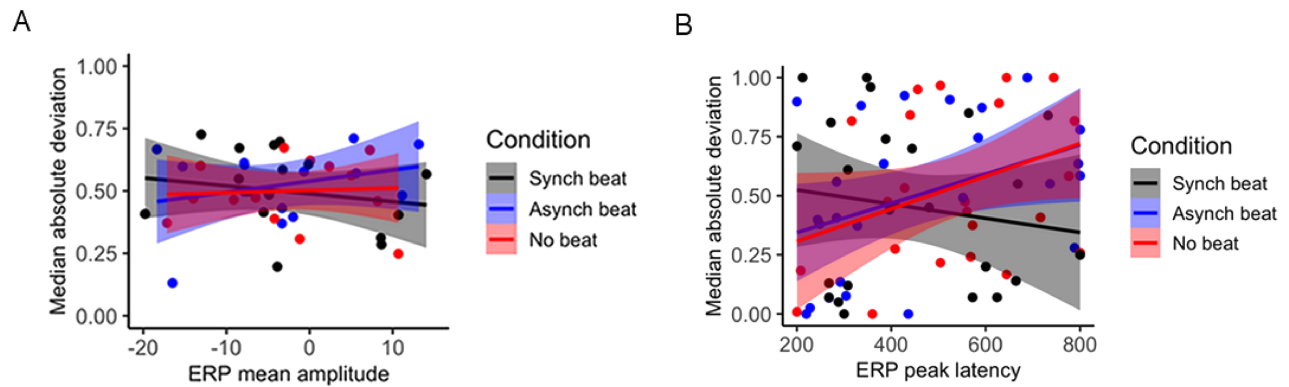


Fig. 3. Relationships Between Mean ERP Amplitude and Peak ERP Latency and Measures of Trial-By-Trial ERP Variability in the 200–800 ms Time Window by Participant for Beat Gesture Relative to Pitch Accented CWs: (A) Normalized MAD of Peak Amplitude by ERP Mean Amplitude; (B) Normalized MAD of Peak Latency by ERP Peak Latency.

secondary (IO) analyses, we switched the names of the synchronous and asynchronous beat conditions in this analysis. Thus, beat gesture temporally asynchronous with the IO in the secondary analysis (previously temporally synchronous with the CW in the primary analysis) is referred to as the asynchronous beat condition. Conversely, beat gesture temporally synchronous in the current analysis (previously temporally asynchronous with the CW in the primary analysis) is referred to as the synchronous beat condition. Because the apices of beat gestures in the temporally asynchronous condition occurred asynchronously with IO onsets, we expected that the effects observed in the asynchronous beat gesture condition in the primary analyses would be reproduced in the asynchronous beat condition in the secondary analyses.

2.2.1. Across-trial amplitude and latency (standard ERP analysis)

Our main goal in the secondary analysis was to verify that ERPs during the N400 time window relative to IO onset differed as a function of the temporal relationship between beat gesture and pitch accent. To achieve this goal, we conducted a standard ERP analysis. In this analysis, we compared the mean amplitudes and peak latencies of grand average ERPs in the left and right posterior ROIs between the synchronous beat, asynchronous beat, and no beat gesture conditions. Grand average ERPs averaged across channels comprising each ROI as

Table 6
Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of Mean Amplitudes of Event-Related Potentials by Condition and Region of Interest for 200–800 ms Time Window Relative to IO Onset (Observations = 2100).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	−0.28	0.73	−0.38	0.71
Condition 1 (synch beat vs. no beat)	−1.35	0.43	−3.15	0.002**
Condition 2 (asynch beat vs. synch + no beat)	−1.01	0.43	−2.35	0.02*
ROI (left vs. right)	−1.27	0.43	−2.97	0.003**
Condition 1 vs. ROI	−0.06	0.61	−0.10	0.92
Condition 2 vs. ROI	1.36	0.61	2.24	0.03*
Random effect				
				s ²
Participant				3.30
Channel				0.01

well as scalp maps are presented by condition in Fig. 4. Parameter estimates for the models for mean amplitudes and peak latencies are displayed in Tables 6 and 7, respectively.

For mean amplitudes in the 200–800 ms time window, a main effect

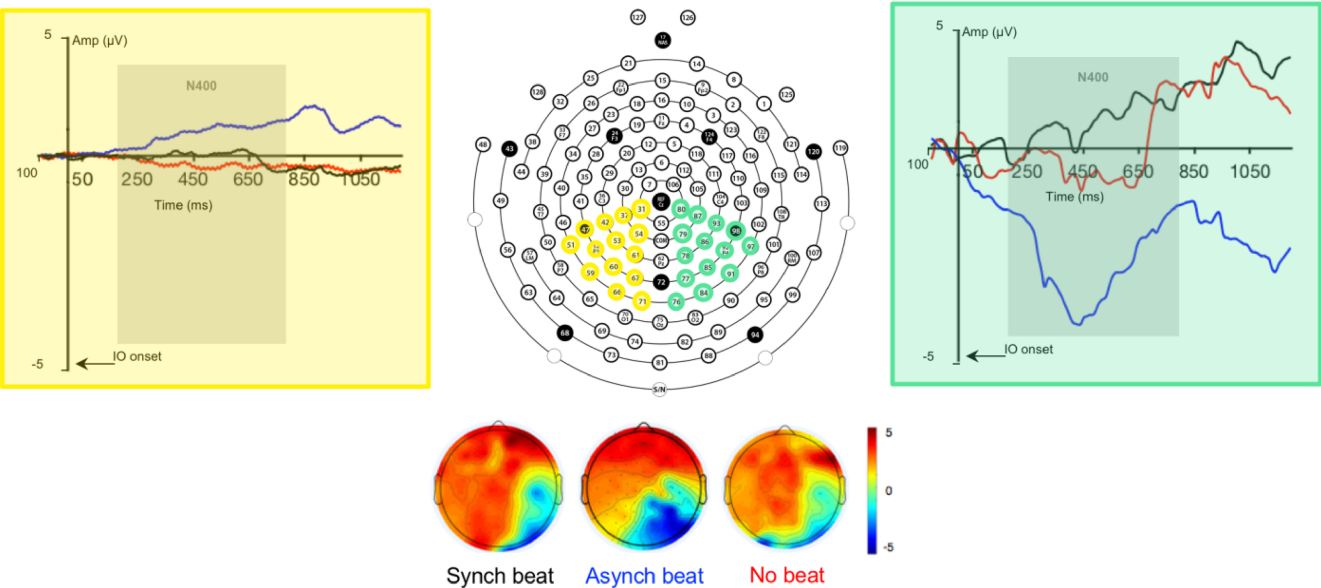


Fig. 4. 128-Channel Montage Used for EEG Recording in This Study with Channels in Left Posterior (Yellow) and Right Posterior (Green) ROIs Highlighted with Grand Average ERPs Evoked by Temporal Synchrony and Presence of Beat Gesture Relative to Pitch Accented IO Onset Averaged Across Channels and Scalp Voltage Topographies During the 200–800 Time Window.

Table 7

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of Peak Latencies of Event-Related Potentials by Condition and Region of Interest for 200–800 ms Time Window Relative to IO Onset (Observations = 2100).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	525.18	19.77	26.57	< 0.001***
Condition 1 (synch beat vs. no beat)	22.54	18.86	1.20	0.23
Condition 2 (asynch beat vs. synch + no beat)	−3.25	16.33	−0.20	0.84
ROI (left vs. right)	16.11	7.70	2.09	0.04*
Condition 1 × ROI	−91.65	9.43	−9.72	< 0.001***
Condition 2 × ROI	52.33	8.17	6.41	< 0.001***
Random effect				σ^2
Participant				96.94
Channel				0.01

of condition was observed for both comparisons (see Table 6). Tukey HSD-adjusted planned comparisons indicated that mean amplitudes in the no beat and asynchronous beat conditions were more negative (larger) than those in the synchronous beat condition ($B = -3.07$, $SE = 0.30$, $t = -10.11$, $p < .001$; $B = -2.03$, $SE = 0.30$, $t = -6.69$, $p < .001$). Furthermore, mean amplitudes in the asynchronous beat condition were more negative than those in the no beat condition ($B = -1.04$, $SE = 0.30$, $t = 3.42$, $p = .002$). Moreover, a main effect of ROI was observed, indicating that mean amplitudes were more negative in the right than the left posterior ROI (see Table 6). Crucially, a two-way interaction between the synchronous vs. asynchronous beat conditions and ROI was observed (see Table 6). Tukey HSD-adjusted planned comparisons indicated that mean amplitudes in the asynchronous beat condition were more negative than mean amplitudes in the synchronous beat condition in the right posterior ROI ($B = -2.71$, $SE = 0.43$, $t = -6.32$, $p < .001$), but not the left posterior ROI ($B = 1.01$, $SE = 0.43$, $t = 2.35$, $p = .18$). These results provide additional evidence that, relative to synchronous beat gesture, asynchronous beat gesture elicited a right-lateralized negative-going potential during the N400 time window, whereas the absence of gesture did not.

For peak latencies in the 200–800 ms time window, no main effect of condition was observed (see Table 7). A main effect of ROI was observed, however, indicating that peak amplitudes across conditions occurred earlier in the right than the left posterior ROI (see Table 7). Moreover, positive two-way interactions between the synchronous and no beat and the asynchronous vs. synchronous and no beat conditions and ROI were observed (see Table 7). These interactions indicate that this effect was driven by earlier peak amplitudes in the no beat and asynchronous beat than the synchronous beat conditions in the right posterior ROI. Tukey HSD-adjusted planned comparisons confirmed that peak amplitudes in the asynchronous beat condition occurred earlier in the right posterior ROI than they did in the left posterior ROI ($B = -102.16$, $SE = 13.3$, $t = -6.03$, $p < .001$). Moreover, they indicated that peak amplitudes in the asynchronous beat condition occurred earlier than they did in the no beat condition in the right posterior ROI ($B = -80.38$, $SE = 13.3$, $t = -7.66$, $p < .001$). No other planned comparisons reached significance. These results provide further evidence that the right-lateralized negative ERP elicited by the asynchronous beat gesture condition during the N400 time window peaked earlier than the right-lateralized ERPs elicited by the synchronous and no gesture conditions.

2.2.2. Trial-by-trial variability

Our secondary goal was to determine whether trial-by-trial variability of the N400 ERP relative to IO onsets differed between the synchronous beat, asynchronous beat, and no beat conditions. To achieve this goal, we compared MAD of peak amplitude by trial (Fig. 5A) and by

sample (Fig. 5B) and MAD of peak latency by trial (Fig. 5C) within the 200–800 ms time window across these conditions using linear mixed effect models. Parameter estimates of these models for MAD of peak amplitude by trial and sample and MAD of peak latency by trial are displayed in Tables 8–10, respectively.

For MAD of peak amplitude by trial, a main effect of condition was observed for the synchronous beat vs. no beat comparison (see Table 8 and Fig. 5A). Tukey HSD-adjusted paired comparisons indicated that MAD of peak amplitude by trial was greater in the asynchronous beat condition than in the synchronous beat condition ($B = 0.02$, $SE = 0.01$, $t = 2.05$, $p = .03$). MAD of peak amplitude by trial did not differ between the asynchronous beat and no beat conditions, however ($B = 0.01$, $SE = 0.01$, $t = 1.65$, $p = .12$), nor did it differ between the no beat and synchronous beat conditions ($B = 0.01$, $SE = 0.01$, $t = 0.11$, $p = .99$). For MAD of peak amplitude by sample, a main effect of condition was observed for both comparisons (see Table 9 and Fig. 5B). Tukey HSD-adjusted paired comparisons indicated that MAD of peak amplitude by sample was greater in the asynchronous and no beat conditions than the synchronous beat condition ($B = 0.03$, $SE = 0.01$, $t = 12.62$, $p < .001$; $B = 0.03$, $SE = 0.01$, $t = 11.91$, $p < .001$). No difference in MAD of peak amplitude by sample was found between the asynchronous beat and no beat conditions, however ($B = 0.01$, $SE = 0.01$, $t = 0.71$, $p = .76$). Finally, for MAD of peak amplitude by trial, a main effect of condition was also observed for both comparisons (see Table 10 and Fig. 5C). Tukey HSD-adjusted paired comparisons indicated that that MAD of peak latency by trial was greater in the asynchronous beat and no beat conditions than it was in the synchronous beat condition ($B = 0.03$, $SE = 0.01$, $t = 4.04$, $p < .001$; $B = 0.03$, $SE = 0.01$, $t = 4.88$, $p < .001$). MAD of peak latency by trial did not differ between the asynchronous beat and no beat conditions, however ($B = 0.01$, $SE = 0.01$, $t = 0.78$, $p = .71$). Considered as a whole, these results provide further evidence that temporal asynchrony and absence of beat gesture relative to pitch accent is associated with higher trial-by-trial variability of ERP amplitude and latency during the N400 epoch than temporal synchrony of beat gesture relative to pitch accent.

2.2.3. Across trial – trial-by-trial ERP correlations

Finally, we examined whether trial-by-trial variability of amplitude and latency during the N400 time window relative to IO onset correlates with ERP amplitude and latency, as well as whether this relationship differs based on temporal synchrony of beat gesture with pitch accent. To do so, we regressed MAD of N400 peak amplitude and latency onto mean N400 amplitude and peak N400 latency both across and within conditions. Relationships between these factors are presented in Fig. 6. No overall relationship was observed between MAD of peak amplitude and ERP mean amplitude, $B = -0.01$, $SE = 0.01$, $t = -0.11$, $p = 0.91$, nor did this relationship differ by condition (see Fig. 6A). Moreover, no overall relationship was observed between MAD of peak latency and peak N400 latency, $B = -0.01$, $SE = 0.01$, $t = -1.04$, $p = 0.30$, nor did this relationship differ by condition (see Fig. 6B). These results suggest that the positive correlation between MAD of peak amplitude and peak ERP amplitude observed during the 200–800 ms time window in the CW analysis may be relatively circumscribed to contexts in which the temporal linkage between beat gesture and pitch accenting is extremely stable.

3. Discussion

The current study manipulated the timing of beat gesture relative to pitch accent to examine the N400 ERP as a neural signature of the temporal relationship between them. The results revealed that beat gesture temporally asynchronous with pitch accent elicited a larger and earlier N400 averaged across trials in the right posterior ROI than beat gesture temporally synchronous with pitch accent, as well as the absence of beat gesture. This finding suggests that temporal asynchrony

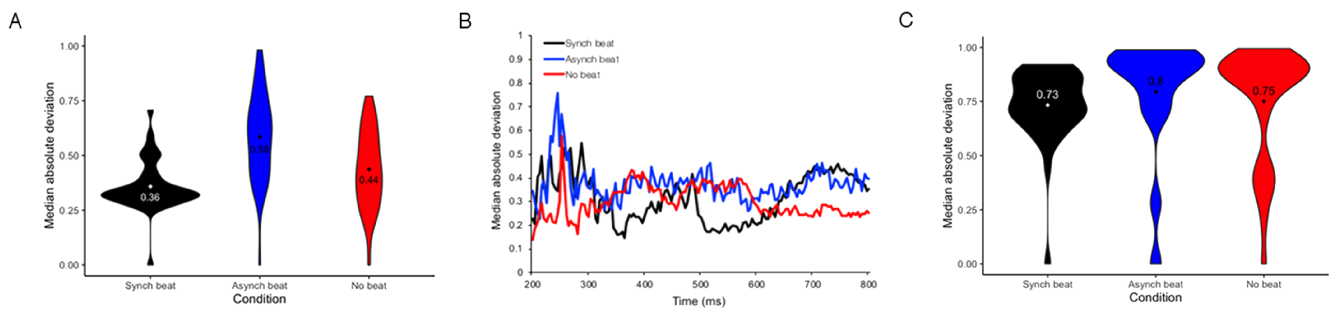


Fig. 5. Estimates of Trial-By-Trial ERP Variability at 200–800 ms for Beat Gesture Relative to Pitch Accented IOs: (A) Normalized MAD of Peak Amplitude by Trial; (B) Normalized MAD of Peak Amplitude by Sample; (C) Normalized MAD of Peak Latency by Trial.

Table 8

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of MAD of Peak Amplitudes of Event-Related Potentials in the Right ROI by Condition for 200–800 ms Time Window Relative to IO Onset by Trial (Observations = 3782).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.50	0.02	20.54	< 0.001***
Condition 1 (synch beat vs. no beat)	−0.02	0.01	−2.05	0.04*
Condition 2 (asynch beat vs. synch + no beat)	0.01	0.01	1.03	0.30
Random effect	s^2			
Participant	0.11			
Trial	0.09			

Table 9

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of MAD of Peak Amplitudes of Event-Related Potentials in the Right ROI by Condition for 200–800 ms Time Window Relative to IO Onset by Sample (Observations = 11325).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.52	0.02	25.00	< 0.001***
Condition 1 (synch beat vs. no beat)	−0.03	0.01	−12.62	< 0.001***
Condition 2 (asynch beat vs. synch + no beat)	0.01	0.01	6.47	< 0.001***
Random effect	s^2			
Participant	0.10			
Sample	0.02			

Table 10

Fixed Effect Estimates (Top) and Variance Estimates (Bottom) for Multi-Level Logit Model of MAD of Peak Latencies of Event-Related Potentials in the Right ROI by Condition for 200–800 ms Time Window Relative to IO Onset by Trial (Observations = 3782).

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.77	0.02	34.67	< 0.001***
Condition 1 (synch beat vs. no beat)	0.27	0.01	4.04	< 0.001***
Condition 2 (asynch beat vs. synch + no beat)	−0.02	0.01	−3.22	0.001**
Random effect	s^2			
Participant	0.06			
Trial	0.16			

between beat gesture and pitch accent hinders processing of focused information. Moreover, greater trial-by-trial variability in peak amplitude and latency were observed during the N400 time window when beat gesture was temporally asynchronous or absent relative to pitch

accent than when beat gesture was temporally synchronous relative to pitch accent. Finally, the positive correlation between across-trial and trial-by-trial peak latency during the N400 time window was affected by the temporal synchrony between beat gesture relative to pitch accent in the critical word analysis. Together, these results demonstrate that the temporal relationship between beat gesture and pitch accent is reflected in both across-trial and trial-by-trial N400 amplitude and latency, indicating that the timing of beat gesture and pitch accent relative to one another affects processing of focused information. Moreover, they reveal that temporal asynchrony of beat gesture relative to pitch accent is reflected in increased trial-by-trial variability in N400 latency. This increased trial-by-trial variability is positively correlated with earlier across-trial N400 latency, suggesting that increased trial-by-trial variability in N400 latency is attributable to earlier N400s to temporally asynchronous stimuli rather than spontaneous fluctuations in neural activity during multisensory processing. Thus, these results reveal that across-trial differences and trial-by-trial variation in N400 latency both reflect temporal integration of beat gesture and pitch accent, which in turn affects processing of focused information.

The current study is the first to show that across-trial N400 amplitude and latency reflect the temporal relationship between beat gesture and pitch accent. Specifically, the results revealed that, in the right posterior region, across-trial mean N400 amplitude was larger and peak N400 amplitude was earlier when beat gesture was temporally asynchronous relative to pitch accent than when beat gesture was temporally synchronous relative to pitch accent. Notably, temporal asynchrony of beat gesture relative to pitch accent affected N400 amplitude and latency more than the absence of beat gesture relative to pitch accent. Strictly speaking, these results are consistent with Wang and Chu's (2013) observation of independent—but not interactive—effects of beat gesture and pitch accent on the N400. Unlike Wang and Chu (2013), however, the current study held pitch accenting constant and manipulated the timing rather than the motion of beat gesture. The larger amplitude and earlier peak of the across-trial N400 observed when beat gesture was temporally asynchronous than when it was absent relative to pitch accent suggests that participants were less tolerant of temporal asynchrony of beat gesture relative to pitch accent than the absence of beat gesture relative to pitch accent. This finding is consistent with recent proposals that beat gesture should be defined with respect to its temporal synchrony with speech prosody (Prieto et al., 2018; Shattuck-Hufnagel et al., 2016). Conversely, the smaller across-trial N400 amplitude observed when beat gesture was temporally synchronous than when it was temporally asynchronous relative to pitch accent suggests that temporal synchrony of beat gesture relative to pitch accent facilitates processing of focused information. This interpretation is consistent with research demonstrating that the degree of temporal synchrony between representational gestures and their lexical affiliates is related to their semantic relationship (Habets et al., 2011; Morrel-Samuels & Krauss, 1992; Obermeier & Gunter, 2014). Taken together, the results of the current study suggest that temporal asynchrony between beat gesture and pitch accent may disrupt

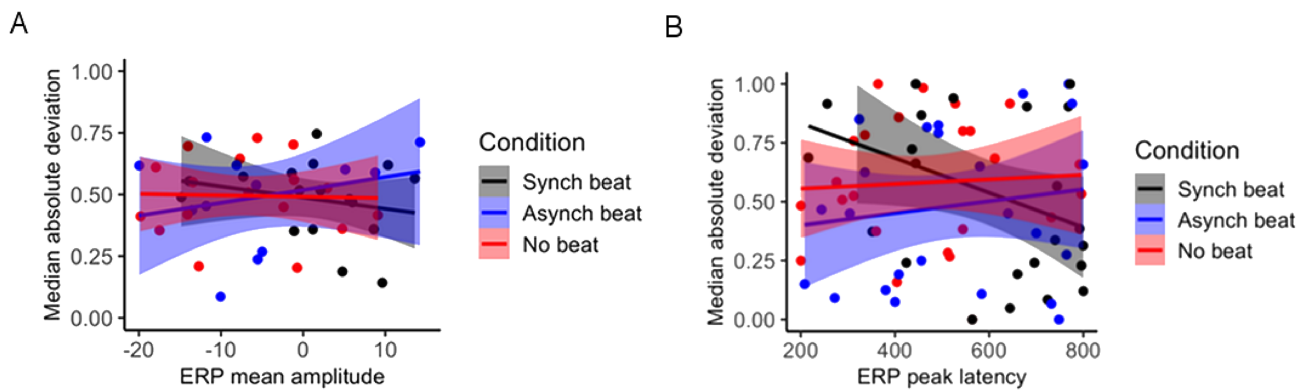


Fig. 6. Relationships Between Mean ERP Amplitude and Peak ERP Latency and Measures of Trial-By-Trial ERP Variability in the 200–800 ms Time Window by Participant for Beat Gesture Relative to Pitch Accented IOs: (A) Normalized MAD of Peak Amplitude by ERP Mean Amplitude; (B) Normalized MAD of Peak Latency by ERP Peak Latency.

processing of focused information even more than the absence of beat gesture relative to pitch accent where expected.

It is worth noting that the morphology and scalp topography of the N400 observed in the current study differ somewhat from those observed in Wang and Chu (2013), which may be due to differences in N400 time locking between the two studies. In the current study, the N400 was time-locked to the onset of the critical word, whereas in Wang & Chu (2013), the N400 was time-locked to the onset of beat gesture. Topographically, the N400 observed in the current study was posterior and right-lateralized, unlike the N400 observed in Wang and Chu (2013), which was posterior but slightly left lateralized for no movement compared to canonical beat gesture and control movement. However, Wang and Chu (2013) also found that canonical beat gesture elicited a smaller N400 than control movement over the right posterior region. In light of work showing a right posterior N400 for non-focused words with superfluous pitch accenting (Dimitrova et al., 2012), this finding suggests that right-lateralized N400s may reflect non-canonicity of beat gesture or incongruence of emphasis cues relative to one another. Nevertheless, the larger right posterior N400 elicited by beat gesture temporally asynchronous relative to pitch accent than by beat gesture temporally synchronous and absent relative to pitch accent in the current study is consistent with extension of the integrated systems hypothesis to the temporal relationship between beat gesture and pitch accent. In particular, this finding suggests that temporally asynchronous beat gesture relative to pitch accent disrupts processing of focused information similarly to semantically incongruent representational gesture relative to co-occurring lexical affiliates, whereas no such disruption of focused information processing is observed when gesture is absent in either case (Skipper, 2014).

The current study is also the first to show that trial-by-trial evoked response variability differs based on the temporal relationship between beat gesture and pitch accent. Specifically, the median absolute deviations (MAD) of both peak N400 amplitude and latency were higher when beat gesture was temporally asynchronous relative to pitch accent than when beat gesture was temporally synchronous or absent relative to pitch accent. This finding indicates that trial-by-trial variability in N400 amplitude and latency reflects the temporal relationship between beat gesture and pitch accent. Importantly, MAD of peak latency was correlated with average N400 peak latency only when beat gesture was temporally asynchronous relative to pitch accent, and this correlation was found only within the critical word analysis. This finding indicates that trial-by-trial and across-trial N400 timing are most closely related in the asynchronous beat condition and that the relationship between these measures is only discernable within contexts in which the timing of beat gesture relative to contrastive accenting is highly stable. This suggests that increased trial-by-trial variability in N400 latency is attributable to earlier N400s to temporally asynchronous beat gesture

relative to pitch accenting rather than spontaneous fluctuations in neural activity during multisensory processing. Thus, this finding provides evidence consistent with temporal asynchrony of beat gesture relative to pitch accenting eliciting a variety of reactions affecting trial-by-trial variability, as well as extension of neural noise theories of multisensory processing to focus processing in healthy populations. It is worth noting, however, that the trial-by-trial variability analyses of the N400 conducted in the current work were exploratory and that no other work has examined them to date. Therefore, future research should confirm the results of these analyses by examining trial-by-trial N400 variability while manipulating temporal synchrony of other multimodal cues during language processing, as well as by examining additional measures of trial-by-trial variability besides MAD.

Although the current study provides important insight into how the N400 reflects temporal synchrony of beat gesture relative to pitch accent during focus processing, it has some limitations. First, the current study did not directly test bi-directional influences of gesture and speech on one another given that only beat gesture was manipulated with respect to pitch accent and that beat gesture always followed pitch accent (and not vice versa). Second, it is unclear whether the N400 observed in the current study is specific to beat gesture. This was the case because we couldn't manipulate the timing of meaningless hand movement relative to pitch accent such that it wouldn't be interpreted as beat gesture, in light of recent proposals that beat gesture is identified by its temporal synchrony with speech prosody (Prieto et al., 2018; Shattuck-Hufnagel et al., 2016) and observations of similar N400s for beat gesture and analogously-timed control hand movement (Wang & Chu, 2013). Third, in the current study, EEG data were epoched relative to critical word onset rather than beat gesture onset, as in Wang and Chu (2013), precluding direct comparison of results between these studies. However, in the current study, the N400 amplitude and latency results observed in the main critical word analysis generalized to the secondary indirect object analysis; thus, none of these limitations negate the central conclusions. Fourth, the results cannot provide insight into whether trial-by-trial variability in the N400 ERP contributes to—or results from—temporal synchrony of focus cues. Despite this limitation, the findings provide evidence that trial-by-trial evoked response variability reflects temporal multisensory integration during language processing in healthy populations, supporting extension of neural noise theories to them. Nevertheless, replication of the results in future research addressing these limitations would strengthen our conclusions.

In conclusion, this work provides the first evidence that across-trial and trial-by-trial N400 amplitude latency reflects temporal synchrony of beat gesture and pitch accent during processing of focused information in spoken language. The results suggest that these measures of the N400 may have potential as biomarkers of atypical temporal

integration of gesture and speech, which has been observed in autism spectrum disorder (de Marchena & Eigsti, 2010; Morett et al., 2016). More generally, the findings of this work illuminate how the N400 reflects the temporal relationship between beat gesture and pitch accent, providing insight into how it is represented—and processed—in the brain, as well as how it affects language processing.

4. Methods and materials

4.1. Participants

Thirty-four native English speakers (age range: 18–35 yrs.; 29 females, 11 males) were recruited and compensated with \$25 in cash for their participation. Data from four participants were excluded due to the presence of artifacts in more than 50% of trials. Thus, the final sample consisted of thirty participants (see EEG Recording and Data Analysis section below for artifact rejection details for included participants). All participants were right-handed and had normal or corrected-to-normal hearing and vision, and none had any speech, language, or neurological disorders. Informed consent was provided by all participants prior to participation.

4.2. Stimuli

English translations of 80 of the Dutch sentences from Wang and Chu (2013) were used in experimental trials. An additional 12 sentences from that study were used as fillers to obscure the purpose of the experiment, and 6 sentences were used in practice trials to acclimate participants to the experiment. All sentences included a subject noun, a verb, a direct object, and an indirect object, as in (1) below. In experimental sentences, the direct object (italicized), which was defined as the critical word (CW), was always consciously pitch accented. In filler sentences, either the subject noun or indirect object was consciously pitch accented.

- (1) Yesterday, Anne brought colorful *lilies* to the room. Audio and video stimuli were recorded separately and subsequently combined. Audio stimuli, which consisted of audio recordings of sentences said by a female talker, were digitized at a sample frequency of 44.1 KHz and were normalized to the average sound pressure level (70 dB SPL) using Praat (Boersma & Weenink, 2016). Descriptive statistics for intensity, duration, f0 mean, f0 standard deviation, and root mean square (rms) amplitude of stressed syllables and entire CWs and IOs are listed in Table 11. The average duration of sentences was 3.26 s.

Video stimuli, which consisted of video recordings of a different female talker saying each sentence, were filmed with a digital camera at 33 ms/frame and edited using Adobe Premiere Pro. In each video clip, the talker was standing, and only her torso and limbs were visible. Three conditions were created: synchronous beat gesture, asynchronous

beat gesture, and no beat gesture (see Fig. 7 for an illustration of these conditions). Beat gestures consisted of flipping the right hand forward downward with an open palm, which is one of the most common types of beat gestures produced concurrently with spontaneous speech (McNeill, 1992). These gestures consisted of four components: the hand and forearm lifting up (preparation), the hand and forearm moving down rapidly (stroke), the hand and arm held briefly with palm upwards and fingers loosely extended (apex), and the hand and arm returning to the initial resting position (retraction). The average duration of each of these components is listed in Fig. 7.

Subsequent to recording, audio and video stimuli were combined using Adobe Premiere Pro. For the synchronous beat condition, the apex was synchronized with the onset of the pitch accented stressed syllable of CWs to reflect the tight temporal alignment between the gestural apex and the pitch accent peak on the stressed syllable (Leonard & Cummins, 2011; Loehr, 2012). For the asynchronous beat condition, the gestural apex was shifted to 500 ms after the pitch accent peak on the stressed syllable of CWs. The orientation and duration of this misalignment was selected to be reliably detectable based on previous behavioral research (Leonard & Cummins, 2011). Audio tracks were removed from video stimuli and were replaced with audio stimuli, and video stimuli were cropped to match the length of audio stimuli (see Fig. 7).

This experiment was structured to replicate Wang and Chu (2013) as closely as possible, with the exception that a partial factorial design was used to vary the timing of beat gesture in relation to pitch accented CWs. Table 12 provides a summary of the current study's experimental design with all independent variables and trial counts. All trials had pitch accenting, and the timing and presence of beat gesture relative to pitch accent was varied in two separate blocks, each of which was preceded by a brief practice block featuring the same manipulations. In the timing block, the beat gesture apex was temporally synchronous with the pitch accented CW in half of experimental trials, and the beat gesture apex was temporally asynchronous with the pitch accented CW in the other half. In the presence block, the beat gesture apex was temporally synchronous with the pitch accented CW in half of experimental trials, and beat gesture was absent in conjunction with the pitch accented CW in the other half. This blocking and distribution of trials across conditions was employed because we were originally planning to compare across trial types within each block in two separate analyses rather than across all three trial types in a single analysis. Each block contained a small number of interleaved filler trials in which the timing or presence of the beat gesture apex was manipulated relative to the pitch accented subject noun or IO in the same manner as it was manipulated relative to the pitch accented CW in experimental trials (timing block: synchronous vs. asynchronous; presence block: present vs. absent), with half of filler trials assigned to each condition. In both blocks, trials were presented in a pseudo-random order such that no more than three successive trials occurred in the same condition. Using four lists for each block, block order, sentence assignment by condition, and sentence presentation order were counterbalanced across participants in practice, experimental, and filler trials. Thus, an equal number of participants completed each block first and received each list.

4.3. Procedure

Each participant sat facing a computer screen located 80 cm away. Video stimuli were presented on the computer screen at a 1920 × 1080 resolution with masking, such that they were 10 cm in height and 11.8 cm in width, subtending a 72.82 degree visual angle. Accompanying audio stimuli were presented through loudspeakers placed behind and above the screen at a pre-specified volume level. Each trial consisted of three consecutive elements: A fixation cross with a jittered duration between 500 and 800 ms, a video clip, and a 1000 ms black screen. Participants were instructed simply to watch video stimuli and listen to the accompanying audio stimuli. Like Wang

Table 11
Mean Acoustic Measurements for Stressed Syllables and Entire Critical Words (CWs) and Indirect Objects (IOs) (Standard Deviation in Parentheses).

	CW		IO	
	Stressed syllable	Entire word	Stressed syllable	Entire word
Intensity (dB)	45.05 (2.52)	44.09 (2.41)	39.52 (2.76)	36.85 (2.39)
Duration (ms)	0.29 (0.07)	0.63 (0.12)	0.24 (0.08)	0.55 (0.13)
f0 mean (Hz)	242.69 (48.88)	220.93 (33.89)	185.65 (77.36)	195.26 (59.40)
f0 SD (Hz)	53.13 (39.95)	67.44 (40.02)	53.03 (68.21)	82.97 (62.84)
Amplitude rms	0.04 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)

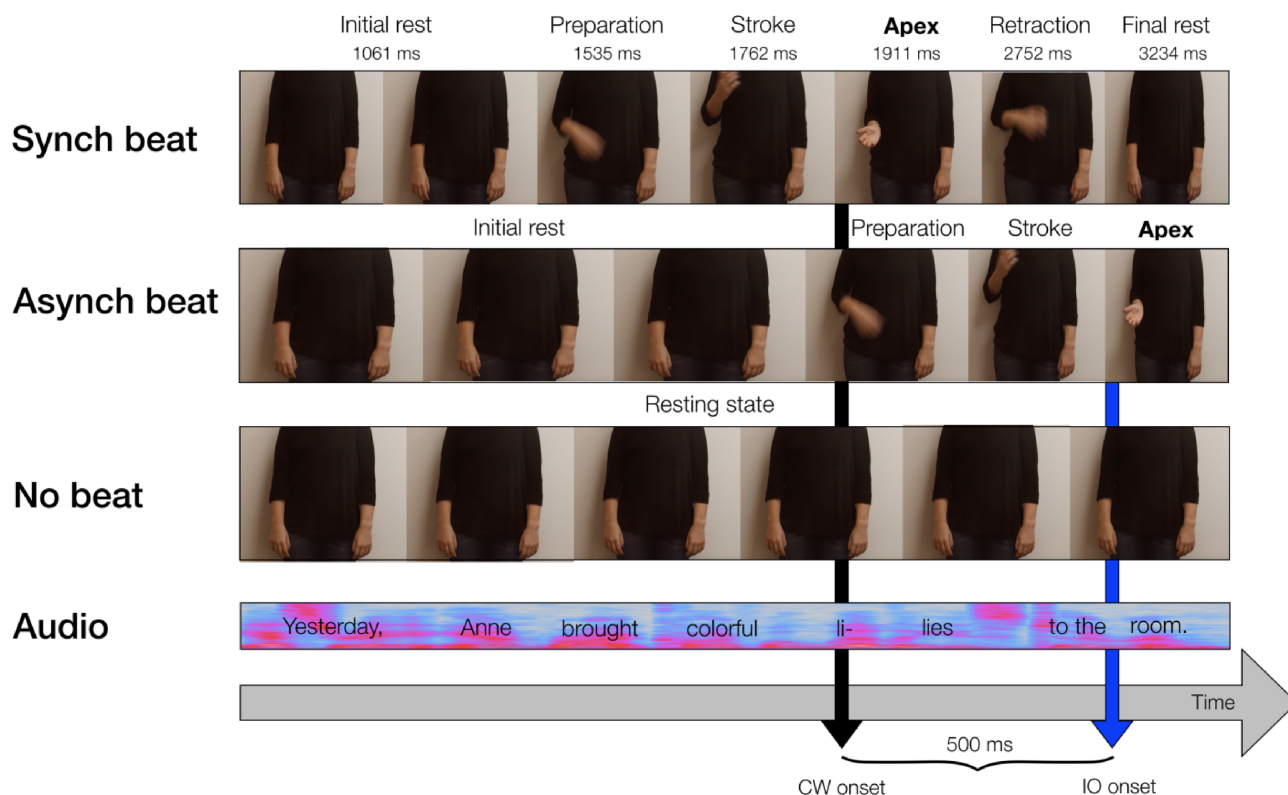


Fig. 7. Schematic of Synchronous Beat, Asynchronous Beat, and No Beat Conditions.

Table 12
Summary of Experimental Design.

Block	Trial type	Manipulation	Trials
Timing Practice	Experimental	Synchronous beat	1
Timing Practice	Experimental	Asynchronous beat	1
Timing Practice	Filler	Synchronous beat	1
Timing Practice	Filler	Asynchronous beat	1
Timing	Experimental	Synchronous beat	40
Timing	Experimental	Asynchronous beat	40
Timing	Filler	Synchronous beat	6
Timing	Filler	Asynchronous beat	6
Presence Practice	Experimental	Synchronous beat	1
Presence Practice	Experimental	No beat	1
Presence Practice	Filler	Synchronous beat	1
Presence Practice	Filler	No beat	1
Presence	Experimental	Synchronous beat	40
Presence	Experimental	No beat	40
Presence	Filler	Synchronous beat	6
Presence	Filler	No beat	6

and Chu (2013) and Dimitrova et al. (2016), we chose this passive listening task to avoid specific task-induced and undesired strategic comprehension processes and to minimize the effect of attention on ERP components. Prior to the experimental task and between blocks, impedances nearing or exceeding the 40 kΩ voltage threshold were reduced. Participants were allowed to pause for as long as desired between blocks. All participants opted to pause for no longer than one minute, however. The experimental task lasted approximately 12 min.

4.4. EEG recording and data analysis

EEG data were recorded via a 128-channel Hydrocel Geodesic sensor net (Electrical Geodesics, Inc., Eugene, OR, USA) with electrodes placed according to the international 10/20 standard. EEG signals were recorded using NetStation 4.5.4 with a NetAmps 300 Amplifier. The online reference electrode was Cz and the ground electrode had a

centroparietal location. EEG data were sampled at 1000 Hz with an anti-aliasing low-pass filter of 4000 Hz. In accordance with standard practice, only EEG data collected during experimental trials was analyzed.

EEG data were pre-processed and analyzed offline using EEGLab (Delorme & Makeig, 2004) and ERPLab (Lopez-Calderon & Luck, 2014). Continuous EEG data were high-pass filtered at 0.1 Hz to minimize drift and re-referenced to the online average of all electrodes. Subsequently, excessively noisy or flat channels and data from between-block breaks were removed. Continuous data were then downsampled to 250 Hz, low-pass filtered at 30 Hz, and segmented into epochs. Epoching was conducted relative to CW onset rather than beat gesture onset because this is the point at which pitch accenting was most pronounced and at which the onset of the gestural apex occurred in the synchronous condition. Thus, we expected that the N400 would show a greater difference relative to this point than relative to beat gesture onset. Epoch data were then screened for artifacts and abnormalities using a simple voltage threshold of 500 μV and a moving-window peak-to-peak threshold with 1000 ms windows, a 100 ms step function, and a 500 μV threshold. Across included participants, 8.2% of trials were rejected (mean = 14; range = 0–44), with rejections equally distributed across conditions ($t < 1$). Finally, trials were classified by condition and averaged across subjects for across-trial ERP analyses. Because beat gesture was temporally synchronous with pitch accented words in half the trials of both the timing and presence blocks, half of the temporally synchronous trials in each block were randomly selected and were collapsed together into a single synchronous beat condition consisting of the same number of trials as the other two conditions.

4.5. Across-trial (standard) ERP data analysis

Following Wang and Chu (2013), the 200–800 ms time window was selected for statistical analysis. In the synchronous beat condition, this time window was relative to beat gesture apex by 200–800 ms. In the asynchronous gesture condition, the apex of the beat gesture occurred

500 ms after the onset of the pitch accented word. Because the earliest hand movement began 2270 ms prior to CW onset and the latest hand movement ended 1200 ms after it, data segmentation started 2270 ms before and ended 1200 ms after CW onset, and the preceding 100 ms (-2370 – -2270 ms) was used as the baseline in all trials.² Mean amplitudes recorded during each condition (synchronous beat, asynchronous beat, no beat) were averaged across two lateralized “regions of interest” ROIs based on inspection of scalp voltage topographies and previous N400 research (Baptista et al., 2018; Kutas & Federmeier, 2011; Regel et al., 2011): left posterior, consisting of channels 31, 37, 42, 47, 51, 52, 53, 54, 59, 60, 61, 66, 67, and 71; and right posterior, consisting of channels 76, 77, 78, 79, 80, 84, 85, 86, 87, 91, 92, 93, 97, and 98.

ERP data were analyzed using linear mixed effect models, which account for random effects of both participant and trial or sample within a single model and therefore don't require aggregation across these factors and separate analyses with each of them as a random effect. Mean ERP amplitudes and peak ERP latencies within the 200 ms – 800 ms. time window were each entered into separate linear mixed effect models with fixed effects of condition (synchronous beat, asynchronous beat, no beat) and ROI (left, right) and crossed random intercepts of participant and trial or sample. Prior to entry into these models, all fixed effects were coded using weighted mean-centered (Helmert) contrast coding in order of the levels mentioned, such that the first level mentioned was the most negative and the last level mentioned was the most positive. For all effects reaching significance for factors with more than two levels, Tukey HSD post-hoc tests were conducted using the *emmeans* package (Lenth, 2019) to test for differences between levels.

4.6. Trial-by-trial evoked response variability analyses

All evoked response variability analyses were conducted on averaged evoked responses for the right ROI, in which the largest N400 differences by condition were found. For each participant, trial-by-trial variability in the expression of event-related potentials within the N400 time window (200–800 ms) was estimated via the median absolute deviation (MAD) by computing peak N400 amplitude and latency by trial and dividing it by the median. In addition, N400 amplitude was computed for each sample (time point) within the 200–800 ms time window for each trial and participant. Due to the inherent difficulty of calculating coefficients of variation when the central tendency is close to zero, amplitude values were normalized prior to computing MAD estimates by converting all of the data to z-scores, as in Milne (2011).

For analysis, MAD of peak amplitude and MAD of peak latency within the 200 ms – 800 ms. time window were each entered into separate linear mixed effect models with a fixed effect of condition (synchronous beat, asynchronous beat, no beat) and crossed random intercepts of participant and trial or sample. Prior to entry into these models, the fixed effect was coded using weighted mean-centered (Helmert) contrast coding in order of the levels mentioned, such that the first level mentioned was the most negative and the last level mentioned was the most positive. For all effects reaching significance, Tukey HSD post-hoc tests were conducted using the *emmeans* package (Lenth, 2019) to test for differences between levels.

² Although this baseline does not immediately precede the analyzed epoch, as is typical in ERP studies, it immediately preceded the onset of the earliest preparation phase, when the hand began to move, consistent with similar studies examining ERPs to gesture (e.g., Wang & Chu, 2013). Due to a limitation of our analysis software (EEGLab), it was not possible to vary baseline onset and offset on a trial-by-trial basis, necessitating application of the same baseline period in all trials.

4.7. ERP – variability correlation analyses

Two correlational analyses of mean ERPs and measures of ERP variability were conducted: mean ERP amplitude vs. MAD of peak ERP amplitude and peak ERP latency vs. MAD of peak ERP latency. For each participant and condition, the average ERP amplitude across trials in the N400 time window was computed and paired with the average MAD of peak ERP amplitude across trials. Likewise, for each participant and condition, the average peak ERP latency across trials in the N400 time window was computed and paired with the average MAD of peak latency across trials. Pearson product-moment correlation coefficients between each set of values were computed overall and by condition and tested using *t* tests with 95% confidence intervals.

CRediT authorship contribution statement

Laura M. Morett: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Nicole Landi:** Methodology, Writing - review & editing. **Julia Irwin:** Methodology, Writing - review & editing. **James C. McPartland:** Conceptualization, Methodology, Resources, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N., 2010. Neural timing is linked to speech perception in noise. *J. Neurosci.* 30 (14), 4922–4926. <https://doi.org/10.1523/JNEUROSCI.0107-10.2010>.
- Baptista, N.I., Manfredi, M., Boggio, P.S., 2018. Medial prefrontal cortex stimulation modulates irony processing as indexed by the N400. *Soc. Neurosci.* 13 (4), 495–510. <https://doi.org/10.1080/17470919.2017.1356744>.
- Bastiaansen, M., Magyari, L., Hagoort, P., 2009. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J. Cognit. Neurosci.* 22 (7), 1333–1347. <https://doi.org/10.1162/jocn.2009.21283>.
- Biau, E., Soto-Faraco, S., 2013. Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124 (2), 143–152. <https://doi.org/10.1016/j.bandl.2012.10.008>.
- Biau, E., Fromont, L.A., Soto-Faraco, S., 2018. Beat gestures and syntactic parsing: an ERP study. *Language Learn.* 68, 102–126. <https://doi.org/10.1111/lang.12257>.
- Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., Soto-Faraco, S., 2015. Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex* 68, 76–85. <https://doi.org/10.1016/j.cortex.2014.11.018>.
- Biau, E., Fernández, L.M., Holle, H., Avila, C., Soto-Faraco, S., 2016. Hand gestures as visual prosody: BOLD responses to audio–visual alignment are modulated by the communicative nature of the stimuli. *Neuroimage* 132, 129–137. <https://doi.org/10.1016/j.neuroimage.2016.02.018>.
- Birch, S.L., Rayner, K., 1997. Linguistic focus affects eye movements during reading. *Memory Cognition* 25 (5), 653–660. <https://doi.org/10.3758/BF03211306>.
- Birch, S.L., Albrecht, J.E., Myers, J.L., 2000. Syntactic focusing structures influence discourse processing. *Discourse Processes* 30 (3), 285–304. https://doi.org/10.1207/S15326950dp3003_4.
- Bock, J.K., Mazzella, J.R., 1983. Intonational marking of given and new information: some consequences for comprehension. *Memory Cognition* 11 (1), 64–76. <https://doi.org/10.3758/BF03197663>.
- Boersma, P., & Weenink, D., 2016. Praat: Doing phonetics by computer (6.0.39) [Computer software].
- Bornkessel, I., Schlesewsky, M., Friederici, A.D., 2003. Contextual information modulates initial processes of syntactic integration: the role of inter-versus intrasentential predictions. *J. Exp. Psychol. Learn. Mem. Cogn.* 29 (5), 871–882. <https://doi.org/10.1037/0278-7393.29.5.871>.
- Borovsky, A., Elman, J.L., Kutas, M., 2012. Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learn. Dev.* 8 (3), 278–302. <https://doi.org/10.1080/15475441.2011.614893>.
- Bredart, S., Modolo, K., 1988. Moses strikes again: focalization effect on a semantic illusion. *Acta Psychol.* 67 (2), 135–144. [https://doi.org/10.1016/0001-6918\(88\)90009-1](https://doi.org/10.1016/0001-6918(88)90009-1).

- Cowles, H.W., Kluender, R., Kutas, M., Polinsky, M., 2007. Violations of information structure: an electrophysiological study of answers to wh-questions. *Brain Lang.* 102 (2), 228–242. <https://doi.org/10.1016/j.bandl.2007.04.004>.
- Cutler, A., Fodor, J.A., 1979. Semantic focus and sentence comprehension. *Cognition* 7 (1), 49–59. [https://doi.org/10.1016/0010-0277\(79\)90010-6](https://doi.org/10.1016/0010-0277(79)90010-6).
- Dahan, D., Tanenhaus, M.K., Chambers, C.G., 2002. Accent and reference resolution in spoken-language comprehension. *J. Memory Language* 47 (2), 292–314. [https://doi.org/10.1016/S0749-596X\(02\)00001-3](https://doi.org/10.1016/S0749-596X(02)00001-3).
- Davidson, D.J., Indefrey, P., 2007. An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Res.* 1158, 81–92. <https://doi.org/10.1016/j.brainres.2007.04.082>.
- de Marchena, A., Eigsti, I.M., 2010. Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. *Autism Res.* 3 (6), 311–322. <https://doi.org/10.1002/aur.159>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Dick, A.S., Goldin-Meadow, S., Hasson, U., Skipper, J.I., Small, S.L., 2009. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain Mapp.* 30 (11), 3509–3526. <https://doi.org/10.1002/hbm.20774>.
- Dimitrova, D., Stowe, L.A., Redeker, G., Hoeks, J.C.J., 2012. Less is not more: neural responses to missing and superfluous accents in context. *J. Cognit. Neurosci.* 24 (12), 2400–2418. https://doi.org/10.1162/jocn_a.00302.
- Dimitrova, D., Chu, M., Wang, L., Özyürek, A., Hagoort, P., 2016. Beat that word: how listeners integrate beat gesture and focus in multimodal speech discourse. *J. Cognit. Neurosci.* 28, 1255–1269. https://doi.org/10.1162/jocn_a.00963.
- Dinstein, I., Heeger, D.J., Lorenzi, L., Minshew, N.J., Malach, R., Behrmann, M., 2012. Unreliable evoked responses in autism. *Neuron* 75 (6), 981–991. <https://doi.org/10.1016/j.neuron.2012.07.026>.
- Dinstein, I., Heeger, D.J., Behrmann, M., 2015. Neural variability: friend or foe? *Trends Cognitive Sci.* 19 (6), 322–328. <https://doi.org/10.1016/j.tics.2015.04.005>.
- Drijvers, L., Özyürek, A., Jensen, O., 2018. Hearing and seeing meaning in noise: alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Hum. Brain Mapp.* 39 (5), 2075–2087.
- Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., Sejnowski, T.J., 2002. Single-trial variability in event-related BOLD signals. *NeuroImage* 15 (4), 823–835. <https://doi.org/10.1006/nimg.2001.1049>.
- Esteve-Gibert, N., Prieto, P., 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech, Language, Hearing Res.* 56 (3), 850–864. [https://doi.org/10.1044/1092-4388\(2012\)12-0049](https://doi.org/10.1044/1092-4388(2012)12-0049).
- Esteve-Gibert, N., Prieto, P., Pons, F., 2015. Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behav. Dev.* 38, 126–129. <https://doi.org/10.1016/j.infbeh.2014.12.016>.
- Garrett, D.D., Kovacevic, N., McIntosh, A.R., Grady, C.L., 2011. The importance of being variable. *J. Neurosci.* 31 (12), 4496–4503. <https://doi.org/10.1523/JNEUROSCI.5641-10.2011>.
- Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15 (4), 511. <https://doi.org/10.1038/nn.3063>.
- Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., Kircher, T., 2009. Neural integration of iconic and unrelated verbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30 (12), 3309–3324. <https://doi.org/10.1002/hbm.20753>.
- Habets, B., Kita, S., Shao, Z., Özyürek, A., Hagoort, P., 2011. The role of synchrony and ambiguity in speech-gesture integration during comprehension. *J. Cognit. Neurosci.* 23 (8), 1845–1854. <https://doi.org/10.1162/jocn.2010.21462>.
- Haigh, S.M., Heeger, D.J., Dinstein, I., Minshew, N., Behrmann, M., 2015. Cortical variability in the sensory-evoked response in autism. *J. Autism Dev. Disorders* 45 (5), 1176–1190. <https://doi.org/10.1007/s10803-014-2276-6>.
- Hancock, R., Pugh, K.R., Hoeft, F., 2017. Neural noise hypothesis of developmental dyslexia. *Trends Cognitive Sci.* 21 (6), 434–448. <https://doi.org/10.1016/j.tics.2017.03.008>.
- He, Y., Gebhardt, H., Steines, M., Sammer, G., Kircher, T., Nagels, A., Straube, B., 2015. The EEG and fMRI signatures of neural integration: an investigation of meaningful gestures and corresponding speech. *Neuropsychologia* 72, 27–42.
- He, Y., Steines, M., Sommer, J., Gebhardt, H., Nagels, A., Sammer, G., Kircher, T., Straube, B., 2018. Spatial-temporal dynamics of gesture-speech integration: a simultaneous EEG-fMRI study. *Brain Struct. Funct.* 223 (7), 3073–3089.
- He, Y., Luell, S., Muralikrishnan, R., Straube, B., Nagels, A., 2020. Gesture's body orientation modulates the N400 during semantic integration of gesture and visual sentence. *BioRxiv*.
- Holle, H., Gunter, T.C., 2007. The role of iconic gestures in speech disambiguation: ERP evidence. *J. Cognit. Neurosci.* 19 (7), 1175–1192. <https://doi.org/10.1162/jocn.2007.19.7.1175>.
- Holle, H., Gunter, T.C., Rüschmeyer, S.-A., Hennenlotter, A., Iacoboni, M., 2008. Neural correlates of the processing of co-speech gestures. *NeuroImage* 39 (4), 2010–2024. <https://doi.org/10.1016/j.neuroimage.2007.10.055>.
- Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A.D., Ward, J., Gunter, T.C., 2012. Gesture facilitates the syntactic analysis of speech. *Front. Psychol.* 74. <https://doi.org/10.3389/fpsyg.2012.00074>.
- Holler, J., Levinson, S.C., 2019. Multimodal language processing in human communication. *Trends Cognitive Sci.* 23 (8), 639–652.
- Hornickel, J., Kraus, N., 2013. Unstable representation of sound: a biological marker of dyslexia. *J. Neurosci.* 33 (8), 3500–3504. <https://doi.org/10.1523/JNEUROSCI.4205-12.2013>.
- Hubbard, A.L., Wilson, S.M., Callan, D.E., Dapretto, M., 2009. Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30 (3), 1028–1037. <https://doi.org/10.1002/hbm.20565>.
- Kelly, S.D., Kravitz, C., Hopkins, M., 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89 (1), 253–260. [https://doi.org/10.1016/S0093-934X\(03\)00335-3](https://doi.org/10.1016/S0093-934X(03)00335-3).
- Kelly, S.D., Manning, S.M., Rodak, S., 2008. Gesture gives a hand to language and learning: perspectives from cognitive neuroscience, developmental psychology and education. *Language Linguistics Compass* 2 (4), 569–588. <https://doi.org/10.1111/j.1749-818X.2008.00067.x>.
- Kelly, S.D., Creigh, P., Bartolotti, J., 2009. Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cognit. Neurosci.* 22 (4), 683–694. <https://doi.org/10.1162/jocn.2009.21254>.
- Kelly, S.D., Özyürek, A., Maris, E., 2010. Two sides of the same coin speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21 (2), 260–267. <https://doi.org/10.1177/0956797609357327>.
- Kielar, A., Meltzer, J.A., Moreno, S., Alain, C., Bialystok, E., 2014. Oscillatory responses to semantic and syntactic violations. *J. Cognit. Neurosci.* 26 (12), 2840–2862. https://doi.org/10.1162/jocn_a.00670.
- Krahmer, E., Swerts, M., 2007. The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57 (3), 396–414. <https://doi.org/10.1016/j.jml.2007.06.005>.
- Kushch, O., Prieto, P., 2016. The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. In: Barnes, J., Brugos, A., Shattuck-Hufnagel, S., Veilleux, N. (Eds.), *Proceedings of the 8th International Conference on Speech Prosody*. International Speech Communication Association, pp. 922–925.
- Kushch, O., Igualada, A., Prieto, P., 2018. Prominence in speech and gesture favour second language novel word learning. *Language, Cognition Neurosci.* 33 (8), 992–1004. <https://doi.org/10.1080/23273798.2018.1435894>.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Ladd, D.R., 2008. *Intonational Phonology*. Cambridge University Press.
- Lau, E.F., Holcomb, P.J., Kuperberg, G.R., 2013. Dissociating N400 effects of prediction from association in single-word contexts. *J. Cognit. Neurosci.* 25 (3), 484–502.
- Lenth, R., 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means (1.4.3) [Computer software]. <https://CRAN.R-project.org/package=emmeans>.
- Leonard, T., Cummins, F., 2011. The temporal relation between beat gestures and speech. *Language Cognitive Processes* 26 (10), 1457–1471. <https://doi.org/10.1080/01690965.2010.500218>.
- Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borrás-Comes, J., Prieto, P., 2018. Beat gestures help preschoolers recall and comprehend discourse information. *J. Exp. Child Psychol.* 172, 168–188. <https://doi.org/10.1016/j.jecp.2018.02.004>.
- Loehr, D.P., 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonol.* 3 (1), 71–89. <https://doi.org/10.1515/lp-2012-0006>.
- Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8. <https://doi.org/10.3389/fnhum.2014.00213>.
- Luo, Y., Zhang, Y., Feng, X., Zhou, X., 2010. Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience* 169 (2), 654–664. <https://doi.org/10.1016/j.neuroscience.2010.05.032>.
- Malins, J.G., Pugh, K.R., Buis, B., Frost, S.J., Hoeft, F., Landi, N., Mencl, W.E., Kurian, A., Staples, R., Molfese, P., 2018. Individual differences in reading skill are related to trial-by-trial neural activation variability in the reading network. 0907-17. *J. Neurosci.* 38 (12). <https://doi.org/10.1523/JNEUROSCI.0907-17.2018>.
- McNeill, D., 1985. So you think gestures are nonverbal? *Psychol. Rev.* 92 (3), 350–371. <https://doi.org/10.1037/0033-295X.92.3.350>.
- McNeill, D., 1992. *Hand and Mind*. University of Chicago Press.
- McNeill, D., 2005. *Gesture and thought*. University of Chicago Press.
- McNeill, D., 2006. *Gesture: a psycholinguistic approach*. *Encyclopedia Language Linguistics* 58–66.
- Milne, E., 2011. Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Front. Psychol.* 2 (51). <https://doi.org/10.3389/fpsyg.2011.00051>.
- Morett, L.M., Fraundorf, S.H., & McPartland, J. C. (under review). Eye see what you're saying: Contrastive use of beat gesture and pitch accent affects their online interpretation in spoken discourse.
- Morett, L.M., Roche, J.M., Fraundorf, S.H., & McPartland, J. C. (under review). Contrast is in the eye of the beholder: Infelicitous beat gesture increases cognitive load during online spoken discourse comprehension.
- Morett, L.M., Fraundorf, S.H., 2019. Listeners consider alternative speaker productions in discourse comprehension and memory: evidence from beat gesture and pitch accenting. *Memory Cognition* 47 (8), 1515–1530. <https://doi.org/10.3758/s13421-019-00945-1>.
- Morett, L.M., O'Hearn, K., Luna, B., Ghuman, A.S., 2016. Altered gesture and speech production in ASD detract from in-person communicative quality. *J. Autism Dev. Disord.* 46 (3), 998–1012. <https://doi.org/10.1007/s10803-015-2645-9>.
- Morrel-Samuels, P., Krauss, R.M., 1992. Word familiarity predicts temporal asynchrony of hand gestures and speech. *J. Exp. Psychol. Learn. Mem. Cogn.* 18 (3), 615–622. <https://doi.org/10.1037/0278-7393.18.3.615>.
- Nooteboom, S.G., Krut, J.G., 1987. Accents, focus distribution, and the perceived distribution of given and new information: an experiment. *J. Acoustical Soc. Am.* 82 (5), 1512–1524. <https://doi.org/10.1121/1.395195>.
- Nunez, P.L., Srinivasan, R., 2006. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, USA.
- Obermeier, C., Gunter, T.C., 2014. Multisensory integration: the case of a time window of

- gesture–speech integration. *J. Cognit. Neurosci.* 27 (2), 292–307. <https://doi.org/10.1162/jocn.a.00688>.
- Obermeier, C., Holle, H., Gunter, T.C., 2011. What iconic gesture fragments reveal about gesture–speech integration: when synchrony is lost, memory can help. *J. Cognit. Neurosci.* 23 (7), 1648–1663. <https://doi.org/10.1162/jocn.2010.21498>.
- Özyürek, A., Willems, R., Kita, S., Hagoort, P., 2007. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cognit. Neurosci.* 19 (4), 605–616. <https://doi.org/10.1162/jocn.2007.19.4.605>.
- Power, A.J., Mead, N., Barnes, L., Goswami, U., 2013. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 7. <https://doi.org/10.3389/fnhum.2013.00777>.
- Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., Vilà-Giménez, I., 2018. Deconstructing beat gestures: a labelling proposal. In: 9th International Conference on Speech Prosody 2018, pp. 201–205. <https://doi.org/10.21437/SpeechProsody.2018-41>.
- Regel, S., Gunter, T.C., Friederici, A.D., 2011. Isn't it ironic? An electrophysiological exploration of figurative language processing. *J. Cognitive Neurosci.* 23 (2), 277–293. <https://doi.org/10.1162/jocn.2010.21411>.
- Roustan, B., Dohen, M., 2010a. Co-production of contrastive prosodic focus and manual gestures: temporal coordination and effects on the acoustic and articulatory correlates of focus. In: Hasegawa-Johnson, M. (Ed.), *Speech Prosody 2010*. International Speech Communication Association.
- Roustan, B., Dohen, M., 2010b. Gesture and speech coordination: the influence of the relationship between manual gesture and speech. In: Hasegawa-Johnson, M. (Ed.), *Speech Prosody 2010*. International Speech Communication Association.
- Rusiewicz, H.L., Shaiman, S., Iverson, J.M., Szuminsky, N., 2013. Effects of prosody and position on the timing of deictic gestures. *J. Speech, Language, Hearing Res.* 56 (2), 458–470. [https://doi.org/10.1044/1092-4388\(2012/11-0283\)](https://doi.org/10.1044/1092-4388(2012/11-0283)).
- Rusiewicz, H.L., Shaiman, S., Iverson, J.M., Szuminsky, N., 2014. Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Commun.* 57, 283–300. <https://doi.org/10.1016/j.specom.2013.06.004>.
- Schumacher, P.B., Baumann, S., 2010. Pitch accent type affects the N400 during referential processing. *NeuroReport* 21 (9), 618–622. <https://doi.org/10.1097/WNR.0b013e328339874a>.
- Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J.D., Dehaene, S., 2015. Cortical activity is more stable when sensory stimuli are consciously perceived. *Proc. Natl. Acad. Sci.* 112 (16), 2083–2092. <https://doi.org/10.1073/pnas.1418730112>.
- Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., Demuth, K., 2016. Non-referential gestures in adult and child speech: are they prosodic? In: Barnes, J., Brugos, A., Shattuck-Hufnagel, S., Veilleux, N. (Eds.), *Proceedings of the 8th International Conference on Speech Prosody*. International Speech Communication Association, pp. 836–839.
- Shattuck-Hufnagel, S., Ren, A., 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Front. Psychol.* 9. <https://doi.org/10.3389/fpsyg.2018.01514>.
- Silverman, L.B., Bennetto, L., Campana, E., Tanenhaus, M.K., 2010. Speech-and-gesture integration in high functioning autism. *Cognition* 115 (3), 380–393. <https://doi.org/10.1016/j.cognition.2010.01.002>.
- Simon, D.M., Wallace, M.T., 2016. Dysfunction of sensory oscillations in autism spectrum disorder. *Neurosci. Biobehav. Rev.* 68, 848–861. <https://doi.org/j.neurobiorev.2016.07.016>.
- Skipper, J.I., 2014. Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. B: Biol. Sci.* 369 (1651), 1–19. <https://doi.org/10.1098/rstb.2013.0297>.
- Skipper, J.I., Goldin-Meadow, S., Nusbaum, H.C., Small, S.L., 2007. Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101 (3), 260–277. <https://doi.org/10.1016/j.bandl.2007.02.008>.
- Ter Bekke, M., Drijvers, L., & Holler, J., 2020. The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech.
- Van Berkum, J.J.A., Hagoort, P., Brown, C.M., 1999. Semantic integration in sentences and discourse: evidence from the N400. *J. Cognitive Neurosci.* 11 (6), 657–671. <https://doi.org/10.1162/089992999563724>.
- Wang, L., Chu, M., 2013. The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia* 51 (13), 2847–2855. <https://doi.org/10.1016/j.neuropsychologia.2013.09.027>.
- Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., Hagoort, P., Bastiaansen, M., 2012. Beta oscillations relate to the N400m during language comprehension. *Hum. Brain Mapp.* 33 (12), 2898–2912. <https://doi.org/10.1002/hbm.21410>.
- Wang, L., Li, X., Yang, Y., 2014. A review on the cognitive function of information structure during language comprehension. *Cogn. Neurodyn.* 8 (5), 353–361. <https://doi.org/10.1007/s11571-014-9305-1>.
- Willems, R.M., Özyürek, A., Hagoort, P., 2007. When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17 (10), 2322–2333. <https://doi.org/10.1093/cercor/bhl141>.
- Willems, R.M., Özyürek, A., Hagoort, P., 2009. Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage* 47 (4), 1992–2004. <https://doi.org/10.1016/j.neuroimage.2009.05.066>.
- Wu, Y.C., Coulson, S., 2005. Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42 (6), 654–667. <https://doi.org/10.1111/j.1469-8986.2005.00356.x>.
- Wu, Y.C., Coulson, S., 2007. How iconic gestures enhance communication: an ERP study. *Brain Lang.* 101 (3), 234–245. <https://doi.org/10.1016/j.bandl.2006.12.003>.
- Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J.A., Poldrack, R.A., 2010. Greater neural pattern similarity across repetitions is associated with better memory. *Science* 330 (6000), 97–101. <https://doi.org/10.1126/science.1193125>.
- Zellin, M., Pannekamp, A., Toepel, U., van der Meer, E., 2011. In the eye of the listener: pupil dilation elucidates discourse processing. *Int. J. Psychophysiol.* 81 (3), 133–141. <https://doi.org/10.1016/j.ijpsycho.2011.05.009>.
- Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J.I., Vigliocco, G., 2020. More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. *BioRxiv*.