

# Beat that Word: How Listeners Integrate Beat Gesture and Focus in Multimodal Speech Discourse

Diana Dimitrova<sup>1,2</sup>, Mingyuan Chu<sup>3,4</sup>, Lin Wang<sup>5</sup>, Asli Özyürek<sup>1,3</sup>, and Peter Hagoort<sup>1,3</sup>

## Abstract

■ Communication is facilitated when listeners allocate their attention to important information (focus) in the message, a process called “information structure.” Linguistic cues like the preceding context and pitch accent help listeners to identify focused information. In multimodal communication, relevant information can be emphasized by nonverbal cues like beat gestures, which represent rhythmic nonmeaningful hand movements. Recent studies have found that linguistic and nonverbal attention cues are integrated independently in single sentences. However, it is possible that these two cues interact when information is embedded in context, because context allows listeners to predict what information is important. In an ERP study, we tested this hypothesis and asked listeners to view videos capturing a dialogue. In the critical sentence, focused and nonfocused

words were accompanied by beat gestures, grooming hand movements, or no gestures. ERP results showed that focused words are processed more attentively than nonfocused words as reflected in an N1 and P300 component. Hand movements also captured attention and elicited a P300 component. Importantly, beat gesture and focus interacted in a late time window of 600–900 msec relative to target word onset, giving rise to a late positivity when nonfocused words were accompanied by beat gestures. Our results show that listeners integrate beat gesture with the focus of the message and that integration costs arise when beat gesture falls on nonfocused information. This suggests that beat gestures fulfill a unique focusing function in multimodal discourse processing and that they have to be integrated with the information structure of the message. ■

## INTRODUCTION

In conversation, speech partners exchange large amounts of information in a limited amount of time. For this communication process to be successful, listeners need to correctly identify what pieces of information are important and to pay more attention to them than to pieces of less important information. A listener’s search for relevant information is facilitated when the speaker follows the principles of information structure and highlights important information (focus) but leaves less important information (nonfocus) unmarked (for a review, see Wang, Li, & Yang, 2014; Arnold, Kaiser, Kahn, & Kim, 2013). In this article, we define “focus” as the element of the utterance that contributes new, nonderivable, or contrastive information and that receives the most prominent pitch accent. To highlight focus, speakers can use various linguistic focusing cues, such as pitch accents (Birch & Clifton, 1995; Nooteboom & Kruyt, 1987; Cutler & Fodor, 1979), focus particles (Sudhoff, 2010; Jacobs, 1986), or syntactic constructions (Birch, Albrecht, & Myers, 2000) as well as nonverbal focusing cues like beat gestures (Biau & Soto-Faraco, 2013; Wang & Chu, 2013; Holle et al., 2012; McNeill, 1992). To date, the role of linguistic and nonverbal focus-

ing cues on speech processing has been studied separately. It has been shown that speech processing is facilitated by both pitch accent (Wang, Bastiaansen, Yang, & Hagoort, 2011; Heim & Alter, 2006; Hruska & Alter, 2004) and gesture (Kelly, Manning, & Rodak, 2008). However, natural communication is multimodal, and in real-life conversations, listeners need to identify important information across multiple modalities. Surprisingly, few studies have addressed the integration of beat gesture and speech during speech comprehension (Biau & Soto-Faraco, 2013, 2015; Wang & Chu, 2013; Holle et al., 2012; Leonard & Cummins, 2011; Roustan & Dohen, 2010; Hubbard, Wilson, Callan, & Dapretto, 2009; Krahmer & Swerts, 2007). Findings of these studies suggest that beat gesture and accentuation have independent effects on the processing of information in single sentences. In natural communication, people often process sentences embedded in context. Therefore, this ERP study addresses the processing of sentences in context: Do beat gesture and accented focus interact in dialogue, where listeners can unambiguously predict the focus of the upcoming message?

## Linguistic Cues for Focus Capture Listeners’ Attention during Speech Comprehension

In the spoken language modality, focused information is usually highlighted by pitch accent. When a nonfocused

<sup>1</sup>Radboud University Nijmegen, <sup>2</sup>University of Cologne, <sup>3</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, <sup>4</sup>University of Aberdeen, <sup>5</sup>Chinese Academy of Sciences

word is accented, listeners consider the sentence unacceptable (Nooteboom & Kruyt, 1987). Mismatches between focus and pitch accent cause processing difficulties (for a review, see Dimitrova, Stowe, Redeker, & Hoeks, 2012). Importantly, for this study, pitch accent draws the listener's attention to the accented word (Birch & Clifton, 1995; Cutler & Fodor, 1979). In EEG studies, this increased attention to accented information is reflected in an early anterior positivity (Dimitrova et al., 2012; Heim & Alter, 2006). Similar early positivities have been found in the written modality and have been attributed to the P3b component for the attentive processing of focused information (for a review, see Nieuwenhuis, Aston-Jones, & Cohen, 2005; Picton, 1992). For example, the it-cleft construction in English (Cowles, Kluender, Kutas, & Polinsky, 2007) and word order alternations in German (Bornkessel, Schlesewsky, & Friederici, 2003) both trigger early anterior P3b-like positivities, which have been related to the integration of focused information in the discourse. The early positivity is independent of the processing modality and likely reflects the attentive processing of focus. The enhanced attention to focus leads in turn to more elaborate processing: Listeners detect semantic and syntactic violations on accented focus more often than those on unaccented focused elements (Wang et al., 2011; Wang, Hagoort, & Yang, 2009). Further evidence for the link of information structure and attention comes from an fMRI study by Kristensen, Wang, Petersson, and Hagoort (2013) who found that accented information recruits attention areas in the brain.

### **Beat Gestures Modulate Online Speech Processing and Serve as Attention Cues**

In multimodal communication, speech is accompanied by gestures in the nonverbal modality, and most studies have focused on the semantic integration of speech and iconic gestures, which mimic the meaning of the words they co-occur with. When the meaning of iconic gestures is incongruent with the meaning of the words they accompany, ERP studies have found a modulation of the N400 component for semantic processing (Holle & Gunter, 2007; Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2005, 2007; Kelly, Kravitz, & Hopkins, 2004). In contrast, very little is known about the role of beat gestures in speech processing. Unlike iconic gestures, beat gestures do not carry any semantic meaning. Beat gestures are rapid rhythmic movements of the hand, which place emphasis on the words they accompany. Theoretically, beat gestures have been claimed to accompany new and contrastive information and to signify its importance (McNeill, 1992).

Recent ERP evidence suggests that beat gestures facilitate speech processing at various linguistic levels and serve as attention cues during the comprehension of natural speech (Biau & Soto-Faraco, 2013). In the ERP signal, words accompanied by beat gestures give rise to two pos-

itive effects, which have been attributed to early sensory processing (before 100 msec after target onset), as well as to the phonological analysis of the target word (P2 component around 200 msec after target onset). Biau and Soto-Faraco (2013) suggested that beat gestures serve as highlighters, which modulate the attentional state of the listener and guide their attention to relevant information in the speech signal. However, the study focused on the role of beat gestures without examining the acoustic aspects of words accompanied by beat gestures. As a result, the outcomes did not provide insights into the possible interplay between beat gestures and speech and, in particular, focus. Another ERP study demonstrated that beat gestures facilitate the disambiguation of syntactically ambiguous sentences in German (Holle et al., 2012). In object-relative clauses like "that the men the woman greeted," listeners need to disambiguate the second noun phrase "woman" toward a non-preferred subject-reading; this process gives rise to a P600 effect. Holle et al. (2012) found that the P600 disappeared when the second noun phrase of the object-relative clause (here, "woman") was accompanied by a beat gesture, suggesting that beat gestures facilitate syntactic processing. Importantly, when the same noun phrase was accompanied by an unrelated visual emphasis cue (moving red dot on the computer screen), object-relative clauses still triggered a P600 effect. The authors concluded that the modulation of the P600 by beat gestures was not because of pure visual attention but rather because of their communicative function. An ERP study by Wang and Chu (2013) found that beat gestures and pitch accents both facilitate semantic processing in single Dutch sentences. Words accompanied by either a beat gesture or a pitch accent showed a reduced N400 component than words that were not emphasized. The effects of beat gesture and accent were independent, presumably because of the use of sentences without context where any word could be emphasized by either cue. In summary, the ERP evidence suggests that beat gestures are integrated with phonological, semantic, and syntactic aspects of the speech signal; however, their effects are independent of the effects of pitch accent. Importantly, none of these studies addressed whether and how beat gestures are integrated with the information structure of the message.

### **Information Structure Determines what Information Is Focused on in Discourse**

In natural situations, sentences are embedded in a discourse context, which sets up the common ground and allows listeners to anticipate new and important information. Hence, based on the context, listeners predict what information will be highlighted by a focusing cue like pitch accent and are sensitive to mismatches between focus and accent (e.g., Nooteboom & Kruyt, 1987). The question arises whether listeners also anticipate the

marking of focused information by a beat gesture. If the theoretical claim that beat gestures signal the importance of new/contrastive information (McNeill, 1992) is correct, we would expect listeners to form expectations about which information the speaker can emphasize with a beat gesture, namely, the focused information. This leads to the prediction that listeners should integrate beat gestures with focused elements more easily than beat gestures with nonfocused elements, because the latter entail an emphasis mismatch and should elicit processing costs. In this study, we explicitly test this theoretical claim experimentally.

Context is crucial for the identification of focus: In context, new information is usually focused, whereas old information remains in the background (i.e., non-focused). The newness difference between focus and non-focus affects semantic processing costs (Schumacher & Baumann, 2010). In this study, we decided to disentangle the effects of focusing attention, which elicit a P300, and the effects of semantic processing, which modulate the N400, because they might temporarily overlap. To this end, we preactivated the meaning of target words by introducing them in a question context such as “Did the student buy the books or the magazines via Amazon? – He bought the BOOKS via Amazon” (accented focus in capitals). Both the focused and accented information “books” and the nonfocused unaccented information “Amazon” in the answer were semantically activated by the question context. Importantly, only the focused word “books” in the answer formed a contrastive relationship with a word in the preceding context, namely, “magazines.” By embedding sentences in context, we go beyond previous studies on single sentences and experimentally test the theoretical claims that beat gesture serves as a cue for focused information in context.

### **Selecting an Appropriate Control Hand Movement Condition to Beat Gesture Is Important**

The current study design has several merits over previous research and allows us to unambiguously investigate the interplay of beat gesture and information structure. First and foremost, we carefully selected an appropriate control condition to the beat gesture condition. The appropriate control condition allows us to examine if beat gesture has any specific function other than enhancing visual attention because of making a hand movement. Previous ERP studies do not provide unambiguous evidence that a particular ERP effect is specific for beat gesture but not for other types of hand movement. For example, Biau and Soto-Faraco (2013) did not use a control gesture, and the attention effect of beat gesture they reported could be the pure result of visual emphasis caused by moving the hand, rather than the result of a specific function of beat gesture. Holle et al. (2012) added a moving dot on the screen as a control condition to beat gesture. Although the dot mimics the trajectory of

the gesture movement, it is very dissimilar from trials showing a gesturing person. Wang and Chu (2013) chose too strict a control gesture, which was very tightly matched to their beat gesture (a vertical palm movement for beat gesture vs. a horizontal palm movement for control gesture). The authors reported similar effects for the processing of beat and control gestures, which could reflect that listeners processed them as too similar. In this study, we sought a solution for the control condition problem. After careful consideration, we chose grooming hand movements (i.e., adjusting one’s shirt) as a control condition to beat gestures. This is because, similar to beat gestures, grooming hand movements occur in natural conversation and induce a prominent change in the visual scene. Both hand movements are unrelated to the semantic content of the message. Unlike beat gestures, grooming hand movements are not used for emphasis or with any linguistic or rhythmic function and are not integrated with the content of speech.

### **Selecting Appropriate Beat Gestures from Natural Stimuli**

A second advantage of this study is the careful preselection of naturally valid beat gestures. Before the ERP experiment, we performed behavioral pretests to determine which types of beat gestures Dutch speakers use in natural situations. After we identified the most common forms of beat gestures, we tested further parameters such as gesture speed and gesture–speech alignment. We asked participants to rate the naturalness and emphasis of beat gestures and grooming hand movements and made sure that there is no functional overlap between the two hand movements. For the ERP experiment, we selected three types of beat gesture and grooming hand movements. Our pretests (see Methods) show that beat gestures were rated as having a natural shape and as being more emphatic than grooming hand movements. Our study thus uses valid beat and grooming hand movements that are naturally used by Dutch speakers and recognized as such by Dutch listeners. On the basis of multiple pretests, we selected natural gesture–speech synchronization where beat gestures start 520 msec before the onset of the target word (Pretest 4).

### **The Present Study**


This study investigates whether beat gestures serve as nonverbal emphasis cues that draw attention and whether listeners integrate beat gestures with focused information in speech. To disassociate sensory from integration effects related to gesture processing, we compare beat gestures with grooming hand movements. To this end, we embedded sentences like “She received an email from the teacher” in dialogues where a preceding question determined the information structure of the answer sentence (see Table 1). Focused elements were always accented,

**Table 1.** Examples of Experimental Conditions

*Focus condition (F)*

Did she receive an email or a letter from the teacher?


She received an **EMAIL** from the teacher.

- 
- No gesture (NG)
  - Beat gesture (BG)
  - Grooming hand movement (GG)

*Nonfocus condition (NF)*

Did she receive an email from the teacher or from the rector?

She received an **email** from the TEACHER.

- 
- No gesture (NG)
  - Beat gesture (BG)
  - Grooming hand movement (GG)

There were two focus conditions (focus vs. nonfocus), and the target word was always the direct object. Focused targets were always accented (in capitals), whereas nonfocused targets were always unaccented. Each focus and nonfocus condition was combined with one of the three hand movement conditions (no gesture, beat gesture, or grooming hand movement).

Target words are presented in **bold**.

and nonfocused elements were always unaccented; both could be accompanied by a beat gesture, a grooming hand movement, or no gesture. Combining focused and non-focused items with a beat gesture allowed us to test if listeners expect beat gestures to occur on focused constituents. Previous studies have shown that placing information in focus increases attention (Dimitrova et al., 2012; Cowles et al., 2007; Bornkessel et al., 2003). This led to the prediction that, in trials without a gesture, focused elements would capture attention and elicit a P300-like positivity (Dimitrova et al., 2012; Cowles et al., 2007; Bornkessel et al., 2003). Because beat gestures also capture attention (Biau & Soto-Faraco, 2013; Wang & Chu, 2013; Holle et al., 2012), we hypothesized that they would modulate the early sensory stages of visual processing and give rise to an N1 and/or a P2/P300 component. Third, we hypothesized that the processing of beat gestures may interact with information structure. That is, if beat gestures serve as focusing cues, listeners would expect them to accompany focused items. If beat gestures accompany nonfocused items instead, listeners should perceive an emphasis mismatch where less relevant information is highlighted; this should cause processing difficulties. In the ERP signal, general difficulties in integrating beat gesture with nonfocused information might take the form of an N400 effect, which has been shown to reflect diffi-

culty in integrating semantic meaning in general (Kutas & Federmeier, 2011). For example, it has been shown that the N400 is elicited when less salient information (non-focus) is highlighted by a pitch accent (e.g., Bögels, Schriefers, Vonk, & Chwilla, 2010; Hruska & Alter, 2004). Moreover, the difficulty to integrate beats with nonfocused information could also take the form of a late positivity, which has been reported for the integration of incongruently accented focus in context (Dimitrova et al., 2012; Schumacher & Baumann, 2010). Finally, we hypothesized that grooming would only elicit sensory effects upon the execution of the hand movement but will otherwise be processed similarly to trials with no gestures and will not be integrated with speech.

## METHODS

### Participants

Thirty healthy Dutch native speakers (18 women; age = 18–32 years, mean = 21.8 years), with normal or corrected-to-normal vision and no hearing, language, or neurological problems, were paid to participate in the EEG experiment. All participants filled out a written informed consent form in accordance with the Declaration of Helsinki and were debriefed after the experiment. The data of two participants were excluded because of an insufficient number of trials per condition (more than 50% trials were rejected in the independent component analysis). Statistical analysis was performed on the remaining 28 participants.

### Materials

We constructed 180 dialogue sets consisting of a question and an answer (see Table 1). In target trials, a yes/no question introduced a choice between two contrastive objects (“Did she receive a letter or an email from the teacher?”). In the answer, one of the contrasted elements was selected and served as the focus of the message. Focus was always accented (in capitals: “She received an EMAIL from the teacher”). The target word was always the direct object in the answer (“email”). Depending on the preceding context, targets could occur in two conditions: accented focused targets or unaccented nonfocused targets (see Table 1). Each target sentence occurred in one of three hand movement conditions: no gesture, beat gesture, or grooming hand movement.

The audio and video files of the materials were recorded separately and combined at a later stage. First, two speaker pairs (i.e., four speakers) each recorded half of the stimuli. In one speaker pair, a male speaker asked a question, and a female speaker replied. In the other speaker pair, a female speaker asked a question, and a male speaker replied. All stimuli were recorded in a soundproof booth in Adobe Audition and were digitalized at a sample rate of 44.1 kHz. All recordings were

segmented, and their amplitude was normalized to a default root mean square volume level of  $-12$  dB. The onset of all target words and prepositional objects was marked in Praat (Boersma & Weenink, 2014). The videos for the hand movement conditions were filmed by two actor pairs (two men and two women) using a digital camera (JVC-GY HM100E) with 40 msec per frame. The actors were recorded in a standing position, and their faces were blurred after recording (see Figure 1). Audio and video files were combined and edited in Adobe Premiere Pro CS5.5. Gesture characteristics and the audio–video combinations were pretested behaviorally (see Gesture pretests).

Questions were combined with videos displaying both speech partners standing still and facing each other, and answers displayed only the speaker as he or she executed a beat gesture or a grooming hand movement or stood still. To increase stimulus variability, each speaker produced three types of beat and grooming hand movements (see Figure 1). All gestures started 520 msec before the onset of the target word (Pretest 4, Figure 2) and were produced without a holding phase after the apex (Pretest 3).

### Experimental Design

A full factorial within-participant design was created with two factors: Hand movement (beat gesture: BG, grooming hand movement: GG, no gesture: NG) and Focus (focus: F, nonfocus: NF). Stimuli consisted of 180 dialogue items (2 speaker pairs  $\times$  90 dialogue items) and

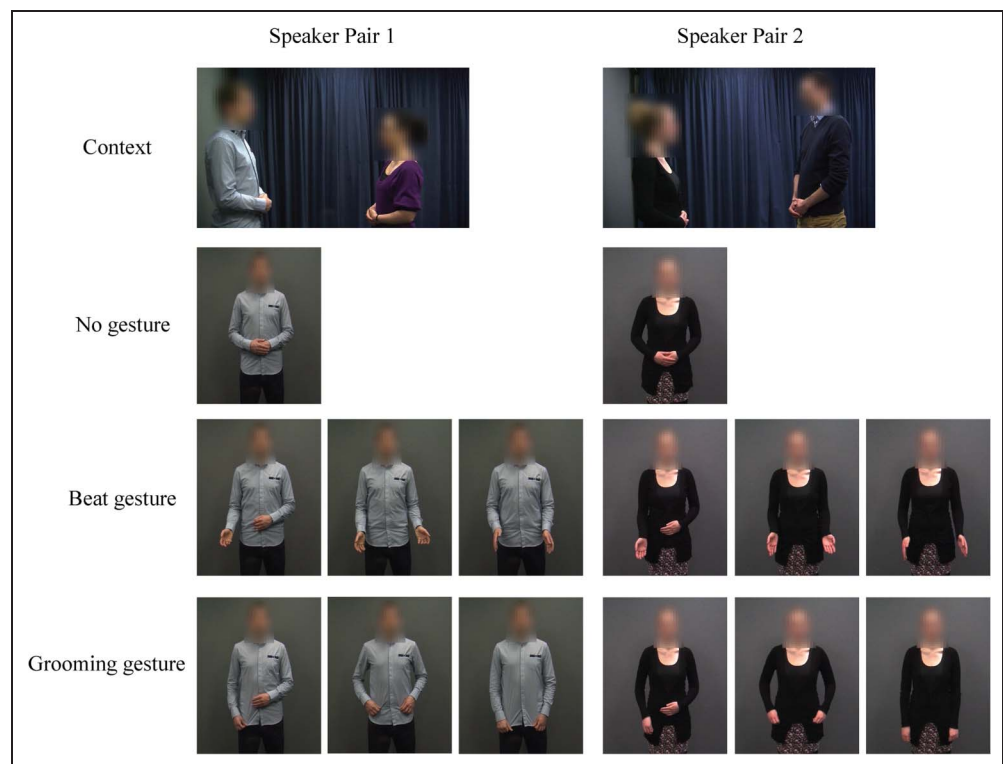
were distributed across six experimental conditions (2 focus types (F, NF)  $\times$  3 hand movement types (BG, GG, NG)  $\times$  30 dialogues per condition). The beat gesture and the grooming hand movement condition each consisted of three different forms (see Figure 1). To create the target stimuli, we combined the recordings of all 180 dialogues with the seven hand movement types (three beat gestures, three grooming hand movements, and one no gesture), which resulted in 1260 videos. To reduce the predictability of gesture position in the answer, 120 filler dialogues were added (60 yes/no questions like the experimental items and 60 yes/no questions of a different type: “Did you know that Peter bought a book? I think JAN bought a book.”). In the fillers, gestures occurred on the subject (“Jan or I”) and on the prepositional object (“book”). All 120 filler dialogue audio files were combined with the seven hand movement types, resulting in 840 filler videos. Overall, we created 2100 video files. Finally, target and filler stimuli were distributed across 12 experimental lists of 300 dialogues each, according to a Latin Square procedure such that no participant watched more than one version of each dialogue. Videos were presented in a pseudorandom order with no more than three consecutive repetitions of the same condition.

### Gesture Pretests

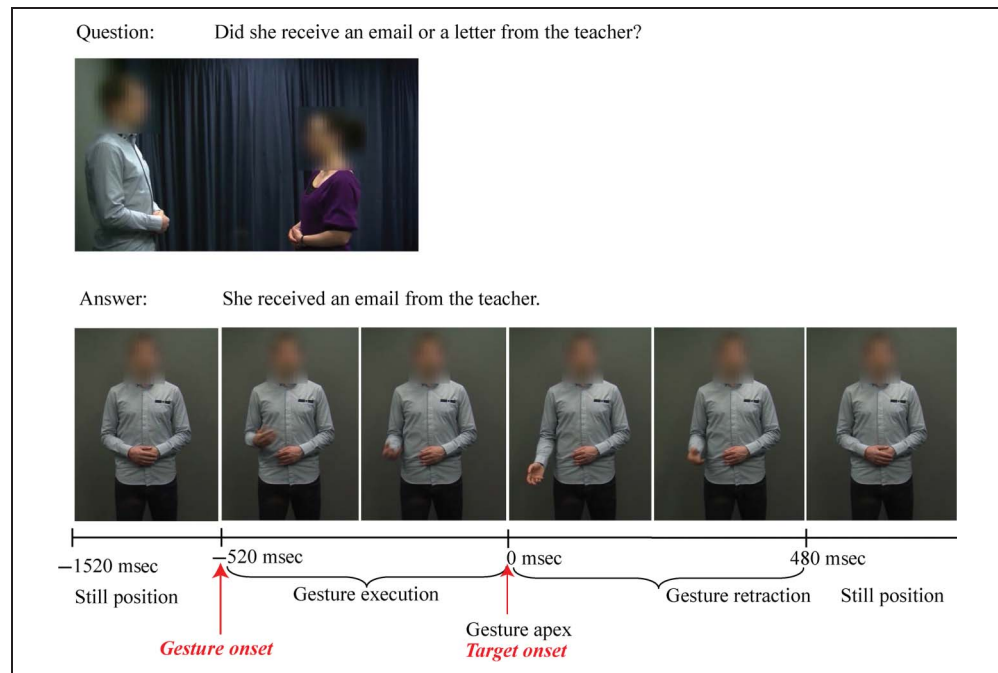
#### *Pretest 1: Selecting Natural Beat Gestures*

The goal of Pretest 1 was to elicit spontaneous beat gestures from native Dutch speakers. Thirty-three Dutch native speakers (age range = 18–29 years, mean = 22 years;

**Figure 1.** Hand movement types used in the ERP experiment. The figure displays frames extracted from the videos to the corresponding hand movement conditions. Two speaker pairs were used. During the presentation of the question context, a still video frame was shown with the two speech partners facing each other (here, “context”). In the no gesture condition, the actors did not move. In the beat gesture and the grooming hand movement conditions, the actors performed three types of hand movements with either the right hand or two hands. The presented frames depict the apex of the gesture, which is the maximal extension of the hand. The male and female actors performed the same three types of beat gestures and grooming hand movements.



**Figure 2.** Trial structure. Each trial starts with a question, which represents a still frame of the two speech partners facing each other for the entire question duration. Then, the person who gives the answer is shown facing the camera. After a still frame (1000 msec), the answering person gives the answer and simultaneously executes a hand movement.



15 men) gave informed written consent and were paid for participation. Four participants were excluded because of technical problems (no sound recording).

Participants listened to prerecorded questions like “Did mum buy tomatoes or onions at the market?” while they saw the question on a computer screen. After the question, a word appeared on the screen (“tomatoes”), and the participant was asked to answer the question using this word in a full sentence (“Mum bought tomatoes at the market.”). To make the experimental setting more natural, the experimenter was present in the testing room, and participants were asked to direct their answers to her. The experiment consisted of two sessions. In the first session, participants were asked to simply answer the questions without any explicit instruction to gesture. In the second session, participants were explicitly asked to perform a hand gesture to emphasize the important information in their answer. No instructions regarding the type of gesture or its alignment with speech were given. All recordings were inspected visually by two independent coders. They identified 66 beat gestures that were not semantically related to the accompanying speech and that were produced to emphasize the target word “tomatoes.” Video clips of participants’ answers containing those 66 gestures were cut out, and these clips were used in Pretest 2.

#### Pretest 2: Selecting the Most Appropriate Beat Gestures

In Pretest 2, we investigated how participants interpret beat gestures from Pretest 1 and whether they perceive them as natural and emphatic. Eighteen Dutch native speakers (age range = 18–28 years, mean = 22 years;

two men) participated after giving informed written consent and were paid for participation. The experiment consisted of two parts. In Part 1, participants viewed 66 video clips from Pretest 1. In each trial, participants first heard a question and then saw a video of a speaker answering the question and making a beat gesture. The face of each speaker was covered. Participants judged whether question and answer matched on a 7-point scale (1 = *no match*, 7 = *very good match*) and then wrote down why in their opinion the speaker gestured. In Part 2, participants viewed the same 66 video clips and judged (1) to what extent the gesture was used to highlight information (1 = *no emphasis at all*, 7 = *very strong emphasis*) and (2) how natural the hand gesture was (1 = *not natural at all*, 7 = *very natural*). Participants rated the question–answer pairs as highly matching (mean rating = 6.64, *SD* = 0.76). The mean emphasis rating was 4.46 (*SD* = 0.56), and the mean naturalness rating was 4.55 (*SD* = 0.72). The five beat gestures with the highest emphasis and naturalness ratings were recorded by a male actor and a female actor and were used in Pretest 3.

#### Pretest 3: Naturalness and Emphasis of Beat and Grooming Hand Movements

A male speaker and a female speaker acted the five best-rated beat gestures from Pretest 2. In addition, the actors were asked to perform five grooming hand movements, which closely matched the beat gestures’ kinematic trajectory, distance, and speed. For example, the actors scratched themselves and pulled or adjusted their shirt (see Figure 1). Grooming hand movements served as a control condition in the experiment. All hand movements

were aligned with speech in the same way: The maximal extension of the gesture (the gesture apex) occurred at the acoustic onset of the target word. The goal of Pretest 3 was to examine the naturalness and emphasis of beat and grooming hand movements. In addition, we tested whether participants perceive gestures as more natural if the hand stays for a prolonged period in the apex position (holding) versus if the hand is immediately retracted to the original position (no holding). All beat gesture and grooming hand movements were acted in two versions: (1) “no-hold” version where the hand was immediately retracted to the resting position after completing the stroke and (2) “with-hold” version where the hand was held still for 480 msec between the completion of the stroke and the initiation of the retraction.

Sixteen Dutch native speakers (age range = 18–26 years, mean = 22 years; eight men) gave informed written consent and were paid for participation. Participants viewed 160 video clips showing answers of the dialogue stimuli. Participants rated the naturalness and emphasis of both no-hold and with-hold versions of each gesture and the naturalness of their form and speed. In Part 1, participants indicated how natural each gesture was on a 7-point scale (1 = *not natural at all*, 7 = *very natural*) and whether the speaker used the gesture to emphasize information (1 = *no emphasis at all*, 7 = *very strong emphasis*). In Part 2, participants viewed the same 160 video clips again and judged how natural the form and speed of each gesture was (1 = *not natural at all*, 7 = *very natural*). Ratings of overall naturalness, overall emphasis, speed naturalness, and form naturalness were submitted to four ANOVA tests, with gesture length (no-hold vs. with-hold) and hand movement type (beat gesture vs. grooming) as independent variables. Beat gestures were rated as more natural, more emphatic, and with more natural form and speed than grooming hand movements (all  $p$ s < .01). No-hold gestures were rated as more natural in speed than with-hold gestures ( $p$  < .01). There were no interactions between gesture length (hold vs. no-hold) and hand movement type (beat vs. grooming) for any of the four ratings.

On the basis of the results, we excluded the beat gesture with the lowest scores in all four ratings and the corresponding grooming hand movement. We selected the no-hold version of all remaining four beat and grooming hand movements for Pretest 4. Furthermore, although the beat gestures and the grooming hand movements we used in this pretest are all taken from natural speech, beat gestures were rated as more natural than grooming hand movements. We hypothesized that the difference in naturalness might arise from the precise alignment of grooming hand movements and speech, because speakers rarely synchronize their grooming hand movements with speech in natural conversation. Therefore, we decided to test how participants perceive beats and grooming without interference from the speech signal.

#### *Pretest 4: Gesture–Speech Timing*

The goal of Pretest 4 was twofold. First, we wanted to test how participants rated beat gestures and grooming hand movements without any influence from speech. To this end, videos were played without sound. Second, we wanted to determine the alignment of gesture relative to the onset of the target word. To this end, we constructed videos where gestures started at 0, 200, 320, and 520 msec before the onset of the target word. Because it took 520 msec for the beat gestures and the grooming hand movements to reach their apex, they were aligned at –520, –320, –200, or 0 msec relative to the onset of the target word. Sixteen Dutch native speakers (age range = 18–27 years, mean = 21 years; six men) gave informed written consent and were paid for participation. In Part 1, participants viewed 32 silent videos, consisting of four beat and four grooming hand movements acted by both speakers twice [(4 beat gestures + 4 grooming hand movements) × 2 speaker pairs × 2 views]. As in Pretest 3, participants rated the overall naturalness, overall emphasis, and the naturalness of form and speed of all hand movements on a 7-point scale. Participants rated beat gesture and grooming hand movements as having a similar overall naturalness and form naturalness. Beat gestures were rated as overall more emphatic than grooming hand movements ( $p$  < .01). The speed of beat gestures was rated as more natural than the speed of grooming hand movements ( $p$  < .01).

In Part 2, participants viewed 128 video clips with the corresponding audio file where the gesture–speech alignment was manipulated. Participants rated the naturalness of alignment on a 7-point scale (1 = *not natural at all*, 7 = *very natural*). The most natural alignment for beat and grooming hand movements was when they were executed at 520 msec before the onset of the critical word. With this alignment, the apex (the most forceful part of the hand movement) was aligned with the onset of the critical word. We selected this alignment in the stimuli for the ERP experiment.

#### *Summary of Pretests*

In summary, based on the results of four pretests, we selected three top-rated beat gestures and three corresponding control grooming hand movements, which matched their kinematic trajectory, distance, and speed. All gestures were executed with either the right hand or two hands. Hand movements started 520 msec before the target word onset, reached the apex at target onset, and were retracted for 520 msec to the original hand position (Figure 2). Pretests focusing on these selected beat gestures and grooming hand movements showed that, (1) when viewed in isolation, beat gestures ( $M = 4.99$ ,  $SD = 0.89$ ) and grooming hand movements ( $M = 4.64$ ,  $SD = 0.98$ ) did not differ in overall naturalness ( $t(15) = 0.294$ ). However, beat gestures ( $M = 5.09$ ,  $SD = 0.87$ ) were

perceived as more emphatic than grooming hand movements ( $M = 2.47$ ,  $SD = 0.94$ ;  $t(15) = 6.57$ ,  $p < .001$ ); (2) when combined with speech, beat gestures ( $M = 5.58$ ,  $SD = 1.04$ ) were rated as more natural than grooming hand movement ( $M = 4.62$ ,  $SD = 1.44$ ;  $t(15) = 3.52$ ,  $p = .003$ ). Moreover, beat gestures ( $M = 4.12$ ,  $SD = 1.12$ ) were perceived as more emphatic than grooming hand movements ( $M = 2.25$ ,  $SD = 1.12$ ;  $t(15) = 6.81$ ,  $p < .001$ ).

## EEG Procedure

After electrode application, participants were seated in a soundproof room and watched video clips presented on a computer screen while listening to the speech presented auditorily via loudspeakers. Each trial (see Figure 2) started with a question clip with the two dialogue partners shown from the side and facing each other (see Figure 1). After the question (average duration = 3180 msec), a silent frame was displayed only containing the answering person, facing the camera and holding his or her hands in the still position (1000 msec), followed by the answer sentence (average duration = 2760 msec) and a silent frame of the answering person in still position (500 msec). The interstimulus interval was 500 msec. On catch trials (20% of all trials), after the answer, a single word was visually displayed on the screen (e.g., “school”), and participants judged by button press whether it was semantically related to the preceding sentence (e.g., “She received an EMAIL from the teacher”). Yes and no responses were counterbalanced. Participants were encouraged to blink naturally but to avoid blinking during the answer. Participants were familiarized with the experiment in a practice session and then continued with the actual experiment. Stimuli were divided into 10 blocks of 30 stimuli, and each block lasted approximately 5 min. In total, the experiment lasted for approximately 2 hr, including EEG preparation, instructions, practice, and debriefing.

## EEG Recordings

The EEG was recorded in an electromagnetically shielded cabin with 64 surface active electrodes (Acticap; Brain Products, Herrsching, Germany), placed in an equidistant montage. A forehead electrode served as the ground; and the left mastoid, as the reference electrode. The horizontal and vertical EOG was administered by three electrodes placed at the left and right canthi of each eye and below the left eye. Impedances were kept below 5 k $\Omega$ . The EEG was digitalized at a rate of 500 Hz.

The raw EEG data were preprocessed in the MATLAB toolbox Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). The data were rereferenced to the algebraic average of both mastoid electrodes. In the first step, trials were segmented at the onset of each hand movement, using a 2-sec pregesture and a 2.5-sec postgesture onset window. The segmented data were filtered using a high-pass filter of 0.5 Hz and a band-pass filter, removing fre-

quencies between 49 and 51 Hz and 99 and 101 Hz. Then, we performed an independent component analysis to remove components capturing eye blinks and horizontal eye movements. The clean data were segmented from 200 msec before gesture onset, which was used as the baseline, until 2 sec after gesture onset. A low-pass filter of 30 Hz was applied. Finally, we computed averages of each condition for each participant and used this average for further statistical analysis.

## ERP Data Analysis

To test the statistical difference between conditions, we performed cluster-based random permutation tests (Maris & Oostenveld, 2007), which are implemented in the MATLAB toolbox Fieldtrip (Oostenveld et al., 2011). This approach controls the Type 1 error rate, which arises because of multiple comparisons in large data sets as in ERP studies, involving multiple electrodes and time windows. The analysis is performed as follows. First, the entire data set (all channels, all electrodes) is entered at once into the analysis. Each data sample (each electrode) is tested with a simple dependent  $t$  test. If a set of electrodes, which are spatially adjacent, exceeds the defined significance level of 5%, these electrodes are grouped into clusters. Second, a cluster-level test statistic is performed using the sum of the  $t$  statistics of each electrode. Third, the conditions of each participant are randomly assigned to one of two sets, assuming no difference between conditions. This creates a null distribution, which is calculated in 5000 randomization steps. In the last step, the actually observed cluster-level statistics are compared against the null distribution. All clusters falling within the highest or lowest 2.5% are considered as cases where the null hypothesis can be rejected. Importantly, positive and negative clusters are treated separately at a significance level of  $p < .025$  after a Bonferroni correction is applied for the two individual tests performed. Then, the probability is multiplied with a factor of 2, which results in  $p$  values corresponding to a parametric probability of  $p < .05$ . In other words, the significance level for the overall cluster-based permutation test is  $p < .05$ .

Using cluster-based random permutation tests, we first tested the main effect of Focus by a pairwise comparison ( $t$  test) of all focus conditions (F) to all nonfocus conditions (NF), collapsing over Hand movement type ( $F_{BG + GG + NG}$  vs.  $NF_{BG + GG + NG}$ ). The main effect of Hand movement was tested by an  $F$  test with all three levels collapsed over the focus condition ( $NG_F + NF$  vs.  $BG_F + NF$  vs.  $GG_F + NF$ ). If the main effect of Hand movement was significant, we split the data and performed pairwise comparisons using  $t$  tests with Bonferroni correction (e.g.,  $NG_F + NF$  vs.  $BG_F + NF$ ,  $NG_F + NF$  vs.  $GG_F + NF$ ,  $BG_F + NF$  vs.  $GG_F + NF$ ). We tested the  $3 \times 2$  interaction between Hand movement and Focus by an  $F$  test. If there was a significant  $3 \times 2$  interaction, we computed the difference waveforms along the Focus dimension (F minus NF) for



each Hand movement type ( $BG_F - NF$  vs.  $GG_F - NF$  vs.  $NG_F - NF$ ). In addition, we compute a planned comparison of an interaction of Beat gesture  $\times$  Focus by comparing their difference waveforms ( $BG_F - NF$  vs.  $NG_F - NF$ ) to test our hypothesis that beat gestures would be easier to integrate with focused words but would cause processing costs when they occur on non-focused words. If a significant Beat gesture  $\times$  Focus interaction was found, we further split the difference waveforms by Hand movement type and compared the effects of focus within each hand movement type (e.g.,  $F_{BG}$  vs.  $NF_{BG}$ ,  $F_{NG}$  vs.  $NF_{NG}$ ).

On the basis of previous findings (Biau & Soto-Faraco, 2013; Wang & Chu, 2013; Dimitrova et al., 2012; Cowles et al., 2007), we selected four time windows for analysis: (1) 100–200 msec after gesture onset (–400 to –300 msec before target onset), to test for visual N1-attention effects because of hand movement; (2) 200–500 msec after gesture onset (–300 to 0 msec before target onset), to test for sensory effects because of the visual processing of the gesture; (3) 700–900 msec after gesture onset (200–400 msec after target onset), to test for attention effects

related to the processing of focus; and (4) 1100–1400 msec after gesture onset (600–900 msec after target onset), to test for effects of integrating focus and gesture. The mean amplitudes of all four time windows of all 59 electrodes (five eye and reference electrodes were removed) were entered into the analysis.

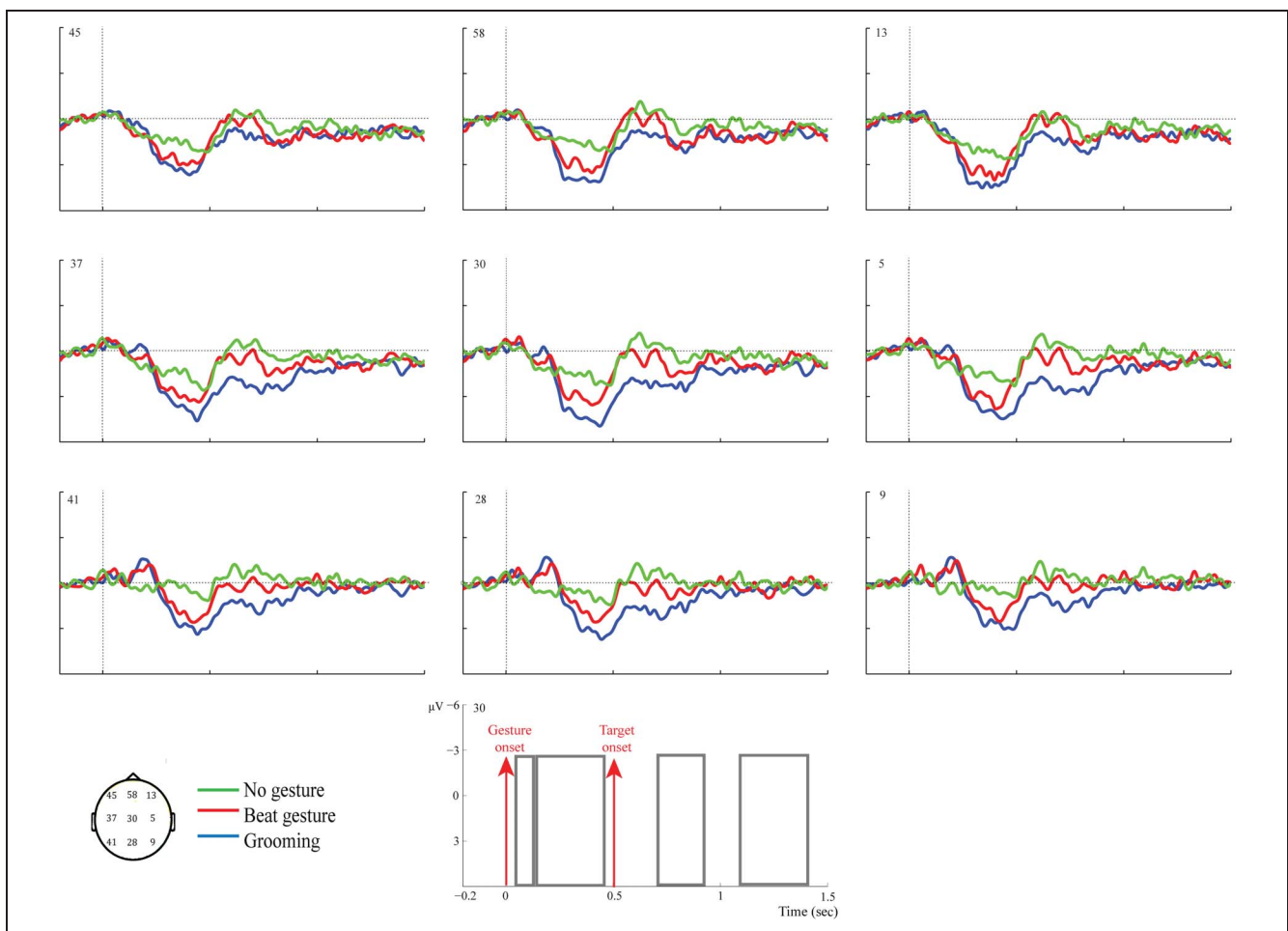
## RESULTS

### Behavioral Results

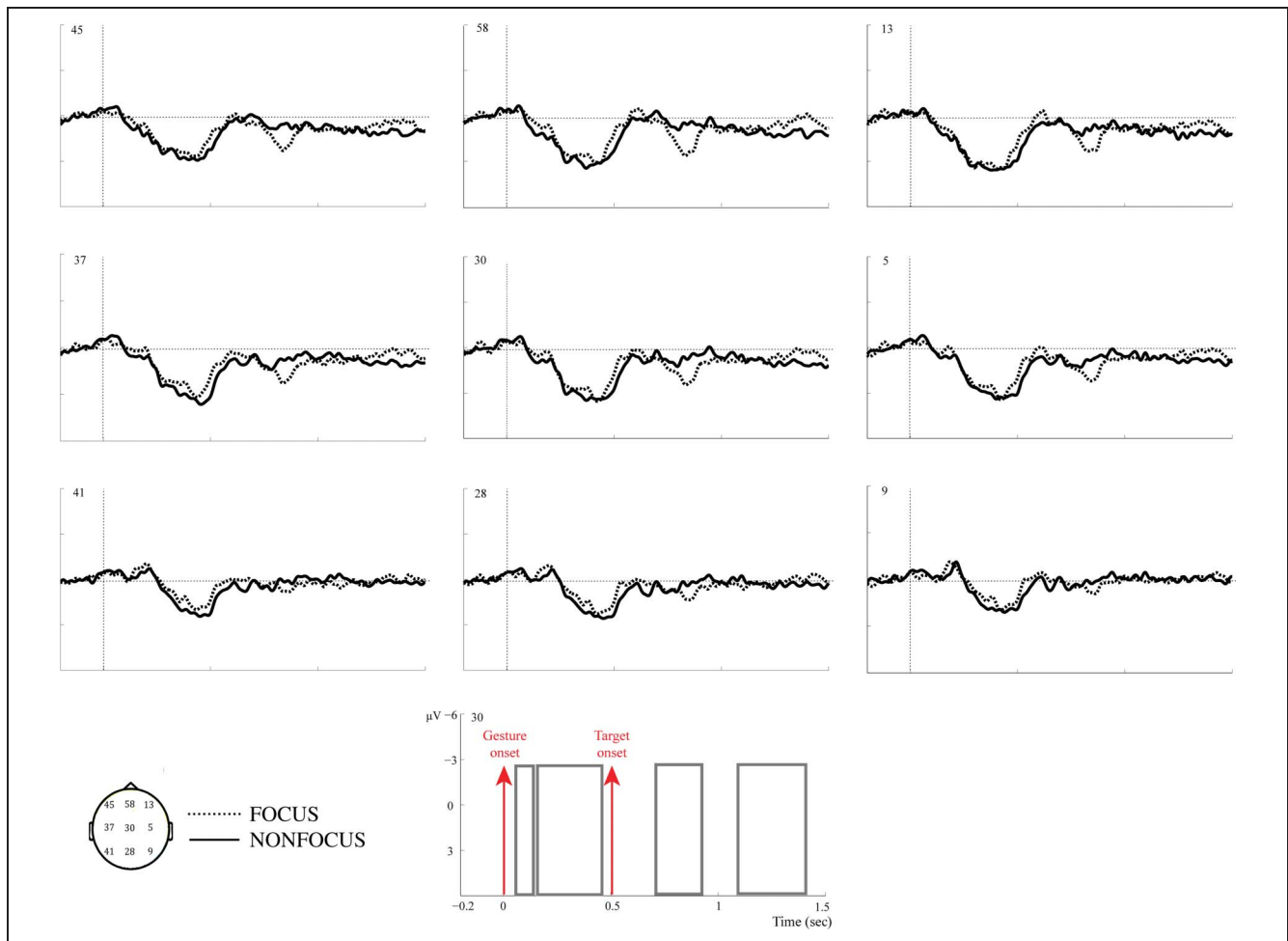
In catch trials, participants judged whether a probe word, which was presented on the screen after 10% of all dialogues, was semantically related to the meaning of the preceding dialogue. The response accuracy was 88.33%, suggesting that participants attended to the stimuli.

### ERP Results

Grand-averaged ERP waveforms were time-locked to the gesture onset. Figures 3–5 present ERPs using selected electrodes (frontal: 45/58/13, central: 37/30/5,



**Figure 3.** Main effect of hand movement. ERPs are shown for the three hand movement conditions: no gesture (NG, green lines), beat gesture (BG, red lines), and grooming hand movement (GG, blue lines). Hand movement conditions are collapsed over the focus conditions (F + NF). Time windows (indicated by squares) were computed relative to gesture onset: 100–200, 200–500, 700–900, and 1100–1400 msec.



**Figure 4.** Main effect of focus. ERPs are shown for the focus condition (F, dotted line) and the nonfocus condition (NF, solid line) and are collapsed over the hand movement conditions (BG + GG + NG). We tested four time windows relative to hand movement onset (indicated here as squares): 100–200, 200–500, 700–900, and 1100–1400 msec.

posterior: 41/28/9). We report the mean difference for each statistical effect ( $M$ , in microvolts [ $\mu\text{V}$ ]) as well as the  $SEM$  and the actual  $p$  value from the cluster statistics where significant results correspond to  $p < .05$ .

#### 100–200 msec after Gesture Onset (–400 to –300 msec before Target Onset)

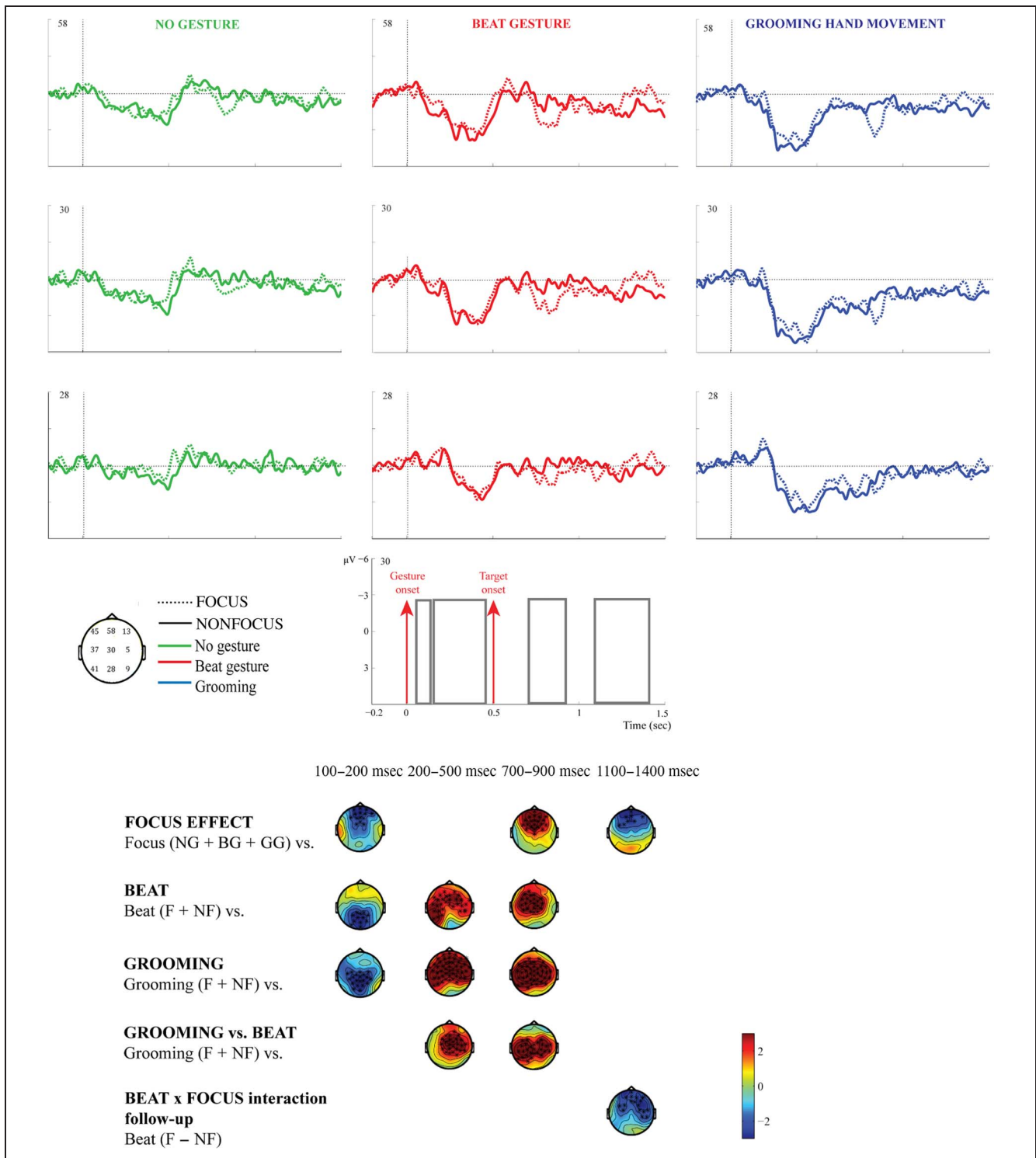
We found a main effect of Focus (Figure 4): Focused targets elicited a negativity relative to nonfocused targets ( $F_{\text{NG}} + \text{BG} + \text{GG}$  vs.  $\text{NF}_{\text{NG}} + \text{BG} + \text{GG}$ ,  $p = .025$ ,  $M = -0.59$ ,  $SEM = 0.21$ ). The main effect of Hand movement was marginally significant ( $\text{NG}_{\text{F}} + \text{NF}$  vs.  $\text{BG}_{\text{F}} + \text{NF}$  vs.  $\text{GG}_{\text{F}} + \text{NF}$ ,  $p = .08$ ). As Figure 3 shows, targets with a beat gesture and a grooming hand movement tended to show a negativity relative to targets with no gesture. There was no difference between the main effects of beat gesture and grooming hand movement ( $\text{GG}_{\text{F}} + \text{NF}$  vs.  $\text{BG}_{\text{F}} + \text{NF}$ , no clusters). We did not find a  $3 \times 2$  interaction of Hand movement  $\times$  Focus or an interaction of Beat gesture  $\times$  Focus.

#### 200–500 msec after Gesture Onset (–300 to 0 msec before Target Onset)

There was no main effect of Focus, but we found a main effect of Hand movement ( $\text{NG}_{\text{F}} + \text{NF}$  vs.  $\text{BG}_{\text{F}} + \text{NF}$  vs.  $\text{GG}_{\text{F}} + \text{NF}$ ,  $p < .001$ ). As shown in Figure 3, relative to targets with no gesture, targets with a beat gesture elicited a positive cluster ( $\text{BG}_{\text{F}} + \text{NF}$  vs.  $\text{NG}_{\text{F}} + \text{NF}$ ,  $p = .024$ ,  $M = 0.91$ ,  $SEM = 0.29$ ). There was a positive cluster for targets with a grooming hand movement versus targets with no gesture ( $\text{GG}_{\text{F}} + \text{NF}$  vs.  $\text{NG}_{\text{F}} + \text{NF}$ ,  $p = .001$ ,  $M = 1.4$ ,  $SEM = 0.33$ ) and versus targets with a beat gesture ( $\text{GG}_{\text{F}} + \text{NF}$  vs.  $\text{BG}_{\text{F}} + \text{NF}$ ,  $p = .025$ ,  $M = 0.81$ ,  $SEM = 0.25$ ). We did not find a  $3 \times 2$  interaction of Hand movement  $\times$  Focus or an interaction of Beat gesture  $\times$  Focus.

#### 700–900 msec after Gesture Onset (200–400 msec after Target Onset)

We found a main effect of Focus (Figure 4), showing a positivity for focused targets relative to nonfocused



**Figure 5.** Main effects and interactions. ERPs display the effect of focus within each hand movement condition and the interaction of hand movement and focus. The time windows (indicated by squares) were computed relative to hand movement onset: 100–200, 200–500, 700–900, and 1100–1400 msec. The topographic plots display significant clusters to the main effects of hand movement, the main effect of focus, and the interaction of beat gesture and focus.

targets ( $F_{NG + BG + GG}$  vs.  $NF_{NG + BG + GG}$ ,  $p = .01$ ,  $M = 0.77$ ,  $SEM = 0.22$ ). There was also a main effect of Hand movement ( $NG_F + NF$  vs.  $BG_F + NF$  vs.  $GG_F + NF$ ,  $p < .001$ ). Figure 3 shows a positivity for targets with a beat

gesture relative to targets with no gesture ( $BG_F + NF$  vs.  $NG_F + NF$ ,  $p = .02$ ,  $M = 0.90$ ,  $SEM = 0.26$ ). Targets with a grooming hand movement also elicited a positivity, both relative to targets with no gesture ( $GG_F + NF$  vs.  $NG_F + NF$ ,

$p < .001$ ,  $M = 1.56$ ,  $SEM = 0.27$ ) and relative to targets with a beat gesture ( $GG_{F+NF}$  vs.  $BG_{F+NF}$ ,  $p = .016$ ,  $M = 1.1$ ,  $SEM = 0.28$ ). We did not find a  $3 \times 2$  interaction of Hand movement  $\times$  Focus or an interaction of Beat gesture  $\times$  Focus.

#### *1100–1400 msec after Gesture Onset (600–900 msec after Target Onset)*

We found a main effect of Focus ( $F_{NG+BG+GG}$  vs.  $NF_{NG+BG+GG}$ ,  $p = .04$ ,  $M = -0.45$ ,  $SEM = 0.16$ ), which showed that nonfocused targets elicited a positivity relative to focused targets (Figure 4). There was no main effect of Hand movement ( $NG_{F+NF}$  vs.  $BG_{F+NF}$  vs.  $GG_{F+NF}$ , no clusters). The  $3 \times 2$  interaction of Hand movement  $\times$  Focus was not significant. The planned comparison of a Beat gesture  $\times$  Focus interaction was significant ( $BG_{F-NF}$  vs.  $NG_{F-NF}$ ,  $p = .04$ ,  $M = -1.1$ ,  $SEM = 0.41$ ). Follow-up tests showed a difference between focus and nonfocus in the Beat gesture condition ( $F_{BG}$  vs.  $NF_{BG}$ ,  $p = .025$ ,  $M = -0.79$ ,  $SEM = 0.3$ ) but not in the No gesture condition ( $F_{NG}$  vs.  $NF_{NG}$ , no clusters). As Figure 5 shows, beat gestures on nonfocused words elicited a positivity relative to beat gestures on focused words.

## DISCUSSION

This ERP study investigated the theoretical claim that beat gestures serve as nonverbal focusing cues that draw attention to relevant information. We examined if beat gestures are integrated with the focus of the message and whether they behave differently from control hand movements like grooming. Our results provide evidence in support of this claim as evident in an interaction of beat gesture and focus. That is, listeners incurred additional costs to integrate beat gestures with nonfocused information, which was reflected in a late positivity (1100–1400 msec after gesture onset/600–900 msec after target onset). The late positivity effect was unique for beat gestures as grooming hand movements did not interact with focus. In addition, focused words elicited an anterior positivity compared with nonfocused words (700–900 msec after gesture onset/200–400 msec after target onset). Words accompanied by a hand movement tended to elicit an early parietal negativity (100–200 msec after gesture onset) and triggered a sustained positivity from 200 msec after gesture onset.

### **Beat Gestures Function as Nonverbal Focusing Cues in Multimodal Speech**

The results of this study demonstrate that beat gestures function as nonverbal cues for focus and that they are integrated with the information structure of a message during multimodal speech comprehension. Beat gestures accompanying nonfocused information gave rise to a late anterior positivity (1100–1400 msec after gesture onset/

600–900 msec after target onset) relative to beat gestures accompanying focused information (Figure 5). Interestingly, no such difference between focus and nonfocus was found in trials with grooming or no hand movements. The timing, latency, and polarity of this effect resemble the characteristics of a late positivity. Prior ERP studies on multimodal comprehension have also found late positivities when listeners integrate multiple information sources from the visual and semantic domains (e.g., Kuperberg, 2007; Sitnikova, Kuperberg, & Holcomb, 2003). The late positivity in our study may thus reflect increased computation costs needed to arrive at a coherent interpretation of the message when beat gesture emphasizes nonfocused information, which should not be highlighted. Alternatively, the effect can be viewed as a negativity for focused elements with a beat gesture. Late anterior negativities have been reported for working memory processes during sentence comprehension (Vos, Gunter, Kolk, & Mulder, 2001; Münte, Schiltz, & Kutas, 1998). The late negativity in our study may suggest that listeners engage more resources to store information in working memory when information is highlighted by two emphasis cues such as focus and beat gesture. In contrast, information that is only highlighted by beat gesture engages less memory resources. The interpretation of this effect as a negativity because of increased working memory load can be addressed in future studies in which listener's memory is compared in sentences with and without a beat gesture.

The late positivity has implications for theories of gesture–speech integration in comprehension. According to the integrated systems hypothesis (Kelly et al., 2008; McNeill, 1992), gesture and speech represent a coupled system and mutually and obligatorily interact during comprehension. Although we did not directly test bidirectional influences of gesture and speech but only examined gesture effects on speech comprehension, our findings provide relevant support for this hypothesis. Our results show that beat gesture behaves similarly to pitch accent in that it highlights focus and is expected to align with relevant information in discourse. Importantly, our study is the first to demonstrate that the two systems interact in context and that only beat gesture, but not grooming hand movements, is integrated with the focus of the message. In a previous ERP study using single sentences, Wang and Chu (2013) reported independent effects of beat gesture and pitch accent on semantic processing, presumably because, in the absence of context, any information can be highlighted by either a beat gesture or a pitch accent. We found that, if the two systems work in harmony, emphasis by beat gesture is associated with contextually and accentually salient information; this facilitates processing. If the two systems are in conflict, adding emphasis by beat gesture to contextually less salient unaccented information causes integration difficulties, which supports the integrated systems hypothesis.

It is important to note that, in this study, beat gestures could fall on either focus or nonfocus, which might have

caused listeners to consider them less reliable as emphasis cues. In contrast, focused words were always accented, and pitch accent might have been viewed as a more reliable cue. Through this manipulation, we may have diminished the potential strength of beat gesture. However, this tendency corresponds to real-life situations where highlighting focus by gesture is not obligatory. Interestingly, despite the potential low reliability of nonverbal cues, beat gestures still affected speech processing and interacted with focus. If anything, we hypothesize that the observed effect of beat gesture would be stronger if a future study would manipulate the reliability of gesture and accent similarly.

### **Accented Focused Elements Grab the Listeners' Attention**

Our study replicates the finding that focused words increase listeners' attention, as evident from an anterior positivity around 200 msec after the onset of the focused word. Similar effects have been found for the processing of focus and of accented words and have been attributed to the P3 component (Dimitrova et al., 2012; Cowles et al., 2007; Bornkessel et al., 2003). We extend this result, which was obtained separately in reading and listening paradigms, to speech processing in a multimodal paradigm. That is, when viewing videos of a speaker who gestures, listeners keep track of contextually important information and direct their attention to focus. Because of the anterior distribution of the positivity (Figure 4), we attribute it to the "novelty" P3a component that signifies the allocation of attention to relevant information in speech. Because focused words were always accented in this study, we cannot unambiguously interpret the underlying mechanism of the P3a component; it could reflect attention allocated at focused and/or accented information.

One argument in favor of a focus mechanism interpretation is that, in addition to the P3a effect, focused words elicited an early anterior negativity relative to non-focused words, which started well before the actual focused word was encountered (100–200 msec after gesture onset/–400 to –300 msec before target onset). Although this effect is unexpected, it suggests that listeners likely anticipate focus guided by contextual expectations. The prestimulus negativity for focus may be related to the N1 component for attentive processing (for a review, see Näätänen & Picton, 1987). Alternatively, the effect may belong to the contingent negative variation, which is an anterior negativity related to the cognitive preparation for an upcoming stimulus (Walter, Cooper, Aldridge, McCallum, & Winter, 1964).

Although focus and beat gesture did not interact in the P3 time window, we explored the focus effect across the different hand movement conditions and found that the P3 effect for focus was largest in trials with a beat gesture. This numerical difference suggests that listeners'

overall attention to focus is increased when the speaker produces a beat gesture. The lack of an interaction between focus and beat gesture in this time window supports the view that focus and beat gesture have initially independent contributions to attentive processing, presumably serving as linguistic and nonlinguistic highlighters. The independent effects are consistent with the results of Wang and Chu (2013). In contrast to that study, however, we show that, when a contextual constraint is added, focus and beat gesture interact in late time windows (1100–1400 msec after gesture onset/600–900 msec after target onset).

### **Beat Gestures and Grooming Modulate Speech Processing Differently**

Adding a gesture to the speech signal had an effect on visual processing: Beat gestures and grooming hand movements tended to elicit an early parietal negativity (100–200 msec after gesture onset) and gave rise to a strongly pronounced early positivity (200–500 msec after gesture onset). The early negativity did not differ for beat gestures versus grooming hand movements and might be related to the N1 component for the attentive processing and discrimination of visual stimuli (e.g., Vogel & Luck, 2000). The early positivity is in line with previous reports of early positivities elicited by beat gestures (Biau & Soto-Faraco, 2013; Wang & Chu, 2013), which were also interpreted as enhanced attention to the visually prominent hand movement. We suggest that the early positivity reflects the attentive processing of a salient change in the visual scene (e.g., hand movement) and as such belongs to the P300 component for attentive processing. Importantly, the positivity lasted until the gesture was retracted (until 1000 msec after gesture onset) and was significantly more pronounced for grooming than for beat gestures. The sustained positivity suggests that nonverbal signals modulate the entire processing of the sentence, increasing the listener's overall attention.

Interestingly, the sustained positivity for the control grooming condition was more enhanced than the positive effect for beat gestures. We speculate that this might reflect that grooming hand movements are perceived as less natural than beat gestures, as the results of Pretest 3 suggest. Unnatural grooming hand movements might distract listeners from the content of speech as they introduce visual information, which is irrelevant for the speech signal. As a result, listeners might try to inhibit unrelated information from grooming, and this process will likely impose higher attentional demands on processing, both for grooming hand movements on focus and nonfocus items. On the other hand, seeing an unnatural grooming hand movement might lead to increased difficulties in comprehending information from the speech signal, again irrespective of the information status of the word.

Prior neuroimaging studies support the latter hypothesis: Adding grooming to an experimental paradigm

affects the overall reliability of gestural information. For example, Holle and Gunter (2007) showed that adding irrelevant grooming movements (scratching or rubbing) to an ERP paradigm weakens the facilitation effect of iconic gestures on semantic disambiguation. In an fMRI study, Skipper, Goldin-Meadow, Nusbaum, and Small (2009) found that grooming increased semantic retrieval demands, whereas iconic gestures facilitated semantic processing. In line with these findings, the stronger sustained positive effect of grooming hand movements relative to beat gestures in this study might indicate an integration difficulty. In addition, adding grooming hand movements might have diminished the strength of the effect of beat gestures. Nonetheless, we found that beat gestures affected processing and were integrated with speech. Because grooming is widely found in daily life, our study provides ecologically valid insights into the role of gestures in speech comprehension. Our data suggest that, even if people groom, listeners pay attention to their beat gestures.

## Conclusion

The current ERP study provides innovative insights into the role of beat gesture in multimodal speech processing. Beat gestures serve as nonverbal emphasis cues and enhance listeners' attention to focused information in discourse. Whereas, in single sentences, beat gestures can highlight any word (Wang & Chu, 2013), in sentences with context, beat gestures are expected to accompany only the focus of the message. If they fall on nonfocused information instead, listeners incur processing difficulties. This study provides empirical support for the theoretical hypothesis that gesture and speech form an integrated system in comprehension. We conclude that beat gesture, but not other types of hand movements like grooming, functions as nonverbal cues for focus and facilitates speech processing.

Reprint requests should be sent to Mingyuan Chu, School of Psychology, University of Aberdeen, Aberdeen, AB24 3FX, United Kingdom, or via e-mail: mingyuan.chu@abdn.ac.uk.

## REFERENCES

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: Linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 403–413.

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*, 143–152.

Biau, E., & Soto-Faraco, S. (2015). Synchronization by the hand: The sight of gestures modulates low-frequency activity in brain responses to continuous speech. *Frontiers in Human Neuroscience*, *9*, 527.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*, 365–391.

Birch, S. L., Albrecht, J. E., & Myers, J. L. (2000). Syntactic focusing structures influence discourse processing. *Discourse Processes*, *30*, 285–304.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer. 7 <http://www.fon.hum.uva.nl/praat/>.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject- and object-control verbs. *Journal of Cognitive Neuroscience*, *22*, 1036–1053.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter- versus intrasentential predictions. *Journal of Experimental Psychology*, *29*, 871–882.

Cowles, H. W., Kluender, R., Kutas, M., & Polinsky, M. (2007). Violations of information structure: An electrophysiological study of answers to wh-questions. *Brain and Language*, *102*, 228–242.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*, 49–59.

Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. J. (2012). Less is not more: Neural responses to missing and superfluous accents in context. *Journal of Cognitive Neuroscience*, *24*, 2400–2418.

Heim, S., & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, *66*, 55.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*, 1175–1192.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*, 74.

Hruska, C., & Alter, K. (2004). Prosody in dialogues and single sentences: How prosody can influence speech perception. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp. 221–226). Berlin, Germany: de Gruyter.

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. *Human Brain Mapping*, *30*, 1028–1037.

Jacobs, J. (1986). The syntax of focus and adverbials in German. In W. Abraham & S. de Mey (Eds.), *Topic, focus, and configurationality* (pp. 103–127). Amsterdam: John Benjamins.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*, 253–260.

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 and Linguistics Compass*, *2*, 569–588.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*, 396–414.

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex*, *23*, 1836–1848.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621.
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*, 1457–1471.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*, 71–73.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus—Norepinephrine system. *Psychological Bulletin*, *131*, 510.
- Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *The Journal of the Acoustical Society of America*, *82*, 1512–1524.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 156869.
- 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. *Journal of Cognitive Neuroscience*, *19*, 605–616.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456–479.
- Roustan, B., & Dohen, M. (2010). Gesture and speech coordination: The influence of the relationship between manual gesture and speech. In *11th Annual Conference of the International Speech Communication Association 2010 (Interspeech 2010)*.
- Schumacher, P. B., & Baumann, S. (2010). Pitch accent type affects the N400 during referential processing. *NeuroReport*, *21*, 618–622.
- Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, *40*, 160–164.
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology*, *19*, 661–667.
- Sudhoff, S. (2010). Focus particles and contrast in German. *Lingua*, *120*, 1458–1475.
- Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*, 190–203.
- Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, *38*, 41–63.
- Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384.
- Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*, 813–820.
- Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, *51*, 2847–2855.
- Wang, L., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, *1282*, 50–56.
- Wang, L., Li, X., & Yang, Y. (2014). A review on the cognitive function of information structure during language comprehension. *Cognitive Neurodynamics*, *8*, 353–361.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*, 654–667.
- Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*, 234–245.