

The role of beat gesture and pitch accent in semantic processing: An ERP study



Lin Wang^a, Mingyuan Chu^{b,*}

^a Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, China

^b Department of Neurobiology of Language, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 11 July 2013

Received in revised form

5 August 2013

Accepted 11 September 2013

Available online 20 September 2013

Keywords:

Beat gesture

Pitch accent

Semantic processing

Speech comprehension

N400

ABSTRACT

The present study investigated whether and how beat gesture (small baton-like hand movements used to emphasize information in speech) influences semantic processing as well as its interaction with pitch accent during speech comprehension. Event-related potentials were recorded as participants watched videos of a person gesturing and speaking simultaneously. The critical words in the spoken sentences were accompanied by a beat gesture, a control hand movement, or no hand movement, and were expressed either with or without pitch accent. We found that both beat gesture and control hand movement induced smaller negativities in the N400 time window than when no hand movement was presented. The reduced N400s indicate that both beat gesture and control movement facilitated the semantic integration of the critical word into the sentence context. In addition, the words accompanied by beat gesture elicited smaller negativities in the N400 time window than those accompanied by control hand movement over right posterior electrodes, suggesting that beat gesture has a unique role for enhancing semantic processing during speech comprehension. Finally, no interaction was observed between beat gesture and pitch accent, indicating that they affect semantic processing independently.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Successful face-to-face communication requires integration of information from multiple sensory modalities. People derive meaning not only from speech, but also from visual cues such as lip movements, facial expression, body posture, and hand gestures. An important question then is how a listener combines information from auditory (e.g., speech) and visual channel (e.g., gesture) to comprehend a message.

To date, most studies on the speech and gesture integration have focused on representational gestures, which depict a concrete or abstract semantic meaning with the shape or motion of the hand/s. Behavioral studies demonstrated that speech comprehension is enhanced by accompanying representational gestures (Beattie & Shovelton, 1999; Holler, Shovelton, & Beattie, 2009; Kelly, Barr, Church, & Lynch, 1999; but see Kelly & Goldsmith, 2004; Krauss, Dushay, Chen, & Rauscher, 1995 for conflict findings). Evidence from neuroimaging studies suggests that the left inferior frontal gyrus (Willems, Özyürek, & Hagoort, 2007, 2009) and the left posterior temporal lobe (Holle, Gunter, Rueschemeyer, Hennenlotter, & Iacoboni, 2008; Holle, Obleser, Rueschemeyer, & Gunter, 2010) are crucially involved in the integration of the semantic information provided by

representational gesture and speech. Several event-related potential (ERP) studies reported that semantically incongruent gesture–speech pairings elicited larger N400 amplitudes than congruent pairings (Cornejo et al., 2009; Holle & Gunter, 2007; Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2005). In general, the N400 amplitude is larger when it is more difficult to integrate the semantic meaning of a word into previous context than when it is easier (Kutas & Hillyard, 1980). Thus, these ERP data suggest that people integrate semantic information from representational gestures into speech. Furthermore, Kelly, Creigh, and Bartolotti (2009) demonstrated that when participants were presented with gesture and speech, they automatically integrated the two modalities even when that integration was not required in the task given to the participants. In sum, existing evidence shows that gesture and speech are automatically integrated during language comprehension.

Speakers not only produce representational gestures, but also beat gestures in communication. A beat gesture is a rapid movement of the hand, usually up and down, produced with the rhythm of the concurrent speech (McNeill, 1992). It does not convey semantic content (Alibali, Heath, & Myers, 2001). It indexes the significance of its accompanying word or phrase and is often used to highlight new or contrastive information (McNeill, 1992). Although beat gesture is a common type of gestures in communication (McNeill, 1992), it has received much less attention in the literature than representational gestures, presumably because

* Corresponding author. Tel.: +31 24 352 1307; fax: +31 24 352 1213.

E-mail addresses: mingyuan.chu@mpi.nl, mingyuan_chu@hotmail.com (M. Chu).

beat gesture does not convey any semantic meaning in itself. The present study aimed to examine how seeing a beat influence listeners' processing of the accompanying speech.

To our knowledge, there is so far only one fMRI study and two ERP studies that investigated how the brain integrates beat gesture with speech during language comprehension. In the fMRI study, Hubbard, Wilson, Callan, and Dapretto (2009) found greater activation in non-primary auditory cortex (e.g., bilateral posterior superior temporal gyrus) when speech was accompanied by beat gestures than when speech was presented alone. They also found stronger activations in left superior temporal gyrus/sulcus when speech was accompanied by beat gestures than when speech was accompanied by nonsense hand movements (i.e., non-iconic American Sign Language movements unknown to participants). The authors concluded that these brain areas are crucial for the integration of beat gesture and speech. The ERP study by Holle et al. (2012) demonstrated that beat gesture had an impact on syntactic analysis. This study focused on the P600 component, a positive-going deflection of the ERP peaking around 600 ms after the onset of the critical word. A larger P600 is often elicited by less preferred syntactic structures in ambiguous sentences (Osterhout & Holcomb, 1992). Holle et al. (2012) showed that the P600 effect disappeared when the subject (e.g., *the men*) of the non-preferred syntactic structure (OSV structure: *The woman_[object] the men_[subject] have greeted.*) was accompanied by a beat gesture. This suggests that the visual emphasis provided by beat gestures increased the plausibility of the non-preferred syntactic structure. The other ERP study by Biau and Soto-Faraco (2013) showed that beat gesture had an impact on speech processing at the sensory/phonological level. In this study, words accompanied by beat gesture elicited a positive shift at an early sensory stage as well as an enhanced P2, compared to words accompanied by no hand movement. The authors concluded that beat gestures facilitate early speech analysis by allocating listeners' attention towards important information. Taken together, these studies have shown that seeing beat gesture has an effect on speech processing.

Given the relatively small number of studies on the integration of beat gesture and speech in the brain, several research questions remain unclear. First, little is known on whether beat gesture has an impact on speech processing at the semantic level. Although beat gesture does not convey any semantic information, it may affect semantic processing by indexing the saliency of its accompanying word (McNeill, 1992; Holle et al., 2012; Biau & Soto-Faraco, 2013). Previous evidence shows that linguistic devices that are used to highlight information (e.g., question context: Wang, Hagoort, & Yang, 2009; pitch accent: Swerts, Krahmer, & Avesani, 2002, Krahmer & Swerts, 2007; syntactic structure: Cowles, Kluender, Kutas, & Polinsky, 2007) can modulate semantic processing. For instance, a smaller N400 was found when information was marked to be focus than when it was not, suggesting that focused information was easier to be integrated into context compared to non-focused information (Wang et al. 2009; Wang, Bastiaansen, Yang, & Hagoort, 2011). Furthermore, in an fMRI study, Kristensen, Wang, Petersson, & Hagoort (2012) showed that words expressed with pitch accent activated a fronto-parietal attention network to a larger degree than words expressed without pitch accent. Also, semantically violating words activated left inferior frontal gyrus (which is sensitive to semantic violations) only when they were expressed with pitch accent. The authors concluded that accented information receives more attentional resources and more elaborate semantic processing relative to unaccented information. Therefore, it is rational to expect that beat gesture, as a visual cue to highlight information, will also have an effect on semantic processing.

Another open question is that if beat gesture could facilitate semantic processing of concurrent speech, what would be the

underlying mechanism. One possibility is that the presence of beat gesture, or any other hand movements increases the general attention level and enable deeper processing of the concurrent speech than when no hand movement is present. A second possibility is that beat gesture has a unique role in modulating semantic processing due to its conventional use to emphasize information in communication. That is, rather than simply capturing attention as a visual signal, beat gesture serves as a special communication signal (Grice, 1975). This implies that other, non-communicative visual signals (such as non-sense hand movements) should not facilitate semantic processing, or at least not as much as beat gesture. Of course, the two possibilities are not mutually exclusive. Beat gesture may facilitate semantic processing through both a general attention capture mechanism and its unique role in emphasizing information in face-to-face communication.

In addition, although beat gesture and pitch accent are closely related in time and function (Leonard & Cummins, 2010; Krahmer & Swerts, 2007), little is known about whether they modulate the semantic processing of speech interactively or independently. Krahmer and Swerts (2007) found that the production of beat gesture enhanced the acoustic prominence of the simultaneously produced speech, and in return, seeing beat gesture increased the perceived prominence of the gesture-accompanied word. Thus, beat gesture and pitch accent may facilitate semantic processing interactively. Beat gesture may only affect semantic processing of a word when the word is expressed with pitch accent. When beat gesture is presented without pitch accent, it may not facilitate semantic processing or may even increase the difficulty of semantic processing due to a violation of the expectation on the co-occurrence of beat gesture and pitch accent. Alternatively, the effect of beat gesture and pitch accent on semantic processing of speech words may be independent of each other, because beat gesture and pitch accent highlight information via different modalities and both of them are prominent cues to highlight the saliency of information expressed in speech.

To sum up, the present study aimed to address three research questions: (1) Does beat gesture facilitate semantic processing of its accompanying word? (2) If beat gesture facilitates semantic processing, is this effect specific to beat gesture, or do other non-beat-like hand movements have the same effect? (3) Do beat gesture and pitch accent affect semantic processing interactively or independently? To answer these questions, we presented participants with short video clips containing spoken sentences. A critical word in each sentence was accompanied by a beat gesture, a control movement, or no hand movement. The control movement was produced by the same hand, at the same starting and ending location, with the same rhythmic properties, but with a non-beat-like moving trajectory as the beat gesture. We pretested the emphasis function and the likelihood of daily use of the beat gesture and the control movement. The critical word was either accented or not. The ERP responses to the critical words in the different conditions were compared. Specifically, the N400 effect, which reflects the difficulty of semantic integration (for a review see Kutas & Federmeier, 2011), was measured.

We made the following predictions. First, if beat gesture reduces the difficulty of a word's semantic processing, there should be a smaller N400 when a word is accompanied by a beat gesture than when it is not. Secondly, if beat gesture has a unique role in facilitating semantic processing due to its conventional use as a communicative signal (which were manifested in the pretest), there should be a smaller N400 when a word is accompanied by beat gesture than when it is accompanied by control movement. Third, if beat gesture and pitch accent affect semantic processing interactively, words accompanied by a beat gesture will only elicit smaller N400s than the words accompanied with no hand movement when they are expressed with pitch accent. When pitch accent is

not present, words accompanied by beat gesture should elicit similar or even larger N400s than those accompanied by no hand movement (because the hypothesized expectation of co-occurrence of beat gesture and pitch accent is violated). Alternatively, if beat gesture and pitch accent exert independent influence on semantic processing, they should each induce smaller N400s for the critical words regardless of the presence of one in the other modality.

2. Methods

2.1. Participants

Twenty-five healthy native speakers of Dutch (7 males) were paid to participate in the experiment. They were all right-handed, with normal or corrected to normal vision, and were aged 18–31 years (mean age 21 years). None of them had dyslexia, a history of hearing problems, or neurological disorder. A consent form according to the Declaration of Helsinki was signed before they started the experiment. The data of one participant (a female) were excluded because of excessive artifacts. The final sample therefore consisted of 24 participants.

2.2. Stimuli

2.2.1. Stimuli construction

Two hundred and forty video clips of a woman uttering Dutch sentences were used in the present study. The videos were filmed by a digital camera (JVC GY-HM100E) with 40 ms/frame and edited using Adobe Premiere Pro CS5.5. In each video clip, the actress was seated in a chair, with only the torso and limbs visible in the video clips. Three conditions of hand movements were created: beat gesture, control movement and no hand movement (see Fig. 1 for an illustration of the two hand movement conditions).

The beat gesture and control movement were always executed by her right hand, starting and ending at the same location on the lap. The beat gesture and control movement consisted of four stages: the hand and forearm lifted up (beat gesture) or moved to the right (control movement) rapidly with a duration of 280 ms, the hand and forearm moved down (beat gesture) or to the left (control movement) rapidly with a duration of 240 ms, the hand and arm held perpendicular to the lap with a duration of 560 ms (beat gesture and control movement), and the hand and arm returned to the initial resting position on the lap with a duration of 640 ms (beat gesture and control movement). In order to make the beat gesture as natural as possible, the actress made beat gesture while she accented a word in an example sentence. She then made control movement while she accented the same word in the same sentence so that the beat gesture and control movement matched as much as possible except their form. The audio was removed during the editing of the video. We used this form of beat

gesture because McNeill (1992) defined beat gestures as rapid hand/finger flicks with two phases of movement (usually up and down) and meaningless in itself. In addition, we only used one beat gesture because the same person tended to use the same form of beat gesture for emphasis. The use of only one beat gesture form can also be found in the two beat gesture-related ERP studies (Biau & Soto-Faraco, 2013; Holle et al., 2012). In the no hand movement condition, the actress's hands remained still in the resting position. A no hand movement video was made by duplicating the frame of the initial position.

Each of the 240 sentences in the video clips contained two constituents: an agent and a patient (see Table 1). We defined the patient as the critical word (CW). Each sentence was spoken in two different ways: either the CW was accented or no particular word in the sentence was accented.

The sentences were recorded separately from the video shooting. They were digitized at a sample frequency of 44.1 KHz and were normalized to the same average sound pressure level using Praat 4.0 (Boersma & Weenink, 2002). To ensure that the accented words were acoustically different to the unaccented words, we performed five pair-wise *t*-tests for the CWs, with Pitch accent (accented versus unaccented) as an independent variable. The dependent variables were CW duration, CW intensity, mean F0 (fundamental frequency) of each CW, standard deviation (SD) for the F0 of each CW and root mean square (rms) for the amplitude of each CW. The accented CWs had significantly longer durations, higher intensity, higher mean F0, larger F0 SD, larger rms of amplitude than the unaccented CWs (see Table 2). The average duration of the sentences was 2855 ms.

The audio recordings of the sentences were combined with the video recordings using the Adobe Premiere Pro CS5.5. In both beat gesture and control movement video clips, the apex of the hand movement (beat gesture: the end of the downward moving phase; control movement: the end of the right-to-left moving phase) was synchronized with the onset of the primary stress of the CW, as it has been shown that there is a tight alignment of the gestural apex with the peak of pitch accent on the stressed syllable (Leonard & Cummins, 2010; McNeill, 1992). Specifically, the hand movement started 520 ms earlier than the CW onset and lasted till 1200 ms after the CWs onset. Additional still frames with the hands in the resting position were added to the remaining parts of the sentences.

Table 1

An example of the stimulus materials.

(1) Accented condition

Gisteren kocht Anne kleurrijke LELIES om de kamer mee te versieren.
(Yesterday Anne bought colorful LILIES to the room to decorate.)

(2) Unaccented condition

Gisteren kocht Anne kleurrijke lelies om de kamer mee te versieren.
(Yesterday Anne bought colorful lilies to the room to decorate.)

Note. The examples were originally in Dutch. Literal translations in English are given in brackets. The critical words (CWs) are underlined, and the words with pitch accent are in capitals.

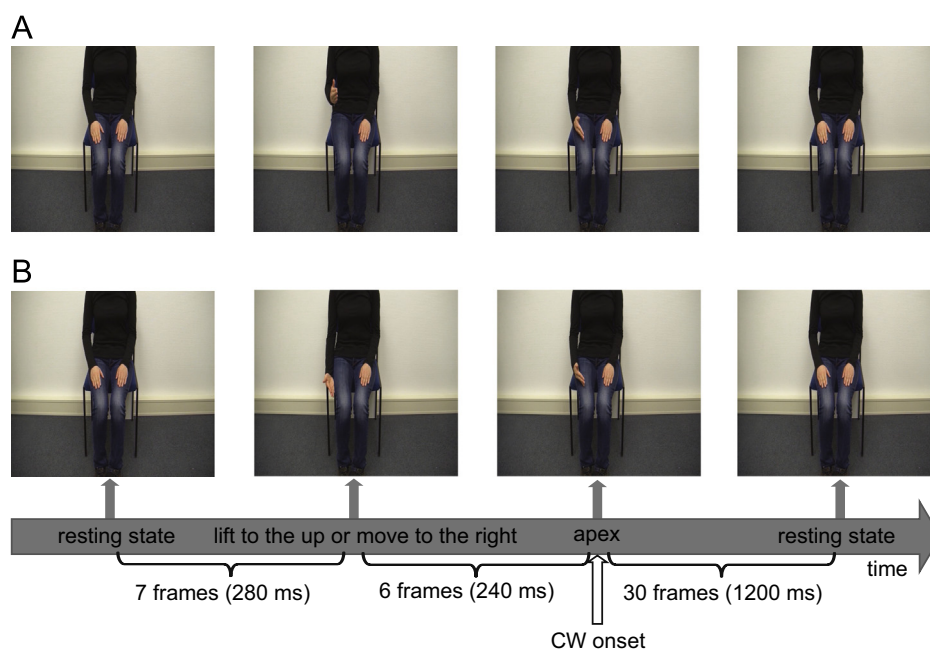


Fig. 1. An illustration of the (A) beat gesture and (B) control movement. Four phases were displayed: the resting state; the hand was lifted to the up or moved to the right; the hand moved back to the lap; the hand went back to the resting position. CW: critical word.

Table 2
Acoustic measurements of critical words in sentences.

	Accented	Unaccented	$t_{(239)}$
Duration (ms)	423 (137)	396 (133)	9.70***
Intensity (dB)	73 (2)	66 (3)	44.45***
F0 mean (Hz)	261 (27)	185 (8)	352***
F0 SD	42 (15)	11 (9)	26.44***
Amplitude rms	.10 (.03)	.04 (.01)	36.65***

*** Significance at the .001 level. Means and standard deviations (in brackets) are presented. F0: fundamental frequency. SD: standard deviation; rms: root mean square.

Overall, a full factorial design was created for the CWs, with a combination of two factors: Hand Movement (Beat gesture: BG; Control movement: CM; No hand movement: NM) and Pitch accent (Accented: A; Unaccented: UA). Consequently, six conditions were created for each item (See supplementary materials for examples). The six conditions of the 240 item sets were distributed across six lists according to a Latin square procedure, with each list containing equal number of items (40 items) per condition. In this way, no single participant watched the video from the same item set more than once, and each video was presented across the six experimental lists. In order to obscure the purpose of the experiment, 36 sets of filler items were constructed. These filler items had similar sentence structures to the experimental items, but pitch accent and hand movement occurred in non-critical word positions of sentences. Consequently, there were 276 videos in each experimental list (240 experimental videos, 36 filler videos). Another six lists were created based on these six lists with a reversed trial order. The 12 lists were equally distributed across the 24 participants. The video clips were presented in a pseudorandom order, such that no more than three items of the same condition were presented in succession.

2.2.2. Pretest on beat gesture and control movement

The pretest assessed whether participants treated the beat gesture and control movement as hand movements used for emphasizing information and whether the beat gesture was a more conventionally used hand gesture for emphasizing information than the control movement. Sixteen native Dutch speakers (three males, mean age=20 years) participated in the pretest. The pretest consisted of two runs. In each run, we presented them with two blocks of eight video clips containing either beat gesture or control movement (the order of blocks was counterbalanced across participants). In each video, the hand movement was combined with different spoken sentences in which the critical words were accented.

In the first run, we asked the participants an open question on their view of the function of the hand movement after each block. We found that 14 participants thought both the beat gesture and control movement were used for emphasizing or stressing information. The other two participants failed to give any answer to the open questions (one did not answer and the other one answered 'do not know').

In the second run, we presented the same two blocks of video clips used in the first run, we then asked them which of the two types of hand movement was used in the real life to emphasize information. Eleven out of sixteen participants indicated that the beat gesture was more likely to be used in the real life to emphasize information than the control movement.

To sum up, both beat gesture and control movement were treated as hand movements used for emphasizing information. However, more than 2/3 of the participants believed that the beat gesture is a more conventional gesture used for emphasizing information in the real life.

2.3. Procedure

Each participant sat in a soundproofing room, facing a computer screen approximately 80 cm away. The video clips were presented on the computer screen. The size of the video frame was 10 cm in height and 12 cm in width. Speech from the video clips was presented through loudspeakers. Each trial consisted of 4 consecutive elements: a 3000 ms fixation cross, a 500 ms blank screen, a video clip, and then a 500 ms blank screen. The participants were instructed to carefully listen to the speech and watch the video clips without a specific task. We chose this passive task in order to avoid specific task-induced and undesired strategic comprehension processes by our participants. The absence of a secondary task during language comprehension studies has been used in many ERP studies (e.g., Hagoort and Brown, 2000; Wang et al. 2011), and is generally accepted to be a valid approach to studying language comprehension processes. The participants were asked to blink during the presentation of the fixation cross and not to move or blink during the video presentation.

The 276 items in one list were divided into 12 blocks (23 trials per block), and each block lasted about three minutes. The whole experiment took about 1.5 h, including subject preparation and instructions. There was a short practice block consisting of 10 items before the experimental trials.

2.4. EEG recordings and data analysis

The EEG was recorded in an electromagnetically shielded cabin, with 60 surface active electrodes (Acticap, Brain Products, Herrsching, Germany) placed in an equidistant montage. The left mastoid electrode served as the reference, and a forehead electrode served as the ground. Vertical and horizontal eye movements were monitored by three electrodes placed in the cap and one electrode placed below left eye. All electrode impedances were kept below 20 k Ω during the experiment (which is well below the recommended impedances for active electrodes). EEG data were digitized at a rate of 500 Hz with a 100 Hz high cut-off filter and a 0.016 Hz low cut-off filter.

Brain Vision Analyzer software 2.0 (Brain Products) was used to preprocess the raw EEG data. The EEG data were re-referenced off-line to the average of both mastoids, and band pass filtered at 0.1–30 Hz (24 dB/oct slope). Since the gesture apex was synchronized with the onset of the CW, the hand started moving 13 frames (i.e., 520 ms) before the CW onset and remained still for 40 frames (i.e., 1200 ms) after the CW onset. We took the 100 ms time window preceding the start of the hand movement as the baseline. Thus, the segmentation of the data started 620 ms before the onset of the CW and ended 1200 ms after the onset of the CW (which covered the whole movement period). A semi-automatic artifact rejection procedure was then applied. The data of one participant were removed due to excessive artifacts (more than 50% of trials were rejected). For the remaining participants, 7% of all trials were rejected on average, with rejections being equally distributed across the eight conditions ($F < 1$). Finally, trials were averaged in each condition for each subject, and this average was used for further statistical analysis.

2.5. ERP data analysis

A cluster-based random permutation test was used to assess the statistical difference between two conditions (Maris & Oostenveld, 2007). It was implemented in the Matlab toolbox Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). This approach controls the Type-1 error rate that involves multiple comparisons (one comparison for each electrode and each time point). In the analysis, a simple dependent-samples t test is performed for every data sample (for each electrode in the current analysis). All spatially adjacent data samples exceeding a preset significance level (5% here) are grouped into clusters. For each cluster the sum of the t statistics is used in the cluster-level test statistic. Then a null distribution is created with an assumption of no difference between conditions. This distribution is obtained by 1000 times randomly assigning the conditions in participants and then the largest cluster-level statistic is calculated for each randomization. Finally, the actually observed cluster-level test statistics are compared against the null distribution, and clusters falling in the highest or lowest 2.5th percentile are considered significant.

Since this statistical analysis only allows for pair-wise comparison, we tested the main effect of Pitch accent by the comparison of A vs. UA (collapsed over Hand movement conditions). Similarly, the main effect of Hand movement was tested by the comparisons of each of the two Hand movement conditions that collapsed over Pitch accent conditions: BG_(A+UA) vs. CM_(A+UA), NM_(A+UA) vs. BG_(A+UA), as well as NM_(A+UA) vs. CM_(A+UA). In addition, the interaction between Pitch accent and Hand movement was acquired by comparing the difference waveforms between the A and UA conditions (A–UA) between each of the two Hand movement conditions: BG_(A–UA) vs. CM_(A–UA), NM_(A–UA) vs. BG_(A–UA), and NM_(A–UA) vs. CM_(A–UA). If any of the interaction was significant, further simple effect analysis would be conducted. After visual inspection of the data, three time windows were selected for the statistical analysis. (1) A positivity between –270 and 200 ms relative to the CW onset (corresponding to the time window of 250–720 ms relative to the hand movement onset). (2) The N400 time window: 200–800 ms relative to the CW onset. (3) A positivity in the time window of 800–1200 ms relative to the CW onset. The mean amplitudes within the three selected time windows of all 59 electrodes (the reference electrode was excluded from the analysis) were entered into the analysis.

3. Results

The grand average waveforms elicited by different conditions at nine representative electrodes (45/43/41, 59/30/27, 13/11/9, encircled in the head model in Fig. 2) are presented in Fig. 2 (collapsed across Pitch accent conditions) and Fig. 3 (collapsed across Hand movement conditions) separately. See the supplementary Figs. 1 and 2 for the effects of Hand movement on the Accented and Unaccented words respectively. Note that the waveforms showed no clear N1-P2 complex because that all the waveforms were time locked to the hand movement onsets, and the duration of auditory CWs varied across trials.

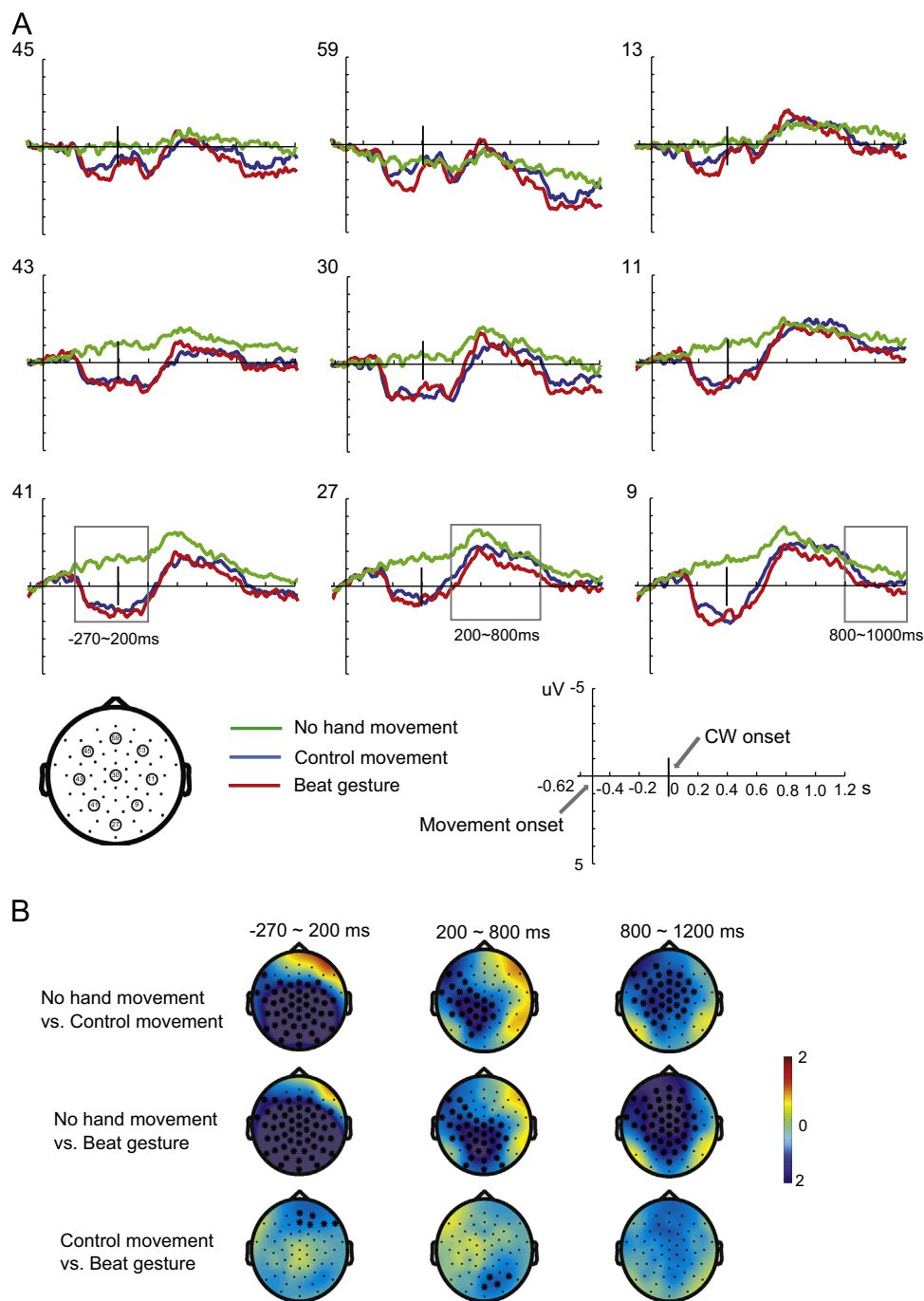


Fig. 2. Hand movement effects. (A) Grand averaged waveforms evoked by the utterance as a function of Hand movement at nine representative electrodes, which were circled in the head model. The green lines represent the No hand movement condition, the blue lines represent the Control movement condition, while the red lines represent the Beat gesture condition. Waveforms are time-locked to the onset of the hand movements, and the CW onset was marked as zero. The time windows in which statistical analysis were performed were marked in gray boxes. CW: critical word (ERP waveforms showing Hand movement effect). (B) Topographies showing the average voltage differences for the different contrasts for the indicated time intervals. The electrodes that showed significant differences were marked by* (Scalp topographies showing Hand movement effects). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. The positivity between –270 and 200 ms relative to the CW onset

In this time window, the BG and CM conditions elicited larger positivities than the NM condition over the whole scalp (both $p_s < .001$), which was larger for the BG than for the CM condition over the right anterior region ($p = .05$; see Fig. 2B for the distribution).

No significant difference was found between the Accented and Unaccented conditions ($p = .149$). Also, no significant interaction

was found between Pitch accent and any of the three Hand movement conditions: $BG_{(A-UA)}$ vs. $CM_{(A-UA)}$; $p = .118$; $NM_{(A-UA)}$ vs. $BG_{(A-UA)}$; no significant cluster; $NM_{(A-UA)}$ vs. $CM_{(A-UA)}$; no significant cluster.

3.2. The negativity in the N400 time window between 200 and 800 ms

The waveforms in this time window showed a broader morphology than the classical N400s obtained in studies where stimuli were

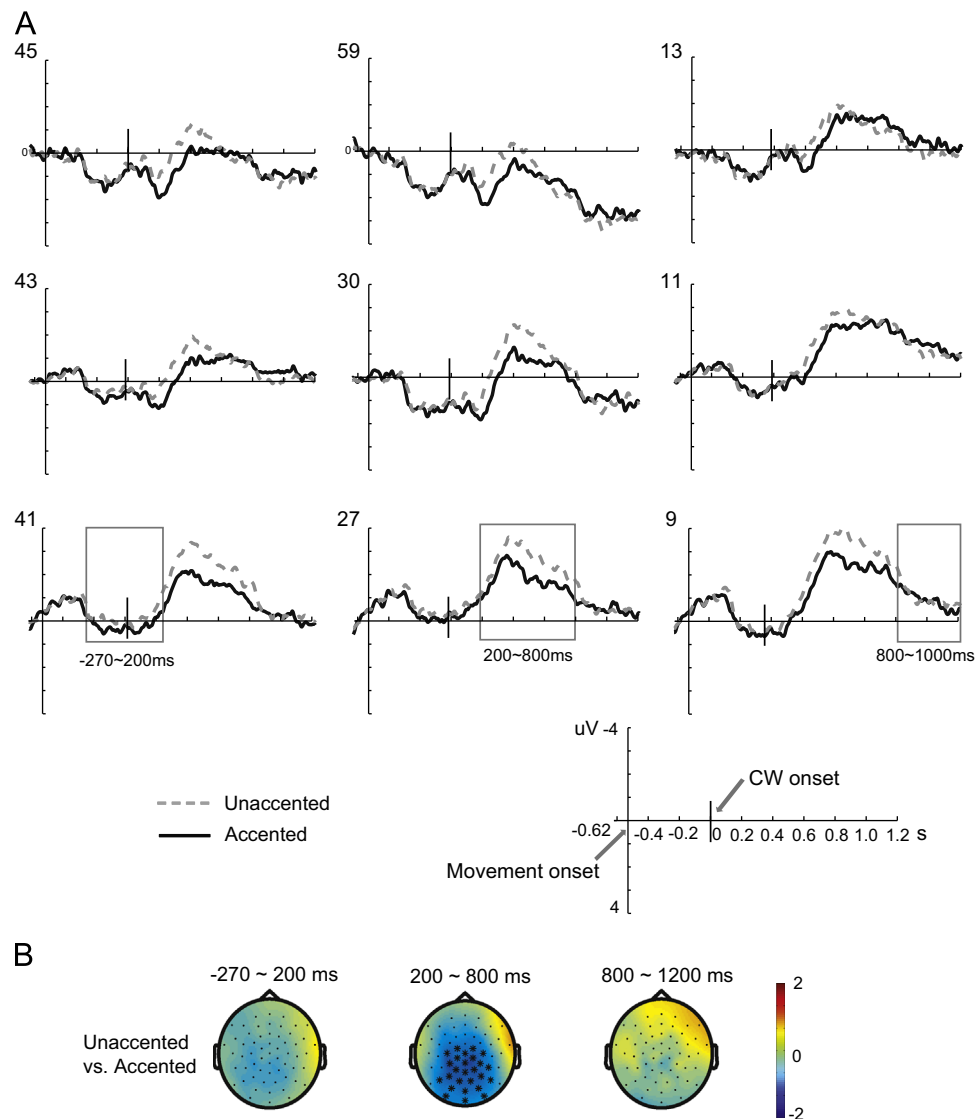


Fig. 3. Pitch accent effects. (A) Grand averaged waveforms evoked by the utterance as a function of Pitch accent at nine representative electrodes. The solid black lines represent the Accented condition, while the dotted gray lines represent the Unaccented condition. Waveforms are time-locked to the onset of the hand movements, and the CW onset was marked as zero. The time windows in which statistical analysis were performed were marked in gray boxes. CW: critical word (ERP waveforms showing Pitch accent effect). (B) Topographies showing the average voltage differences for the contrast of Unaccented vs. Accented, for the indicated time intervals. The electrodes that showed significant differences were marked by * (Scalp topographies showing Pitch accent effects).

presented visually (Kutas & Federmeier, 2011). This might be related to the auditory presentation of the stimuli, where the exact duration of auditory stimuli varies across trials. Such broad N400s have also been reported in other auditory studies (e.g., Hagoort & Brown, 2000; Li, Hagoort, & Yang, 2008; Wang et al., 2011).

In this time window, both the BG and CM conditions elicited smaller negativities than the NM condition over slightly left lateralized central and posterior regions ($p < .001$ and $p = .002$, respectively for the BG and CM conditions). Moreover, the BG condition elicited a smaller negativity than the CM condition over the right posterior region ($p = .05$; see Fig. 2B for the distribution).

As for the main effect of Pitch accent, we found a significantly smaller negativity for the Accented compared to the Unaccented condition over the central and posterior regions ($p = .01$; see Fig. 3B for the distribution). No significant interaction was found between Pitch accent and any of the three Hand movement conditions: $BG_{(A-UA)}$ vs. $CM_{(A-UA)}$: no significant cluster; $NM_{(A-UA)}$ vs. $BG_{(A-UA)}$: no significant cluster; $NM_{(A-UA)}$ vs. $CM_{(A-UA)}$: no significant cluster.

3.3. The time window of 800–1200 ms

In the time window of 800–1200 ms, the BG and CM conditions elicited larger positivities than the NM condition ($p < .001$ and $p = .001$ respectively for the BG and CM conditions). However, no significant difference was found between the BG and CM conditions ($p = .168$).

In addition, no difference was found between the Accented and Unaccented conditions ($p = .115$). Neither was the interaction between Pitch accent and any of the three Hand movement conditions significant: $BG_{(A-UA)}$ vs. $CM_{(A-UA)}$: $p = .101$; $NM_{(A-UA)}$ vs. $BG_{(A-UA)}$: $p = .189$; $NM_{(A-UA)}$ vs. $CM_{(A-UA)}$: no significant cluster.

4. Discussion

This study examined the effect of beat gesture and pitch accent on semantic processing during language comprehension. In the time window of 250–720 ms after the initiation of hand movements (from –270 to 200 ms relative to the onset of critical

words), both beat gesture and control movement elicited larger positivities than no hand movement over the whole scalp, and beat gesture yielded a larger positivity than control movement over right anterior region. In the time window of 200–800 ms after the critical word onset, both beat gesture and control movement elicited smaller negativities than no hand movement condition over left-lateralized central and posterior regions. Interestingly, beat gesture elicited smaller negativities than control movement over right posterior region. In addition, the accented words elicited smaller N400s than the unaccented words over central and posterior regions. Finally, no interaction was found between beat gesture and pitch accent in any region in any time window.

4.1. How does beat gesture facilitate semantic processing

The critical words accompanied by beat gesture and control movement elicited reduced negativities in the N400 time window relative to the words paired with no hand movement over the left-lateralized central and posterior regions. Although the distribution of the N400 effect (i.e. the difference in N400 amplitudes between conditions) varied with different types of stimuli, the N400 was generally shown to vary as a function of the ease of integration of words into context (for a review see [Kutas & Federmeier, 2011](#)). Therefore, our results indicate that both beat gesture and control movement facilitated semantic processing. The results are in line with a previous study where question context was used to mark a certain word to be focus or not ([Wang et al. 2009](#)). It was found that compared to the non-focused words, the focused words elicited smaller N400s, which suggests that the focused words were easier to be integrated into context than the non-focused words. In the current study, the pretest indicated that both beat gesture and control movement were perceived as hand movements used for signaling the saliency of the accompanying words, so they both facilitated the semantic integration of the critical words into context, probably by attracting more attentional resources to the accompanying words.

The idea that beat gesture modulates attentional resources was also advocated by [Biau and Soto-Faraco \(2013\)](#). [Biau and Soto-Faraco \(2013\)](#) proposed that seeing beat gesture helps the listener to direct attention to important information. However, they only reported an effect of beat gesture on the accompanying words at early auditory processing stage prior to semantic processing stage (they only analyzed the time window of –100ms to 500 ms relative to critical words). In the current study, we found both early (the P300 effects relative to hand movement onset) and late (the N400 effects for the accompanying words) influences of beat gesture. The influence of beat gesture on the higher-level speech processing (e.g., syntactic analysis) was also reported in [Holle et al., \(2012\)](#).

Furthermore, the words accompanied by beat gesture elicited a smaller N400 than those paired with control movement over the right posterior region. This result suggests that although rhythmic movements, such as the beat gesture and the control movement used in the present study, can capture attention and enhance semantic processing to a certain extent, beat gesture has a unique role in highlighting information presumably due to its conventional use as a gesture for emphasizing information. This is also reflected in our pretest that although the participants' rating on the emphasis did not differ between the beat gesture and control movement, more than 2/3 of the participants indicated that the beat gesture was used in the real life to emphasize information.

As for the specific role of beat gesture comparing to non-beat movements, our result is in line with the study of [Hubbard et al. \(2009\)](#), in which the authors found stronger activations in the left

superior temporal gyrus/sulcus when speech was accompanied by beat gesture than when speech was accompanied by meaningless hand movements. However, our results were not consistent with the ERP study by [Holle et al. \(2012\)](#). They found that unlike beat gesture, moving dots that followed the exact trajectory of the beat gesture had no effect on the syntactic structure analysis. This discrepancy may arise from the different types of control movements used in the different studies. In both the present study and the study of [Hubbard et al. \(2009\)](#), the control movement was executed by the same actor in the video clip, so the participants could actively attend to information from both the auditory and the visual modalities. In contrast, in the study of [Holle et al. \(2012\)](#), the control movement condition differed from the beat gesture condition from the beginning of the spoken sentence because a red moving dot rather than an actor was presented. Thus, participants might have ignored information from the visual modality during speech processing.

4.2. Beat gesture and pitch accent modulate semantic processing independently

In addition to beat gesture, pitch accent also elicited an N400 effect, with a smaller N400 for the accented than for the unaccented words. This finding is in line with other studies on the role of pitch accent in semantic processing ([Hruska & Alter, 2004](#); [Li et al. 2008](#); [Magne et al. 2005](#); [Toepel, Pannekamp, & Alter, 2007](#); [Li & Ren, 2012](#)). The reduced N400 triggered by pitch accented words might indicate easier semantic processing as the accented words obtained more attentional resources ([Kristensen et al., 2012](#)).

No interaction was found between beat gesture and pitch accent during semantic processing. Therefore, although there is a tight temporal alignment between beat gesture and pitch accent ([Leonard & Cummins, 2010](#)), they seem to influence semantic processing independently. The lack of interaction between them in the current study suggests that the visual (gesture) and auditory (pitch accent) cues trigger the attentional system separately for semantic processing. Moreover, the results suggest that there is no one to one mapping between beat gesture and pitch accent in unconstrained sentences. The results might be due to the fact that beat gesture and pitch accent do not have to co-occur to provide emphasis. That is, visual emphasis (e.g., beat gesture) is not necessarily accompanied by auditory stress (e.g., pitch accent), and vice versa.

Finally, it is worth pointing out that beat gesture and pitch accent were manipulated in isolated sentences in the present study. Since beat gesture and pitch accent often occur within multiple-sentence contexts, their relationship might be modulated by the constraint of previous context. This remains to be tested in future studies.

4.3. The role of beat gesture in early visual processing

After the initiation of hand movements, both beat gesture and control movement elicited larger positivities than no hand movement over the whole scalp. We identified them as P300 effects based on their morphology (positive shifts), latency (250–720 ms) and scalp distributions (parietal–central maximum; [Nieuwenhuis, De Geus, & Aston-Jones, 2011](#)). The P300 effect is typically elicited by salient, unexpected, novel and task-relevant stimuli (for a review, see [Nieuwenhuis et al., 2011](#)). In the current study, the participants saw no hand movement in some trials, and in other trials they saw a hand movement at different positions of the sentences. Therefore, participants could not predict when and where a hand movement would appear. The appearance of a hand movement led to increased P300s. This saliency detection might

have involved the activation of motion-sensitive brain regions, as it has found that motor related regions are often activated in the observation of others' action (Hari et al. 1998).

In addition, beat gesture elicited a larger P300 effect than control movement over the right anterior region. Since beat gesture and control movement differ in their form from the initiation of their movements (beat gesture: moving up and down; control movement: moving from right to left), the ERP difference between them around –270 ms relative to the CW onset (corresponding to 250 ms relative to the movement onset) may indicate that even before the critical word started, beat gesture may play a special role in signaling significant stimuli due to its conventional use for highlighting information (as indicated in the pretest).

Since the critical word onset was aligned with the gesture apex, different ERPs (the P300 effects) were already present between the hand movement and no hand movement conditions before the onset of the critical words. Therefore, the observed N400 effects might overlap with the preceding P300 effects. Since the overlapping components reflected the summed effects of hand movements on audio-visual integration, the negative effects were the effects of our interest. Interestingly, although the absolute amplitudes of the negativities were smaller for the two movement conditions than for the no movement condition, visual inspection indicates that the onset-to-peak amplitude of the negativities was larger for the two movement conditions. As the ERPs largely reflect time- and phase-locked EEG signals to the eliciting event (see Bastiaansen, Mazaheri, & Jensen, 2012 for a more detailed discussion), the larger onset-to-peak amplitudes might indicate a process of phase-resetting under the influence of hand movements. Thorne, De Vos, Viola, and Debener (2011) have shown that when visual input reliably predicted the following auditory input, the visual input could prepare the auditory system for subsequent auditory processing by resetting the phase of neuronal oscillatory activity in the auditory cortex. Therefore, the hand movements might also facilitate semantic processing by resetting the phase of brain oscillations to prepare for the incoming salient information.

5. Conclusions

We provided ERP evidence that beat gesture facilitates semantic processing of a word during single sentence comprehension. The lack of an interaction between beat gesture and pitch accent suggests that they exert their influence independently. In addition, although both beat gesture and control movement with similar rhythmic properties could capture listeners' attention and reduce N400, beat gesture has a unique role in facilitating semantic processing, probably because it is a more conventionally used cue for highlighting speech information.

Acknowledgments

This research was supported in part by the Natural Science Foundation of China [Grant number: 31200849] to Lin Wang. We would like to thank Peter Hagoort, Louise Schubotz, Lucy Foulkes, Judith Holler, and Marcel Bastiaansen for their valuable comments on our paper.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.neuropsychologia.2013.09.027>.

References

- Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.
- Bastiaansen, M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In: S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31–50). New York: NY: Oxford University Press.
- Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, 18, 438–462.
- Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, 124(2), 143–152.
- Boersma, P., & Weenink, D. (2002). *Praat 4.0: a system for doing phonetics with the computer [computer software]*. Amsterdam: Universiteit van Amsterdam.
- Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., et al. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition*, 70, 42–52.
- 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(3), 228–242.
- Grice, H. P. (1975). Logic and conversation. In: P. C. a. J. L. Morgan (Ed.), *Speech acts: Syntax and semantics*. New York: Academic Press.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38(11), 1531–1549.
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: a neuro-magnetic study. *Proceedings of the National Academy of Sciences*, 95(25), 15061–15065.
- 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., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage*, 39, 2010–2024.
- 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, 1–12.
- Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, 49, 875–884.
- Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, 33, 73–88.
- 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: Walter 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.
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40, 577–592.
- Kelly, S. D., Creigh, P., & Bartolotti, J. (2009). Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*, 22, 683–694.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253–260.
- Kelly, S. D., & Goldsmith, L. H. (2004). Gesture and right hemisphere involvement in evaluating lecture material. *Gesture*, 4(1), 25–42.
- 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.
- Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, 31(6), 533–552.
- Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2012). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex* (Advance online publication).
- 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(1), 621–647.
- Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205.
- Leonard, T., & Cummins, F. (2010). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26, 1457–1471.
- Li, X.-q., & Ren, G.-q. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study. *Neuropsychologia*, 50, 1882–1894.
- Li, X., Hagoort, P., & Yang, Y. (2008). Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese. *Journal of Cognitive Neuroscience*, 20, 906–915.

- Magne, C., Astesano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of “pop-out” words in spoken French dialogues. *Journal of Cognitive Neuroscience*, 17, 740–756.
- 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.
- Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, 48, 162–175.
- 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.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806.
- Özyürek, A., Willems, R. M., 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.
- Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics*, 30, 629–654.
- Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *The Journal of Neuroscience*, 31, 3853–3861.
- Toepel, U., Pannekamp, A., & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions*, 3, 53.
- 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., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, 1282, 50–56.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, 17, 2322–2333.
- 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, 1992–2004.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42, 654–667.