



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Action-speech and gesture-speech integration in younger and older adults: An event-related potential study

Kim Ouwehand^{a,*}, Jacqueline de Nooijer^{a,b}, Tamara van Gog^{a,c}, Fred Paas^{a,d}

^a Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, the Netherlands

^b Rotterdam School of Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, the Netherlands

^c Department of Education, Utrecht University, Heidelberglaan 1, 3584 CS, Utrecht, the Netherlands

^d School of Education/Early Start, University of Wollongong, Northfields Avenue, Wollongong NSW, 2522, Australia

ARTICLE INFO

Keywords:

Action-speech integration
 Gesture-speech integration
 Age
 N400

ABSTRACT

In daily communication, speech is enriched with co-speech gestures, providing a visual context for the linguistic message. It has been shown that older adults are less sensitive to incongruencies between context (e.g., a sentence) and target (e.g., a final sentence word). This is evidenced by a smaller and delayed N400 (in)congruency effect that reflects the difference between the N400 component in response to congruent versus incongruent targets. The present study investigated whether the effect of age on the N400 effect in sentence-final word integration would also arise for verb-gesture/action integration. Assuming that gestures have a tight connection to language these would provide a higher contextual constraint for the action phrase than the literal actions (i. e., an action performed on an object can be understood in isolation, without the action phrase). EEG was recorded from a sample of younger and older participants, while they watched audio-visual stimuli of a human actor performing an action or pantomime gesture while hearing a congruent or incongruent action phrase. Results showed that the N400 (in)congruency effect was less widespread in the older than the younger adults. It seemed that older adults, but not younger adults were less sensitive to the gestural than the action (object) information when processing an action phrase.

Gestures are used in everyday communication and thinking and can take many forms. McNeill (1992) categorized these communicative co-speech gestures into iconic gestures (also called representative gestures), metaphoric gestures, deictic gestures, and beats. Iconic gestures are gestures that represent features of a concrete event, object or action, for example moving the hands following a specific contour to communicate about the shape of an object. Metaphoric gestures refer to abstract concepts and relations, for example raising and lowering your hand with palm down to index a certain hierarchy. Deictic gestures are used to index locations and pathways, such as pointing and tracing with the index finger. Beats are used to put emphasis on a particular word/part of a spoken message, for example, a fist on the table. However, some types of gestures are linked less to speech, such as pantomime gestures. Pantomime gestures, representing an action quite literally, can often be understood in the absence of speech and can be used when speaking is difficult (for example in a noisy environment, see Holle et al., 2010)

Gestures and actions are often linked in the literature, suggesting gestures being simulated actions (Hostetter & Alibali, 2008, 2019)

* Corresponding author.

E-mail address: ouwehand@essb.eur.nl (K. Ouwehand).

<https://doi.org/10.1016/j.jneuroling.2022.101100>

Received 22 September 2021; Received in revised form 13 July 2022; Accepted 22 July 2022

Available online 18 August 2022

0911-6044/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

or predictive actions (Pouw, & Hostetter, 2016). However, not much research has been done to directly compare action-versus gesture-speech processing. There is one study by Kelly et al. (2015) in which this comparison was made in younger adults. In this study, participants were primed with a written action word that was followed by a video in which they saw an actor perform an action on an object or pantomime the action while listening to a spoken action phrase. Half of the participants had to relate the prime to the visually perceived gesture/action and the other half to the heard action phrase. The only difference between the actions and the pantomime gestures was the presence or absence of an object (i.e. in the action condition the action was performed on an object while in the gesture condition the action was pantomimed). In this way, the paradigm allowed for a controlled comparison between actions and gestures. Overall, participants were more accurate in judging the (in)congruency between written word primes and the actions than between the primes and the gestures. However, when they had to judge the (in)congruency between the prime words and the spoken action phrases, observing an incongruent gesture led to worse task performance than an incongruent action, showing that participants were more distracted by incongruent gestures than incongruent actions. The authors suggested that gestures might be perceived as more relevant to the speech they accompany, and are therefore integrated faster with the speech than the actions.

Whether these results would also apply to older adults is questionable, because research on gesture- and action-speech processing in older adults is scarce. However, gesture research suggests that the responsiveness to gestures differs between older and younger adults (Cocks et al., 2011; Montepare et al., 1999; Thompson, 1995; Thompson & Guzman, 1999). For example, Cocks et al. (2011) compared the capability to integrate speech and gestures between younger and older adults. Participants watched videos under three conditions, verbal only, gesture only and a combination. In response to these videos they had to choose one out of four pictures (one matching the verbal message only, one matching the gesture only, one matching the combination and one unrelated picture) that they thought matched the video best. Using a specific algorithm subtracting the preference for visual/gestural information from the average responses to the gesture-speech combinations (for the exact formula and procedure, see Cocks et al., 2011), it was assessed whether gestures in addition to speech were beneficial for sentence comprehension. It was found that the gesture-benefit for comprehending gesture-speech combinations was smaller for older than younger adults, and that older adults did not even seem to attend to the gestures at all. These findings suggest that the beneficial effect of co-speech gestures in sentence comprehension for younger adults may not apply to older adults. In line with this, earlier research by Thompson (1995) also showed that co-speech gestures on immediate recall of heard sentences benefited younger, but not older adults. In another study, using a dichotic listening task in which participants perceived different auditory information in each ear, gestures aided younger, but not older adults (Thompson & Guzman, 1999). Participants received different verbal stories in each ear simultaneously and had to shadow (repeat aloud what was said in) one particular ear. The story that participants had to attend to was accompanied by a video showing an actor doing nothing, or making articulatory movements with the face or making articulatory movements with the face combined with iconic gestures. While the shadowing improved in the articulatory and gesture condition for the younger participants, this effect was absent for the older adults. That older adults have more trouble gleaning information from accompanying movements, also became apparent in a study by Montepare et al. (1999) who showed that older adults were less able than younger adults to infer emotional states from body movements.

When studying the online processing of gestures, the event-related potential (ERP) called the N400 component is often measured (Kelly et al., 2007; Wu & Coulson, 2005; Özyürek et al., 2007). The *N400 component* is a negative ongoing wave peaking approximately 400 ms after stimulus onset (Luck, 2014). Early explanations of the N400 component state that it reflects integrative processes (Holcomb, 1993); the more effort the integration (i.e., between different features of a stimulus or between different stimuli) takes, the more negative the N400 component. The N400 component is well-studied in language comprehension studies comparing congruent and incongruent words in a sentence context (Kutas and Federmeier, 2011). The *N400 effect* reflects the difference between the mean amplitudes of the N400 in response to congruent compared to incongruent (compound) stimuli. The *N400 onset* or *N400-effect onset* reflects the onset of the negative waveform or the onset of the difference wave (congruent – incongruent), respectively.

Several factors can influence the strength and the timing (onset) of the N400- component and effect, such as age of the participants and characteristics of the stimuli. For example, older adults typically show a delayed *onset* of the N400- component and effect (e.g., Federmeier et al., 2010; Federmeier & Kutas, 2005) and a smaller N400 effect (e.g., Federmeier et al., 2010; Federmeier & Kutas, 2005) compared to younger adults. These findings indicate that older adults are less sensitive and slower to detect the difference between the (in)congruency or are less able to use (integrate) context in sentence processing (Payne & Federmeier, 2018). In addition, the onset of the N400 effect occurs somewhat later than the traditional N400-effect repeatedly shown for sentence processing, when studying actions (N420, Proverbio & Riva, 2009) and (iconic) gestures (N450, Wu & Coulson, 2005). This might be because information from two modalities (i.e., multimodal information) need to be integrated which might take some more time.

The strength of (in)congruency could stem from a purely semantic fit of the target words in the sentence context (i.e., plausibility), but also the concreteness of the sentence context. For example, Federmeier and Kutas (2005) showed that the N400 effect is larger for unexpected final words in sentences that provide a highly constraint (concrete) context (e.g., She went for a long run on the beach on her new ...) than a weakly constraint (more abstract) context (e.g., She was walking down the street when she saw a ...). Besides purely linguistic sentence processing, the N400 effect has also been investigated in the context of action and gesture processing, where it has been found to occur with (in)comprehensible actions (Proverbio & Riva, 2009), (in)congruent actions (De Nooijer et al., 2016), functional and orientational (mis)matches between actions and objects (Bach et al., 2009), and (in)congruency between co-speech (iconic) gestures and speech (Holle & Gunter, 2007; Kelly et al., 2010; Özyürek et al., 2007).

The present study compared the processing of multimodal action-speech and (manual) gesture-speech stimuli in younger and older adults by looking at the N400 effect. Pantomime gestures are the gestures used in the present study because they allow for a controlled comparison with actions (with the presence or absence of an object as the only difference). We suggest that on the one hand, perceiving pantomimes provides a less concrete/constraint context for action phrases (less predictive power), than perceiving the literal action

(on an object). Based on this suggestion, a larger N400 effect for the action-speech versus gesture speech combinations would be a logical expectation. On the other hand, looking at the study of Kelly et al. (2015), in terms of reaction times, young adults were slower in response to incongruent gesture-speech combinations as compared to action-speech combinations. This indicates a higher interference of the incongruity between gestures and speech than actions and speech. The authors explain these findings by saying that gestures are perceived as more communicatively intended and therefore receive more attention when combined with speech. Following these outcomes it would be more logical to expect a larger N400 effect for the gesture-speech versus action speech combinations. Because of the special relation of gestures to speech and an assumed automatic integration of gestures and speech (Kelly et al., 2010), this second expectation is followed in the present study. More support for this expectation is provided by Pouw and Dixon (2019) who also found evidence for a strong tendency of the human cognitive system to integrate gestures with speech. In their experiment the authors invoked a delayed auditory feedback of speech of 150 ms in participants retelling a cartoon they had just watched. Such delayed auditory feedback makes speaking difficult, leading to slurred or intermittent speech. It was found that delayed feedback versus natural undelayed feedback resulted in gesture strokes that adapted according to the timing of the peak pitch of co-gesture speech, even more so than the undelayed condition. The researchers conclude that this stable gesture-speech synchrony when speech is made difficult shows a tight and continuous interrelation between gestures and speech.

In the present study we conducted a controlled direct comparison between the processing of gestures and actions for younger and older adults, which has (to the best of our knowledge) not been done yet. As in the study of Kelly et al. (2015) we used pantomime gestures because these simulate the actions with the only difference of an object not being present, allowing for a very controlled comparison between the conditions. Furthermore, it has been said that pantomime gestures could be understood in isolation just as actions and might be processed differently than iconic gestures that are more often used in communication. Indeed, Willems et al. (2009) showed that pantomime gestures and iconic gestures elicit partly differential neural activity. However, reviewing the studies that have been conducted on iconic gestures and the N400, there appears to be an overlap between what is defined as iconic or pantomime. For example, the figures in Holle et al. (2010) show an actor make a cheese grating gesture that is very similar to the actual movement that needs to be made to grate cheese. Similarly, in Holle et al. (2008), gestures made for touching a mouse (either the animal or the device) looked very similar to the action movement that would be performed on the device/animal. Comparing these examples with our example in Fig. 1 we feel that a fair number of our stimuli are quite comparable to the ones described above. Therefore, these previous findings together with the literature on aging, were used to formulate the following hypotheses:

H1. Because of the tight interrelationship between gestures and the linguistic message they accompany, we expected a larger N400 effect for gesture-speech combinations than action-speech combinations (i.e., an interaction between type of movement and congruency effect).

H2. Based on previous findings showing a smaller N400 effect in older versus younger adults (for a review, see Wlotko et al., 2010), we expected that older adults would also show a smaller N400 effect than younger adults in general (i.e., main effect of age).

H3. Based on the literature on aging and the N400 effect (Federmeier & Kutas, 2005; Payne & Federmeier, 2018) and gestures and the N400 effect (Kelly et al., 2007) we expected the difference between N400 effects of the action-speech and gesture-speech combinations to be smaller for older adults than younger adults (i.e., interaction between age, type of movement and congruency effect).

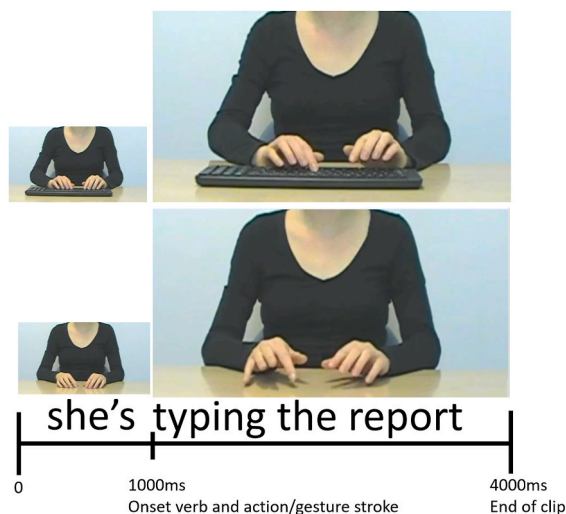


Fig. 1. Schematic Representation of a Trial Showing Snap Shots of one of the Action- and Gesture Movie Aligned with the Presentation of the Sentence.

1. Method

This study was conducted in accordance with the protocol of the ethical committee at the Department of Psychology, Education and Child studies, at a Dutch university.

1.1. Participants and design

Participants were 41 younger adults (22 women; $M_{age} = 21.7$ years, $SD = 2.5$) and 39 older adults (24 women; $M_{age} = 65.9$ years, $SD = 4.5$). The younger adults were University students who participated in this study as part of a course requirement. The older adults were recruited via advertisements in local newspapers and a special course program for older adults at our university (higher education for older adults program) and community centers. Within the sample of older adults, there was more variety in educational levels (14 participants obtained a Master's degree 13 a Bachelor's degree, two completed pre-university education, one completed higher general secondary education and six lower general secondary education). All participants were native speakers of Dutch, had normal or corrected-to-normal vision, no known neurological disorders, and were naïve to the experimental conditions.

In a within-subjects design, participants were presented with spoken action phrases combined with videos that were (in)congruent with the action phrase in two ways: (1) The *action* performed by the actor was congruent (i.e., hearing "s/he is typing the report" and seeing the actor typing on a keyboard) or incongruent with the content of the action phrase (i.e., hearing "s/he is typing the report" but seeing the actor with a knife "slicing a tomato"). (2) The pantomime *gesture* was congruent or incongruent with the content of the action phrase in the same way. This resulted in four conditions; (1) congruent action-speech combinations, (2) incongruent action-speech combinations, (3) congruent gesture-speech combinations, (4) incongruent gesture-speech combinations.

1.2. Pre-tests

A separate sample of participants ($N = 30$, $M_{age} = 35.6$, $SD = 15.2$) were presented with the videos of the gestures and actions in isolation and had to indicate which action verb they recognized. Subsequently, these videos were presented with the congruent action/gesture speech combinations and they had to indicate on a five-point scale how well they thought the action/gesture matched the action verb. Action and gesture-videos were considered acceptable for use in the present study when more than 60% of the participants indicated the correct action and at least a score of three was given for the match between action/gesture and action verb. This resulted in a selection of 40 videos of the actions and 40 videos of the corresponding gestures for the final materials. For this set, the worst recognized video had 65% of the participants with a correct indication of the action depicted and the lowest matching score was 3.30 (on a five-point Likert scale).

Next, another separate sample of participants ($N = 42$) rated all the compound stimuli consisting of the videos of the actions and gestures paired with the congruent and the incongruent action phrases used for the experiment. Unfortunately due to a programming error, age was not logged. However, all participants were from the same pool as the younger adults in the experiment, so we assume age is comparable to that sample. Participants were asked to rate the semantic congruency of all pairs. A 2 x 2 repeated measures ANOVA was conducted with type of movement (action vs. gesture) and congruency (congruent vs. incongruent) as within-subject factors. A main effect was found for movement, $F(1, 41) = 18.84$, $p < .001$, $\eta_p^2 = 0.32$, congruency $F(1, 41) = 2660.58$, $p < .001$, $\eta_p^2 = 0.99$ and an interaction between movement and congruency, $F(1, 41) = 48.46$, $p < .001$, $\eta_p^2 = 0.54$. These results indicate the congruency effect was smaller for the gesture stimuli than the action stimuli.

1.3. Materials

In total 54 verbs were used to create 108 videos, with half of the videos (i.e., 54) showing actions depicting the verbs and the other half showing gestures depicting the verbs. In the videos the actor was sitting behind a table or standing and executed or pantomimed the denoted action. All videos were recorded in a sound-attenuated room, spoken at a normal rate by a native Dutch speaker. To eliminate any effects of the gender of the actor, two versions were created: (1) a version with a male voice and a male actor and (2) a version with a female voice and a female actor. The videos were edited in Final Cut Pro in which the length of the videos was controlled; all videos had a duration of four seconds and the onset of the verb was aligned with the onset of the action/gesture stroke starting exactly 1000 ms after the onset of the video. Similarly to the study of Kelly et al. (2015) all movements were carefully created in that the movement in the action and gesture videos were as similar as possible, with the only difference that in the action videos the actions were performed on objects and in the gesture videos no objects were present. For an example of the stimuli see Fig. 1.

The experiment consisted of four blocks (two blocks with action-speech combinations and two blocks with gesture-speech combinations). Each block consisted of 40 trials (twenty congruent combinations, twenty incongruent combinations) with each video being presented once in each block. In the congruent action-speech and gesture-speech combinations, the action or gesture was coupled with a matching action phrase, but for the incongruent combinations, the actions/gestures were coupled with a mismatching action phrase.

Incongruent action-speech combinations and gesture-speech combinations were unique in that each incongruent stimulus appeared once so that each action/gesture was coupled with a different action phrase (e.g., if the mismatching video for the sentence: "S/he is typing the report" was "slicing a tomato", then the mismatching video for the sentence "S/he is slicing the tomato" would not be "typing the report"). This procedure is similar to the one used in e.g., Bach et al. (2009) and Friedrich and Friederici (2004) and was used to prevent participants from guessing what the mismatching video would be. Sentences always consisted of the same four elements; the third-person pronoun "he" or "she" followed by a manual action verb, a definite article and an object (e.g., "Z/Hij typt het

verslag” meaning “*S/he is typing the report*”). Given that Dutch is an SVO (subject-verb-object) language, the verb always appeared in the second position. This resulted in video stimuli with a duration of 4 s in which the critical stimulus onset was 1000 ms after the video onset (see Fig. 1). Verbs had an average of 5.1 phonemes ($SD = 1.30$). In the end this resulted in 80 action-speech pairs (40 congruent and 40 incongruent) and 80 gesture-speech pairs (40 congruent and 40 incongruent) from which we made four blocks of 20 congruent and 20 incongruent stimuli. Each block had a duration of approximately 3 min (40×4 s videos and + 584 ms average response time younger and 784 ms average response time older adults, see Table 1 for all means and standard deviations). After each block, participants could take a short break and were asked to press a key when they were ready to continue. Recording sessions took approximately 15–20 min depending on individual response times and duration of the self-paced breaks.

Within blocks, types of movement and gender of the actors were not intermixed, to keep the task presentation as simple as possible. Rationale behind this is that we feared this could influence performance, hence our results in an aversive manner, especially for the older adults. For example, when randomly facing different genders and types of movement, participants need to update information irrelevant for the main task. This reasoning is in line with research showing that older adults have trouble updating their (working) memory (e.g., De Beni & Palladino, 2004) and are more susceptible to irrelevant information (easier distracted, especially under arousal, e.g., Gallant et al., 2020).

1.4. Procedure

The EEG experiment was implemented in E-Prime (Schneider et al., 2002). During EEG recording, participants were seated in a comfortable chair and were instructed to minimise movement and eyeblinks during the videos. Task instructions were built in the E-prime file and preceded the practice phase. While reading these instructions, the experimenter was present and checked with the participant whether they understood what they had to do (i.e. index whether video and the audio presented the same action). During the practice phase we checked the sound and asked the participants explicitly whether they could hear and understand the stimuli clearly. All participants stated that they could hear the spoken stimuli clearly. The practice phase consisted of eight different critical action phrases and actions and gestures from the ones used in the main part of the experiment. Also, the experimenters observed the participants during the practice phase and watched their responses to see whether the task was correctly performed. If too many errors were made, practice phase was repeated until the task was performed correctly. Participants were randomly assigned to the experimental version with the male or female actor. In each version, they received four experimental blocks (two blocks with action-speech combinations and two blocks with gesture-speech combinations). Note that the experiment consisted of 40 trials per condition (action congruent, action incongruent, gesture congruent and gesture incongruent). In each block 20 congruent and 20 incongruent trials were presented randomly, but type of movement was not varied within block (see Table 2 for an overview of the block/stimuli presentation). Whether participants started with the two action blocks or the two gesture blocks and were presented with a male or female actor and voice, was counterbalanced across four lists, resulting in eight versions of the experiment (four for the male and four for the female version). Versions were: male/female a) action block 1-action block 2-gesture block 1-gesture block 2, b) gesture block 1-gesture block 2-action block 1-action block 2, c) action block 2-action block 1-gesture block 2-gesture block 1, d) gesture block 2-gesture block 1-action block 2-action block 1. Within all blocks, the action/gesture-speech combinations were presented in a random order. The videos had a duration of four seconds after which the question: “Does the action in the video match the content of the spoken sentence?” appeared. Participants were instructed to indicate as fast as possible whether or not the content of the sentence matched the video by responding with the left index finger on the “q” or the right index finger on the “p” on a QWERTY-keyboard (e.g. for “q” = yes and “p” = no). Left/right and yes/no couplings were counterbalanced between participants. Participants were instructed to keep their index fingers on the “q” and “p” during the experiment, attentively watch the video and to try to only blink their eyes in between videos. Reaction times were computed counting from the end of the sentence. Triggers were placed in the E-prime software in order to mark the audio onset of the verb.

1.5. EEG recording

EEG was recorded from 32 electrode sites across the scalp using active Ag/AgCl electrodes (BioSemi, Amsterdam, the Netherlands, ActiveTwo amplifier system) placed in an elastic cap according to the 10/20 system. Electrodes were placed on standard sites (Fz, Cz, Pz, Oz, AF3, AF4, Fp1, Fp2, F3, F4, F7, F8, FC1, FC2, FC5, FC6, T7, T8, C3, C4, CP1, CP2, CP5, CP6, P3, P4, P7, P8, O1, O2, PO3, PO4). To be able to filter out artifacts due to eye movements, horizontal eye movements (HEOG) were measured by placing electrodes on the left and right outer canthi of the eye. Vertical eye movements (VEOG) were measured with electrodes above and below the left eye. Lastly, reference electrodes were placed on the left and right mastoid. Recordings were amplified using an ActiveTwo amplifier system

Table 1
Means and Standard Deviations of Accuracy (%) and Reaction Times (ms) of the behavioural task.

| | | Younger adults ($n = 40$) | | Older adults ($n = 32$) | |
|---------|-------------|-----------------------------|----------|---------------------------|----------|
| | | % correct | RT in ms | % correct | RT in ms |
| Action | Congruent | 97.99 (3.91) | 542(226) | 99.22 (1.61) | 761(250) |
| | Incongruent | 99.19 (1.99) | 530(209) | 99.45 (1.23) | 717(214) |
| Gesture | Congruent | 95.16 (11.41) | 594(181) | 97.58 (3.21) | 721(203) |
| | Incongruent | 99.17 (3.05) | 582(180) | 98.13 (2.38) | 696(162) |

Table 2
Content of the blocks and the block order presentation per version.

| | | Action Congruent | Action Incongruent | Gesture Congruent | Gesture Incongruent |
|-------------------------|-----------|------------------|--------------------|-------------------|---------------------|
| Content of the blocks | Block 1 | Item 1 - 20 | Item 21 - 40 | | |
| | Block 2 | Item 21 - 40 | Item 1 - 20 | | |
| | Block 3 | | | Item 1 - 20 | Item 21 - 40 |
| | Block 4 | | | Item 21 - 40 | Item 1 - 20 |
| Block order per version | Version 1 | Block 1 | Block 2 | Block 3 | Block 4 |
| | Version 2 | Block 2 | Block 1 | Block 4 | Block 3 |
| | Version 3 | Block 3 | Block 4 | Block 1 | Block 2 |
| | Version 4 | Block 4 | Block 3 | Block 2 | Block 1 |

and sampled at 512 Hz.

1.6. Data analysis

1.6.1. Behavioural data

Accuracy was determined by mean percentage accurate per participant per condition and reaction times were computed by mean reaction time in milliseconds (ms) per participant per condition (Type of Movement x Congruency). Analyses of the behavioural data were conducted on the sample that was used for the EEG analyses.

1.6.2. ERP data

EEG data were rereferenced offline to the linked mastoids. A first segmentation was done from 200 ms before until 1800 ms after video onset to be able to perform a baseline correction. A second segmentation has been done from 800 to 1800 ms after video onset, because the critical stimulus (action/gesture-speech combination) occurred 1000 ms after video onset, leaving 800 ms poststimulus for the analysis of the N400. Epochs were filtered with a 0.01–40 Hz band filter and corrected for eye movements using the algorithm of Gratton et al. (1983). Segments were only analysed when the correct answer was provided by the participant. With 40 action/gesture

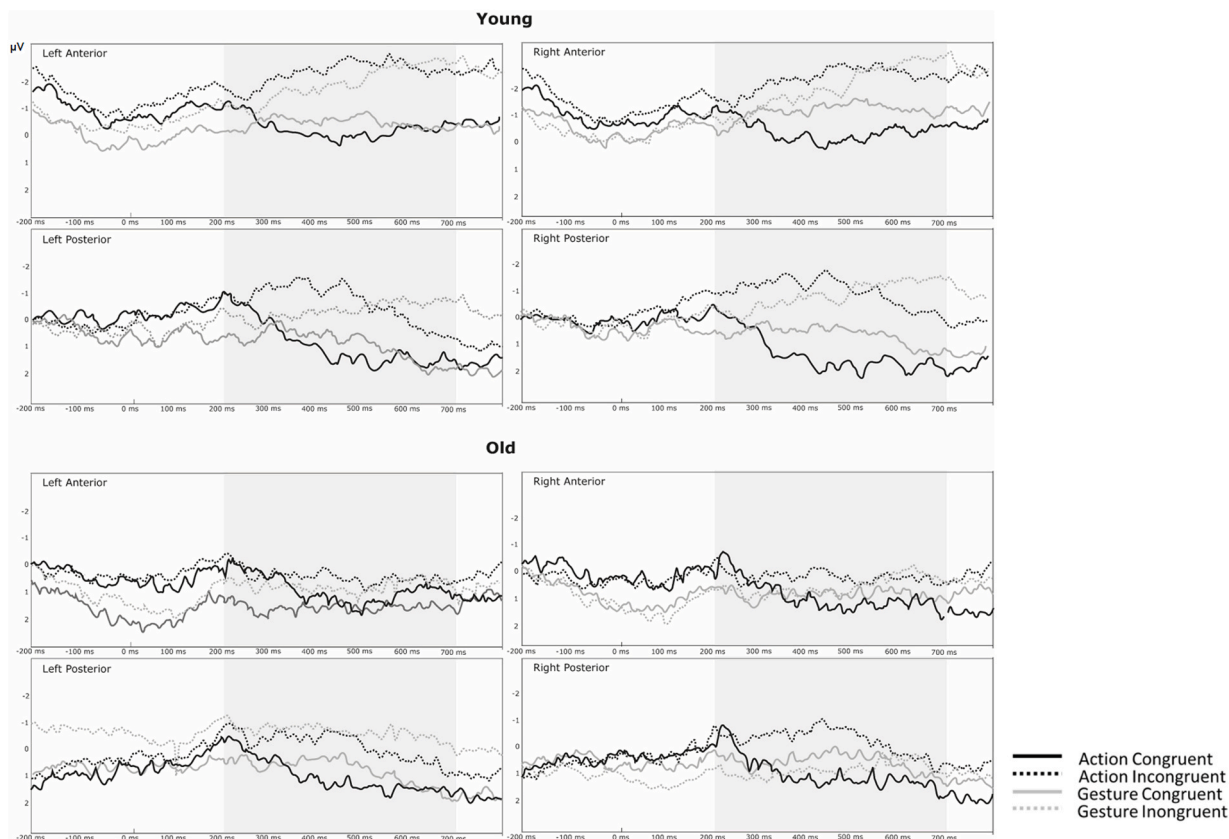


Fig. 2. N400 Waveforms in µV for the Younger and Older Adults.

speech combinations in each condition, the maximum number of data segments per participant per condition was 40. If in one of the conditions, fewer than 20 segments remained after artifact rejection, the participant was excluded from further analysis (one younger adult and seven older adults), leaving 40 younger adults and 32 older adults for the EEG analysis. For the younger adults, 1.56% of the segments were rejected for congruent action-speech combinations, 1.00% for incongruent action-speech, 2.32% for congruent gesture-speech and, 2.69% for incongruent gesture-speech combinations. For the older adults, 4.06% of the segments were rejected for congruent action-speech combinations, 4.30% for incongruent action-speech, 3.67% for congruent gesture-speech and, 2.42% for incongruent gesture-speech combinations. Segments were normalised on the basis of the 200 ms baseline preceding the onset of the video (the critical stimulus was presented 1000 ms post video onset) and ICA correction. ERPs were calculated for each participant by averaging trials for each electrode and condition separately (i.e., action congruent, action incongruent, gesture congruent and gesture incongruent). To be able to detect where the N400 effect is strongest, we created four quadrants, just like De Nooijer et al. (2016). All recorded channels are included in these sections: left anterior (F3, F7, FC1, FC5, C3 AF3, FP1, T7), right anterior (F4, F8, FC2, FC6, C4, AF4, FP2, T8), left posterior (CP1, CP5, P3, P7, O1, PO3) right posterior (CP2, CP6, P4, P8, O2 PO4).

1.7. N400 time-window

The selected time window for the present study (200–700 ms) differs from the classic time window of 300–500 ms. Three reasons underly this decision. Firstly, in older adults the N400 (effect) can be delayed (e.g., Federmeier et al., 2010; Federmeier & Kutas, 2005). Secondly, the processing of actions (N420, Proverbio & Riva, 2009) and gestures (N450, Wu & Coulson, 2005) occurs later than the classical N400 effect. For example, Kelly et al. (2007) specified the time window for gesture-speech stimuli to be between 350 and 700 ms post critical stimulus. Thirdly, studies looking at object recognition with matching or mismatching contexts, specify an N300 effect that is structurally similar to the N400 effect, but with an earlier onset, requiring a time window between 200 and 500 ms (Truman & Mudrik, 2018). Note that for the present study in the action condition, objects were present before the actions were performed and might have elicited similar early responses as found by Truman and Mudrik (2018). Indeed, from visual inspection of our waves (see Fig. 2) it can be seen that our data seem to confirm both the findings of Kelly et al. (2007) and Truman and Mudrik (2018).

2. Results

A significance level of 0.05 was used for the main analyses. On follow-up analyses a Bonferroni correction was applied. Partial eta-

Table 3

Test Statistics for the 2 x 2 x 2 x 2 x 2 ANOVA with Within-Subject Factors for Congruency (C: Congruent vs. Incongruent), Type of Movement (M: Action vs. Gesture), Front-Back (FB: Anterior vs. Posterior Quadrants) and Laterality (L: left vs. Right Quadrants) and a Between-Subjects Factor for Age (A: Younger vs. Older Adults).

| Factor | <i>F</i> | <i>P</i> | η_p^2 |
|--------------------|--------------|-----------------|------------|
| A | 0.76 | .387 | .01 |
| C | 11.93 | <.001 | .15 |
| M | 0.40 | .531 | .01 |
| FB | 7.11 | .010 | .09 |
| L | 2.86 | .095 | .04 |
| C * A | 1.81 | .182 | .06 |
| M * A | 0.46 | .500 | .01 |
| FB * A | 13.29 | <.001 | .16 |
| L * A | <0.01 | .990 | <.01 |
| M * C | 1.79 | .185 | .03 |
| FB * C | 0.21 | .649 | <.01 |
| L * C | <0.01 | .970 | <.01 |
| FB * M | 0.01 | .913 | <.01 |
| L * M | 2.55 | .115 | .04 |
| FB * L | 0.35 | .556 | .01 |
| M * C * A | 0.19 | .663 | <.01 |
| FB * M * A | <.01 | .981 | <.08 |
| L * M * A | 0.90 | .346 | .01 |
| FB * L * A | 0.44 | .834 | <.01 |
| FB * C * A | 0.01 | .910 | .01 |
| L * C * A | 0.87 | .354 | .01 |
| FB * L * M | 0.79 | .378 | .01 |
| FB * L * C | 3.05 | .085 | .05 |
| FB * M * C | 2.83 | .097 | .04 |
| L * M * C | 2.06 | .156 | .03 |
| FB * L * M * A | 1.31 | .256 | .02 |
| FB * L * C * A | 2.49 | .119 | .03 |
| FB * M * C * A | 4.37 | .040 | .06 |
| L * M * C * A | 0.01 | .941 | <.01 |
| FB * L * M * C | 0.25 | .616 | <.01 |
| FB * L * M * C * A | 0.45 | .503 | .01 |

squared (η_p^2) was calculated as a measure of effect size for F -values, with values of 0.01, 0.06, and 0.14, characterising small, medium, and large effect sizes, respectively (Cohen, 1988). Cohen's d was calculated as a measure of effect size for t -values, with values of 0.20, 0.50, and 0.80, characterising small, medium, and large effect sizes, respectively (Cohen, 1988).

2.1. Behavioural data

The behavioural data (accuracy and reaction times) were analysed by mixed $2 \times 2 \times 2$ ANOVAs with congruency (congruent vs. incongruent) and type of movement (action vs. gesture) as within-subject factors and age group (younger vs. older adults) as between-subjects factor. Means and standard deviations of the descriptive data are given in Table 1.

For accuracy, a main effect of type of movement, $F(1, 70) = 9.10, p = .004, \eta_p^2 = 0.12$, was found, reflecting higher accuracy for the action-speech than the gesture-speech combinations. A marginal effect was found for congruency, $F(1, 70) = 3.95, p = .051, \eta_p^2 = 0.05$, reflecting a trend towards higher accuracy for the incongruent than the congruent stimuli. For age group no effect was found, $F(1, 70) = 2.29, p = .135, \eta_p^2 = 0.03$. No interactions were found between type of movement and age group, $F(1, 70) = 0.15, p = .701, \eta_p^2 < 0.01$, congruency and age group, $F(1, 70) = 1.86, p = .177, \eta_p^2 = 0.03$, type of movement and congruency, $F(1, 70) = 1.14, p = .289, \eta_p^2 = 0.02$, or congruency, type of movement and age group, $F(1, 70) = 0.57, p = .454, \eta_p^2 < 0.01$.

For reaction times, a main effect of congruency, $F(1, 70) = 7.21, p = .009, \eta_p^2 = 0.09$, was found, reflecting longer reaction times for the congruent than the incongruent stimuli. This is consistent with the accuracy data showing higher accuracy for the incongruent than the congruent stimuli. In addition a main effect of age group $F(1, 70) = 14.61, p < .001, \eta_p^2 = 0.17$, was found, reflecting longer reaction times for the older than the younger adults. No effects were found for movement, $F(1, 70) = 0.44, p = .508, \eta_p^2 < 0.01$. No interactions were found between type of movement and age group, $F(1, 70) = 3.53, p = .064, \eta_p^2 = 0.05$, congruency and age group, $F(1, 70) = 1.64, p = .204, \eta_p^2 = 0.02$, type of movement and congruency, $F(1, 70) = 0.40, p = .530, \eta_p^2 < 0.01$, or congruency, type of movement and age group, $F(1, 70) = 0.36, p = .549, \eta_p^2 < 0.01$.

2.2. ERP data

2.2.1. N400 amplitude

A mixed $2 \times 2 \times 2 \times 2 \times 2$ ANOVA was conducted with congruency (congruent vs. incongruent), type of movement (action vs. gesture), laterality (left vs. right quadrants), and front-back (anterior vs. posterior quadrants) as within-subjects factors and age group (younger vs. older adults) as between-subjects factor. For readability purposes, only significant results from the omnibus analysis are described in the text. All test statistics are presented in Table 3.

A significant main effect of congruency was found, reflecting a more negative mean N400 amplitude for the incongruent than the congruent trials (i.e. a significant N400 effect). In addition, a main effect of front-back was found, reflecting an overall more negative mean amplitude in the anterior areas compared to the posterior areas. A two-way interaction was found between front-back and age group. No three-way interactions were found. A four-way interaction was found between congruency, type of movement, front-back and age group. No five-way interaction was found.

The two-way interaction between front-back and age group was followed up by a paired t -test for each age group comparing the overall N400 component between the anterior and posterior areas. Results showed that in the younger adults, the N400 component was more negative in the anterior areas than in the posterior areas, $t(39) = -4.64, p < .001, d = -0.73$, while no such difference was present for the older adults, $t(31) = 0.60, p = .278, d = 0.11$.

The four-way interaction between congruency, type of movement, front-back and age group was followed up by analyses inspecting the congruency effect for the types of movement \times front-back within and between age group. See Fig. 3 for a visual display of the interaction. First, to inspect the congruency effects within age groups, four paired t -tests were performed for each age group separately for the N400 (incongruency) effects for the (1) action stimuli in the anterior areas, (2) action stimuli in the posterior areas, (3) gesture stimuli in the anterior areas and, (4) gesture stimuli in the posterior areas. Significance levels were adjusted according the Bonferroni method, resulting in a cutoff of $\alpha = 0.006$ ($05/8$). Results of the younger adults showed a significant N400 effect in nearly all contrasts,

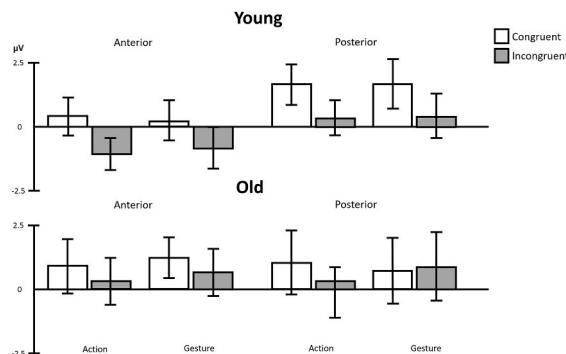


Fig. 3. Interaction effect between congruency, movement, front-back and age group.

action-anterior, $t(39) = 3.60, p < .001, d = 0.57$, action-posterior, $t(39) = 3.32, p < .001, d = 0.53$, gesture – anterior, $t(39) = 2.60, p = .007, d = 0.41$, and, gesture – posterior, $t(39) = 2.66, p = .006, d = 0.42$. However, for the older adults the N400 effect was only present in one of the four contrasts, 1) action-anterior, $t(31) = 1.28, p = .011, d = 0.23$, 2) action-posterior, $t(31) = 0.48, p = .006, d = 2.46$, 3) gesture-anterior, $t(31) = 1.22, p = .012, d = 0.22$, and, 4) gesture-posterior, $t(31) = -0.17, p = .434, d = -0.03$. These results indicate that in the young adults the incongruency effect is present for gestures and actions, in the anterior and posterior areas, while in the older adults only for the actions in the posterior area.

Second, to inspect the congruency effects (type of movement x front-back) between age groups, four independent sample T-tests were performed on the difference scores (congruent-incongruent) for the 1) action stimuli in the anterior areas, 2) action stimuli in the posterior areas 3) gesture stimuli in the anterior areas and, 4) gesture stimuli in the posterior areas. Results showed no significant effects for action-anterior, $t(70) = 1.46, p = .075, d = 0.35$, action-posterior, $t(70) = 0.27, p = .394, d = 0.06$, gesture - anterior, $t(70) = 0.86, p = .198, d = 0.20$, and, gesture - posterior, $t(70) = 1.42, p = .080, d = 0.34$. These results indicate that the congruency effects for the action and gesture stimuli in anterior and posterior areas did not significantly differ between the age groups.

3. Discussion

Aim of the present study was to compare the processing of multimodal action-speech and gesture-speech combinations between younger and older adults, by looking at the N400 (in)congruency effect. Our first hypothesis that gesture-speech combinations would elicit larger N400 effects than action-speech combinations in both age groups, was not supported by the results. No interaction between type of movement and congruency effect was found. Our second hypothesis that older adults would have an overall smaller N400 effect than younger adults was also not supported by the data. No interaction between age group and congruency effect was found. The third hypothesis that the difference between the N400 effect for action-speech and gesture-speech combinations would be smaller for older adults than younger adults was partly supported. No interaction between congruency, type of movement and age group was found. However, we did find an interaction between type of movement, congruency, age group and front-back. This interaction showed that the N400 effect was present in younger adults for both the action and gesture stimuli in the anterior and posterior areas. In contrast, older adults only a significant N400 effect for the action stimuli in the posterior areas was found.

That the N400 effect is less pronounced and widespread in older adults can be explained by cognitive aging literature on working memory decline (e.g., [Salthouse, 1994](#)). A recent study of [Zuber et al. \(2019\)](#) pointed out that specifically processes concerned with updating and inhibition of information are targeted by age-related declines. These specific age-related declines are also mentioned in the review of [Wlotko et al. \(2010\)](#) who showed that older adults have more trouble than younger adults to revise an initially activated but incorrect idea. However, this does not explain our finding that the N400 effect remained present for specifically action-speech stimuli in the posterior areas. In a recent review by [Kandana Arachchige et al. \(2021\)](#) methodological issues regarding gesture research are discussed. The authors mention that type of gesture, linguistic context (i.e. the gesture/verb presented in isolation or embedded in a sentence), other ERPs taking place simultaneously, and characteristics of the participants (native or non-native speakers, children) can influence outcomes. For example, it is mentioned that stimuli embedded in sentences elicited more pronounced N400 components in anterior sites, while single word stimuli were more pronounced in centro-parietal sites. In line with these findings, our data (based on verbs within sentences) also showed that the N400 component was more negative in the anterior sites than the posterior sites. [Kandana Arachchige et al. \(2021\)](#) also suggest that the anterior regions process the stimuli at a more global (sentence) level, while the more centro-parietal regions deal with the more local level (word level). Connecting this to the aging literature, we suggest that processing on a global level might be challenged in older adults because processing whole sentences requires more working memory capacity than processing single words. The finding that the N400 effect only remained significant for the action-stimuli (in the posterior areas) can be explained by earlier studies suggesting the decreased sensitivity to gestures in older adults, as discussed in the introduction ([Cocks et al., 2011](#); [Montepare et al., 1999](#); [Thompson, 1995](#); [Thompson & Guzman, 1999](#)).

4. Limitations

The present study was for an important part inspired by the study of [Kelly et al. \(2015\)](#) who was able to make a controlled comparison between literal actions and pantomime gestures. However, pantomime gestures could be understood in the absence of speech which is not necessarily the case for co-speech gestures used more often in daily communication. Furthermore, [Willems et al. \(2009\)](#) showed that pantomime gestures and iconic gestures elicit partly similar (activation in posterior superior temporal sulcus and mid temporal gyrus) and differential neural activity (activation in the left inferior frontal gyrus for only co-speech gestures). This means that results of the present study should not be (completely) associated with other research outcomes considering other (co-speech) gestures. However, we do expect that in a more narrative context, pantomime gestures can assist comprehension compared to a context with only speech, for example when the verbal information is not clear. Indeed it has been found that when the verbal message is degraded ([Holle et al., 2010](#)), or ambiguous ([Holle & Gunter, 2007](#)) gestures can increase the understanding of the message. As already mentioned in the introduction, although the gestures in these studies are called iconic gestures, the examples presented in the articles of [Holle et al. \(2008, 2010\)](#) and [Holle and Gunter \(2007\)](#) look quite similar to the ones used in this study. It seems that although pantomime gestures and iconic gestures are theoretically distinct, clearer boundaries are needed for the operationalisation of stimuli to apply this distinction in research.

A second limitation of this study is that in the action videos, the objects were visible during the whole video, while in the gesture videos no objects were present. As the critical compound stimuli (i.e. the gesture-speech or action-speech combinations) were presented 1 s after onset of the videos, the objects already presented part of the stimuli in the action condition, but not in the gesture

condition. It might have been the case that the visual information presented by the object in the action condition needed additional computational processes (when compared to the gesture condition in which no object is present). However, because this object information preceded the critical stimuli by a full second, we assume this information was fully processed at the onset of the critical stimulus and not hindered the processing of the critical stimuli. On the contrary, it might have facilitated the processing of the critical stimuli, especially for the older adults. As mentioned earlier, online (transient) information processing (Salthouse, 1994), updating (Zuber et al., 2019) and revising (Wlotko et al., 2010) is challenged in older adults. In addition, processing speed deteriorates as a result of cognitive aging (Albinet et al., 2012; Kerchner et al., 2012). Because the object was present one 1 s the critical stimulus, participants had plenty of time to already process part of the information related to the critical stimulus. Because at the onset of the critical stimulus, less new information was presented, this might have compensated for the reduced processing speed in the older adults.

Future studies should investigate whether the difference between the action-speech and gesture-speech N400 effect is a result of the preview of the object. To control for this, the object could be primed (e.g., shortly present picture of the objects) before the onset of the action- or gesture video. Another interesting idea for future research would be to investigate whether the effect of age on the integration of multimodal information would also influence learning and memory. This is increasingly important considering the emphasis on life-long learning beyond formal education (Vera-Toscano et al., 2017). If older adults indeed are less responsive to gestures (which has been evidenced also by Cocks et al., 2011; Montepare et al., 1999; Thompson, 1995; Thompson & Guzman, 1999), this might affect memory and learning negatively, as gestures can complement a spoken message. In a learning situation this means that they might encode the information less deeply which would make it harder to remember later on. In addition, if incompatibilities in information are not sufficiently detected at encoding for example because attention in one modality is dominant, incongruent (incorrect) information may be stored as correct information which can lead to false memories. This could be tested by adding a test phase to an experimental paradigm similar to the present one in which the participants need to remember the action/gesture-speech combinations (either congruent or incongruent). If the accuracy on the memory test correlates with the size of the N400 effect in younger and older adults, this might provide evidence for the explanation that older adults have trouble integrating information presented in multiple modalities.

Another interesting follow-up question would be whether a certain modality or strategy is preferred. Note that, a frontal N400 effect has been associated with semantic integration of visual stimuli such as videos of real-world events (Sitnikova et al., 2003), gestures (Wu & Coulson, 2005) pictures of objects (Holcomb & McPherson, 1994) sentence (global level) processing (Kandana Arachchige et al., 2021) and a more posterior N400 effect has associated with semantic integration of linguistic input: written (Kutas & Van Petten, 1990) as well as spoken (Hagoort & Brown, 2000) and more word-level (local level) processing (Kandana Arachchige et al., 2021). Evidence suggests that functions related to lexical (e.g., Laver, 2009; Waters & Caplan, 2005) and semantic processing (Federmeier et al., 2003) remain relatively intact with aging, while all kinds of sensory (e.g., Federmeier et al., 2003; Madden, 2007; Voss et al., 2008) and episodic processes (Trott et al., 1999) show age-related changes. It is well possible that the performance of older adults might improve when they could use a strategy in which they can rely on the lexical aspects of the stimuli, while younger adults might respond best to the sensory aspects of the stimuli.

A final limitation of the study is that the age groups were not comparable on educational level. In the older adults, more variance was present as 14 participants obtained a Master's degree, 13 a Bachelor's degree, two completed pre-university education, one higher general secondary education and 6 lower general secondary education. However, because the majority of the younger adult group were first year Psychology students, and most of them graduated from pre-university education (high school) the year before and had no Ba or MSc yet, they were comparable to most of the older adults having (pre)university or higher secondary education. We do acknowledge that six older adults did have a different educational level. However, the task used for the present study (judging the match between common action verbs and gestures/actions) was not complicated and we carefully practiced the task with all participants. Therefore, we feel that this has not influenced our outcomes in a problematic manner.

5. Conclusion

In conclusion, the present study found some indication that older adults are less sensitive to the appropriateness (congruency effect was less widespread) of visual context information in the processing of linguistic information. However, older adults were more sensitive to the concreteness of the visual information (actions but not gestures elicited a congruency effect in some areas). This is partly in line with earlier evidence (Federmeier et al., 2003, 2005, 2010). Furthermore, the present study adds to the current literature by showing that the age-related decline in sensitivity for contextual information also applies to multimodal language-related input.

Author statement

Ouwehand: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Original draft preparation, Writing - Review & Editing, Visualization, Supervision, Project administration, Writing - Review & Editing.

De Noijer: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing- Original draft preparation, Review & Editing.

Van Gog: Writing- Original draft preparation.

Paas: Original draft preparation, Supervision, Writing - Review & Editing.

Acknowledgments

We thank the Erasmus Behavioural Lab staff for their technical assistance and Charly Eielts for assistance with the graphics.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jneuroling.2022.101100>.

References

- Albinet, C. T., Boucard, G., Bouquet, C. A., & Audiffren, M. (2012). Processing speed and executive functions in cognitive aging: How to disentangle their mutual relationship? *Brain and Cognition*, 79(1), 1–11. <https://doi.org/10.1016/j.bandc.2012.02.001>
- Bach, P., Gunter, T. C., Knoblich, G., Prinz, W., & Friederici, A. D. (2009). N400-like negativities in action perception reflect the activation of two components of an action representation. *Social Neuroscience*, 4(3), 212–232. <https://doi.org/10.1080/17470910802362546>
- Cocks, N., Morgan, G., & Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture*, 11(1), 24–39. <https://doi.org/10.1075/gest.11.1.02coc>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- De Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory*, 12(1), 75–89. <https://doi.org/10.1080/09658210244000568>
- De Nooijer, J. A., Gootjes, L., Van Gog, T., Paas, F., & Zwaan, R. A. (2016). Picturing meaning: An ERP study on the integration of left or right-handed first-person perspective pictures into a sentence context. *Language, Cognition and Neuroscience*, 31(2), 167–178. <https://doi.org/10.1080/23273798.2015.1062901>
- Federmeier, K. D., & Kutas, M. (2005). Aging in context: age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133–141. <https://doi.org/10.1111/j.1469-8986.2005.00274.x>
- Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149–161. <https://doi.org/10.1016/j.bandl.2010.07.006>
- Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: age-related changes across levels of language processing. *Psychology and Aging*, 18(4), 858–872. <https://doi.org/10.1037/0882-7974.18.4.858>
- Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16(8), 1465–1477. <https://doi.org/10.1162/0898929042304705>
- Gallant, S. N., Durbin, K. A., & Mather, M. (2020). Age differences in vulnerability to distraction under arousal. *Psychology and Aging*, 35(5), 780–791. <https://doi.org/10.1037/pag0000426>
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for offline removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55, 468–484. [https://doi.org/10.1016/0013-4694\(83\)90135-9](https://doi.org/10.1016/0013-4694(83)90135-9)
- 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. [https://doi.org/10.1016/S0028-3932\(00\)00053-1](https://doi.org/10.1016/S0028-3932(00)00053-1)
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology*, 30, 47–61. <https://doi.org/10.1111/j.1469-8986.1993.tb03204.x>
- Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, 24(2), 259–276. <https://doi.org/10.1006/brcg.1994.1014>
- Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19(7), 1175–1192. <https://doi.org/10.1162/jocn.2007.19.7.1175>
- Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage*, 39(4), 2010–2024. <https://doi.org/10.1016/j.neuroimage.2007.10.055>
- Holle, H., 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(1), 875–884. <https://doi.org/10.1016/j.neuroimage.2009.08.058>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15, 495–514. <https://doi.org/10.3758/PBR.15.3.495>, 2008.
- Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 26, 721–752. <https://doi.org/10.3758/s13423-018-1548-0>, 2019.
- Kandana Arachchige, K. G., Simoes Loureiro, I., Blekic, W., Rossignol, M., & Lefebvre, L. (2021). The role of iconic gestures in speech comprehension: An overview of various methodologies. *Frontiers in Psychology*, 12, Article 634074. <https://doi.org/10.3389/fpsyg.2021.634074>
- Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a Strooplike task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*, 22(4), 683–694. <https://doi.org/10.1162/jocn.2009.21254>
- Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, 22(2), 517–523. <https://doi.org/10.3758/s13423-014-0681-7>
- Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, 101(3), 222–233. <https://doi.org/10.1016/j.bandl.2006.07.008>
- Kerchner, G. A., Racine, C. A., Hale, S., Wilhelm, R., Laluz, V., Miller, B. L., & Kramer, J. H. (2012). Cognitive processing speed in older adults: relationship with white matter integrity. *PLoS One*, 7(11), Article e50425. <https://doi.org/10.1371/journal.pone.0050425>
- 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, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Van Petten, C. (1990). Electrophysiological perspectives on comprehending written language. *Electroencephalography and Clinical Neurophysiology*, 41(4), 155–167. <https://doi.org/10.1016/b978-0-444-81352-7.50020-0>
- Laver, G. D. (2009). Adult aging effects on semantic and episodic priming in word recognition. *Psychology and Aging*, 24(1), 28–39. <https://doi.org/10.1037/a0014642>
- Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT press.
- Madden, D. J. (2007). Aging and visual attention. *Current Directions in Psychological Science*, 16(2), 70–74. <https://doi.org/10.1111/j.1467-8721.2007.00478.x>
- McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. University of Chicago Press.
- Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, 23(2), 133–152. <https://doi.org/10.1023/A:1021435526134>
- 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(4), 605–616. <https://doi.org/10.1162/jocn.2007.19.4.605>

- Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research*, 1687, 117–128. <https://doi.org/10.1016/j.brainres.2018.02.021>
- Pouw, W., & Dixon, J. A. (2019). Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cognitive Science*, 43(3), Article e12721. <https://doi.org/10.1111/cogs.12721>
- Pouw, W., T. J. L., & Hostetter, A. (2016). Gesture as predictive action. *Invited contribution for Special issue of Reti, Saperi, Linguaggi: Italian Journal of Cognitive Sciences*, 3, 57–80. <https://doi.org/10.12832/83918>
- Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, 459(3), 142–146. <https://doi.org/10.1016/j.neulet.2009.05.012>
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8(4), 535–543.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh: Psychology Software Incorporated.
- Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, 40(1), 160–164. <https://doi.org/10.1111/1469-8986.00016>
- Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging*, 10(2), 215–228. <https://doi.org/10.1037/0882-7974.10.2.215>
- Thompson, L. A., & Guzman, F. A. (1999). Some limits on encoding visible speech and gestures using a dichotic shadowing task. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 54B(6), 347–349. <https://doi.org/10.1093/geronb/54b.6.p347>
- Trott, C. T., Van Petten, C., Schwartz, T. J., & Kutas, J. G. (1999). Episodic priming and memory for temporal source: Event-related potentials reveal age-related differences in prefrontal functioning. *Psychology and Aging*, 14(3), 390–413. <https://doi.org/10.1037/0882-7974.14.3.390>
- Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*, 117, 222–232. <https://doi.org/10.1016/j.neuropsychologia.2018.06.004>
- Vera-Toscano, E., Rodrigues, M., & Costa, P. (2017). Beyond educational attainment: The importance of skills and lifelong learning for social outcomes. Evidence for Europe from PIAAC. *European Journal of Education*, 52(2), 217–231. <https://doi.org/10.1111/ejed.12211>
- Voss, M. W., Erickson, K. I., Chaddock, L., Prakash, R. S., Colcombe, S. J., Morris, K. S., ... Kramer, A. F. (2008). Dedifferentiation in the visual cortex: an fMRI investigation of individual differences in older adults. *Brain Research*, 1244, 121–131. <https://doi.org/10.1016/j.brainres.2008.09.051>
- Waters, G., & Caplan, D. (2005). The relationship between age, processing speed, working memory capacity and language comprehension. *Memory*, 13, 403–413. <https://doi.org/10.1080/09658210344000459>
- Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, 47(4), 1992–2004. <https://doi.org/10.1016/j.neuroimage.2009.05.066>
- Wlotko, E. W., Lee, C. L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass*, 4(8), 623–638. <https://doi.org/10.1111/j.1749-818X.2010.00224.x>
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654–667. <https://doi.org/10.1111/j.1469-8986.2005.00356.x>
- Zuber, S., Ihle, A., Loaiza, V. M., Schnitzspahn, K. M., Stahl, C., Phillips, L. H., ... Kliegel, M. (2019). Explaining age differences in working memory: The role of updating, inhibition, and shifting. *Psychology & Neuroscience*, 12(2), 191–208. <https://doi.org/10.1037/pne0000151>