A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration

Christian Obermeier,[1] Spencer D. Kelly,[2] and Thomas C. Gunter[1]

[1]Max-Planck-Institute for Human Cognitive and Brain Sciences, Department of Neuropsychology, Leipzig, Germany and [2]Colgate University, Department of Psychology, NY 13346, Hamilton, USA

In face-to-face communication, speech is typically enriched by gestures. Clearly, not all people gesture in the same way, and the present study explores whether such individual differences in gesture style are taken into account during the perception of gestures that accompany speech. Participants were presented with one speaker that gestured in a straightforward way and another that also produced self-touch movements. Adding trials with such grooming movements makes the gesture information a much weaker cue compared with the gestures of the non-grooming speaker. The Electroencephalogram was recorded as participants watched videos of the individual speakers. Event-related potentials elicited by the speech signal revealed that adding grooming movements attenuated the impact of gesture for this particular speaker. Thus, these data suggest that there is sensitivity to the personal communication style of a speaker and that affects the extent to which gesture and speech are integrated during language comprehension.

Keywords: indexical cues; individual differences; gesture-speech integration; disambiguation; N400

INTRODUCTION

In face-to-face communication humans are able to extract enormous amounts of information from their social environment, and this information greatly impacts the comprehension of language in real time. This is an important characteristic, particularly when confronted with multiple interlocutors. Think of a conversation in a commuter train. Although you do not know the people in the compartment, you start making small talk about the weather. When the others start talking back, you receive important information that goes beyond the actual content of the utterances. The speech signal itself contains valuable non-phonemic information like prosody or acoustic correlates of talker identity (important when you enter the tunnel), which additionally could give significant insight regarding information like sex, age, social status, etc. of the interlocutor. These so-called 'indexical cues' can even give us a hint what kind of topics the speaker is more likely to address (semantics: Lattner and Friederici, 2003; Van Berkum et al., 2008), or whether phonemic or syntactic flaws need to be taken into account (syntax errors in speakers of a second language: Hanulikova et al., 2012). If the conversation in the train lingers on and you get to know the other passengers a bit better, you could even assess what kind of speech/communication style to expect from a particular talker (use of irony: Regel et al., 2010).

In communicative situations like the one discussed earlier, the visual domain can also give us relevant information concerning the communicative constellation. Trivially, talker identity can be inferred by face, position in space, etc. but there are also visual counterparts of the acoustic indexical cues discussed earlier. Movement of the body, for instance, can give us important information regarding the emotional state of the talker (Wallbott, 1998), whereas eye-gaze informs us about where the attention of the talker is directed (cf. communicative intent: Senju and Johnson,

2009). Even seemingly irrelevant hand-movements, such as scratching the chin or rubbing the nose, so-called grooming movements, disclose important talker information. DePaulo et al. (2003), for instance, showed that excessive grooming can cause a speaker to appear less trustworthy. Interestingly, grooming also has an impact on the efficiency of other visual information—such as co-speech gestures[1]—as a valuable cue during language processing (Holle and Gunter, 2007).

Co-speech gestures are ubiquitous in most face-to-face interactions, and they convey useful information about a speaker's intent and meaning (McNeill, 1992). Although research has shown that interlocutors are sensitive to these hand movements (for a review, see Hostetter, 2011), little is known about how gestures are processed in the context of other types of non-communicative hand movements. In one of the few studies on the topic, Holle and Gunter (2007) showed that the communicative power of gestures is affected by the presence of trials involving grooming instead of gestures. This study used ambiguous words, in particular so-called unbalance homonyms like 'ball', which have a more frequent dominant meaning (e.g. ball as in a game) and a lesser frequent subordinate meaning (e.g. ball as a kind of formal dance event). During the processing of a homonym both of these meanings are activated in working memory (Swinney, 1979), and the comprehension system can use two types of information to select the appropriate meaning. Typically, on the basis of the foregoing context one of the meanings will be selected and the other inhibited. When however, the context does not provide any information, word meaning frequency determines the meaning selection. Thus, the selection of what a homonym refers to depends on how strong the context is. Once the contextual constraints become weak, the comprehension system makes increased use of other sources of information like word meaning frequency. Holle and Gunter (2007) used gestures as a contextual constraint and showed that their strength depended on how clear the gesture channel was. Whereas in an experimental design without grooming trials, a homonym was clearly disambiguated by gestures related to the dominant or subordinate meaning, this disambiguation was present only for gestures related to the subordinate

[1] In the rest of the paper, when referring to gestures, we explicitly mean co-speech gestures.

meaning when one-third of grooming trials were added to the experimental design (i.e. a mix of 33% grooming movements and 66% gestures trials). In this mixed design, the reduced probability of observing a hand movement which conveys meaning caused the listener to put less weight on gestural information and more on other sources of information (i.e. word meaning frequency) during the meaning selection process at the homonym. This research suggests that non-verbal communication style may modulate how people integrate speech and gesture during language comprehension. To explore this issue, we ask whether this grooming-effect also occurs in situations where there is more than one speaker, and whether addressees use such variability in gesturing as an indexical cue for the importance of gestural information within a particular speaker? Before exploring this question, it is necessary to review research on variability in gesturing.

Clearly, people gesture in a highly variable way and the actual frequency of gesture use varies greatly between individuals (McNeill, 1992). Recently, such individual differences in spontaneous gesture production have been related to cognitive abilities such as working memory, spatial transformation and conceptualization of a particular gesturer (lower ability is associated with higher gesture rates: Gillespie *et al.*, 2014; Chu *et al.*, 2014). This suggests that speakers could, for instance, use gesturing for reducing their working memory load (cf. Marstaller and Burianová, 2013) or cognitive load in general (for a review, see Pouw *et al.*, 2014). Note, however, that only part of the individual differences in gesturing can be explained by the drive of gesturers to compensate their cognitive abilities. Another part relates to the possibility that speakers adjust their gestures to facilitate communicative processing. For example, Chu *et al.* (2014) showed that a higher level of empathy of a particular speaker is correlated with higher frequency of more salient and interactive gestures. These findings suggest that speakers' tailor their gestures specifically for their addressees (see also Kelly *et al.*, 2011), but it is not clear whether addressees are sensitive to these subtle adjustments.

The present study, therefore, explores the fundamental question of whether individual differences in gesture style are perceived/stored and selectively coupled to a particular interlocutor. To put it differently, in a situation where there is more than one speaker, does a perceiver experience a *speaker-specific* impact of the gesture style on an utterance of that speaker, or is the impact of gesture style distributed across the whole group? To answer this question, participants were presented with an adapted version of the disambiguation paradigm of Holle and Gunter (2007) in which the stimuli were produced by a mix of two gesturing speakers who had subtle differences in communication style. One of the speakers, the groomer, produced in addition to the gesture-trials also a substantial amount of trials containing grooming movements thereby weakening the impact of the gestures. The other speaker, the non-groomer, produced only trials containing gestures, thus strengthening the impact of the gestures. Both speakers uttered experimental sentences that contained an unbalanced homonym in the initial part of the sentence (e.g. *She controlled the ball*... where ball could mean an object of an event). The subsequent clause contained a target word related to either the high frequent dominant or the low frequent subordinate meaning of the homonym (...*which during the game*...*vs*...*which during the dance*...). Coincident with the initial part of the sentence, the speakers produced a gesture related either to the dominant or the subordinate meaning of the homonym (so-called dominant and subordinate gestures, see Figure 1) thereby disambiguating the ambiguous word (for an example, see later). In one-third of the trials, the groomer did not produce a meaningful gesture but a grooming movement instead. During the experiment, trials of the groomer and non-groomer were randomly presented to the participants. That is, in this experiment, gesture style was a within subjects variable whereas in the original Holle and Gunter (2007) study it was a

between experiment manipulation. If our participants are able to keep track of the gesture style of the two speakers, one would expect that the disambiguating impact of gesture in the groomer will be reduced compared with the non-groomer.

To measure the effects of gesture on speech comprehension as in Holle and Gunter (2007), event-related potentials (ERPs) taken from the Electroencephalogram (EEG) were measured as a dependent variable. The ERPs computed on the target words (i.e. *game/dance* in the earlier example) provide an indirect index of the disambiguating impact of the gestures expressed in the amplitude of the N400 component[2]. The N400 is a negativity that peaks roughly 400 ms after the onset of a potentially meaningful stimulus such as words or pictures (for a review, see Kutas and Federmeier, 2011). Traditionally, the N400 amplitude is interpreted to reflect the ease of semantic integration of a word into a context (Brown and Hagoort, 1993). The easier this semantic integration is the smaller the N400. The Holle and Gunter (2007) study showed that the N400 elicited by target words is small when the homonym-gesture combination disambiguated toward the meaning of the target word and large when the homonym-gesture combination disambiguated into the wrong direction. Thus, in the example given earlier, when a target word (i.e. *game*) is expected on the basis of the foregoing context (i.e. the homonym *ball* together with a dominant gesture, suggesting a ball as in a game), its semantic integration is easy. This leads to a smaller N400 compared to when the context does not set up that expectation (i.e. *ball* together with a subordinate gesture, suggesting a type of formal dance event). Thus, only when there is an N400 disambiguation effect[3] at a particular target word, the gestures disambiguated the homonym successfully. This means that the N400 at the target word constitute an indirect measure of disambiguation of the homonym via gesture information.

Holle and Gunter (2007) showed that when communicative gestures were the only source of visual information, they influenced the interpretation of the homonyms such that the N400-disambiguation effect was present for both the dominant and subordinate target words. Thus, when observed hand movements systematically contain relevant information, this information will be used exclusively for the disambiguation process thereby discarding other information like word meaning frequency. However, when non-communicative grooming trials are added to the experimental design, other information like word meaning frequency starts to play a role in the disambiguation process. As a result, the dominant word meaning is always activated after weak contexts, even if the context biased the subordinate meaning (Martin *et al.*, 1999). In these cases, only the *subordinate* target words show the N400-disambiguation effect (the SS *vs* DS conditions in Figure 1), whereas the dominant target words do not (the DD *vs* SD condition), suggesting that the comprehension system assumes the high frequent dominant meaning as default disambiguation and only takes the subordinate gestures into account as noteworthy gestural information. This grooming related difference in N400-effects was used in the present experiment to explore whether our participants can selectively track the gesture style of the groomer and the non-groomer when the utterances of these two speakers are randomly mixed on a trial-by-trial basis.

We therefore hypothesized that if a particular gesture style has a speaker-specific impact on language processing, gestures of the *non-groomer* will show a clear N400-disambiguation effect on the target word for dominant (DD *vs* DS) as well as subordinate (SS *vs* DS) target words, i.e. a strong impact of

---

[2] Note, that the participants in Holle and Gunter (2007) did have a shallow task that did not relate to the homonym, the gestures or the grooming (see Method section). That is, the task does not require integrating gesture and speech and is only carried out in order to be sure that attention was given to the audio-visual stimuli.
[3] This is the difference in N400 amplitude when a target word was preceded by a congruent or incongruent homonym-gesture combination.

Fig. 1 Stimulus material. The introduction was identical for all trial types. Panel A shows the four different gesture conditions both presented in the grooming and non-grooming-gesture style condition. The first two columns indicate the conveyed meaning of gesture (Dominant, Subordinate, upper two panels) and target word (Dominant or Subordinate). These four conditions (DD, DS, SD, SS) were used for statistical analysis. Panel B shows the grooming conditions that replaced 33% of the gesture trials in the grooming gesture stale condition (GS replaced DS and SS trials; GD replaced DD and SD trials). The grooming trials per see did not contain any disambiguating information and were only introduced to allow for the gesture style manipulation. Therefore, they were not part of the statistical analysis. Target words in the sentence material are in bold. Literal translation is in italics. Half of the participants were presented with the grooming speaker in the original gesture version and the non-grooming-speaker in the mirrored gesture version, the other half of the participants vice versa.

gesture on speech. In contrast, the gestures[4] of the *groomer* will show the N400-disambiguation effect exclusively for the subordinate target words (SS *vs* SD). If there is no speaker specific impact of gesture style, there are two possibilities: either grooming will make gesture cues less effective across the board (N400-disambiguation effect only for the subordinate target words for

both speakers) or grooming will be totally discarded (N400-disambiguation effect for dominant *and* subordinate target words for both speakers).

METHODS
Participants

Thirty-nine native speakers of German participated in the present study and gave written informed consent according to the guidelines

---

[4] Note that all trials containing grooming movements will be discarded from the analyzes because their sole purpose was to manipulate gesture style.

of the Ethics committee of the University of Leipzig. They were paid for their participation. Three of the participants were excluded because of excessive artifacts in the EEG signal. The remaining 36 participants (18 female; 19–30 years, mean 25.3 years) were right-handed (mean laterality coefficient 90.7; Oldfield, 1971), had normal or corrected-to-normal vision, no known hearing deficits and had not taken part in any previous experiment using the same or similar stimulus material.

Stimuli

Original material

The stimulus material for the present study was taken from Holle and Gunter (2007). In a series of experiments, they used 48 different homonyms with a clear dominant and subordinate meaning (for a more specific description of how dominant and subordinate meanings were determined, see Gunter *et al.*, 2003). For every homonym, 2 two-sentence utterances were constructed, which either contained a target word related to the dominant or the subordinate meaning of the homonym. Each utterance consisted of a short introductory sentence introducing a person, which was followed by a longer complex sentence describing an action by the respective person. The complex sentence was composed of a main clause containing the homonym and a subsequent sub-clause containing the disambiguating dominant or subordinate target word. Prior to the target word, the sentences for the dominant and subordinate versions were identical (see example sentences in Figure 1).

Complete set of materials

A professional actress was video-recorded while uttering the sentences. In a first session, she was asked to simultaneously perform a gesture that supported the sentence context which could either relate to the dominant or subordinate meaning of the homonym. In a second session, she was asked to produce meaningless hand-movements instead of gestures, so-called self-adaptors or grooming, while uttering the sentence stimuli. Dominant gestures, subordinate gestures and grooming were spontaneously created by the actress and performed coinciding with the initial part of the complex sentence that contained the homonym (for stimulus timing see Table1). To minimize the impact of facial cues and mimicking, the face of the actress was covered with a nylon stocking. The sentence material was combined with the gesture videos resulting in six different conditions (Dominant gesture-Dominant target word, Dominant gesture—Subordinate target word, Subordinate gesture—Subordinate target word, Grooming-Dominant target word, Grooming-Subordinate target word, see also Figure 1). Each of these six conditions (DD, DS, SD, SS, GD, GS) contained 48 stimuli resulting in a full stimulus set of 288 items. For more details about the recording scenario and preparation of the original stimuli, please see Holle and Gunter (2007).

Stimulus material of the present study

The goal of our study was to test how an addressee integrates gestural information when engaging with two communicational partners using different gesture styles. For this purpose, we had to create stimulus material, which on the one hand distinguished the two different gesture styles, but, on the other hand, should be as identical and comparable as possible. In order to obtain this goal, we applied the following stimulus manipulations. First, we created a horizontally flipped version for the complete set of gesture stimuli used by Holle and Gunter (2007). By using this procedure, all important characteristics, including movement characteristics, of the original video and flipped video versions were similar. Second, to create the illusion of two different communicative partners, the original speech material was subjected to a pitch manipulation using the Praat 4.5.1.0 software

Table 1 Stimulus properties

Hand movement	Target word	Gesture stroke onset	Gesture stroke offset	Homonym onset	Target word onset	Target word offset
D	D	2.07 (0.46)	2.91 (0.48)	2.84 (0.40)	3.78 (0.38)	4.16 (0.38)
D	S	2.07 (0.46)	2.91 (0.48)	2.84 (0.40)	3.80 (0.38)	4.17 (0.38)
S	S	2.17 (0.52)	3.01 (0.51)	2.84 (0.40)	3.80 (0.38)	4.17 (0.38)
S	D	2.17 (0.53)	3.01 (0.51)	2.84 (0.40)	3.78 (0.38)	4.16 (0.38)
G	D	2.16 (0.49)	2.96 (0.50)	2.84 (0.40)	3.78 (0.38)	4.16 (0.38)
G	S	2.16 (0.49)	2.96 (0.50)	2.84 (0.40)	3.80 (0.38)	4.17 (0.38)
Mean		2.13 (0.49)	2.96 (0.50)	2.84 (0.40)	3.79 (0.38)	4.17 (0.38)

Mean values of the time points are in seconds relative to the onset of the introductory sentence (*SD* in parenthesis).

(Boersma and Weenink, 2005). In one version, pitch of the speech was tuned upwards by two semitones, creating the perception of a high female voice. In a second version, pitch was tuned down by four semitones, creating the perception of a deep female voice. Thus, the general duration and timing of speech was kept identical for both high- and low-pitched versions of the sentences[5]. Next, we manipulated gesture style.

For the *non-grooming gesture style* condition, we used the four gesture conditions of the original stimulus material by Holle and Gunter (2007), i.e. the DD, DS, SD and SS conditions. Overall, the non-grooming speaker condition contained 192 stimuli (4 × 48 stimuli). Each homonym was presented four times. For the *grooming gesture style condition*, we used all the same four gesture conditions as in the non-grooming speaker condition (DD, DS, SD and SS), but sticking to the rationale of Holle and Gunter (2007) replaced 33% of all trials, i.e. 16 of 48 trials in each condition, with grooming trials (GD instead of DD and SD; GS instead of DS and SS). By using this rationale, we obtained a set that also contained 192 stimuli as in the non-grooming speaker condition, but contained six different conditions (DD, DS, SD, SS, GD, GS, with 32 trials per condition).

Additionally, we kept the number of repetitions (i.e. four) per homonym in the grooming speaker condition identical to the number of repetitions in the non-grooming speaker. To obtain both our goal of four repetitions per homonym, we had to balance out all grooming speaker conditions across six experimental lists, respectively participants.

In a last step, we coupled a particular speaker identity (voice and position on the screen) with a specific gesture style. To do so, the original and flipped versions of the stimuli were distributed across participants such that for half of them the original version was the non-grooming gesture style condition, and the flipped version was the grooming gesture style condition, and vice versa for the other half of the participants. In each of these two different participant subgroups, half of the participants heard the speech in the non-grooming gesture style condition in the high-pitched version and the speech in the grooming gesture style condition in the low-pitched version and vice versa for the other half. Thus, all mentioned video and speech manipulations were completely counterbalanced across participants for both gesture style conditions.

Therefore, any effects related to the difference in gesturing styles (non-grooming *vs* grooming) cannot be explained by perceptual differences in the video or auditory material.

Summing up, we presented participants with video stimuli of two easily discriminable speakers who had a different gesture styles

---

[5] Note that this manipulation only worked because the face of our actress was masked. Otherwise our participants would have immediately uncovered the illusion of two different gesturers.

(non-grooming *vs* grooming). A total of 192 trials were presented for each of the speakers. In the grooming gesture style condition, 67% of the stimuli contained gestures (DD, DS, SD, SS) and 33% contained grooming (GD, GS). Each of these six conditions consisted of 32 trials. In the non-grooming gesture style condition, 100% of the stimuli contained gestures (DD, DS, SD, SS), with each of these trial types consisting of 48 trials. Thus, overall, 384 items were presented. Note that although the number of trials per trial type is different for the two gesture styles this does not pose a problem. Since, we were interested in the impact of gestures on the addressee (grooming *per se* is non-informative and was only used for manipulating gesture style), we only used the identical 32 gesture trials (DD, SD, DS, SS) from both gesture style conditions for statistical analyzes. This selection ensured similar signal-to-noise ratios for the gesture trials of both gesture style conditions. The experimental design for our statistical analysis, therefore, was a $2 \times 2 \times 2$ design with gesture style (non-grooming, grooming), gesture (dominant, subordinate) and target word (dominant, subordinate) as within subject factors. Each condition contained 32 trials.

Procedure

Participants were seated in a dimly lit, soundproof chamber, facing a computer screen. They received the following task instructions: 'In this experiment, you will see a number of short videos with sound. More specifically, you will see two video screens located to left and the right of the middle of the screen. However, in each trial you will only see one of two different speakers during each trial. This speaker will appear in one of the video screens, whereas the other only shows the empty video background. During these videos, the speaker moves her arms. After some videos, you will be presented with a probe video of a movement or a probe word and asked whether you saw this movement or heard this word in the previous video'. They were additionally instructed to equally attend to the movement in the video and to the accompanying speech of both speakers.

A trial started with a fixation cross which was presented for 2000 ms, followed by the video presentation. The two video screens were placed on a black background to the left and right of the center of the monitor and each extended for 10° visual angle horizontally and 8° vertically. A visual prompt cue was presented after each video had ended. In 87.5% of the trials of both gesture style conditions (i.e. 168 of 192 trials), the cue was 'next video' and the next trial started. In 12.5% of the cases (or 24 videos) the cue prompted the participants to perform a task either related to the seen movements of the heard speech. Note that this task was introduced to control for the attention of the participants during the course of the experiment, and it was not used to measure the effects of gesture on speech comprehension. Task trials were equally distributed across all gesture and target word conditions for the grooming and non-grooming gesture style. In half of the task trials, the cue 'movement' indicated that the task was related to the gesture movement of the previously seen trial. After this cue, a short video of a movement, either being part of the previously seen gesture or not was presented. In the other half of the task trials, the cue 'word' indicated that the task was related to the speech of the previous trial. After this cue, a word was presented for 1500 ms which was either part of the heard speech or not. Then a question mark prompted participants to answer with a button press within 2000 ms after which feedback was given for incorrect and missed responses ('Wrong'/'Respond faster').

A session was divided into eight blocks of ~9 min each. For all blocks, the presentation order of the two speakers and items of different conditions was varied in a pseudo-randomized fashion. Key assignment was counterbalanced across participants. An experimental session lasted for ~90 min.

ERP recording

The EEG was recorded from 59 Ag/AgCl electrodes (Electrocap International). It was amplified using a PORTI-32/MREFA amplifier (DC to 135 Hz) and digitized at 500 Hz. Electrode impedances were kept below 5 kΩ. The left mastoid served as a reference. Vertical and horizontal electrooculogram (EOG) was measured for artifact rejection purposes.

Data analysis

All trials containing grooming (GS, GD) were removed from the data analysis, because we were only interested in the impact of the gestures (DD, DS, SD, SS). Additionally, only those gesture trials that were presented in both gesture style conditions for a particular participant were used for statistical analysis (i.e. 32 trials for both the non-grooming and grooming gesture style) to ensure comparability and interpretability of gesture style impact (i.e. S/N-ratio of the ERPs). EEG data were rejected offline by applying automatic artifact rejection using a 200 ms sliding window on the EOG ($\pm$ 30 µV) and EEG channels ($\pm$ 40 µV). After the selection and rejection procedure, ~25% of the data were excluded from further analysis resulting in 24 trials on average per gesture trial type. No offline data filtering was applied. Single-subject averages were calculated for every gesture trial type at the target word position.

Epochs were time-locked to the onset of the target word and lasted from 200 ms prior to the onset to 1000 ms afterwards. A 200 ms pre-stimulus baseline was applied. Ten regions of interest (ROI) were defined: anterior left (AL): AF7, F5, FC5; anterior center-left (ACL): AF3, F3, FC3; anterior center (AC): AFZ, FZ, FCZ; anterior center-right (ACR): AF4, F4, FC4; anterior right (AR): AF8, F6, FC6; posterior left (PL): CP5, P5, PO7; posterior center-left (PCL): CP3, P3, PO3; posterior center (PC): CPZ, PZ, POZ; posterior center-right (PCR): CP4, P4, PO4; posterior right (PR): CP6, P6, PO8. A time window ranging from 300 to 600 ms was used to analyze the N400-disambiguating effect of gesture on speech. A repeated measures analysis of variance (ANOVA) was performed using gesture style (grooming, non-grooming), gesture (D, S), target word (D, S), ROI (1, 2, 3, 4, 5) and region (anterior, posterior) as within-subject factors. Only effects which involve the crucial factors gesture style, gesture and target word will be reported. In all statistical analyzes, the Greenhouse-Geisser correction (Greenhouse and Geisser, 1959) was applied if necessary. In such cases, the uncorrected degrees of freedom (*df*), the corrected *P* values, and the correction factor ε are reported. To enhance the graphical presentation of the ERPs depicted in Figure 2, the curves were smoothed using a low-pass filter of 10 Hz.

RESULTS
Behavioral data: attentional control

Overall, the behavioral performance of the participants was good (mean percentage correct: 93.4%), indicating that the participants paid attention to our stimulus material. A repeated measures ANOVA with the factors gesture style (grooming, non-grooming) and task (movement task, word task) only resulted in a significant main effect of task (movement task: 95.0% correct, word task: 91.8% correct, $F(1,35) = 6.41$, $P = 0.016$, partial-eta-square $= 0.16$), but importantly did not reveal a main effect of gesture style [$F(1,35) = 0.90$, $P = 0.35$], nor an interaction of gesture style and task [$F(1,35) = 0.01$, $P = 0.90$]. Therefore, although our participants were slightly better in the movement task, they equally paid attention to both grooming and non-grooming speakers indicating that any ERP differences related to gesture style cannot be attributed to attentional differences.
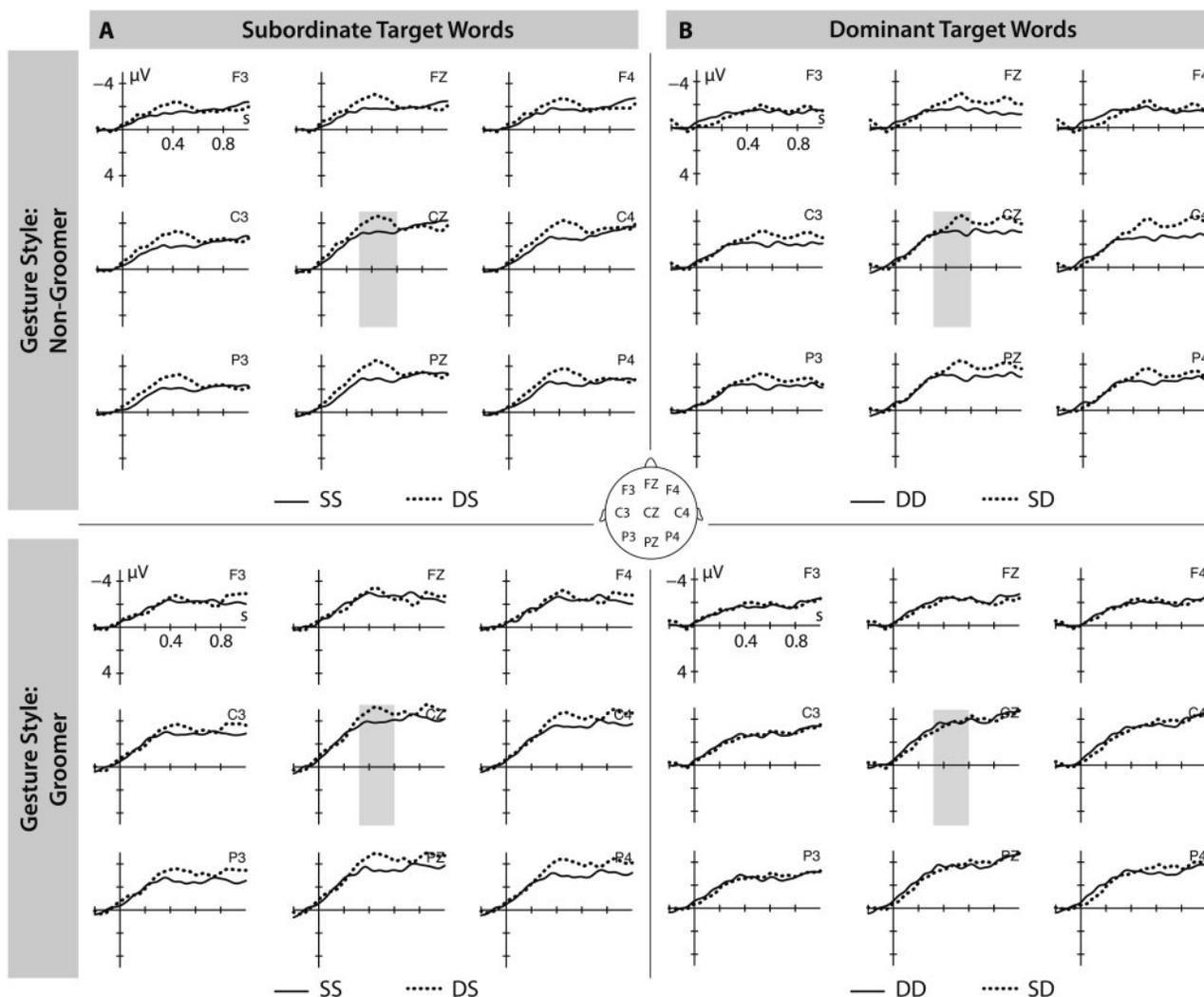
Fig. 2 ERP results: the left panel shows a significant N400 effect for gesture-target word congruency at subordinate target words for both the non-grooming (upper panel) as well as the grooming gesture style condition (lower panel). The right panel shows a significant N400 effect for gesture-target word congruency at dominant target words for the non-grooming gesture style condition (upper panel) whereas there is no such effect for the grooming gesture style condition (lower panel). The latter result indicates that gestures were treated as strong gesture cues for the non-grooming speaker, but as weak gesture cues for the grooming speaker. Gray areas denote the time window for the statistical analysis of the N400 effect (300-600 ms).

ERP data: gesture disambiguation

As can be seen in Figure 2, there is a clear N400 disambiguation effect for both subordinate and dominant target words in the non-grooming speaker. In the grooming speaker, the disambiguation effect is restricted to the subordinate target only. A repeated measures ANOVA with gesture style (grooming, non-grooming), gesture (D, S), target word (D, S), ROI (1, 2, 3, 4, 5) and region (anterior, posterior) as within-subject factors revealed a significant main effect of gesture style $[F(1,35) = 5.16, P = 0.029$, partial-eta-square $= 0.13]$, a significant interaction of gesture and ROI $[F(4,140) = 3.17, P = 0.046$, partial-eta-square $= 0.013, \varepsilon = 0.52]$, a significant interaction of gesture and target word $[F(1,35) = 34.55, P < 0.000$, partial-eta-square $= 0.50]$, and most importantly a significant interaction of the factors gesture style, gesture and target word $[F(1,35) = 4.77, P = 0.035$, partial-eta-square $= 0.12]$.

Subsequent step-down analyzes resulted in a main effect of gesture for *subordinate target words* irrespective of gesture style $[F(1,35) = 25.85, P < 0.0001$, partial-eta-square $= 0.43]$, i.e. for both the grooming and non-grooming speaker, gestures incongruent to the target word elicited a larger N400 than gestures congruent to the target word.

In contrast, the step-down analyzes for the *dominant target word* not only revealed a significant main effect of gesture $[F(1,35) = 5.39, P = 0.026$, partial-eta-square $= 0.16]$, but importantly also a significant interaction of gesture style and gesture $[F(1,35) = 6.54, P = 0.015$, partial-eta-square $= 0.16]$. Resolving this interaction, we found a significant main effect of gesture for the non-grooming speaker condition {paired-$t(35) = 3.412$; 95% CI of difference $[-1.62$ to $-0.41]$, $P = 0.002$, Cohen's $d = 0.60$} but not for the grooming speaker condition {paired-$t(35) = 0.384$; 95% CI of difference $[-0.48$ to $0.70]$, $P < 0.70$, Cohen's $d = 0.06$}.

Thus, only for the non-grooming speaker there was an effect of gesture-speech congruency (i.e. the N400 disambiguation effect) at dominant target words, suggesting that our participants indeed distinguished between the grooming and non-grooming gesture style in their gesture-speech integration.

DISCUSSION

For the non-grooming speaker, combining an ambiguous word with a gesture lead to a reduced N400 at the expected target words presented downstream the sentence indicating an easier integration with the

foregoing context. This N400-effect suggests that participants used gesture to disambiguate speech thereby replicating several ERP studies (Holle and Gunter, 2007; Obermeier *et al.*, 2011, 2012; Obermeier and Gunter, 2015). Although Holle and Gunter (2007) showed that the presence of grooming movements impacts the N400-disambiguation effect, the present experiment shows for the first time that this influence is speaker specific. That is, for the non-grooming speaker, the N400-disambiguation effect was present for both the dominant and subordinate target words, whereas for the groomer this effect was only present for the subordinate targets. The lack of an N400-effect at dominant targets suggests that the addition of grooming movements weakened the impact of gesture for this particular speaker and that word meaning frequency was the primary constraint used for disambiguating the homonym. Thus, it is not just seeing grooming movements that impacts the integration of gesture and speech generally (as in Holle and Gunter, 2007); rather, it impacts the integration specifically for the speaker who is producing those grooming movements. In other words, these ERP-data show that there is sensitivity to the personal communication style of a speaker and that affects the extent to which gesture and speech are integrated during language comprehension.

In this experiment, a speaker was identified by means of pitch height of the voice and the approximate position on the screen. The individual gesture style gave information regarding how to integrate gesture and speech of a particular speaker. Therefore, we regard gesture style as a visual indexical cue since it must be considered as a learned association with a particular speaker. It appears that just as a particular accent indexes social status and influences how speech is interpreted, gesture style is a cue to how relevant gesture is to a person's speech, and this also changes how listeners (viewers) interpret that speech. Interestingly, the present indexical cue relates to how a particular cognitive process (here, gesture-speech integration) has to be executed, a situation similar to that of for instance Hanulikova *et al.* (2012) where syntactic processing was affected by an acoustic indexical cue. Such a situation is clearly distinct from the more typical situation where an indexical cue like voice pitch or prosody, relates to a semantic association. In gesture terms, such a semantic situation could occur in the case of emblems[6]. When, for instance, either a football player or a hooligan (visual identification of a person) would use the raised fist emblem, they clearly mean two completely different things. To put things in perspective, it seems that independent of the modality of the indexical cue, the cue can be associated with particular semantic/world-knowledge information, but also with specific processing strategies. This observation is important because the literature on talker information typically explores the semantic perspective, leaving the processing strategy perspective relatively unexplored.

One can think of an indexical cue as relevant add-on information after the selection of the speaker has been accomplished. Thus knowing the associated gesture style of a particular speaker selectively impacts how gesture and speech influence each other. This impact seems to be relatively automatic in nature because the task did not require the participants to integrate gesture and speech explicitly[7]. This contrasts with the more controlled nature of processing acoustic indexical cues[8]. Whether or not this automaticity specifically relates to multisensory natural communication remains a question for further research.

Importantly, the present findings cannot be explained by physical differences between the gesturing/moving of the two speakers because of our mirroring manipulation. When comparing the gesture-speech integration of two different speakers, physical differences in gesturing can be an important confound. Because in the present experiment the gestures actually came from the same person, and we only mirrored the movements, the grooming effect must be related to processing and not to physical differences in the gestures *per se.*

In conclusion, this study showed that the specific way in which an interlocutor gestures selectively impacts how gesture and speech are integrated. It suggests that listeners (viewers) use gesture style as a visual cue to adjust their strategy for how much they should process gesture and speech produced together by another person. This finding highlights the importance of considering how larger contextual variables modulate the processing of gestures and speech at the utterance level (see also Holler *et al.*, 2014). We hope that this broader contextual approach will pave the way toward understanding how other individual patterns in producing gesture—and bodily expressions more generally—impact language comprehension.

Conflict of Interest

None declared.

REFERENCES

Boersma, P., Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.5.01) [Computer program], Amsterdam: Institute of Phonetic Sciences. Available: http://www.praat.org/ (accessed 28 October 2006).

Brown, C., Hagoort, P. (1993). The processing nature of the N400—evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1), 34–44.

Chu, M., Meyer, A., Foulkes, L., Kita, S. (2014). Individual differences in frequency and saliency of speech accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, 143(2), 694–709.

Chwilla, D.J., Brown, C.M., Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32, 274–85.

Creel, S.C., Tumlin, M.A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, 65, 264–85.

DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74–118.

Gillespie, M., James, A.N., Fedemeier, K.D., Watson, D.G. (2014). Verbal working memory predicts co-speech gesture: evidence from individual differences. *Cognition*, 123, 174–80.

Greenhouse, S.W., Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112.

Gunter, T.C., Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters*, 372, 52–6.

Gunter, T.C., Wagner, S., Friederici, A.D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: a difficult case for activation theories. *Journal of Cognitive Neuroscience*, 15(5), 643–57.

Hanulikova, A., Alphen, P.M., van Goch, M.M., Weber, A. (2012). When one person's mistake is anther's standard usage: the effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–87.

Holle, H., Gunter, T.C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19(7), 1175–92.

Holler, J., Beattie, G. (2003). Pragmatic aspects of representational gestures—do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, 3(2), 127–54.

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S.D., Özyürek, A. (2014). Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social Cognitive and Affective Neuroscience*, 10(2), 255–61.

Hostetter, A.B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137, 297–315.

Kelly, S.D., Byrne, K., Holler, J. (2011). Raising the ante of communication: evidence for enhanced gesture use in high stakes situations. *Information*, 2, 579–93.

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, 621–47.

Lattner, S., Friederici, A.D. (2003). Talker's voice and gender stereotype in human auditory sentence processing—evidence from event-related brain potentials. *Neuroscience Letters*, 339(3), 191–94.

Marstaller, L., Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin and Review*, 20, 496–500.

---

[6] Emblems are conventionalized hand postures ('thumbs-up') that have a clear regional meaning. Some neuroscience studies show that emblems have the equivalent ERP-pattern as words (Gunter and Bach, 2004).

[7] A common way to explore the nature of a particular cognitive process is to manipulate the depth of processing needed to perform an additional task. If effects related to the cognitive process under scrutiny only show up when an explicit task is used this is typically seen as evidence for a controlled nature of the process. When however, effects are present when a shallow, 'irrelevant' task is used (as the one in the present experiment) this is typically seen as evidence for an automatic nature of a particular cognitive process (Chwilla *et al.*, 1995).

[8] Creel and Tumlin (2011), for instance, showed that acoustic talker information is stored only when attended to or when it is highly relevant for the context.

Martin, C., Vu, H., Kellas, G., Metcalf, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *Quarterly Journal of Experimental Psychology, Series A*, 52, 813–39.

McNeill, D. (1992). *Hand and Mind—What Gestures Reveal About Thought*. Chicago, IL: The University of Chicago Press.

Obermeier, C., Gunter, T.C. (2015). Multisensory integration: the case of a time window of gesture-speech integration. *Journal of Cognitive Neuroscience*, 27(2), 292–307.

Obermeier, C., Dolk, T., Gunter, T.C. (2012). The benefit of gestures during communication: evidence from hearing and hearing impaired persons. *Cortex*, 48(7), 857–70.

Obermeier, C., Holle, H., Gunter, T.C. (2011). What iconic gesture fragments reveal about gesture-speech integration: when synchrony is lost, memory can help. *Journal of Cognitive Neuroscience*, 23(7), 1648–63.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.

Pouw, W.T.J.L., de Nooijer, J.A., van Gog, T., Zwaan, R.A., Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, 5, 359.

Regel, S., Coulson, S., Gunter, T.C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, 1311, 121–35.

Senju, A., Johnson, M.H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, 13, 127–34.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)Consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–59.

Van Berkum, J.J., van den Brink, D., Tesink, C.M., Kos, M., Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580–91.

Wallbott, H.G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28(6), 879–96.