Integrating Co-Speech Gestures into Sentence Meaning Comprehension

Ludmila Reimer (ludmila.reimer@rub.de)
Department of Philosophy II, Ruhr University Bochum, Universitätsstraße 150, 44870 Bochum, Germany

Maria Spychalska (maria.spychalska@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

Markus Werning (markus.werning@rub.de)
Department of Philosophy II, Ruhr University Bochum, Universitätsstraße 150, 44870 Bochum, Germany

Abstract

To investigate how co-speech gestures modulate linguistic understanding, we conducted an EEG experiment exploring the amplitude changes in the N400 component. We used videos of a person uttering underspecified action sentences which either featured no gesture or an iconic co-speech gesture that represented a more specific action. The following target sentence contained an instrument noun followed by its required action verb; these could either match the action represented in the previously seen gesture or mismatch it. We measured ERPs for both the nouns and the verbs and found an N400 effect for mismatching target words as well as a sustained positivity effect for both gesture conditions.

Keywords: iconic co-speech gestures; sentence meaning composition; enacted cognition; N400; EEG

Introduction

Everyday human communication is multimodal; we employ spoken language, facial expressions, and various gestures to express ourselves. These various sources of information streams seem to be smoothly understood by conversation partners, yet it is not fully uncovered, how they are integrated on a cognitive level. Gestures interact with speech in many ways: they facilitate language comprehension, production, and acquisition (Holler, 2022; Iverson & Goldin-Meadow, 2005; Özyürek, 2014); they are found to prime or activate internal action representations (Goldin-Meadow & Beilock, 2010; Krauss et al., 2000; Pouw et al., 2014).

Up to this date, McNeill (1992) is still widely cited in gesture research. His analysis of gesture types and gesture phases, and his assumptions about iconic co-speech gestures (gestures accompanying speech and resembling concrete actions or outlining the shape of an object) and their (presumably semantic) content are still present in modern research, as is exemplarily evident in the extensive review on gesture research methodologies by Arachchige et al. (2021). More precisely speaking, for McNeill (1992), co-speech gestures, carry non-reducible meaning. Utterances and co-speech gestures convey information in two separate streams, but those two streams are not separable from one another and thus cannot be analyzed meaningfully each on their own (McNeill, 2016). However, theoretical approaches often assume co-speech gestures to have meaning or content similar to single words and phrases that can be analyzed on its own: Ebert (2014a, 2014b) argues that gestures enter truth-conditions differently than speech, and contribute only non-at-issue content unless their contribution to the at-issue content of a sentence is explicitly indicated; phrased differently, speech is the main information stream processed in conversation (and thus is at-issue) and gestures provide secondary information that is not entering the conversation unless explicitly marked, e.g., by adding phrases like "in such a way" while performing a gesture. Both positions share the assumption, though, that gestures carry meanings and that these meanings interact with spoken utterances. In the case of co-speech gestures, there seems to be no part of the linguistic utterance, neither syntactically overt, nor syntactically covert, that explicitly integrates, embeds or imports the gestural content into the semantic content of the sentence. The only relationship between gestures and utterances seems to be that they are part of the same act of communication, i.e., utterance and gesture are performed (more or less) simultaneously by the speaker, perhaps with a coordinated intention to evoke a certain understanding in the listener.

We assume that gestures, even though very ambiguous on their own, contain information that can become very specific when combined with speech. They can add more nuance to the shape of a spoken about object, add details to a described action (such as speed, orientation), and even indicate which tool was used during an action. We are interested in this latter kind, i.e., iconic co-speech gestures that accompany action verbs and provide detailed information on actions and tools used in comparison to the spoken utterances on their own. We aim to explain how iconic co-speech gestures influence speech processing and to produce a series of experiments to test our models and hypothesis. In this paper, we will present one of our EEG studies investigating the influence of action representing iconic co-speech gestures on sentence meaning comprehension. To illuminate how the different relationships between iconic co-speech gesture and utterance modulate semantic comprehension during communication, we investigate the semantic predictions generated by the speaker in the listener.

The N400 Effect & Gesture

Predictive processing is widely acknowledged in cognitive and neuroscience as a general mechanism by which the

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

subject at every point in time generates the most probable prediction of the next event on the basis of ongoing perceptual input and learned statistical regularities (Clark, 2013; Hohwy, 2013). To directly investigate whether gestures modulate the expectations of listeners, we will use event-related brain potentials (ERPs), focusing especially on the N400 component. ERPs are scalp-recorded voltage changes time-locked to trigger events, such as spoken or written word. The N400 is a negative shift in the ERP waveform starting at ca. 200ms and peaking around 400ms post-stimulus onset over the centro-parietal scalp sites. (Kutas & Federmeier, 2011; Swaab et al., 2012). It is modulated by the predictability of the stimulus, e.g., it tends to be larger for words that are semantically less appropriate or less expected in the context (Kutas & Federmeier, 2000; Kutas & Hillyard, 1980; Kutas & Van Petten, 1994, 1994) The size of the N400 is inversely correlated with the cloze probability of the triggering word, i.e., the percentage of individuals who would continue a given sentence fragment with that word (Federmeier et al., 2007).

Prior EEG research has shown that gestures can modify the N400 component or even elicit a similar effect themselves: In such studies, gestures elicited an N400-like effect when they did not match a preceding video or picture, referred to as N450 (Wu & Coulson, 2005). Moreover, words as well as pictures elicit an N400 effect when they do not match a preceding (silent) gesture, suggesting that gestures can establish expectations about upcoming linguistic input (Wu & Coulson, 2007, 2015). As described above, the N400-effect has been linked to the computation of semantic content; thus, it can be assumed that gestures at least interact with meaning and possibly carry meaning themselves.

In another ERP study, it was shown that gestures help to disambiguate homonyms and that mismatching gesture elicit N400 effects (Holle & Gunter, 2007). However, the same paper reports that the presence of "meaningless gestures", i.e., grooming hand movements, reduce the N400 effects, and also the overall helpfulness of the iconic gestures to disambiguate the homonym. Obermeier et al. (2011) conducted follow up studies, using gesture fragments and manipulated the task and the synchrony of gesture and speech. Their experiments showed that if speech and gesture were off-set and paired with a task that required participants explicitly rating the compatibility of gesture and speech, the effect was similar to a set-up with no overt task but synchronous speech and gesture. However, in another variation without an overt task nor with an asynchrony, an N400 could not be detected. This again is in favor of multiple factors that influence how gestures are processed with speech (or how speech is processed in the presence of gestures).

Experimental Design and Motivation

The goal of the study was to investigate whether co-speech gestures modulate semantic predictions for upcoming content. To this aim, we focus on measuring the modulation of the N400 component on words that are expected or not, based on the gesture used in prior context. More precisely, we wanted to use naturalistic videos of a person uttering an underspecified action sentence, featuring no gesture or an iconic co-speech gesture that represents a more specific action. The following target sentence contained an instrument noun followed by its required action verb; these could either match or mismatch the previously seen gesture.

Table 1:
Solid lines indicate congruent continuations (matches), dashed lines incongruent continuations (mismatches). Target words are underscored.

Context Sentences	Target Sentences
No Gesture *Das Kind ist dabei, die Kekse zu backen.* (The child is baking cookies.)	**Target I** *Es hat sie schon mit dem <u>Förmchen ausgestochen</u>.* (They have already <u>cut</u> them out with the <u>cookie cutter</u>.)
Gesture I *Das Kind ist dabei, die Kekse zu backen.* + g_1 [hand moving down as if stamping, vertically]	**Target II** *Es hat sie schon mit dem <u>Pinsel bestrichen</u>.* (They have already <u>glazed</u> them with <u>the brush</u>.)
Gesture II *Das Kind ist dabei, die Kekse zu backen.* + g_2 [hand moving left to right, horizontally]	

To do so, we used carefully constructed materials in which the linguistic material did not vary in regard to modulating expectations about upcoming words, so any change in the processing would be due to a change in the gesture. To avoid effects caused by gestures representing completely unrelated actions, we made sure that all iconic co-speech gesture were congruent with the simultaneously uttered context sentence.

Such an approach is novel in several regards: First, experiments that measured ERPs elicited by linguistic targets following a gesture used paradigms of isolated and silent gesture videos followed by single probe words. The linguistic targets elicited N400-like effects, however, this depended on the timing between gesture and probe word (Habets et al., 2011; Wu & Coulson, 2007).In contrast, our sentences belong to one discourse and do not solely rely on the simultaneity of gesture and speech. Second, other experiments using full sentences focused on the (mis)match between a co-speech gesture and the simultaneously uttered phrase or word (Hintz et al., 2023; Özyürek et al., 2007), not on the mismatch elicited by one gesture in regards to the following linguistic input, as we do. Some of these studies also varied the linguistic context sentences, adding another factor that might influence the processing of both utterances and co-speech gestures (Hintz et al., 2023), while we kept linguistic input constant and varied only the gestures. Nevertheless, these and other studies (see review by Arachchige et al. (2021)) already showed that iconic co-speech gestures can interact with the semantic processing of linguistic input, even if the task is not requiring participants

to actively assess both inputs (Wu & Coulson, 2007). Such interactions can be taken as evidence that gestures themselves have semantic content (or represented semantic content), as was already claimed by Kendon and McNeill (1988; 1992) However, we are more hesitant to assume that the information provided by gestures is semantic content (representation), as there are many factors that can interact with semantic processing without being themselves semantic in nature. The notion of iconic gestures providing "conceptual information" is more accurate considering the current state of research and is adopted from Özyürek (2007).

To investigating the meaning component of iconic co-speech gestures, we restricted our materials to scenarios using tools and actions requiring these tools. We present underspecified sentences uttered by a speaker and they were either presented without a gesture or combined with one of two possible congruent iconic co-speech gestures.

A target sentence describing a more specific action congruent with the first sentence is presented afterwards; this more specific action is only congruent with one of the gestures used in the context sentence. Thus, we arrive at three conditions: Neutral, Match, and Mismatch. Since we consider iconic co-speech gestures to be useful tools when it comes to disambiguating spoken utterances, we expected the Mismatch conditions to be more surprising and thus elicit an N400 effect. For the Neutral condition, our expectation was that we get an intermediate N400 effect since there is no gesture present that could lower or raise the expectation about the upcoming input. Since we are presenting the tool noun before the corresponding action verb, we additionally expect these effects to be stronger for the tool noun.

Material Preparation

We constructed our German materials with uniform constraints and later assessed the suitability of the material in online surveys. First, we prepared sets of context and target sentences in a systematic way:

a) Context: "The agent x is v_{gen}-ing object y."
b) Target: "The agent x is v-ing it with n."

Note that in German all target sentences only divert in the end, i.e. they use the same words up until the target words are shown (v followed by n), as can be seen in Table 1 (n are underscored once, v twice). The target sentences were constructed in pairs, featuring two different instruments, which are denoted by the nouns n, required by their respective (specific) action, denoted by the verbs v. This results in verbs v_1 and v_2 as well in n_1 and n_2. All four are coherent continuations of the context sentences when those are presented without gestures as well as with gestures corresponding to actions denoted by verbs v. However, they are not coherent continuations for the context sentences presented with the co-speech gesture g that corresponds to the other v and n; i.e., only v_1 and n_1 are coherent continuations of the general context combined with g_1. To use the example in Table 1, we only expect "They have already cut them out with the cookie cutter" after we have seen g_1 representing the movement of using a cookie cutter, i.e., a vertical movement of the hand with the palm down. By using crossed pairs, we aim to average out slight variations in the predictability of the target words.

We made sure to not use verbs representing actions that require the same tool or very similar hand motions. All target pairs, nouns and verbs, were controlled for frequency and for the grammatical gender of nouns n. To ensure that both target sentences were equally likely to follow the context sentences, we determined the GloVe values between: (1) context sentences and target verbs v, (2) context sentences and the nouns n (3) context verb v and target noun n. We only kept target pairs with similar values. We used an implementation of GloVe (Global Vectors, (Pennington et al., 2014)) based on all articles from German Wikipedia and ca. three million news articles from Leipzig Corpora Collection (Goldhahn et al., 2012). Since GloVe is a measure of semantic similarity based on co-occurrences in corpora, we produced more sets than needed , namely 57, and later assessed our material with additional surveys (see section Sentence Ratings).

Video Recordings, Post-Processing, and Assessment

Videos for the context sentences, combined with three gesture conditions (g_1 , g_2 , no gesture) were recorded in full HD resolution (1920x1080 pixel) and a frame rate of 50fps. The speaker was recorded from the knees up, allowing for framing with the hands to be visible at all times. The speaker was not informed about the purpose of the experiment prior to filming; the only instructions received were to (1) read out the first sentence of a sheet while trying to convey additional information present in the second sentence and (2) to read out the same sentence once without gesturing. The gestures were not rehearsed, but spontaneously produced by the speaker.

In post-processing, the speaker was centered. The videos were cut with ca. 600ms (30 frames) before gesture or speech onset and 600ms after offset. Speech on- and offsets were determined by inspecting the audio track, and gesture on- and offsets were defined by the hands leaving and returning to their resting position. This resulted in 3 videos with the same sentence; to minimize pronunciation variation effects across these 3 sentences, all audio was rated in terms of clarity of speech and uniform speed by three German native speakers. The best-scoring audio was used for all 3 videos. Next, the speaker's face was blurred to mask discrepancies between the new audio and mouth movements, and to eliminate the influence of facial expressions of the speaker on the listener.

Next, the recorded materials were assessed with three online surveys. Out of the initial 57 sets, we chose the best 40 sets for use in the following EEG studies. The survey studies are briefly summarized below, detailed methos and findings are reported by Reimer and Werning (2023).

Recognizability To determine how recognizable actions represented by co-speech gestures were in the absence of any linguistic information, we stripped off audio and only showed videos of the gesture material to the participants. They had to rate how well the verb (target) described the shown gesture. Overall, all match target verbs were rated as good

descriptions of the shown gestures, indicating that participants perceived a meaning or semantic component in the gestures and matched them to the target verbs.

Sentence Ratings Since the GloVe values (used to control our linguistic material) are only capturing semantic similarity but not the relationships between the objects expressed by the words (e.g., tool nouns n and their corresponding action verbs), we asked participants to rate our linguistic materials in terms of how likely a sentence (Target $Noun$) was likely to follow after a given context sentence. In addition to the material used in this experiment, we created sentences using the target verbs in place of the more general context verbs. For the second sentence, we modified our Target $Noun$ sentences to be "to do, they used noun n"; this was done to avoid target verbs being in both first and second sentences. This indicates participants were sensitive to the action denoted by the verb and matched the corresponding instruments denoted by the nouns n accordingly.

Video Ratings Participants were shown a video of our speaker uttering the context sentence, either performing a matching, mismatching, or no gesture. They had to rate how likely they thought a displayed sentence (the same as used for the sentence ratings, "to do, they used noun n") would follow the video.

Experiment Conduction

Participants 33 monolingually raised German native speakers were recruited. The participants were right-handed, between 19 and 32 years old, had no neurological impairments, did not take psychoactive medication, and had normal or corrected to normal vision. We excluded 8 participants due to excessive noise in their data or a high error rate in the behavioral task. Out of the remaining 25 participants, 18 identified as female, 6 as male, and one non-binary. The mean age was 24,88 (SD=3,20).

Materials We used the videos of the context sentences with the three gestures conditions (g_1 , g_2 , no gesture) and the $Noun$ target sentences. Context sentences without gestures followed by the target sentences formed the Neutral condition; context sentences with gestures followed by congruent target sentences formed the Match condition; and analogously, if they were followed by the incongruent target sentences, they formed the Mismatch condition. This results in 80 trials per condition.

Procedure The experiment was conducted in the EEG laboratory belonging to the chair of language of cognition at the Ruhr-University Bochum, Germany. Participants were seated inside an electrically isolated and acoustically attenuated cabin, in front of a shielded glass with a computer screen behind it. USB-powered speakers were placed inside the cabin as well as a USB-powered Cedrus response pad with two designated buttons. Every subject signed a written, informed consent of participation. They were informed that their data will be stored and handled in a fully anonymous manner, and that they have the right to withdraw from the experiment at any time. Participants were screened regarding their demographic criteria and handedness, as described above, as well as for their Autism-Spectrum Quotient (Baron-Cohen et al., 2001) and working memory, ensuring a homogenous score across participants to reduce any potential interference during language processing.

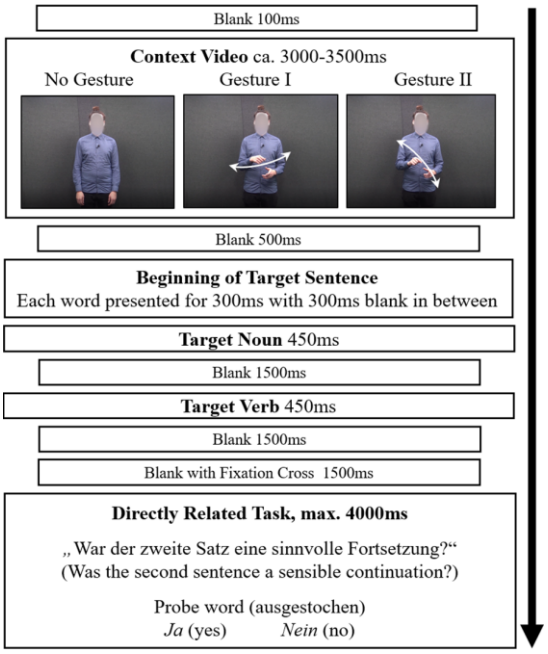

Figure 1: Schematic representation of one trial

The experiment started with an instruction screen and 6 exercise trials, each followed by the task screen. An experimenter was present during the exercise phase to aid the participants if they had any question concerning their task. After the exercises, the experimenter gave the participants feedback regarding the correctness of their task and left. The trials of the main experiment did not receive feedback. The main experiment was shown in 8 blocks, each about 7min longs, with breaks in between, resulting in a total runtime of about 65 to 85 minutes per participant. They were shown 30 trials per block, 240 trials total. A video of a person uttering the context sentence (with and without gestures) was played, then the target sentence was displayed word by word. After that, participants were prompted to answer by pressing a button corresponding to "Yes" or "No", which were displayed on randomly alternating sides, half the time "Yes" was on the left (for details, see Fig. 1).

EEG Recording EEG was recorded with a 64-channel BrainAmp actiCAP EEG system. FCz location was used as the physical reference and AFz as the ground electrode. Four electrodes were relocated and used to measure eye-movement: FT9 (HEOGL) and FT10 (HEOGR) were used for horizontal movements (placed on the right and left temple), FPz (VEOGO) and Iz (VEOGU) for vertical movements (placed above and below the right eye). Impedance was kept below $5k\Omega$. The EEG was recorded with a sampling rate of 1000 Hz, a 10s low cut-off filter and a hardware anti-aliasing filter. The EEG data was processed using the software Brain Vision Analyzer 2.0. An off-line band-pass filter was applied: 0.1–30 Hz (order 4), and the data was down-sampled to 500Hz. Breaks and other periods

Table 2: Overview of Significant Clusters , left for ERPs measured on the noun, right for ERPs measured on the verb

Target Noun				**Target Verb**			
Comparison	Cluster Polarity	Time in ms	p value	Comparison	Cluster Polarity	Time in ms	p value
Match - Neutral	positive	164 - end	<0.001	Match - Neutral	positive	672 - 872	0.008
Mismatch - Neutral	positive	40 - end	<0.001	Match – Neutral	positive	314 - 578	0.010
Mismatch - Match	negative	260 -780	<0.001	Mismatch – Match	negative	182 - 976	<0.001

of noisy signal were excluded manually. Automatic raw data inspection rejected all trials that had an absolute amplitude difference higher than 150μV/150ms or with activity lower than 0.5μV per 100ms intervals. The maximal voltage step was 50μV/ms. Both vertical and horizontal eye-movements were corrected by means of independent component analysis. Data was re-referenced to the average of mastoid electrodes (TP9 and TP10). Segments from 200ms pre-target onset until 1000ms post-onset were extracted for every trial and condition. Baseline correction used the 200ms interval preceding the stimulus onset. Segments with any remaining physical artifacts, including those with the amplitude lower than $-90\mu V$ or higher than $90\mu V$, were excluded and condition averages were calculated for each subject.

We excluded participants' data with less than 65% usable segments or who had an error rate in the behavioral task (more than 40%, indicating that the subjects were not paying attention). The minimal number of segments preserved per subject and condition was 26/40 and an average of 35,2/40.

Statistical Analysis To evaluate whether the recorded ERPs differ significantly between conditions, we performed cluster-based permutation statistical test using Matlab Fieldtrip package (Maris & Oostenveld, 2007; Oostenveld et al., 2011). For each target word, separately, all three conditions were compared pairwise, in epochs of 0-1000ms post-target onset, over all channels. The test used α of 0.025 in the clustering procedure (t-tailed dependent t test) and 10000 permutations for evaluating the p-value.

Results Noun Visual inspection of the Grand Averages in Fig. 2) revealed long-lasting positivity for both gesture conditions (Match and Mismatch) with regard to the Neutral condition, the difference being ca. 3 µV to 4 µV. A cluster-based permutation test revealed these effects to be significant, based on significant, positive clusters spanning from around 100ms post onset until the end of the epoch (see Table 2). In both comparisons, the effect extended over the central-medial and the posterior regions of the scalp. In addition, we observed a negativity for the Mismatch relative to Match condition around the N400 time window. The statistical comparison proved this effect to be significant (see Table 2). This negativity extended over the whole scalp, except for the frontal pole and antero-frontal regions, and amplitude difference in the centro-parietal region appeared to be largest.

Results Verb We also measured the ERPs elicited by the second target word, namely the verb *v* (Grand Averages in Fig. 3). Visual inspection of the Grand Averages indicates that there is a negativity effect for the comparison Mismatch-Match starting around 300ms and extending towards the end of the epoch of ca. 1 µV to 2 µV difference. A negativity effect is present in the comparison Neutral vs. Match, spanning over a similar time window. A cluster-based permutation test revealed the effects to be significant. There was a significant negative cluster for Mismatch vs. Match starting around 200ms post-stimulus onset and extending until the end of the epoch, most prominent over the central-medial and posterior regions. For Match vs. Neutral, the effect is supported by two consecutive cluster: the earlier positive cluster extended over the frontal-central region and the later cluster over the whole scalp (see Fig. 4). The Neutral-Mismatch comparison was not significant.

Discussion

Nouns For the target nouns, we found a strong negativity for Mismatch vs. Match, as expected, in the typical time window and scalp distribution of the N400 effect. However, this N400 negativity overlaps with the time window of the positivity

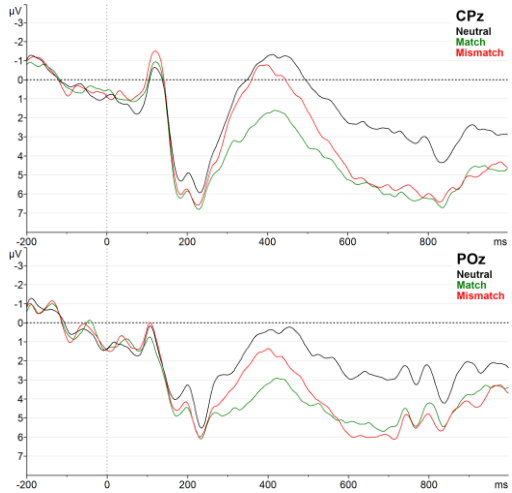

Figure 2: Grand Averages for word 1 (noun)

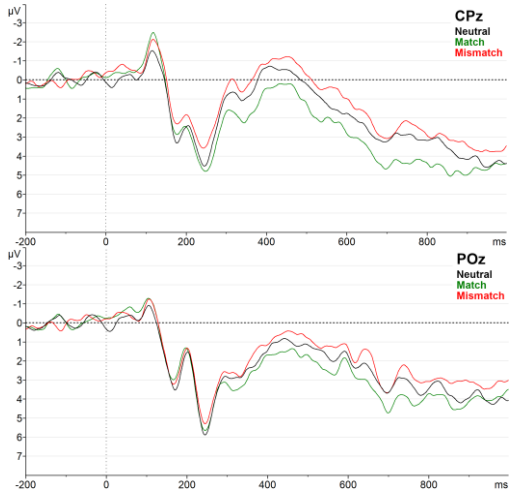

Figure 3: Grand Averages for word 2 (verb)

Mismatch vs. Neutral, suggesting that we observe an overlap of components, with the positivity being masked by the negative going shift of the Mismatch condition.

The sustained positivity for both Neutral-Match and Neutral-Mismatch could be interpreted as a modulation of the P300 and/or the P600 components. Due to the nature of the design, two third of all trials used gestures and only one third did not. The likelihood of encountering a gesture and having to integrate its meaning was thus higher, causing the task demands between these conditions to differ. However, the presence of the gesture itself makes participants attend to the gesture trials more closely, given also that the task is related to the gestures. The P300 effect, a positivity elicited roughly between 250 and 500ms post-stimulus onset, is known to be modulated by the probability of the target as well as by its relation to the task. Especially the P300b subcomponent tends to be larger for items that are task-relevant, at focus and thus awaited in the experiment (Polich, 2003, 2007).

Due to the temporal and topographical overlap of the P300b and the N400 it sometimes can be difficult to dissociate the contribution of the two components to the differences between experimental conditions (Alday & Kretzschmar, 2019; Roehm et al., 2007).

Interestingly, Hintz (2023) report a similar pattern for their experiment that measured ERPs of target words that were temporally overlapping with gestures. Their control or neutral condition employed "meaningless" movements such as scratching, and their target condition used matching iconic co-speech gestures. For the comparison gesture vs. control movement, they also found a positivity extending to the end.

The prolonged positivity effect could indicate a prolonged P300 and/or an additional modulation of the P600 component for conditions with gesture. P600 is a late and often prolonged positive shift in the ERP waveform, posteriorly maximal around 600 ms post-onset (Hagoort et al., 1993; Swaab et al., 2012) and sometimes argued to be closely related to the P300 (Leckey & Federmeier, 2019)). It is modulated by general processing demands and task (Hagoort et al., 1993; Kolk & Chwilla, 2007; Osterhout & Holcomb, 1992)and has been argued to reflect combinatorial aspects of linguistic processing (form-to-meaning mapping) (Brothers et al., 2020; Kuperberg, 2007) or even semantic integration mechanisms(Brouwer et al., 2012). Most importantly, the P600 appears to be triggered in cases of more effortful processing that requires combining both structural and meaning-related analyses(Brothers et al., 2020).

As the integration of the target words may be more demanding after having seen a gesture, this could lead to a more pronounced P600 effect for the gesture conditions: The gesture remains in working memory storage and is retrieved upon encountering the target that confirms or contradicts the participant's previous interpretation of the gesture.

Verbs Given that the verbs always matched the stereotypical actions required to use the tools denoted by the nouns n, i.e. v_1 always followed n_1, and v_2 always followed n_2, we did not expect big differences between the conditions, as it is very likely that the stereotypical verbs follow their instrument

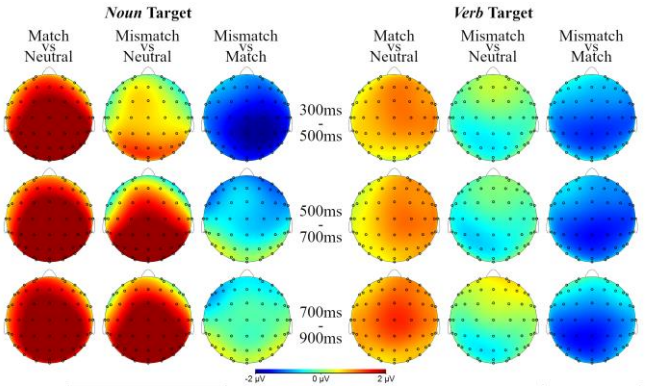

Figure 4: Topographical maps for the significant comparisons, in 200ms intervals

nouns. We considered the possibility that participants might interpret the mismatching tools as either being erroneously uttered or used atypically to perform the actions represented by the gestures, in which case the effect would be also carried onto the verb position. The significant negative cluster in the Mismatch-Match comparison confirms this suspicion. We interpret this as an N400 followed by sustained negativity, since the gesture raised the likelihood of an atypical action performed with the instrument in comparison to the Neutral and Match conditions.

We prefer the interpretation that participants were open to the possibility that the noun n was being atypically used for the action represented in the gesture g. This is in line with studies that modify the context in which a target sentence is uttered, such that a word denoting an unusual or atypical action is more expected than a semantically close word. Most famously reported in the "Peanuts in Love" paper by Niewland and Van Berkum (2006), but also Cosentino et al. (2017) and Werning et al. (2019) investigated this.

The positivity observed for the Neutral vs. Match conditions could indicate on the one hand, an attenuation of the N400 for the Neutral condition, which is in line with our initial hypothesis that without a gesture, there is no additional information that could lower or raise the expectation for the verb. On the other hand, what we observe might be a modulation of the P300, which is less pronounced than the one observed for the noun target, since the task demands are lowered at this stage of sentence processing. The ERP results for the verb target show that information provided by the iconic co-speech gestures remains available to participants throughout the discourse.

Conclusion

We found that an iconic co-speech gesture makes a difference for a listener's probabilistic prediction regarding an upcoming instrument noun and a following action verb, and thus has a semantic effect on linguistic comprehension.

Since our experimental task overtly asks participants to judge the likelihood of the target sentences, we cannot infer whether this process is happening automatic or is only activated in relevance to the task.

References

Alday, P. M., & Kretzschmar, F. (2019). Speed-Accuracy Tradeoffs in Brain and Behavior: Testing the Independence of P300 and N400 Related Processes in Behavioral Responses to Sentence Categorization. Frontiers in Human Neuroscience, 13, 285. https://doi.org/10.3389/fnhum.2019.00285

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31(1), 5–17.

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. Neurobiology of Language, 1(1), 135–160.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. In Brain Research (Bd. 1446, S. 127–143). Elsevier B.V. https://doi.org/10.1016/j.brainres.2012.01.055

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36, 181–253.

Cosentino, E., Baggio, G., Kontinen, J., & Werning, M. (2017). The time-course of sentence meaning composition. N400 effects of the interaction between context-induced and lexically stored affordances. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2017.00813

Ebert, C. (2014a). The non-at-issue contributions of gestures and speculations about their origin. Workshop Demonstration and Demonstratives.

Ebert, C. (2014b). The semantic impact of co-speech gestures. Embodied meaning goes public - gestures, signs, and other visible linguistic effects of simulation processes.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. Brain Research, 18(1146), 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12), 759–765.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. Perspectives on psychological science, 5(6), 664–674.

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. Journal of Cognitive Neuroscience, 23(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and Cognitive Processes, 8(4), 439–483.

Hintz, F., Khoe, Y. H., Strauß, A., Psomakas, A. J. A., & Holler, J. (2023). Electrophysiological evidence for the enhancement of gesture-speech integration by linguistic predictability during multimodal discourse comprehension. Cognitive, Affective, & Behavioral Neuroscience, 23(2), 340–353. https://doi.org/10.3758/s13415-023-01074-8

Hohwy, J. (2013). The Predictive Mind. Oxford University Press.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. Journal of Cognitive Neuroscience. https://doi.org/10.1162/jocn.2007.19.7.1175

Holler, J. (2022). Visual bodily signals as core devices for coordinating minds in interaction. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1859), 20210094. https://doi.org/10.1098/rstb.2021.0094

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. Psychological Science, 16(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

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. https://doi.org/10.3389/fpsyg.2021.634074

Kendon, A. (1988). How Gestures Can Become Like Words. In Cross-cultural perspectives in nonverbal communication.

Kolk, H. H. J., & Chwilla, D. J. (2007). Late positivities in unusual situations. Brain and Language, 100(3), 257–261. https://doi.org/10.1016/j.bandl.2006.07.006

Krauss, R. M., Chen, Y., & Gotfexnum, R. F. (2000). 13 Lexical gestures and lexical access: A process model. Language and gesture, 2, 261.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. Brain Research, 1146(1), 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Science, 4(12), 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology, 62(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. Science, 207(4427), 203–205.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics Electrified: Event-related potential investigations. In M. A. Gernsbacher (Hrsg.), Handbook of Psycholinguistics (S. 83–143). Academic Press.

Leckey, M., & Federmeier, K. D. (2019). Electrophysiological Methods in the Study of Language Processing. In G. I. de Zubicaray & N. O. Schiller (Hrsg.), The Oxford Handbook of Neurolinguistics (S. 41–71).

Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190672027.013.3

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1), 177–190.

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

McNeill, D. (2016). Why We Gesture. In Why We Gesture. https://doi.org/10.1017/cbo9781316480526

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. Journal of Cognitive Neuroscience, 18(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

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), 1648–1663. https://doi.org/10.1162/jocn.2010.21498

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience, 2011. https://doi.org/10.1155/2011/156869

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language, 31(6), 785–806.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130296. https://doi.org/10.1098/rstb.2013.0296

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

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1532–1543.

Polich, J. (2003). Theoretical overview of P3a and P3b. In Polich (Hrsg.), Detection of Change: Event-Related Potential and fMRI Findings (S. 83–98). Kluwer Academic Press: Boston.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology, 118(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

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(APR), 1–14. https://doi.org/10.3389/fpsyg.2014.00359

Reimer, L., & Werning, M. (2023). Modelling the Integration of Co-Speech Gestures into Sentence Meaning Composition. Proceedings of the Annual Meeting of the Cognitive Science Society, 45, 625–631.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. Journal of Cognitive Neuroscience, 19(8), 1259–1274.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-Related ERP Components. In S. J. Luck & E. S. Kappenman (Hrsg.), The Oxford Handbook of Event-Related Potential Components (S. 397–440). Oxford University Press.

Werning, M., Unterhuber, M., & Wiedemann, G. (2019). Bayesian Pragmatics Provides the Best Quantitative Model of Context Effects on Word Meaning in EEG and Cloze Data. Proceedings of the 41th Annual Conference of the Cognitive Science Society.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. Psychophysiology. https://doi.org/10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. Brain and Language, 101(3), 234–245. https://doi.org/10.1016/j.bandl.2006.12.003

Wu, Y. C., & Coulson, S. (2015). Iconic Gestures Facilitate Discourse Comprehension in Individuals With Superior Immediate Memory for Body Configurations. Psychological Science. https://doi.org/10.1177/0956797615597671

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number GRK-2185/1 (DFG Research Training Group Situated Cognition), and 367110651 (PI: MW), within the Priority Program XPrag.de (SPP1727).