The processing of semantic complexity and co-speech gestures in schizophrenia: a naturalistic, multimodal fMRI study

Paulina Cuevas*^{1,2}, Yifei He*^{1,2}, Miriam Steines^{1,2} & Benjamin Straube^{1,2}

¹ Translational Neuroimaging Lab Marburg, Department of Psychiatry and Psychotherapy, Philipps University Marburg, Rudolf-Bultmann-Strasse 8, 35039 Marburg, Germany

² Center for Mind, Brain, and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Hans-Meerwein-Strasse 6, 35032 Marburg, Germany

* Contributed equally

Corresponding address: Yifei He, Department of Psychiatry and Psychotherapy, Philipps University Marburg, Rudolf-Bultmann-Str. 8, 35039, Marburg, Germany

Corresponding email: yifei.he@staff.uni-marburg.de, ORCID: 0000-0002-2826-4230

Abstract

Schizophrenia is marked by aberrant processing of complex speech and gesture, which may contribute functionally to its impaired social communication. To date, extant neuroscientific studies of schizophrenia have largely investigated dysfunctional speech and gesture in isolation, and no prior research has examined how the two communicative channels may interact in more natural contexts. Here, we tested if patients with schizophrenia show aberrant neural processing of semantically complex story segments, and if speech-associated gestures (co-speech gestures) might modulate this effect. In a functional MRI study, we presented to 34 participants (16 patients and 18 matched-controls) an ecologically-valid retelling of a continuous story, performed via speech and spontaneous gestures. We split the entire story into ten-word segments, and measured the semantic complexity for each segment with idea density, a linguistic measure that is commonly used clinically to evaluate aberrant language dysfunction at semantic level. Per segment, the presence of numbers of gestures varied ($n = 0, 1, +2$). Our results suggest that, in comparison to controls, patients showed reduced activation for more complex segments in the bilateral middle frontal and inferior parietal regions. Importantly, this neural aberrance was reduced in segments presented with gestures. Thus, for the first time with a naturalistic multimodal stimulation paradigm, we show that gestures reduced group differences when processing a natural story, probably by facilitating the processing of semantically complex segments of the story in schizophrenia.

Key words

idea density, gesture, naturalistic stimulation, schizophrenia, semantic processing, multimodal

1. Introduction

Schizophrenia is characterized by impairments in social interaction and communication, which are often conducted in a multimodal form including auditory speech and hand gestures — two of the most prominent communicative channels in daily situations (1–6). In schizophrenia, on the one hand, the production and perception of auditory speech at multiple levels of the linguistic hierarchy are impaired (7–12). On the other, it has been suggested that patients with schizophrenia suffer from a range of dysfunctions in the preparation, execution, perception, and understanding of gestures (13–19). As a result, impairments in both speech and gesture processing may directly contribute to deficits during multimodal daily communication. To date, extant neuroscientific studies of schizophrenia have primarily investigated speech and gesture impairments in isolated and strictly-controlled experimental paradigms. It remains unclear how both communication channels are processed (e.g., the observation of gesture, the comprehension of semantics) interactively by patients in an ecologically-valid and naturalistic multimodal context.

1.1. *Schizophrenia: dysfunctional gesture processing in a naturalistic setting?*

Gestures are spontaneous hand movements that deliver either repeating or complementary semantic information in addition to speech. Intact functioning of gesture has a positive impact on daily communication (20,21), but is often impaired in schizophrenia (16). In production, patients tend to produce fewer gestures (22), and they have difficulties imitating and producing pantomimes (13–15,23,24). Meanwhile, dysfunctional gesture processing has been reported. Patients showed impaired functional brain connectivity when integrating speech and gesture (25). At higher-semantic levels, their understanding of gestures is reported to be dependent on language context. For example, when processing gestures in an abstract sentence context (metaphoric gestures), patients showed impaired behavioral performance to match the semantic information of gesture to a corresponding sentence (26,27), even more so for those with more severe positive formal thought disorders (17). Of note, even if the processing of abstract gestures is impaired in schizophrenia (18,28), when the same information is presented bimodally (speech accompanied by gesture), fMRI results suggest that patients showed comparable brain activations in comparison to controls (28). These prior studies, together, may suggest that patients' processing of gesture and its semantics may benefit from a multimodal context, as they would mostly encounter in daily naturalistic situations.

1.2. *Idea density as a window into language deficits in schizophrenia*

A naturalistic approach, which employs naturalistic and ecologically-valid experimental stimulation, has contributed significantly to a sharpened understanding of the neurobiology of language (29–32). However, this novel approach hasn't been sufficiently exploited in clinical neuroscience research (8). Notably, clinicians have commonly used a measure called idea density (ID) to shed light on language dysfunctions in clinical populations. Idea density is a metric of semantic complexity that measures how many propositions are expressed within a specific number of words. Deriving from narratives, its ecological validity enables researchers to investigate cognitive demands and language complexity without interfering directly with the acquisition of experimental data (33). Idea density is proven informative in investigating the relationship between early linguistic abilities and in predicting the development of Alzheimer's disease (34,35), mild cognitive impairment (36), speech impairments of patients with aphasia (37,38), and aging effects of language (39–42). More relevant to the current research, in schizophrenia, idea density has been used to understand patients' higher-order semantic performance (33,42–44). Despite substantial insights, studies on schizophrenia to date have primarily focused on how patients *produce* narratives in either written or oral form; how patients *process* different parts of naturalistic stimuli differing in the degree of semantic complexity has not been studied. Besides, no studies have investigated the neural correlates of the processing of idea density in clinical populations, e.g., in schizophrenia. Moreover, how patients with schizophrenia, when processing a natural story, might be impaired or benefit from a *multimodal* context (i.e., processing both speech and gesture) remains unclear. In a recent study with a healthy college student sample, we examined the processing of a natural, multimodal story with fMRI, and observed a facilitative effect of co-speech gesture when the story is semantically more complex, i.e., when idea density is higher (45). Thus, it would be worthwhile to examine if patients with schizophrenia may benefit from a multimodal context in a natural context, as have been previously shown in more controlled experiments (28)

1.3. Present study

Thus, in the current study, we employed a naturalistic approach in an fMRI study to investigate neural correlates of dysfunctional speech and gesture processing in schizophrenia. We focused on three specific research questions for the processing of this multimodal story: 1) if the observation of gestures is impaired in schizophrenia, 2) if patients are impaired in the processing of semantic complexity of a natural story (as parameterized by idea density), and 3) if

a multimodal context (with vs. without gesture) modulates semantic deficits (if there are any) in schizophrenia.

2. Methods

2.1. Participants

We summarized participants' demographic and clinical characteristics in Table 1. Healthy controls were recruited, matching age, gender, and education to patients. Sixteen patients were recruited at the Department of Psychiatry and Psychotherapy at the Philipps University of Marburg, and were diagnosed according to ICD-10 with schizophrenia (F20.0, n=12, and F20.3, n=1) or schizoaffective disorder (F25.0, n=2, and F25.2, n=1). Participants in both groups are native speakers of German. All except one patient received antipsychotic treatment; six were additionally treated with antidepressant medication. Positive and negative symptoms were assessed with the Scale for the Assessment of Positive Symptoms (SAPS) and the Scale for the Assessment of Negative Symptoms (SANS). Eighteen age-, gender-, and education-matched healthy participants with no history of any mental disorders were recruited from the same area. Exclusion criteria for both groups were brain injury and neurological or other medical diseases affected by brain physiology. In both groups, we conducted neuropsychological tests to assess working memory function, digit span, trail making (TMT), verbal IQ (MWT-B) (46), and gesture production and perception (BAG, Brief Assessment of Gesture (47). These measures are reported in Table 1. All participants had normal or corrected-to-normal vision and hearing. Except for one control and one patient, all other participants are right-handed (Oldfield, 1971). All participants gave written informed consent prior to participation in the experiment and were compensated monetarily. The study was approved by the ethics committee of the School of Medicine, Philipps University Marburg.

Table 1. Demographic, medication, symptom, and neuropsychological measures.

	Patients (n=16)	Controls (n=18)
Age (years)	34 (12.18)	31.94 (10.21)
Gender male/female	13/3	13/5
Education (years)	11.81 (1.83)	12.72 (1.36)
TMT A (seconds)	31.47 (11.08)	26.17 (9.89)
TMT B (seconds)	69.30 (39.01)	52.93 (19.58)

Digit Span forward	7.75 (1.61)	8.05 (2.43)
Digit Span backward	6.00 (1.31)	6.61 (2.50)
Verbal IQ	28.87 (5.42)	28.5 (3.79)
BAG subscores	3.20 (0.53)	3.37 (0.46)
SAPS (global)	15 (6.89)	
SANS (global)	9 (6.02)	
CPZ Equivalent	562.52 (372.63)	

Values are presented as mean (SD). Pairwise t-tests for all reported measures were non-significant (all p-values > 0.12). TMT: trail making test; CPZ: chlorpromazine.

2.2. Stimuli

All participants were presented with 16 video clips depicting an actor (Figure 1) narrating consecutive parts of a slightly modified version of the short story “Der Kuli Kimgun” (48). The story was unfamiliar to all participants. The trained actor narrated the story naturally and performed spontaneous gestures of any kind (iconic, metaphoric, beat, and emblematic) using hands and arms. The actor decided freely the moment and the way to make the gestures, which were all congruent with the semantic content of the story. The presentation of the videos lasted 32:12 minutes, with individual clips lasting between 1:02 and 3:31 minutes (for detailed information, see (45,49)).

Figure 1: Example of stimuli. Seven still frames of the video illustrate how the actor performed the gestures (45). A short segment of the story was translated into English and depicted in the speech bubble for illustrative purposes. The actor moved his fingers in the air while moving down his arm, indicating how the liquid (palm juice) ran down into the pots. Note that in this preprint the face area of the actor is masked. In the experiment, the face of the actor remains visible for all participants.

2.3. *Experimental design and procedure*

The entire story was divided into 330 segments of 10 words each, in order to identify parts of the story with low or high semantic complexity. As there were jitter periods of 6-14 seconds between videos, we thus excluded text segments that included the end of one video and the beginning of the next one from the analysis. Then, the semantic complexity of each remaining 10-word-segment was measured by the ID (for more details about the calculation, see (45)). Every segment of the story had an ID value between two and nine (mean: 5.35 SD: 1.12). The segments with an ID value of five or below were counted as low complexity segments and segments with a value of six or higher were considered as high complexity ones. In addition, each segment included between zero and four gestures (mean: 1.50, SD: 0.82). The number of gestures appearing in each segment did not differ significantly between the low- and high-ID conditions (see (45)). Gestures that occurred between two segments were counted twice: once in the segment where they started, and once in the segment where they ended. Based on the ID value (low vs. high) and the number of gestures per segment, six different conditions were defined: zero gestures (low_noG, high_noG), one gesture (low_1G, high_1G), and two or more gestures (low_2+G, high_2+G). No significant interaction for the number of trials and segment duration was found between the factors *gesture* and *ID*. However, we had fewer trials without gestures and longer duration for the high complexity conditions. This was regarded as unproblematic as our analyses focused on the interaction between both factors (for the complete description of the stimulus material, see (45)).

During the acquisition of fMRI data, participants were asked to attend to the videos (presented through a mirror mounted on the head coil) and to watch and listen to the actor carefully. Air conduction headphones (Siemens) were used for the presentation of verbal information. The loudness of the headphones was kept constant across sessions. In order to make sure that the subjects listened to the narrative, they were asked to answer a questionnaire containing details of the story (e.g., critical moments, narrative shifts, etc.).

2.4. *fMRI data acquisition*

All images were acquired using a 3-Tesla Siemens scanner (Siemens MRT Trio series). The functional images were obtained using a T2*-weighted echo-planar image sequence (TR = 2 s, TE = 30 ms, flip angle = 90°, slice thickness = 4 mm, interslice gap = 0.36 mm, field of view =

230 mm, matrix = 64 × 64, voxel size = 3.6 x 3.6 x 4.0 mm, 30 axial slices orientated parallel to the AC-PC line). 970 functional images were acquired (450 during the first run and 520 during the second run). The first acquisition run lasted 15 minutes and the second one lasted 17 minutes and 20 seconds, due to the varying lengths of individual videos. Simultaneously to fMRI data, additional EEG data were acquired. These data were intended to be used for another research project and will not be further discussed here (50,51). The use of an EEG cap was not expected to affect the quality of the BOLD responses (52).

2.5. Data analysis

The MRI images were analyzed using Statistical Parametric Mapping (SPM12; www.fil.ion.ucl.ac.uk) implemented in MATLAB 2009b (Mathworks Inc. Shevorn, MA). To minimize T1-saturation effects, the first two volumes were discarded from the analysis. Afterward, all images were registered to the first image of the first run and co-registered to the anatomical volume, normalized into MNI space (voxel size = 2 x 2 x 2 mm), and smoothed with an 8 mm isotropic Gaussian filter. A high-pass filter (cut-off period 128 sec) was used. Statistical analysis was performed in a two-level procedure. The design matrix for the modulation of single-subject BOLD responses at the first level comprised the onsets and durations of all six conditions, as well as the six movement parameters of each subject. The hemodynamic response function (HRF) was modeled by the canonical HRF. A flexible factorial second-level analysis with six conditions (low_noG, low_1G, low_2+G, high_noG, high_1G, high_2+G) was performed.

A Monte Carlo simulation of the brain volume of the current study was employed to determine an adequate voxel contiguity threshold (53). It is suggested that this correction provides sensitivity to smaller effect sizes and also corrects for multiple comparisons across the whole brain volume (54). Assuming an individual voxel type 1 error of $p < .001$, a cluster extent of 87 contiguous resampled voxels was indicated as necessary to correct for multiple comparisons at $p < .05$. Thus, clusters with at least 87 voxels and a significance level of $p < .001$ are reported for all contrasts. All described coordinates of activation are located in MNI space. The AAL toolbox (55) was employed for the anatomical localization of the clusters.

2.6. Contrasts of interests

We firstly tested for general group differences in the neural processing of the narrative and group differences in the processing of gestures (group x gesture). We also tested, across

groups, if gesture facilitates the processing of story differing in complexity, as have been observed in (45,49). Then, we focused on the between-group difference in the processing of story segments differing in ID (group x complexity). Lastly, we tested the interaction of group x complexity x gestures to test if group-specific processing of semantic complexity may be modulated by the presence of co-speech gestures.

3. Results

We found no significant main effect of group and no group differences regarding the processing of gestures. Post-hoc tests revealed an activation increase in both groups during the processing of co-speech gestures (*Speech + Gesture [gesture:1&2+] > Speech [gesture:0]*) in the bilateral temporo-occipital regions (Figure 2B, 2D and Table 1s). The activation in this region increased according to the number of gestures presented, similarly across both groups, as suggested by the results from the conjunction analysis between controls and patients (Figure 1s, Table 3s, supplement).

The reverse contrast (*Speech [gesture:0] > Speech + Gesture [gesture:1&2+]*) indicates that gestures resulted in group-independent reduced activation of the left middle temporal gyrus (MTG), the left inferior frontal gyrus (IFG), and the right caudate (see Figure 2A, 2C and Table 1s in the supplement). Again, conjunction analysis suggests that left MTG activation is comparable across groups (Figure 1s, Table 3s, supplement).

Figure 2: Gesture-related activation decrease and activation increase in both groups. (A) Contrast of [*Speech > Speech + Gesture*] as shown in the glass brain (left panel) and boxplots (right panel) showing eigenvariates extracted from the cluster that peaks at the left middle temporal gyrus (left MTG, [-48, -16, -10; k = 444]). (B) Contrast of [*Speech + Gesture > Speech*]

+ Gesture > Speech] as shown in the glass brain (left panel) and boxplots (right panel) showing eigenvariates extracted from the cluster that peaks at the left occipital lobe [-46, -72, -6; $k = 4092$]. For all glass brain figures, the threshold for voxel activations was set at $p < 0.001$ uncorrected, and only clusters larger than 87 voxels have been included (Monte-Carlo cluster-extent corrected at $p < .05$). Color bar indicates the scale of the T-statistics. Speech: Gesture = 0; Speech + Gesture: Gesture = 1 & 2+.

We report group-independent interaction analysis of *gesture x complexity* in the supplement, Table 2s, and Figure 2s. The findings replicated results from our previous studies (45,49), suggesting that gesture reduced the activation in the left temporal lobe only when semantic complexity is high (high-ID).

The contrast *group x complexity* revealed that, in comparison to controls, patients showed reduced activation in the left inferior parietal lobe (left IPL) when ID is high; however, in the low ID passages, the left IPL activation in patients was higher than that of controls (Figure 3 and Table 1s).

Figure 3: (A) Interaction *group x complexity* shown in the glass brain figure. (B) Boxplots showing eigenvariates extracted from the cluster that peaks at the left inferior parietal lobe (left IPL, [-52, -54, 44; $k = 149$]). The threshold for the glass brain figure was set at $p < 0.001$ uncorrected, and only clusters larger than 87 voxels have been included (Monte-Carlo cluster-extent corrected at $p < .05$). Color bar indicates the scale of the F-statistics.

Lastly, the interaction *group x complexity x gesture* (Figure 4, Table 1s) activated the bilateral IPL and the middle frontal gyrus (MFG), and additionally the right IFG. Here, in these regions, the activation increase for higher ID passages (high-ID > low-ID) was observed only when no gestures were presented (Gestures = 0). When the story was presented in a multimodal form (Gestures = 1 & 2+), this high-ID > low-ID difference was normalized for patients.

Figure 4: (A) Interaction *group x complexity x gestures* shown in the glass brain figure. (B) Boxplots showing the difference of eigenvariates between high-ID and low-ID conditions, extracted from the clusters that peak at the left inferior parietal lobe (left IPL, [-50, -56, 46; k = 367]), and the right middle frontal gyrus (right MTG [-34, 12, 58; k = 181]). The threshold for the glass brain figure was set at $p < 0.001$ uncorrected, and only clusters larger than 87 voxels have been included (Monte-Carlo cluster-extent corrected at $p < .05$). Color bar indicates the scale of the F-statistics.

4. Discussion

In the current study, we investigated if patients with schizophrenia are impaired in the neural processing of semantic complexity within a naturalistic multimodal context. Our results suggest that patients, in comparison to controls, showed increased activation of the left IPL during comprehending semantically less complex passages; however, the left IPL activation was reduced for patients when semantic complexity is high. More importantly, the presentation of gestures reduced this group difference, suggesting a facilitative or compensatory role of gestures in this context.

4.1. The processing of gesture is comparable between groups

For both groups, the *Speech + Gesture > Speech* comparison activated the bilateral posterior temporo-occipito-parietal visual regions, thus suggesting that patients are intact in recruiting this region to process gesture when processing naturalistic stimuli, in a similar way to isolated bimodal videos (27,45,51,56). These findings may imply that, despite well-documented gesture dysfunctions in schizophrenia (13,57), when engaged in a high-level context (i.e., presented together with speech in both isolated and naturalistic multimodal settings), the perception of gestures in schizophrenia could appear to be largely intact.

Results from the *Speech > Speech + Gesture* comparison are in accordance with our prior studies (45,49), and suggest that segments without gestures may recruit more neural resources in the left hemispheric speech and language regions (e.g., the left IFG and MTG. Importantly, this facilitative effect of gesture was present for both patients and controls, thus suggesting that gesture may facilitate story comprehension in general in schizophrenia.

4.2. *Aberrant neural processing of semantic complexity in schizophrenia is modulated by gesture*

The focus of this study was on potential impairments when patients with schizophrenia process the naturalistic multimodal story differing in the degree of semantic complexity. Indeed, when collapsing across all gesture conditions (gesture: 0, 1, 2+), patients, as opposed to controls, showed increased activation for the low complexity and decreased activation for high complexity passages in the left IPL. Moreover, when the factor of gesture is considered, we observed that this effect interacts with gesture in the bilateral IPL and additionally the middle frontal regions. Specifically, the neural aberrance in these regions (including the left IPL) was only present when no gesture was presented. In contrast, when the story was presented in a multimodal form (gesture: 1 and 2+), group differences were largely not observable.

The left IPL is an important region that underlies a range of cognitive functions that are functionally relevant to human social communication (58,59). Importantly, the left IPL is crucially involved during language comprehension at the semantic level (60,61), and has been reported to support semantic prediction during comprehension of a naturalistic auditory story (31). Besides, the left IPL is associated with working memory and is sensitive to the cognitive load of the task (62). In schizophrenia, the left IPL activity is reported to be reduced for patients when they are engaged in sentence-level language comprehension tasks (63). In the current study, although patients showed even increased activation in the left IPL in the low-complexity passages, when

they process parts of the story that are semantically more complex, they failed to activate the left IPL as controls did. This neural aberrance, during the processing of a naturalistic narrative, might derive from patients' impaired working memory capacity and semantic processing (7,64–66).

The left middle frontal gyrus (MFG) is associated with the maintenance of verbal information (67). In this regard, participants might recruit this region especially during the complex speech conditions because multiple semantic events have to be kept in mind in order to interpret the entire story. Nevertheless, patients recruit this region already even to a higher degree than controls, during the low complexity speech condition. However, in the high complexity condition, the left MFG activity in patients was reduced. Notably, a study investigating the processing of novel and conventional metaphors by patients with schizophrenia also reported increased activation of the left MFG (68), importantly, there, the activation decreased as the cognitive demands of the task increased (in this case abstract thinking). Similarly, in our study, patients' activation in this region decreased as the demands of the narrative increased, i.e., in high complex passages. On the other hand, the right MFG plays an important role in the temporal sequencing of discourse (69), an ability that is fundamental for interpreting the story.

Together, for the bilateral IPL and MFG, the increased activation by patients during the low complexity condition indicate that patients already need more neural effort (increased BOLD response) for the processing of low complexity segments; when the story is semantically too complicated (high-ID), patients then failed to provide the necessary resources in these regions (reduced BOLD response), partially due to their impaired working memory maintenance or processing of complex semantics.

However importantly, this neural aberrance was only observable when there was no gesture presented together with speech. In contrast, in these regions, whenever the story was presented in a multimodal form (gesture: 1 and 2+), the activation pattern is comparable between groups. A similar facilitative effect of gesture when processing a story has been observed for elder vs. younger participants, i.e., group difference was reduced when gestures were presented together with story segments (49). In line with these findings, for isolated sentences, despite the fact that patients showed neural aberrance when processing abstract-social gestures and concrete-non-social speech, group difference was significantly reduced when these events were presented in a multimodal form (28). In the current study, we extend prior findings by showing a similar facilitative nature of gesture during speech comprehension: In a naturalistic story setting, even if patients neural aberrance may be more likely to derive from their impaired semantic

prediction and working memory capacity, which are not as heavily demanded in isolated situations, nevertheless, a multimodal context is, again, proven beneficial (6,21).

4.3 Limitation

Despite new insights, the current study has clear limitations. Limited by the sample size of this study (n = 16 & 18 for each group), we are unable to test the role of medication for story comprehension, and cannot pinpoint if the observed neural aberrance (as well as neural facilitation) is specific to chronic or first-episode schizophrenia. Therefore, future studies are expected to investigate language processing and gesture–speech interaction in schizophrenia with naturalistic stimulation with larger samples.

5. Conclusion and future research

The following conclusions can be drawn from this naturalistic experiment: First, patients and controls recruit similar regions for the processing of gestures and speech in general. Second, patients, when compared to controls, show reduced activation in the bilateral middle frontal and parietal regions when processing parts of the story that are semantically more complex. Finally, gestures reduced group differences, potentially by facilitating the processing of the narrative. Thus, future studies could examine if the differences in the processing of semantic complexity are reduced by conducting existing training developed to improve verbal communication and social cognition (70,71)

6. Acknowledgments

This work was supported by the International Research Training Group “The Brain in Action” (IRTG 1901) by the German Research Foundation (DFG), the von Behring-Röntgen-Stiftung (vBR59-0002, 64-0001), the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project ‘The Adaptive Mind’), and the individual projects of the DFG (HE8029/2-1). The study was also supported by the Core Facility Brain Imaging, Faculty of Medicine, University of Marburg, Rudolf-Bultmann-Str. 9, 35039, Marburg, Germany.

7. References

1. Kendon A (1980): Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication* 25: 207–227.
2. Goldin-Meadow S (2005): *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press.
3. McNeill D (2008): *Gesture and Thought*. University of Chicago Press.
4. He Y, Luell S, Muralikrishnan R, Straube B, Nagels A (2020): Gesture's body orientation modulates the N400 for visual sentences primed by gestures. *Hum Brain Mapp*. <https://doi.org/10.1002/hbm.25166>
5. Willems RM, Özyürek A, Hagoort P (2007): When language meets action: the neural integration of gesture and speech. *Cereb Cortex* 17: 2322–2333.
6. Holler J, Levinson SC (2019): Multimodal Language Processing in Human Communication. *Trends Cogn Sci*. <https://doi.org/10.1016/j.tics.2019.05.006>
7. Brown M, Kuperberg GR (2015): A Hierarchical Generative Framework of Language Processing: Linking Language Perception, Interpretation, and Production Abnormalities in Schizophrenia. *Front Hum Neurosci* 9: 611.
8. Meyer L, Lakatos P, He Y (2021): Language Dysfunction in Schizophrenia: Assessing Neural Tracking to Characterize the Underlying Disorder(s)? *Front Neurosci* 15: 640502.
9. Covington MA, He C, Brown C, Naçi L, McClain JT, Fjordbak BS, et al. (2005): Schizophrenia and the structure of language: the linguist's view. *Schizophr Res* 77: 85–98.
10. DeLisi LE (2001): Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. *Schizophr Bull* 27: 481–496.
11. Kircher T, Bröhl H, Meier F, Engelen J (2018): Formal thought disorders: from phenomenology to neurobiology. *The Lancet Psychiatry* 5: 515–526.
12. Kiang M, Gerritsen CJ (2019): The N400 event-related brain potential response: A window

- on deficits in predicting meaning in schizophrenia. *Int J Psychophysiol.*
- <https://doi.org/10.1016/j.ijpsycho.2019.04.005> PMID - 31047943
13. Thakkar KN, Peterman JS, Park S (2014): Altered Brain Activation During Action Imitation and Observation in Schizophrenia: A Translational Approach to Investigating Social Dysfunction in Schizophrenia. *Am J Psychiatry* 171: 539–548.
14. Matthews N, Gold BJ, Sekuler R, Park S (2013): Gesture imitation in schizophrenia. *Schizophr Bull* 39: 94–101.
15. Walther S, Stegmayer K, Sulzbacher J, Vanbellingen T, Müri R, Strik W, Bohlhalter S (2015): Nonverbal Social Communication and Gesture Control in Schizophrenia. *Schizophr Bull* 41: 338–345.
16. Walther S, Mittal VA (2016): Why We Should Take a Closer Look at Gestures. *Schizophr Bull* 42: 259–261.
17. Nagels A, Kircher T, Grosvald M, Steines M, Straube B (2019): Evidence for gesture-speech mismatch detection impairments in schizophrenia. *Psychiatry Res* 273: 15–21.
18. Straube B, Green A, Sass K, Kirner-Veselinovic A, Kircher T (2013): Neural integration of speech and gesture in schizophrenia: Evidence for differential processing of metaphoric gestures. *Hum Brain Mapp* 34: 1696–1712.
19. Millman ZB, Goss J, Schiffman J, Mejias J, Gupta T, Mittal VA (2014): Mismatch and lexical retrieval gestures are associated with visual information processing, verbal production, and symptomatology in youth at high risk for psychosis. *Schizophr Res* 158: 64–68.
20. McNeill D (1992): *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.
21. Kelly SD, Özyürek A, Maris E (2009): Two Sides of the Same Coin Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychol Sci.*
22. Lavelle M, Healey PGT, McCabe R (2013): Is nonverbal communication disrupted in interactions involving patients with schizophrenia? *Schizophr Bull* 39: 1150–1158.

23. Walther S, Vanbellingen T, Müri R, Strik W, Bohlhalter S (2013): Impaired pantomime in schizophrenia: Association with frontal lobe function. *Cortex* 49: 520–527.
24. Park S, Matthews N, Gibson C (2008): Imitation, simulation, and schizophrenia. *Schizophr Bull* 34: 698–707.
25. Wroblewski A, He Y, Straube B (2020): Dynamic Causal Modelling suggests impaired effective connectivity in patients with schizophrenia spectrum disorders during gesture-speech integration. *Schizophr Res* 216: 175–183.
26. Schülke R, Straube B (2018): Transcranial Direct Current Stimulation Improves Semantic Speech–Gesture Matching in Patients With Schizophrenia Spectrum Disorder. *Schizophr Bull* 45: 522–530.
27. Straube B, Green A, Sass K, Kircher T (2014): Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia. *Schizophr Bull* 40: 936–944.
28. He Y, Steines M, Sammer G, Nagels A, Kircher T, Straube B (2021): Modality-specific dysfunctional neural processing of social-abstract and non-social-concrete information in schizophrenia. *Neuroimage Clin* 29: 102568.
29. Brennan J (2016): Naturalistic Sentence Comprehension in the Brain. *Lang Linguist Compass* 10: 299–313.
30. Armeni K, Willems RM, Frank SL (2017): Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neurosci Biobehav Rev* 83: 579–588.
31. Willems RM, Frank SL, Nijhof AD, Hagoort P, van den Bosch A (2016): Prediction During Natural Language Comprehension. *Cereb Cortex* 26: 2506–2516.
32. Brennan JR, Hale JT (2019): Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS One* 14: e0207741.
33. Cohen AS, McGovern JE, Dinzeo TJ, Covington MA (2014): Speech deficits in serious mental illness: a cognitive resource issue? *Schizophr Res* 160: 173–179.

34. Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR (1996): Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA* 275: 528–532.
35. Engelman M, Agree EM, Meoni LA, Klag MJ (2010): Propositional density and cognitive function in later life: findings from the Precursors Study. *J Gerontol B Psychol Sci Soc Sci* 65: 706–711.
36. Roark B, Mitchell M, Hosom J-P, Hollingshead K, Kaye J (2011): Spoken Language Derived Measures for Detecting Mild Cognitive Impairment. *IEEE Trans Audio Speech Lang Processing* 19: 2081–2090.
37. Fromm D, Greenhouse J, Hou K, Russell GA, Cai X, Forbes M, *et al.* (2016): Automated Proposition Density Analysis for Discourse in Aphasia. *J Speech Lang Hear Res* 59: 1123–1132.
38. Thorne J, Farooqi-Shah Y (2016): Verb Production in Aphasia: Testing the Division of Labor between Syntax and Semantics. *Semin Speech Lang* 37: 23–33.
39. Kemper S, Greiner LH, Marquis JG, Prenovost K, Mitzner TL (2001): Language decline across the life span: findings from the Nun Study. *Psychol Aging* 16: 227–239.
40. Kemper S, Sumner A (2001): The structure of verbal abilities in young and older adults. *Psychol Aging* 16: 312–322.
41. Farias ST, Chand V, Bonnici L, Baynes K, Harvey D, Mungas D, *et al.* (2012): Idea density measured in late life predicts subsequent cognitive trajectories: implications for the measurement of cognitive reserve. *J Gerontol B Psychol Sci Soc Sci* 67: 677–686.
42. Covington MA, Riedel WJ, Brown C, He C, Morris E, Weinstein S, *et al.* (2009, January): Ketamine and schizophrenic speech: more difference than originally reported. *Journal of Psychopharmacology*, vol. 23. researchgate.net, pp 111–112.
43. Moe AM, Breitborde NJK, Shakeel MK, Gallagher CJ, Docherty NM (2016): Idea density in the life-stories of people with schizophrenia: Associations with narrative qualities and

psychiatric symptoms. *Schizophr Res* 172: 201–205.

44. Covington MA, Riedel WJ, Brown C, He C, Morris E, Weinstein S, *et al.* (2007): Does ketamine mimic aspects of schizophrenic speech? *J Psychopharmacol* 21: 338–346.

45. Cuevas P, Steines M, He Y, Nagels A, Culham J, Straube B (2019): The facilitative effect of gestures on the neural processing of semantic complexity in a continuous narrative. *Neuroimage* 195: 38–47.

46. Lehrl S (1999): Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B: Spitta.

47. Nagels A, Kircher T, Steines M, Grosvald M, Straube B (2015): A brief self-rating scale for the assessment of individual differences in gesture perception and production. *Learn Individ Differ* 39: 73–80.

48. Dauthendey M (2016): *Lingam: Zwölf asiatische Novellen*. BoD – Books on Demand.

49. Cuevas P, He Y, Billino J, Kozasa E, Straube B (2020): Age-related effects on the neural processing of semantic complexity in a continuous narrative: Modulation by gestures already present in young to middle-aged adults. *Neuropsychologia* 151: 107725.

50. He Y, Nagels A, Schlesewsky M, Straube B (2018): The Role of Gamma Oscillations During Integration of Metaphoric Gestures and Abstract Speech. *Front Psychol* 9: 1348.

51. He Y, Steines M, Sommer J, Gebhardt H, Nagels A, Sammer G, *et al.* (2018): Spatial-temporal dynamics of gesture-speech integration: a simultaneous EEG-fMRI study. *Brain Struct Funct* 223: 3073–3089.

52. Iannetti GD, Niazy RK, Wise RG, Jezzard P, Brooks JCW, Zambreanu L, *et al.* (2005): Simultaneous recording of laser-evoked brain potentials and continuous, high-field functional magnetic resonance imaging in humans. *Neuroimage* 28: 708–719.

53. Slotnick SD, Moo LR, Segal JB, Hart J Jr (2003): Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res Cogn Brain Res* 17: 75–82.

54. Slotnick SD (2017, July): Cluster success: fMRI inferences for spatial extent have

acceptable false-positive rates. *Cognitive Neuroscience*, vol. 8. Taylor & Francis, pp 150–155.

55. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, *et al.* (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15: 273–289.

56. He Y, Gebhardt H, Steines M, Sammer G, Kircher T, Nagels A, Straube B (2015): The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. *Neuropsychologia* 72: 27–42.

57. Mittal VA, Walther S (2018): As Motor System Pathophysiology Returns to the Forefront of Psychosis Research, Clinical Implications Should Hold Center Stage. *Schizophr Bull* 73: 1082.

58. Seghier ML (2013): The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist* 19: 43–61.

59. Bzdok D, Hartwigsen G, Reid A, Laird AR, Fox PT, Eickhoff SB (2016): Left inferior parietal lobe engagement in social cognition and language. *Neurosci Biobehav Rev* 68: 319–334.

60. Binder JR, Desai RH (2011): The neurobiology of semantic memory. *Trends Cogn Sci* 15: 527–536.

61. Binder JR, Desai RH, Graves WW, Conant LL (2009): Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19: 2767–2796.

62. Ravizza SM, Delgado MR, Chein JM, Becker JT, Fiez JA (2004): Functional dissociations within the inferior parietal cortex in verbal working memory. *Neuroimage* 22: 562–573.

63. Kuperberg GR, West WC, Lakshmanan BM, Goff D (2008): Functional Magnetic Resonance Imaging Reveals Neuroanatomical Dissociations During Semantic Integration in Schizophrenia. *Biol Psychiatry* 64: 407–418.

64. Lee J, Park S (2005): Working memory impairments in schizophrenia: a meta-analysis. *J*

Abnorm Psychol 114: 599–611.

65. Goldman-Rakic PS (1994): Working memory dysfunction in schizophrenia. *J*

Neuropsychiatry Clin Neurosci 6: 348–357.

66. Li X, Branch CA, DeLisi LE (2009): Language pathway abnormalities in schizophrenia: a review of fMRI and other imaging studies. *Curr Opin Psychiatry* 22: 131–139.

67. Woodward TS, Cairo TA, Ruff CC, Takane Y, Hunter MA, Ngan ETC (2006): Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. *Neuroscience* 139: 317–325.

68. Mashal N, Vishne T, Laor N, Titone D (2013): Enhanced left frontal involvement during novel metaphor comprehension in schizophrenia: evidence from functional neuroimaging. *Brain Lang* 124: 66–74.

69. Knutson KM, Wood JN, Grafman J (2004): Brain activation in processing temporal sequence: an fMRI study. *Neuroimage* 23: 1299–1307.

70. Kurtz MM, Gagen E, Rocha NBF, Machado S, Penn DL (2016): Comprehensive treatments for social cognitive deficits in schizophrenia: A critical review and effect-size analysis of controlled studies. *Clin Psychol Rev* 43: 80–89.

71. Riedl L, Nagels A, Sammer G, Straube B (2020): A multimodal speech-gesture training intervention for patients with schizophrenia and its neural underpinnings—the study protocol of a randomized controlled *Front Psychiatry*. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpsy.2020.00110/full?report=reader>