Multimodal speech-gesture training in patients with schizophrenia spectrum disorder: Effects on quality of life and neural processing☆,☆☆

Lydia Riedl a,b,*, Arne Nagels c, Gebhard Sammer d, Momoko Choudhury a, Annika Nonnenmann a, Anne Sütterlin a, Chiara Feise a, Maxi Haslach a, Florian Bitsch a, Benjamin Straube a,b

a Translational Neuroimaging Lab, Department of Psychiatry and Psychotherapy, Philipps-University, Marburg, Germany
b Center for Mind, Brain and Behavior (CMBB), Philipps-University Marburg and Justus Liebig University, Giessen, Germany
c Department of English and Linguistics, Johannes-Gutenberg-University, Mainz, Germany
d Department of Psychiatry and Psychotherapy, Justus-Liebig-University, Giessen, Germany

ARTICLE INFO

Keywords:
Intervention
fMRI
Neural correlates
Concretism
Metaphor
Social functioning

ABSTRACT

Dysfunctional social communication is one of the most stable characteristics in patients with schizophrenia spectrum disorder (SSD) that severely affects quality of life. Interpreting abstract speech and integrating nonverbal information is particularly affected.

Considering the difficulty to treat communication dysfunctions with usual intervention, we investigated the possibility to apply a multimodal speech-gesture (MSG) training.

In the MSG training, we offered 8 sessions (60 min each) including perceptive and expressive tasks as well as meta-learning elements and transfer exercises to 29 patients with SSD. In a within-group crossover design, patients were randomized to a TAU-first (treatment as usual first, then MSG training) group (N = 20) or a MSG-first (MSG training first, then TAU only) group (N = 9), and were compared to healthy controls (N = 17). Outcomes were quality of life and related changes in the neural processing of abstract speech-gesture information, which were measured pre-post training through standardized psychological questionnaires and functional Magnetic Resonance Imaging, respectively.

Pre-training, patients showed reduced quality of life as compared to controls but improved significantly during the training. Strikingly, this improvement was correlated with neural activation changes in the middle temporal gyrus for the processing of abstract multimodal content. Improvement during training, self-report measures and ratings of relatives confirmed the MSG-related changes.

Together, we provide first promising results of a novel multimodal speech-gesture training for patients with schizophrenia. We could link training induced changes in speech-gesture processing to changes in quality of life, demonstrating the relevance of intact communication skills and gesture processing for well-being.

1. Introduction

Communication represents the fundamental basis of social life and is disturbed on all linguistic levels in SSD (Covington et al., 2005; DeLisi, 2001; Heim et al., 2019; Marini et al., 2008). Two frequently reported abnormalities in schizophrenia that have a particular impact on the social-communicative skills and potentially quality of life are the understanding of figurative language and nonverbal communication (including gesture processing and production) which is impaired in schizophrenia and comprises several interacting fundamental processes including motor behavior, language, and sensory integration (Waldth et al., 2016). However, currently no effective treatment exists for these impairments.

Figurative speech is conveyed frequently through metaphors (Lakoff...
and Johnson, 2008) which patients tend to misinterpret in a concrete way (‘concretism’) (Bergemann et al., 2008; de Bonis et al., 1997; Iakimova et al., 2010; Kircher et al., 2007; Rapp and Schnierer, 2010; Rossetti et al., 2018). The interpretation of metaphors has a tremendous impact on successful social interactions (Kircher and Gauggel, 2008; Lakoff and Johnson, 2008) and represents a real-world communication problem for patients with SSD (Kircher et al., 2007; Rossetti et al., 2018). A second key element in a person’s ability of interpersonal communication is the ability to combine information from multiple sensory modalities (Stevenson et al., 2011): Gestures can disambiguate speech (Driskell and Radtké, 2003; Holle and Gunter, 2007; Kelly et al., 1999, 2010), increase attention (Mariichio et al., 2009) and play a crucial role in grounding people’s mental representations in action (Bavelas et al., 2011; Nathan and Albali, 2011; Beilock and Goldin-Meadow, 2010). In patients with SSD, aberrations in gesture processing are found in production (Goss, 2011; Matthews et al., 2013; Mittal et al., 2006; Troisi et al., 1998; Welther et al., 2013a, 2013b, 2015, 2016) as well as in perception and interpretation (Bennell et al., 1986; Bucci et al., 2008; White et al., 2016). Communication skills (Kuschke, 2019) are put forward to play a crucial role in social integration and can have a serious impact on the psychiatrist-patient communication (McCabe et al., 2013) with further negative consequences for rehabilitation. The limited participation on both personal and professional levels results in a dramatically reduced quality of life (Bambini et al., 2016; Falkai, 2016; Gaebel and Wölwer, 2010).

The conventional medical treatment of schizophrenia is often (Hegarty et al., 1994; Jääskeläinen et al., 2013) successful in treating positive symptoms, but especially symptoms concerning communication remain relatively stable (Dollfus and Pettit, 1995; Gaebel and Wölwer, 2010; Joyal et al., 2016; Lavelle et al., 2014; Wüthrich et al., 2020). Despite ample evidence of dysfunctional social communication in SSD and its association with negative outcomes for patients’ quality of life (Bambini et al., 2016; Falkai, 2016; Gaebel and Wölwer, 2010), there have been only few studies on speech therapy (Allen et al., 1978; Baker, 1971; Bosco et al., 2016); Clegg et al., 2007; Foxx et al., 1988; Kondel et al., 2006; Kramer et al., 2001; Ojeda et al., 2012; Santos et al., 2021; Viani et al., 2014; Wykes, 1998). Results from a first study with focus specifically on concretism (Bambini et al., 2022) give reason to assume beneficial effects of speech language interventions also for quality of life in patients with schizophrenia. However, to our knowledge, speech language interventions have not yet been combined with a specific nonverbal gesture training, addressing also the nonverbal difficulties of SSD patients. We therefore developed a specific multimodal speech-gesture (MSG) training which considers verbal communication problems with focus on concretism and problems in nonverbal communication.

The key element of the training is the production and perception of meaningful arm and hand gestures aligned with corresponding concrete or abstract speech. The integration of concrete gestures (such as forming the shape of a dog’s mouth with a hand while discussing a dog) has been related to processes in the posterior temporal lobe, which might be mainly intact in patients with SSD (Straube et al., 2013a, 2013b). The inferior frontal gyrus (IFG) on the other hand is more involved in complex higher-order integration processes, e.g., when lexical elements are combined and integrated into larger structures (unification) (Dick et al., 2009; He et al., 2015; Straube et al., 2011). This region seems to be relevant for the integration of abstract speech-gesture combinations involving metaphors (such as forming a cup with a hand while discussing an abstract concept such as love) (Choudhury et al., 2021; Kircher et al., 2009; Steines et al., 2021; Straube et al., 2009, 2011, 2013a, 2013b, 2014). Compared to healthy controls, subjects with high risk for schizophrenia (Gupta et al., 2021) and patients with schizophrenia have difficulties with interpreting abstract meaning in gestures that involve metaphors (such as forming a cup with a hand while discussing an abstract concept such as love) (Nagels et al., 2019; Straube et al., 2013a, 2014) and show altered functional activation and connectivity, mostly in fronto-temporal regions, during the integration of metaphoric gestures (Straube et al., 2013a, 2014). However, these deviant neural activation patterns seem to be modifiable. Studies have shown cognitive and functional activation improvements appearing to be positively correlated after cognitive remediation therapy (Penades et al., 2017). Other studies showed reduced group differences of sentences or continuous narratives when accompanied with gestures (comparing SSD patients and a control group) in neural processing (Cuevas et al., 2021; He et al., 2021). Thus, a multimodal training program including gesture training tasks might help to develop, reactivate or promote communication resources in patients with schizophrenia. Gestures that visualize concepts could help patients with SSD to understand the meaning of abstract sentences and integrate it into the sentence context.

The main aim of the current study was to test the feasibility and efficacy of a newly developed MSG training with patients suffering from SSD. We focused on the feasibility regarding drop-out rates and on efficiency regarding changing dysfunctional neural processing and inducing positive effects on quality of life. To get a more comprehensive understanding of the potential training effects, we also explored suitable behavioral outcomes from the training tasks and subjective impressions of the patients and their relatives about nonverbal communication and social life. We expected patients to show a reduced quality of life prior to the training. Furthermore, based on former studies on psychotherapy and gesture related changes in neural activation, we expected the potential improvement in quality of life after the MSG training to correlate with specific training-related increase in neural activation of fronto-temporal brain regions.

2. Methods

2.1. Experimental design

This was a single-center randomized waiting list controlled pilot trial of intensive single MSG training versus wait-list control (TAU) being conducted at Philipps-University Marburg, Department of Psychiatry and Psychotherapy. Main outcomes were measured through pre-post-fMRI and standardized psychological questionnaires (German version of SWLS – Satisfaction with Life Scale (Glaesmer et al., 2011)). The SWLS is a short 5-item instrument with satisfying validity and reliability with cronbach alpha = 0.87 (Pavot and Diener, 2009). Secondary outcomes were the improvement in gesture production during training as well as self-report measures and ratings of relatives after the last training session, regarding community functioning.

Patients were measured in a within-subjects crossover design three times and randomly assigned (computerized random numbers) to a TAU-first group (N = 20). In order to obtain sufficient data on MSG compared to TAU effects) and an MSG-first group (N = 10, in order to explore possible long-term effects), see Fig. 1. The TAU-first group was first measured (ses-pre) before a TAU period and a baseline measurement (ses-bl), then the MSG training and the post measurement (ses-post) followed. The MSG-first group conducted the MSG training right after the very first measurement (ses-pre), then the second (ses-bl) measurement, the TAU period and the post measurement (ses-post) followed. During TAU, the patients did not receive any speech-gesture training. The duration of MSG training and TAU period were the same within a patient, in average three weeks. This approach allowed us to compare MSG training versus TAU with perfectly matching participants and it gave all patients the possibility of benefiting from the training. The comparison of three measurements provides further evidence for possible training-specific effects through comparing intra-individual repetition and training effects.

In order to better interpret the differences in performance and neural activation between sessions, we also measured a control group without a diagnosis of mental disorders. The control group received no training (modification of the study design described in our study protocol (Riedl et al., 2020)).
2.2. Participants

A total of 29 German patients aged between 18 and 62 and with the capability to give informed consent and to participate in a fMRI study were included in the final analysis. The patients were diagnosed by a psychiatric specialist according to the International Classification of Diseases (World Health Organization, 1992) with schizophrenia spectrum disorder (F20.0, n = 13; F20.1, n = 2; F20.3, n = 1; F20.6, n = 1; F25, n = 9; F23.0, n = 1; F23.1, n = 1; F23.2, n = 2; F1x.50, n = 3; F60.1, n = 1).\(^1\) Twenty-two of the patients were treated with atypical antipsychotic medication; six patients received (additional) antidepressants or other psychiatric medical treatment. Five patients were not medicated. Fourteen patients reported a history of illegal substance abuse. In order to exclude large non-training related variations, only outpatients without drug intoxication were recruited for the SSD group.

In the healthy control group, the final sample comprised n = 17. All participants reported German to be their primary language. In both groups, two subjects each were left-handed. The groups were matched for sex, age and education (see Table 1).

The study has been registered at the German Clinical Trials Register and the study plan including the MSG training procedure was described in detail in our study protocol (Riedl et al., 2020).

2.3. The MSG training program

The examiners in our study were trained in detail before they conducted parts of the experiment with the patients. The examiners were unaware of our specific neural and behavioral hypotheses regarding the MSG training effects on the different measures.

We offered eight sessions (60 min each) considering the dose-effect relationship (Howard et al., 1986) of individual standardized training in a 1:1 setting with a high frequency (3–5 training sessions per week).

---

\(^1\) The divergent total number of diagnoses from the number of participating patients results from multiple diagnoses in individual patients.

---

Table 1

|                     | Patients with SSD | Controls | Difference (p) |
|---------------------|-------------------|----------|----------------|
| Age (years)         | 35.48 ±11.07      | 36.53 ±11.77 | 0.774          |
| Sex male/female     | 22/7              | 13/4      | 0.964          |
| Education           | 3/11/15           | 2/7/8     |                |
| SAPS (total)        | 24.54 ±19.26      | –         | –              |
| SANS (total)        | 22.52 ±14.8       | –         | –              |
| TULIA-AST           | 11.36 ±6.93       | –         | –              |
| CPZ equivalent      | 275.03 ±330.27    | –         | –              |
| Duration of illness | 12.24 ±9.05       | –         | –              |
| Age of onset        | 23.29 ±8.36       | –         | –              |
| Smokers (cigarettes/day) | 18 ±23.49       | 1 ±2.82   | (0.71)         |

Values are presented as mean ± standard deviation.

CPZ Equivalent: chlorpromazine equivalent of antipsychotic medication (with 5 patients not being medicated); SAPS (Andreasen, 1984)/SANS (Andreasen, 1981): scale for the Assessment of Positive/Negative Symptoms; SSD: schizophrenia spectrum disorder; TULIA-AST: Apraxia Screen of TULIA (AST) (Vanbellingen et al., 2011).

\(^a\) Education: Number of patients receiving 1: Certificate of secondary education/2: General certificate of secondary education/3: General qualification for university entrance.
for efficiency reasons (Joyal et al., 2016). After a short personal introduction, patients executed four exercises with increasing complexity in every session. Considering speech production being secondary to speech perception (Geschwind, 1970) the training sessions started with two perceptual tasks. One of these tasks was a relatedness task (Schülke and Straube, 2019). The other task was an audiovisual working memory task using speech-gesture videos, because the integration of speech and gesture into communicative context is strongly associated with working memory capacities (Kintsch and Van Dijk, 1978; Rudner, 2018). The perceptive tasks were performed on a computer device and supported by written during the training and productive tasks were video recorded. The perceptual tasks were performed on a computer device and supported by the examiner. The training also included a productive (imitation/mime) and a free productive (gesture fluency (Wende et al., 2012)) task. The latter gesture fluency task is especially suitable to explore training related improvements in gesture production. In addition, patients were provided with background information about gesture and how it is related to language and communication as a meta-learning element. Handouts summarizing the content of this information and an explanation of the new homework were given to patients at the end of each session to encourage transfer of the training effects to daily life routine.

For the purpose of later independent evaluation, protocols were written during the training and productive tasks were video recorded.

2.4. FMRI data acquisition

Imaging data were collected with a 3 T whole body MRI system (SIEMENS MAGNETOM TrioTim syngo MR B17) equipped with a standard head coil. Structural image acquisition consisted of 176 T1-weighted sagittal slices (slice thickness = 1.0 mm; FoV = 256 mm; TR = 1900 ms; TE = 2.26 s). To measure BOLD changes in brain activity during acquisition, T2* weighted gradient echo planar imaging (EPI) with 34 transversal slices covering the whole brain were used (voxel size = 3 × 3 × 4 mm; descending slice acquisition; TR = 1650 ms; TE = 25 ms; flip angle = 70°; FoV = 192 mm; GRAPPA = 2). The slices were aligned to the anterior commissure posterior commissure (AC–PC) line. In the three measurements (se-pre, ses-bl and ses-post), 936 functional images were acquired during the acquisition phase. A gradient echo field map sequence was measured prior to the functional runs to get information for unwarping B0 distortions.

2.5. Stimuli

The videos clips presented during fMRI data acquisition displayed a male actor speaking abstract (abs) or concrete (con) sentences accompanied by gestures (SG) (Fig. 1). Additionally, abstract and concrete unimodal conditions (S: speech only, G: gesture only) were presented. This allowed us to investigate implicit speech and gesture processing in concrete and abstract conditions as well as in multimodal integration (Straube et al., 2011, 2013a). More details about the stimuli are given in Appendix A.

In order to confirm that the subjects had watched the video and paid attention during all conditions, they were instructed once from the examiner and again in written form right before the task started to tap with their left forefinger on the buttons of a response box that was fixated at their left leg.

2.6. FMRI data analysis

Preprocessing was performed using fMRIPrep 20.2.0rc0 (Esteban et al., 2019, 2020a) (RRID:SCR_016216), which is based on Nipype 1.5.1 (Esteban et al., 2020b; Gorgolewski et al., 2011) (RRID:SCR_002502). For resampling we utilized a coarse gridding of 2 mm isotropic of the target template. Additionally, functional images were smoothed using a Gaussian kernel of 6 mm full width at the half maximum in all directions to adjust for inter-subject brain anatomical variations (scripts publicly available: https://github.com/LydiaRiedl/fmrilrep2spm).

FMRI data were analyzed using standard routines for first and second level analyses in SPM (Statistical Parametric Mapping, Wellcome Trust Center for Neuroimaging, London; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/, RRID: SCR_007037) implemented in MATLAB R2017a (version 9.2.0 The MathWorks, Inc.). On the first level, single subjects’ voxel-wise BOLD activity was modeled by a General Linear Model (GLM) (Friston et al., 1995; Worsley and Friston, 1995). The six realignment parameters extracted by fMRIPrep (head motion) were modeled as regressors of no interest. The onset of each event was defined as the integration point (the time when the stroke of the gesture coincides with the keyword of the sentence) (Green et al., 2009). All events were modeled with a duration of 1 s and assigned to one of the conditions.

The single subject level parameter estimates were deployed for a full factorial analysis. The main effect of condition in all three experimental blocks were defined as contrasts of interest for the second level analysis, resulting in baseline contrasts for the two conditions of interest and the four control conditions. In the statistical model, group (patients, controls) was considered as a between-subject factor, session (ses-pre, ses-post), content (abs, con) and modality (SG, S, G) as within-subjects factors, resulting in a 2 × 2 × 2 × 3 design. A Monte-Carlo-Simulation was performed (acquisition matrix: x = 64, y = 64; slices: 34; DIM: xy = 3 mm, z = 4 mm; FWHM = 10.3 mm; DIM resampled = 2 mm; no mask; iterations: 1000) to calculate the minimum voxel contiguity threshold needed to correct for multiple comparisons at p < 0.05, assuming an individual voxel type I error of p < 0.01 (Slotnick, 2017; Slotnick et al., 2003). Based on this calculation, a cluster extent threshold of 221 contiguous resampled voxels at p < 0.05 (whole-brain analysis) was used for all contrasts of interest. Voxel coordinates reported are referenced to the Montreal Neurological Institute brain (MNI) space. For anatomical location, functional data were referenced to the Automated Anatomical Labeling toolbox (AAL) in SPM12 (Rolls et al., 2015; Tzourio-Mazoyer et al., 2002).

2.7. Behavioral data analysis and correlations

For further statistical analyses of neural and behavioral data, R Studio (RStudio, 2021) with ggstatsplot package for R (Patil, 2021) was utilized.

2.8. Main outcome measures

To clarify training-induced changes of multimodal input processing in patients with SSD, we calculated an interaction of neural activation in session X abstractness, contrast (1): (ses-pre(absSG > conSG)) < ses-post(absSG > conSG).

To investigate specific training effects, comparing MSG training and TAU period, we furthermore calculated Three-sessions-interaction X abstractness (F-test), comparing ses-pre, ses-bl and ses-post in the patients’ TAU-first group, contrast (2).

To examine potentially benefits of the MSG training on daily life, subjective quality of life was investigated through a standardized psychological questionnaire (SWLS) at ses-pre and ses-post. To compare the overall quality of life, the summarized SWLS score was compared between groups in ses-pre using a between-subjects t-test and between sessions (ses-pre versus ses-post) for patients using a within-subjects t-test (note: SWLS in the control group was only measured in ses-pre).

In order to examine exploratory the relationship of outcomes in quality of life (SWLS score changes) and neural activation changes, eigenvariates of the multimodal condition regressors (absSG and conSG) were extracted from the significant clusters in contrast (1) and contrast (2). Activation changes (ses-post - ses-pre) in the neural activation of the absSG condition and changes in the SWLS score (ses-post - ses-pre) were correlated via Pearson’s r. Two subjects had to be excluded from the analyses (details are described in Appendix C).
2.9. Secondary outcome measures

For information regarding the speech-gesture fluency performance during training, see a detailed description in Appendix E.

To get an impression of possible performance changes during the MSG training, the performance in the gesture fluency task from the first training session was compared with the performance from the last training session using a within-subjects t-test.

The changes in performance during the MSG training were also correlated with the neural activation changes in contrast (1) and contrast (2) via Pearson’s r.

3. Results

3.1. Quality of life

The group comparison of the quality of life in controls versus patients confirmed the assumption that patients suffer from a significant reduction in quality of life ($t_{\text{Welch}}(35.07) = 2.45; p = 0.020; \hat{g}_{\text{Hedges}} = 0.74$), see Appendix D. The pre-post-training comparison of the SWLS score in patients (Fig. 3) revealed a significant increase (with medium effect) of quality of life ($t_{\text{Student}}(22) = 4.74; p = 9.85 \times 10^{-5}; \hat{g}_{\text{Hedges}} = 0.95$).

3.2. Training

During training, the patients showed an increase in gesture fluency performance. A comparison of the performance from the first to the last training session showed a significant increase in performance ($t_{\text{Student}}(24) = 7.30; p = 1.55 \times 10^{-7}; \hat{g}_{\text{Hedges}} = 1.41$), see Fig. 3.

3.3. FMRI

Contrast (1): For the interaction of session X abstractness, significant activation was found in three clusters including the left parahippocampal gyrus, middle frontal and bilateral temporal regions as well as superior frontal regions. Patients with SSD demonstrate a specific activation increase in these regions for the processing of abstract speech-gesture videos and an activation decrease for the processing of concrete speech-gesture videos after MSG training (see Fig. 2).

Contrast (2): For the interaction (F-test) of session (ses-pre, ses-bl, ses-post) X abstractness, significant activation was found in a large left middle temporal cluster, in a smaller cluster in right temporal gyrus and in the cerebellum. Patients with SSD demonstrate a specific activation increase for the processing of abstract speech-gesture videos and an activation decrease for the processing of concrete speech-gesture videos.

**Fig. 2.** Neural activation session X abstractness in SSD group.
Slice view visualization for the interaction of session X abstractness (ses-pre(absSG>conSG)) < ses-post(absSG>conSG) in the patients (contrast (1)). Activation clusters were thresholded at $p < 0.01$, with a minimum cluster size of 221 voxels, cluster level corrected at $p > 0.05$. Error bars indicate the standard error of the mean (s.e.m.) in all cases. Crosshair point at the right MTG [$x = 52, y = 2, z = -26$]; left PHG [$x = -30, y = -24, z = -18$] and left SFG [$x = -20, y = 18, z = 58$]. MTG = middle temporal gyrus; PHG = parahippocampal gyrus; SFG = superior frontal gyrus; L = left; R = right; absSG = abstract multimodal condition; conSG = concrete multimodal condition; SWLS = Satisfaction With Life Scale; ses-pre = before training measurement; ses-post = after training measurement; SSD = schizophrenia spectrum disorder.
after MSG training (see Fig. 4).

3.4. Correlations

We found neural changes (Table 2) for the processing of abstract speech-gesture videos (absSG ses-post - absSG ses-pre) in contrast (1) to be associated with the changes in quality of life (with $t_{\text{Student}}(21) = 3.51$, $p_{\text{unadjusted}} = 0.002$, $p_{\text{Bonferroni-Holm}} = 0.006$, $p_{\text{Pearson}} = 0.61$ for the right MTG, $t_{\text{Student}}(21) = 2.87$, $p_{\text{unadjusted}} = 0.009$, $p_{\text{Bonferroni-Holm}} = 0.018$, $p_{\text{Pearson}} = 0.53$ for the left PHG and $t_{\text{Student}}(21) = -2.41$. $p_{\text{unadjusted}} = 0.025$, $p_{\text{Bonferroni-Holm}} = 0.025$, $p_{\text{Pearson}} = -0.47$ for the SFG) and for the processing of concrete speech-gesture videos (conSG ses-post - conSG ses-pre) to be associated with the patients’ improvement in the gesture fluency task over training (with $t_{\text{Student}}(23) = 4.01$, $p_{\text{unadjusted}} = 0.001$, $p_{\text{Bonferroni-Holm}} = 0.003$, $p_{\text{Pearson}} = 0.64$ for the right MTG, $t_{\text{Student}}(23) = 2.19$, $p_{\text{unadjusted}} = 0.039$, $p_{\text{Bonferroni-Holm}} = 0.074$, $p_{\text{Pearson}} = 0.41$ for the left PHG and $t_{\text{Student}}(23) = 2.21$, $p_{\text{unadjusted}} = 0.037$, $p_{\text{Bonferroni-Holm}} = 0.074$, $p_{\text{Pearson}} = 0.42$ for the SFG), shown in Fig. 3. For contrast (2) we found neural changes in left temporal regions to be associated with changes in quality of life (with $t_{\text{Student}}(14) = 3.34$, $p = 0.005$, $p_{\text{Pearson}} = 0.67$) and with the patients’ improvement in the gesture fluency task over training (with $t_{\text{Student}}(15) = 2.32$, $p = 0.035$, $p_{\text{Pearson}} = 0.51$), shown in Fig. 5.

3.5. Self-report measures

Our specifically outlined post training questionnaires provide insights into subjective evaluation of training related improvement of community functions: E.g., 94 % of the patients rated the MSG training as useful, 53 % of the patients’ relatives reported an increase of the patients’ social contacts and 74 % reported a gesture related improvement of perceptive, 66 % of expressive communication skills.

4. Discussion

Social-communicative dysfunctions in schizophrenia have received increased interest from the field of clinical neuroscience. Considering the tremendous impact of figurative speech and the fundamental role of gestures in social communication and functioning, we developed a novel MSG training and focused on the training’s effects on quality of life and its neural correlates in schizophrenia spectrum disorder.

This novel training was both feasible and tolerable, as demonstrated by the low drop-out rate in contrast to other successful interventions (Gordon et al., 2018) and the patients’ positive ratings of the training program. The overwhelming majority of the patients and their relatives that we analyzed exploratory, the subjective impressions about the effects of the MSG training seem to be satisfying. Consequently, we evaluated our design and training as well feasible for this group of patients.

As predicted, the quality of life in the patients was reduced before the training but increased significantly after the MSG training, confirming the idea that patients with schizophrenia might benefit from a specific speech-gesture training in their general quality of life (Heim, 2020; Joyal et al., 2016).

Changes in neural activation provide further insights into the working mechanisms of the MSG training program. In the session *X abstractness interaction* for multimodal speech-gesture processing in patients, brain regions in the temporal lobe including (para-)hippocampal regions showed significant changes in activation and therewith seem to play a key role in training induced compensation mechanisms for the processing of abstract multimodal input, as expected (Dick et al., 2009; He et al., 2018; Joue et al., 2020; Kircher et al., 2009; Rossetti et al., 2018; Straube et al., 2011, 2013a).

According to the Memory Unification Control model (Hagoort, 2013; Hagoort et al., 2009; Holler and Levinson, 2019; Willems and Hagoort, 2007), temporal regions are involved in integration processes and lexical-semantic processing. Temporal regions were also reported in former studies to be active during speech-gesture integration (Dick et al., 2014; Green et al., 2009; Joue et al., 2020; Kircher et al., 2009; Straube et al., 2011, 2013a, 2013b, 2014) with the right hemisphere aiding in the integration of speech and gesture information (Heim, 2020; Kircher et al., 2009; Straube et al., 2011). The increasing activation in the abstract multimodal condition (absSG) after the training might provide evidence for some kind of adaptive processes toward the patterns observed in healthy control subjects, as it is reported for cognitive remediation therapy (Penadés et al., 2017). The changes in neural responses might reflect a specific training effect on brain regions relevant for metaphor comprehension (impaired in concretism) or the integration of abstract speech-gesture combinations (Bergemann et al., 2008; de Bonis et al., 1997; Iakimova et al., 2010; Kircher et al., 2007, 2009; Nagels et al., 2019; Rapp and Schmierer, 2010; Straube et al., 2011, 2014).

Strikingly, we found the neural changes in the left and right middle temporal gyrus to be associated with the changes in quality of life as well as with the performance increase over the MSG training (shown in Fig. 3). Although the explanatory power of correlations is limited with this sample size, this links the potential benefits of a specific training of speech-gesture skills with improvement in everyday life in patients with SSD. Reports of relatives support this finding, indicating improved communication skills and increase of social contacts after MSG training in the majority of patients.

Comparing all three fMRI measurements, the pattern of activation in the cluster in left middle temporal gyrus (shown in Fig. 4) shows no significant differences between ses-pre and ses-bl (TAU period), but significant differences between ses-pre and ses-post (MSG training period; again with an increase of activation in absSG and decrease in conSG), providing evidence for the specific effects during the MSG training period on the neural processing of abstract multimodal speech-gesture videos in patients with SSD. Again, the neural activation changes after the MSG training correlate with the increase in quality of life.

Methodological limitations to be considered for a valid interpretation of the results concern the modification of the sample size in the control group, due to recruitment problems during the Covid-19 pandemic. Thus, the control group did not receive the MSG training, so that the effects of the training cannot be directly compared between patients and controls. Because our inclusion criteria were very strict, e. g., in terms of medical contraindications against fMRI measurements, our sample might be an imperfect representative for patients with SSD in general, particularly regarding patients suffering from first episode (Kindler et al., 2019; Newton et al., 2018). This is also true for sex and age, since our patient sample consisted of 7 women and 22 men with a wide age range. While in our within-subject crossover design factors that

---

**Table 2**

| Anatomical region of peak | Cluster extend | MNI coordinates | $t$-Value | No. of voxels |
|---------------------------|---------------|-----------------|-----------|---------------|
| Right MTG                | TMP, ITG      | $x$ = 52, $y$ = -26, $z$ = 4.61 | 414 |
|                           | HIP, INS, STG | $x$ = 38, $y$ = -14, $z$ = 3.48 |   |
|                           | MTG, ITG, STG | $x$ = 58, $y$ = -22, $z$ = 2.24 |   |
| Left PHG                 | HIP, FFG      | $x$ = -30, $y$ = -24, $z$ = 4.30 | 723 |
|                           | MTG, STG, ITG | $x$ = -46, $y$ = -20, $z$ = 4.06 |   |
|                           | MTG, ITG, STG | $x$ = -44, $y$ = -16, $z$ = 4.01 |   |
| Left SFG                 | Left MFG, SMA | $x$ = 20, $y$ = 18, $z$ = 5.18 | 727 |
|                           | Right SFG, SMA, SFGmedial | $x$ = 20, $y$ = 24, $z$ = 56 | 4.01 |
|                           | Right SFG, SMA, SFGmedial | $x$ = 20, $y$ = 12, $z$ = 64 | 3.82 |

FFG, fusiform gyrus; HIP, Hippocampus; INS, Insula; ITG, inferior temporal gyrus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; PHG, parahippocampal gyrus; SFG, superior frontal gyrus; SFGmedial, Superior frontal gyrus, medial; SMA, supplementary motor area; STG, superior temporal gyrus; TMP, temporal pole; SSD, schizophrenia spectrum disorder.
we did not collect, such as IQ and neurocognitive abilities, at least maintained constant across measurement points, we cannot exclude that they represent an important mediator of the effects. Also the MSG training outcomes may have been influenced by other factors such as the examiner-patient relationship.

After this successful results using a waiting list controlled pilot trial,
comparing two active treatment arms would be important to demonstrate the unique contribution of the MSG training intervention. Furthermore, specific conclusions which therapeutic mechanisms caused the outcome have to be examined by further research (e.g., most effective training task or the influence of the trainers’ interpersonal style are yet unknown). After demonstrating feasibility with few dropouts, it also seems feasible to collect other relevant performance data, e.g., via the brief self-rating scale for the assessment of individual differences in gesture perception and production (BAG: Nagels et al., 2015) or the rating scale for the assessment of objective and subjective formal Thought and Language Disorder (TALD: Kircher et al., 2014). Importantly, the results of this exploratory study require further validation in independent, ideally double-blind studies. For this purpose, we provide detailed descriptions of the study design, fMRI paradigms as well as training material on our public project on the open science framework (OSF: DOI 10.17605/OSF.IO/UH4F9), which is freely available under CCBy Attribution 4.0 International license.

5. Conclusion

The overall analyses evaluating the MSG training provide extraordinarily promising results which should be validated and extended in further independent studies. Especially the subjectively reported benefits on everyday life and the associated neural changes in brain regions in the temporal lobe provide evidence for potentially beneficial effects of innovative add-on treatments. Future studies should investigate combinations of speech-gesture training with neural stimulation techniques like transcranial direct current stimulation (Walther et al., 2020) or transcranial direct current stimulation (Schülke and Straube, 2019), which seem to be further promising approaches (Cavelti et al., 2018).

Statement of ethics

This study was carried out in accordance with the recommendations of the local ethics committee (Philips-University Marburg, Department of Medicine, Deanery/Ethics Committee, Reference: R1, Study 01/17) on 28th February 2017 with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the local ethics committee. All information collected is kept confidential, stored securely and archived in accordance with the research governance policy of the university. Participant anonymity is retained by allocating a unique identification number for the trial and any identifiable information stored separately from this.

Data availability statement

Anonymized behavioral and clinical data as well as MSG training material and detailed information about our assessments are freely available under CCBy Attribution 4.0 International license on the Open Science Framework: DOI 10.17605/OSF.IO/UH4F9.

Funding

This work was supported by a grant from the von behring|röntgen|foundation (grant number 64-0001) and supported by “The Adaptive Mind”, funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. BS was supported by the Heinrich Böll Foundation. BS is supported by the ‘Deutsche Forschungsgemeinschaft’ (Project no. DFG: STR 1146/11–2, STR 1146/15–1).

CRediT authorship contribution statement

BS, AN and GS conceived of the presented intervention approach. LR, AN, MC, AS, CF, MH and FB carried out the experiment. AN, MC, AS, CF, and MH conducted the trainings. LR performed the computations and analytic calculations and wrote the manuscript. BS supervised the study and verified the analytical methods. All authors discussed the results, contributed to the final manuscript and agree to be accountable for all aspects of the work.

Declaration of competing interest

The Authors have declared that there are no conflicts of interest in relation to the subject of this study.

Acknowledgements

Special thanks to the Core Facility Brain Imaging Marburg for assistance with data collections and Dr. Christoph Vogelbacher for his help with installing singularity container images for usage on the local server system. We would like to thank Johanna Funk and Katrin Leinweber who rated the patient videos from the training as well as Thomas S. Hartmann, who was a great help in preparing the study materials for publication.

Appendix A. Details about the video stimuli

The videos clips presented during fMRI data acquisition were standardized, extensively evaluated and had been successfully applied in a large number of studies (Choudhury et al., 2021; Green et al., 2009; He et al., 2018; Kircher et al., 2009; Nagels et al., 2019; Schülke and Straube, 2019; Straube et al., 2009, 2011, 2013a, 2014; (Wroblewski et al., 2020)). The video sequences had a duration of five seconds. The videos were displayed on an MRI-compatible screen using the software Presentation® (Version 18.3, Neurobehavioral Systems, Inc.; https://www.neurobs.com/presentation) made visible via a mirror attached to the head coil. The experiment comprised 30 video sequences per condition, 180 video sequences in total, presented in a pseudo-randomized order in three blocks with a duration of ~9 min each. The participants were pseudo-randomly assigned to one of eight different counterbalanced versions of stimuli presentation to avoid sequence effects.

Appendix B. FMRI data quality control

Diverging from the planned preprocessing procedure (Riedl et al., 2020), for quality control, MRIQC was used based on different parameters like co-registration, motion and temporal signal-to-noise (Esteban et al., 2017). No data had to be excluded.
Appendix C. SWLS data quality control

Outcomes (overall SWLS scores) were checked for statistical outliers. Data from two patients\(^2\) had to be excluded from further analysis (in one case due to a massive reported increase in wellbeing and in another case due to worsening of negative symptoms accompanied by temporary discontinuation of medication during the training period). In ses-pre (before the MSG training started), one patient and in ses-post (after the MSG training), three patients withheld information regarding satisfaction with life or filled in the SWLS questionnaire incompletely. As consequence, only incomplete data were available for four patients, so that the session differences could not be calculated and therefore could not be included in the correlation.

Appendix D. SWLS comparison of SSD group versus control group (in ses-pre)

![Fig. D1. Quality of life group differences (control group versus SSD patients).](image)

Group comparison of SSD patients versus control group in ses-pre (before the MSG training) of life quality, examined using the SWLS score. SSD = schizophrenia spectrum disorder; SWLS = Satisfaction With Life Scale; ses-pre = before training measurement.

Appendix E. Speech-gesture fluency task

\(^2\) One of the excluded participants reported an exceptional increase (by 10 points) and another an exceptional decrease (by –9 points) pre to post training in the SWLS questionnaire.
Overview over all eight MSG training sessions: comparison of the patients’ performance in speech-gesture fluency training task (sum of correctly produced speech-gesture pairs in all three categories per session).

The speech-gesture fluency task was a fixed component in every session of the MSG training and therefore suitable to explore training related changes across training sessions. Similar to verbal fluency tasks (Rosenkranz et al., 2019; Wende et al., 2012), the patients were instructed to generate as many words with accompanying suitable gesture as possible for each of the three semantic fields per session, resulting in a total of 24 categories. Time was limited to 1 min per semantic field.

Since the stimulus material in the gesture fluency task was designed with a primary focus on eligibility for therapeutic intervention, the gesture fluency word categories differed in every session in order to achieve the best possible training effects. Thus, a direct comparison between the individual training sessions is difficult to interpret. However, the patients tend to improve from the first to the last training session. Together with the subjective information on the effectiveness of the training, this leads to a positive overall picture of the training’s impact.

It seems that the SSD patients already benefit from the training in the second MSG training session. Previous studies using stimulation techniques to improve gesture performance or speech-gesture perception in patients with schizophrenia have also demonstrated improvement after only one session (Schülke and Straube, 2019; Walther et al., 2020). Although the interpretation of MSG training data is difficult, as mentioned above, it would be an interesting research question for future studies to investigate how many training sessions are necessary to achieve a benefit.

Fig. E1. Improvement of SSD patients in the SG fluency task over all 8 training sessions.

References

Allen, D.J., Antonitis, J.J., Magaro, P.A., 1978. Reinforcing effects of prerecorded words and delayed speech feedback on the verbal behavior of a neologistic schizophrenic. Percept. Mot. Skills 46, 343–346. https://doi.org/10.2466/pms.1978.46.2.343.

Andreasen, Nancy C., 1981. Scale for the assessment of negative symptoms (SANS). University of Iowa Press, Iowa City.

Bambini, V., Arcara, G., Bechi, M., Cavallaro, R., Bosia, M., 2016. The communicative impairment as a core feature of schizophrenia: frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. Compr. Psychiatry 71, 106–120. https://doi.org/10.1016/j.comppsych.2016.08.012.

Bambini, V., Agostoni, G., Buonocore, M., Tonini, E., Bechi, M., Ferri, I., Sapienza, J., Martini, F., Cuoco, F., Cocchi, F., Bischetti, L., Cavallaro, R., Bosia, M., 2022. It is time to address language disorders in schizophrenia: a RCT on the efficacy of a novel training targeting the pragmatics of communication (PragmaCom). J. Commun. Disord. 106196 https://doi.org/10.1016/j.jcomdis.2022.106196.

Bavelas, J.B., Gerwing, G., Allison, M., Sutton, C., 2011. Dyadic evidence for grounding with abstract deictic gestures. In: Stam, G., Ishino, M. (Eds.), Integrating Gestures: The Interdisciplinary Nature of Gesture. John Benjamins Publishing Company, Amsterdam, pp. 49–60.

Beeman, M., 1998. Coarse semantic coding and discourse comprehension. In: Chiarello, C., Beeman, M. (Eds.), Right Hemisphere Language Comprehension: Perspectives from Cognitive Neuroscience. Lawrence Erlbaum Associates Inc, Publishers, Mahwah, New Jersey, pp. 255–284.

Beilock, S.L., Goldin-Meadow, S., 2010. Gesture Changes Thought by Grounding It in Action. Psychological Science 21 (11), 1605–1610. https://doi.org/10.1177/0956797610385353.

Bergemann, N., Parzer, P., Jaggy, S., Axler, B., Munding, C., Maier-Braunleider, S., 2008. Estrogen and comprehension of metaphoric speech in women suffering from schizophrenia: results of a double-blind, placebo-controlled trial. Schizophr. Bull. 34, 1172–1181. https://doi.org/10.1093/schbul/sbm138.

BerrILL, K., von Granich, M., Gruiser, O.-J., 1986. Impairment of perception and recognition of faces, mimic expression and gestures in schizophrenic patients. Eur. Arch. Psychiatr. Neurol. Sci. 235, 282–291. https://doi.org/10.1007/BF00515915.

Bosco, Francesca M., Gabbiatore, Ilaria, Gualano, Luigi, Sacco, Katiuscia, 2016. Communicative-Pragmatic Treatment in Schizophrenia: A Pilot Study. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2016.00166.

Bourton, Isabelle, Altman, Douglas G., Moher, David, Schulz, Kenneth F., Ravaud, Philippe, 2017. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Annals of internal medicine 167 (1), 40–47. https://doi.org/10.7326/M17-0464.

Bucic, S., Startup, M., Wynn, P., Baker, A., Lewin, T.J., 2008. Referential delusions of communication and interpretations of gestures. Psychiatry Res. 158, 27–34. https://doi.org/10.1016/j.psychres.2007.07.004.

Cavelti, M., Kircher, T., Nagels, A., Strik, W., Homan, P., 2018. Is formal thought disorder in schizophrenia related to structural and functional aberrations in the language network? A systematic review of neuroimaging findings. Schizophr. Res. 199, 2–16. https://doi.org/10.1016/j.schres.2018.02.051.
Willems, R.M., Hagoort, P., 2007. Neural evidence for the interplay between language, gesture, and action: a review. Brain Lang. Gesture Brain Lang. 101, 278–289. https://doi.org/10.1016/j.bandl.2007.03.004.

World Health Organization, 1992. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. World Health Organization.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series Revisited—Again. NeuroImage 2, 173–181. https://doi.org/10.1006/nimg.1995.1023.

Wroblewski, Adrian, He, Yifei, Straube, Benjamin, 2020. Dynamic Causal Modelling suggests impaired effective connectivity in patients with schizophrenia spectrum disorders during gesture-speech integration 216, 175–183. https://doi.org/10.1016/S0920-9964(19)30076-6.

Wüthrich, F., Pavlidou, A., Stegmayer, K., Eisenhardt, S., Moor, J., Schöpfl, L., Vanbellingen, T., Bohlhalter, S., Walther, S., 2020. Nonverbal communication remains untouched: no beneficial effect of symptomatic improvement on poor gesture performance in schizophrenia. Schizophr. Res. 223, 258–264. https://doi.org/10.1016/j.schres.2019.12.005.

Wykes, T., 1998. What are we changing with neurocognitive rehabilitation?: illustrations from two single cases of changes in neuropsychological performance and brain systems as measured by SPECT. Schizophr. Res. 34, 77–86. https://doi.org/10.1016/S0920-9964(98)00076-6.