Behavioural Brain Research 412 (2021) 113411

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Deficits in multi-scale top-down processes distorting auditory perception in schizophrenia

Fuyin Yang [a,b], Hao Zhu [b,c], Lingfang Yu [d], Weihong Lu [d], Chen Zhang [d,**], Xing Tian [b,c,*]

[a] Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China

[b] NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, 3663 Zhongshan Road North, Shanghai, 200062, China

[c] Division of Arts and Sciences, New York University Shanghai, 1555 Century Avenue, Shanghai, 200122, China

[d] Schizophrenia Program, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China

ARTICLE INFO

Keywords:
Cognitive control
Source monitoring
Gain
Weighting function
Bayesian integration
Prediction

ABSTRACT

Cognitive models postulate that impaired source monitoring incorrectly weights the top-down prediction and bottom-up sensory processes and causes hallucinations. However, the underlying mechanisms of the interaction, such as whether the incorrectly weighting is ubiquitously on all levels of sensory features and whether different top-down processes have distinct effects in subgroups of schizophrenia are still unclear. This study investigates how multi-scale predictions influence perception of basic tonal features in schizophrenia. Sixty-three schizophrenia patients with and without symptoms of auditory verbal hallucinations (AVHs), and thirty healthy controls identified target tones in noise at the end of tone sequences. Predictions of different timescales were manipulated by either an alternating pattern in the preceding tone sequences (long-term regularity) or a repetition between the target tone and the tone immediately before (short-term repetition). The sensitivity index, d prime (d'), was obtained to assess the modulation of predictions on tone identification. Patients with AVHs showed higher d' when the target tones conformed to the long-term regularity of alternating pattern in the preceding tone sequence than when the target tones were inconsistent with the pattern. Whereas, the short-term repetition modulated the tone identification in patients without AVHs. Predictions did not influence tone identification in healthy controls. Our results suggest that impaired source monitoring in schizophrenia patients with AVHs heavily weights top-down predictions over bottom-up perceptual processes to form incorrect perception. The weighting function in source monitoring can extend to the processes of basic tonal features, and predictions at multiple timescales could differentially modulate perception in different clinical populations. The impaired interaction between top-down and bottom-up processes might underlie the development of hallucination symptoms in schizophrenia.

1. Introduction

Auditory verbal hallucinations (AVHs) are symptoms of hearing voices in the absence of external stimuli [1]. Around 50 %–70 % of people who are diagnosed with schizophrenia experience AVHs [2,3]. Antipsychotic medications, such as olanzapine, risperidone and quetiapine that block the D2-receptors, are effective drugs to treat AVHs [4,5]. However, weight gain and sedation are serious side effects associated with antipsychotic medications [6]. Medications may expose patients to metabolic complications and result in treatment non-adherence. Moreover, about 25 % of patients are resistant to standard antipsychotic treatment [7]. Noninvasive neuro-stimulation techniques have been tested as a new treatment option for AVHs [8]. Repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) are two noninvasive techniques that are recently introduced to treat AVHs [9,10]. The rTMS and tDCS, as promising treatment options [4,11], show a moderate effect size in the reduction of AVHs frequency [12,13]. The fact that the treatment effects depend on the stimulation protocols and cortical targets [12,14] highlights the necessity of understanding AVHs from a cognitive neuroscience perspective.

The cognitive models postulate that AVHs may result from a process in which inner or sub-vocal speech is misidentified as externally caused

* Corresponding author at: New York University Shanghai, 1555 Century Avenue, Shanghai, 200122, China.
** Corresponding author.
E-mail addresses: zhangchen645@gmail.com (C. Zhang), xing.tian@nyu.edu (X. Tian).

https://doi.org/10.1016/j.bbr.2021.113411
Received 8 September 2020; Received in revised form 12 May 2021; Accepted 7 June 2021
Available online 11 June 2021
0166-4328/© 2021 Elsevier B.V. All rights reserved.

[15]. Such source monitoring account of AVHs requires an internally generated source. Prediction has been proposed as an algorithm that induces this internal source [16]. Prediction refers to the set of processes based on the information of memory, knowledge, and belief to generate representations of future events [17–20]. For example, similar representations as perception can be constructed based on retrieving memory of past experiences and regularities without external stimulations [21–23]. These internally constructed representations from prediction could be the critical source in the cognitive monitoring account of AVHs.

By combining the internal source of prediction and the external source of sensory analysis, the computation of source monitoring can be quantified as a weighting function between prediction and perception. Numerous studies have suggested that the top-down prediction interacts with bottom-up sensory processes to shape perception [24–26]. Predictions can balance the cognitive resources of processing the repetitive stimuli and detecting unexpected events [27]. Breakdown in predictive function may cause less attention to cues of upcoming sensory signals [28]. The unbalanced weighting between sources from prediction and sensory input could cause AVHs [29]. For example, participants with severe hallucinations significantly increase gain over predictions in ambiguous perceptual situations [30], suggesting that a relatively higher priority is assigned to top-down factors in determining the final percepts [31]. To an extreme, abnormal top-down prediction processes in patients overwhelm the auditory input [29,32]. The impaired weighting in the source monitoring is consistent with the framework of Bayesian inferences, sharing the assumption that AVHs are induced when sensory predictions are activated without sensory input, or these predictions are not properly deactivated and incorrectly replace the sensory analysis [33].

Two crucial aspects to understand the source monitory of AVHs are perceptual features and temporal characteristics that the weighting function is operating on. First, source monitoring weights and balances multiple levels of features across different sources to establish coherent percepts. Previous studies have demonstrated the influences of top-down prediction at the semantic and phonological levels in healthy subjects [34]. Recently, the effects of top-down prediction on sensory analysis have extended to lower and basic sound attributes, such as pitch and loudness [23,35]. However, whether the weighting function in source monitoring of AVHs can extend to basic level attributes is still debatable. Some studies found that the severity of hallucination-prone was correlated with errors that were induced by semantic priming but not with phonological priming [36]. Other studies provided preliminary evidence of increased top-down influences for tonal stimuli [29]. In this study, we investigated whether schizophrenia patients increased top-down influences in the processing of tones. That is, we aim to answer whether the source monitoring in AVHs only incorrectly weights the higher-level features or has a ubiquitous weighting function that applies to sound attributes of all levels.

Second, prior information is available at multiple timescales and facilitates information processing across time [37]. For example, speech processing may operate at two distinct timescales [38,39]. Multiple levels of prior information could help comprehension of linguistic information ranging from phonemes to words, to sentences and paragraphs [40–42]. The memory of recent events integrates information over milliseconds, seconds, and minutes to form predictions at multiple timescales that continuously support the processing of incoming information [43]. Would the source monitoring of AVHs weight predictions at multiple timescales differently?

Together, this study investigates how multi-scale predictions influence tone perception in schizophrenia patients. Specifically, we hypothesize that AVHs are caused by an incorrect weighting of top-down predictions, distorting the balance between bottom-up and top-down processes. This distorted balance of weighting sources in AVHs may influence the processing of basic speech features such as tones, and predictions at different timescales may modulate the bias differently.

In this study, we manipulated the long-term regularity and immediate repetitions in sequences of tones to investigate the featural and temporal characteristics of the weighting function in source monitoring of AVHs. Sequences of tones were presented either in an alternating pattern (long-term regularity) or randomly with the possibility that the last tone repeated the immediately preceding one (short-term repetition). Three groups of participants, schizophrenia patients with and without AVHs as well as healthy controls, were asked to identify the last tone that was embedded in noise. Perceptual sensitivity, d', was obtained based on the signal detection theory (SDT). According to our hypothesis that patients with AVHs might confuse the sources of their memory-based predictions and the sensory processing of tonal features, we predicted that the d' of tone identification in the group of AVHs would be modulated by the manipulations in the preceding tone sequences. Moreover, the modulation effects would be different for the long-term and short-term predictions across groups.

2. Materials and methods

2.1. Participants

Thirty-two (14 males) patients, who matched a DSM-V diagnosis of schizophrenia and were currently experiencing AVHs (AVHs group) without concomitant hallucinations in other modalities, were recruited from Shanghai Mental Health Center. Furthermore, thirty-one (13 males) patients who met DSM-V diagnosis of schizophrenia and had never experienced AVHs (non-AVHs group) were recruited from the same hospital. Two experienced psychiatrists independently diagnosed each patient, and the diagnosis was confirmed by the Structured Clinical Interview for DSM-V (SCID). All patients were receiving atypical antipsychotic medications and were clinically stable.

Thirty healthy participants (9 males) were recruited as the control group from the local communities and schools in Shanghai (HCs group). A clinical psychiatrist assessed these healthy subjects' current mental status and any personal and family history of mental disorders. Moreover, any subject with potential psychiatric morbidity was excluded from the control group after the psychiatrist's unstructured interviews. None of the healthy subjects had any family history of psychiatric disorders or physical diseases.

All participants were in the age range of 18–45 years old, right-handed, and without any substance abuse records.

The study was approved by the Institutional Review Board of the New York University Shanghai and the Institutional Ethics Committee at Shanghai Mental Health Center. Informed consent was obtained from all participants before they participated in the study.

2.2. Clinical measures

Demographic data were collected from patients and healthy controls. Four psychiatrists, who were blind to the study, assessed the patient's psychopathology. First, two senior psychiatrists diagnosed patients independently. Patients diagnosed by both psychiatrists with schizophrenia were included in a follow-up evaluation. Two other psychiatrists assessed the psychotic symptoms using the Positive and Negative Syndrome Scale (PANSS). The PANSS measures both the presence and severity of positive, negative and general symptoms on a 7-point scale. AVHs severity was rated from the P3 of the PANSS scale, with higher ratings indicating an increase in AVHs severity. Non-AVHs patients had a rating of 1 in the P3 factor score, indicating that the symptom was absent. Additionally, the severity level of AVH was assessed using the 7-item Auditory Hallucinations Rating Scale-AHRS [44]. To ensure consistency and reliability of PANSS and AHRS, paired ratings between two psychiatrists for the same patient assessment were compared at each of the repeated assessments. All paired ratings had a correlation coefficient greater than 0.8 on the PANSS and AHRS total scores.

2.3. Experimental design

2.3.1. Materials

Two Mandarin tones of vowels /a/ (/ā/ and /á/) were synthesized via the NeoSpeech engine [45] with a female voice. Both Mandarin tones were 377 ms in duration and scaled to 75 dB SPL in intensity using Praat software [46]. Two additional stimuli were created by adding white noise to the two tones. The signal-to-noise ratio (SNR) was determined at the individual level during a pre-test. All stimuli were digitized at 44.1 kHz sampling rate and 16-bit bitrate. These auditory stimuli were delivered through Sennheiser HD 280 Pro headphones. The volume was adjusted to a comfortable level for each participant and kept consistent throughout the experiment for all stimuli.

2.3.2. Threshold test procedures

The experimental procedure was composed of two steps. First, participants participated in a pre-test that measured a threshold for detecting two Mandarin tones (/ā/ or /á/) in white noise. This pre-test session consisted of 200 trials. At the beginning of a trial, a visual cue was presented for 500 ms. After the offset of the visual cue, one of the auditory stimuli in noise was presented. White noise was 1000 ms in duration. The auditory stimulus was presented 100 ms after the onset of the white noise and lasted for 377 ms. The intensity of the white noise changed trial by trial given by the Bayesian adaptive "PSI" staircase method [47] while the intensity of the vowels was fixed at 75 dB. The Psi-staircase assumed a log-Weibull (Gumbel) function with a non-zero (2%) attentional lapse rate (Lambda) and a 5% guess rate (Gamma). Two randomly interleaved Psi-staircase objects for the two auditory stimuli were created with 100 trials per staircase. Participants were required to provide a perceptual judgment of the tone in a two-alternative forced-choice (2AFC) task. Participants took a break of a few minutes after every 50 trials. The threshold intensity of noise for each tone was determined by the 75 % accuracy point in the fitted psychometric curves for each participant. The signal-to-noise ratio (SNR) used in the main experiment was determined by the (fixed) intensity of signals to the threshold intensity of noise. Each participant has an SNR for /ā/ or /á/ respectively.

Last, we ran a confirmation test. Participants judged twenty trials (ten for each tone) with the SNR determined in the preceding threshold test. When the number of correct judgments was seven or eight for each tone, participants were considered to have passed the test. If the number of correct judgments was not seven or eight, we adjusted the SNR value and repeated the confirmation test until the number of correct judgments was seven or eight. Only when participants passed the confirmation test (the SNR of 75 % accuracy was confirmed) could proceed to the following procedure. In this case, we did not exclude any participants.

The threshold test procedure and the following main procedure were completed on the same day. After completing the threshold test procedure, participants were given an option of taking a break between 10−25 min before moving on to the main procedure.

2.3.3. Main procedure

After determining an appropriate SNR of each tone, participants proceeded to the main experiment in which they heard a sequence of tones and made judgments about the last tone in noise.

At the beginning of a trial, a fixation appeared in the screen center for 500 ms. After the onset of the visual cue, participants passively heard four to seven clean /ā/ or /á/ Mandarin tone in a sequence. The duration of each tone was 377 ms. The inter-stimulus interval (ISI) was 623 ms. Therefore, the stimulus onset asynchrony (SOA) was 1000 ms. The last stimulus in a trial was always a tone in noise with the individual measured SNR in the pre-test. The target was randomly selected from /ā/ or /á/, and was presented in noise in the same way as in the pre-test –100 ms after the onset of 1000 ms-long white noise and lasted for 377 ms. Participants judged whether the tone in the noise was /ā/ or /á/ by pressing one of two buttons (Fig. 1). The probability of each tone in the first clean tone position, in the last clean tone position, and in the noise was equalized.

We manipulated two parameters in this procedure to investigate how the top-down prediction interacted with the bottom-up sensory processing and influenced the perceptual sensitivity and bias. The first parameter was whether the clean tone sequence was presented in order. The sequence could be a regular pattern (RE) in which two tones were presented in an alternating manner. Alternatively, the sequence could be constructed by randomly presenting the two tones (IR). The second parameter was whether the tone in noise was consistent with the prediction of different time scales. In the RE conditions, the last tone in noise could be consistent with the regularity (RE_{con}) or inconsistent (RE_{inc}). That is, whether the target tone was consistent with the prediction formed by the long-term regularity of preceding tone sequences. Whereas, in the IR conditions, the tone in noise could be the same as (IR_{same}) or different from the last clean tone (IR_{diff}). That is, whether the target tone was a repetition that was consistent with the short-term immediate effect formed by the last clean tone in a random sequence. Therefore, a total of four conditions were included in this experiment. Thirty-two trials were included for each condition, yielding a total of 128 trials. The presentation order of trials was pseudorandomized among all participants.

2.4. Statistical analysis

When computing the measures to quantify tone identification responses, we took the tone /á/ as the target tone. The hit rate was calculated as a proportion of correct response on the /á/ tone, while the proportion of making /á/ responses to the /ā/ tone stimulus was defined as the false alarm rate. Following the Signal Detection Theory, the detection sensitivity (or discrimination ability) can be expressed by calculating the sensitivity index (d') [48].

Statistical analyses were performed using IBM SPSS Statistics version 17.0, GraphPad.Prism 5.02. The normal distribution of data was tested using the Kolmogorov-Smirnov tests. Demographic and clinical continuous variables were subject to one-way analysis of variance (ANOVA) with the factor of group, whereas the categorical values were subject to chi-squares test. Two-way mixed ANOVA was used to assess the performance of different groups in four conditions. Pearson correlation analyses were performed to determine the relationship between clinical variables and behavioral data within the AVHs group. We used stepwise multiple regression analysis with d' as the dependent variable to investigate the impact of age, gender, age of onset, duration of illness, AHRS total scores and PANSS and its subscales. Data are presented as mean (SD). Differences at $p < 0.05$ were considered to be significant.

3. Results

3.1. Demographic and clinical data

Table 1 shows the participants' demographic data and the clinical variables. ANOVA analyses showed a significant age difference ($p < 0.001$) among the three groups, but not in education, height, and weight ($p > 0.05$). The chi-square test showed no significant differences among the three groups about gender ($\chi^2 = 0.71$, $p = 0.496$). Further, P3 subscore, positive and general psychopathology subscores were significantly higher in the AVHs group than those in the non-AVHs group (all $p < 0.01$). Neither the age of onset, duration or the PANSS total score was significantly different between the two groups of patients.

3.2. Performance on the speech tone recognition task

The sensitivity indices d' were firstly subject to a two-way mixed ANOVA with groups (AVHs, non-AVHs, and HCS) as a between-subject factor and conditions (RE_{con}, RE_{inc}, IR_{same}, and IR_{diff}) as a within-subject

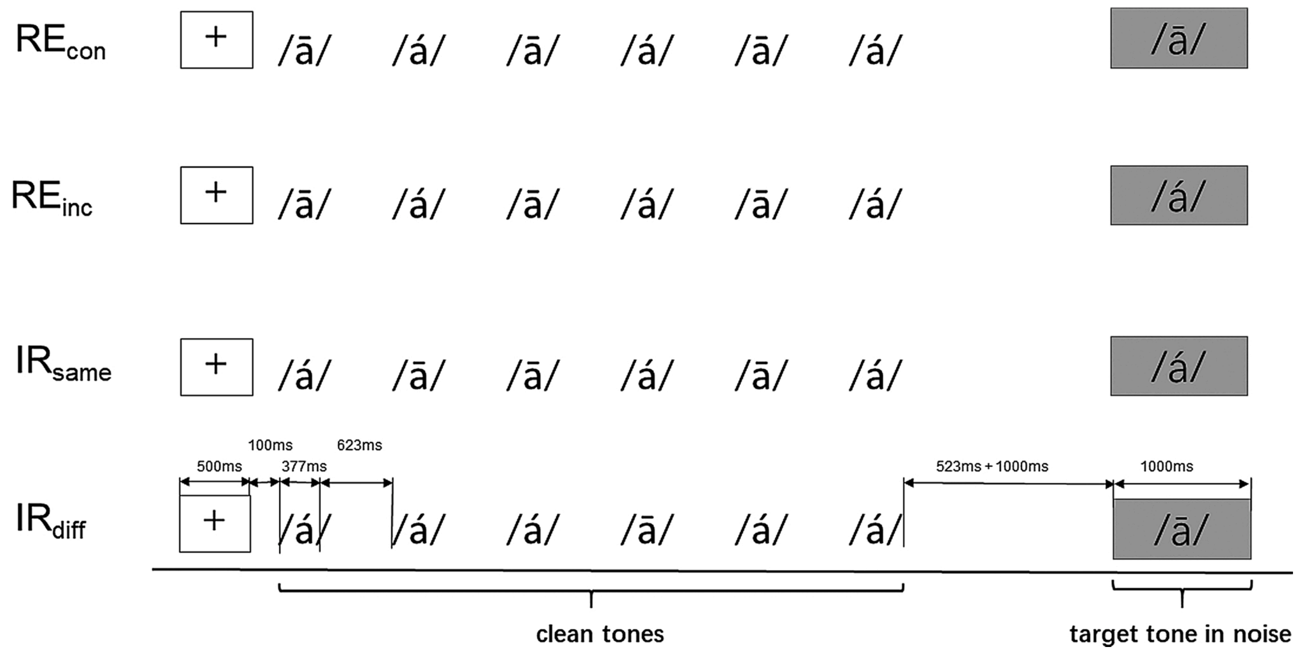

Fig. 1. Schematic description of experimental procedures.
At the beginning of a trial, a fixation appeared in the center of the screen for 500 ms. After the offset of the visual cue, participants passively heard four to seven clean /ā/ or /á/ Mandarin tones in a sequence. The duration of each tone was 377 ms. The inter-stimulus interval (ISI) was 623 ms. Trials with a different number of tones were randomly presented. The tone sequence could be in an alternating pattern (RE conditions in the upper two rows) or was presented randomly without any patterns (IR conditions in the lower two rows). The last stimulus in a trial was always a tone in noise with the individual measured SNR in the pre-test. The target tone was randomly selected from /ā/ or /á/, and was presented in noise in the same way as in the pre-test —100 ms after the onset of 1s-long white noise and the target tone lasted for 377 ms. Participants judged the target tone by pressing one of two buttons. The target tone in the RE conditions was either consistent (RE_{con}) or inconsistent (RE_{inc}) with the alternating pattern in the preceding clean tones. Whereas the target tone in the IR conditions was either the same (IR_{same}) or different (IR_{diff}) from the tone immediately before.

Table 1
Demographics of Schizophrenia patients with and without AVHs and Healthy Controls.

	AVHs Patients (n = 32)	Non-AVHs Patients (n = 31)	Healthy Controls (n = 30)	F or χ^2	*p*-value
Gender (M/F)	14/18	13/18	9/21	0.71	0.496
Age (years)	27.28 (6.37)	26.32 (6.12)	21.67 (2.45)	9.72	<0.001
Education (years)	12.44 (2.47)	12.90 (2.59)	13.43 (0.77)	1.69	0.190
Height (m)	1.68 (8.31)	1.66 (9.07)	1.67 (7.07)	0.31	0.736
Weight(kg)	66.78 (15.43)	62.81 (7.82)	60.37 (8.93)	2.55	0.084
Age of onset (years)	21.09 (6.66)	19.74 (5.09)	N/A	0.90	0.370
Duration (month)	74.05 (77.64)	69.74 (56.43)	N/A	0.25	0.802
PANSS total score	82.03 (8.87)	78.52 (4.27)	N/A	2.01	0.060
P3 subscore	4.75 (0.98)	1.00 (0.00)	N/A	21.56	<0.001
P subscore	23.53 (3.82)	14.96 (2.77)	N/A	10.16	0.001
N subscore	19.78 (3.48)	22.22 (1.60)	N/A	3.56	0.001
G subscore	38.72 (4.09)	41.32 (2.55)	N/A	3.04	0.004
AVHs total score	25.25 (3.85)	N/A	N/A	N/A	N/A

Notes: PANSS: Positive and Negative Syndrome Scale; P subscore: positive symptom subscore; N subscore: negative symptom subscore; G subscore: general psychopathology subscore; N/A: not applicable.

factor. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated ($\chi^2(5) = 11.356$, $p = 0.045$), Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.924$), and Huynh-Feldt estimates of sphericity ($\varepsilon = 0.977$). When $\varepsilon > 0.75$, the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. The main effect of condition was significant ($F(2.931, 263.791) = 17.748$, $p < 0.001$), the main effect of group was not significant ($F(290) = 2.407$, $p = 0.096$), the interaction between condition and group was significant ($F(5.862,263.791) = 10.359$, $p < 0.001$).

To further investigate differences between sub-groups of patients, a two-way mixed ANOVA was carried out with a within-subject factor of condition (RE_{con}, RE_{inc}) and a between-subject factor of group (AVHs, non-AVHs). The main effect of condition was significant ($F(1, 61) = 39.822$, $p < 0.001$), the main effect of group was not significant ($F(1,61) = 0.275$, $p = 0.602$), and the interaction was significant ($F(1,61) = 25.090$, $p < 0.001$). Moreover, a two-way mixed ANOVA was performed with a within-subjects factor of condition (IR_{same}, IR_{diff}) and a between-subject factor of group (AVHs, non-AVHs). The main effect of condition was significant ($F(1, 61) = 8.365$, $p = 0.005$), the main effect of group was not significant ($F(1,61) = 0.056$, $p = 0.813$), the interaction was not significant ($F(1,61) = 0.001$, $p = 0.972$).

To further explain the interaction, we compared tone recognition between conditions within each group to investigate how predictions of different time scales modulated perception in different groups. In the AVHs group, as shown in Fig. 2A, the d' in the RE_{con} condition was significantly higher than that in the RE_{inc} condition ($t_{(1,62)} = 7.45$; $p < 0.001$), suggesting that tone identification was influenced by long-term regularity. However, the d' was not different between the IR_{same} and IR_{diff} conditions, suggesting that the tone recognition was not influenced by the short-term repetition. This observation was consistent with our hypothesis suggesting that the long-term prediction biases perception in patients with AVHs.

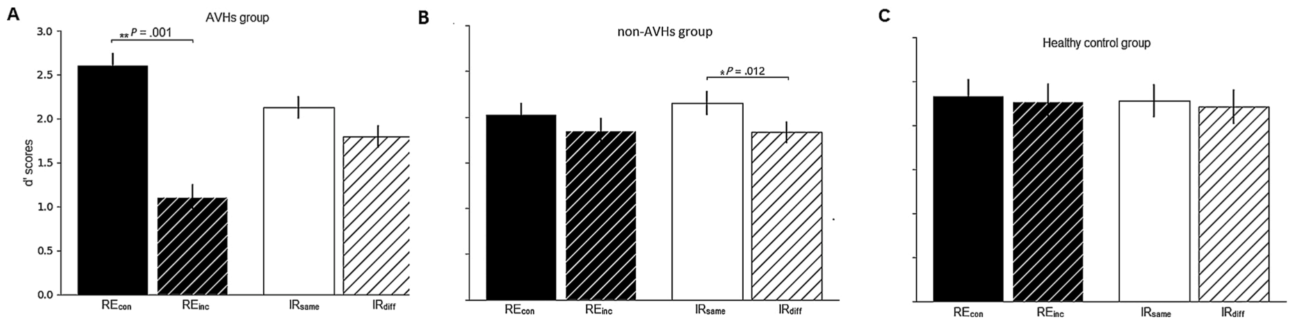

Fig. 2. Tone identification results of three groups.
A) Identification sensitivity index, d' results in four conditions from the group of patients with auditory verbal hallucinations (AVHs).
B) Results from the group of patients without auditory verbal hallucinations (non-AVHs).
C) Results from the group of healthy participants (HCs).
Error bars indicate standard errors. **significant at the level of $p < 0.01$; *significant at the level of $p < 0.05$.

In the non-AVHs group (Fig. 2B), the d' in the conditions with regular sequences showed no significant difference between the RE_{con} and RE_{inc} conditions. Whereas in the IR_{same} condition d' was significantly higher than that in the IR_{diff} condition ($t_{(1,60)} = 2.66$; $p = 0.012$). These results contrast with the results in the AVHs group, suggesting that the perceptual judgment in the non-AVH group was more influenced by the short-term repetition.

In the healthy control group (Fig. 2C), neither the difference between RE_{con} and RE_{inc} nor the difference between IR_{same} and IR_{diff} was significant, suggesting that healthy controls made perceptual judgment without influences from long-term regularity or short-term repetition.

To further explain the effects of RE and IR sequences, we carried out statistical analyses in each condition among groups. In the RE_{con} condition, one-way ANOVA showed a significant difference of d' among three groups($F(2,90) = 3.518, p = 0.034$). Further t-tests revealed that d' in the RE_{con} condition was significant different between AVHs and non-AVHs ($p = 0.009$), between AVHs and HCs was not significant ($p = 0.216$), between non-AVHs and HCs was not significant ($p = 0.173$). In the RE_{inc} condition, one-way ANOVA showed that d' was significantly different among the three groups ($F(2,90) = 13.078, p < 0.001$). Further t-tests revealed that d' in the RE_{inc} condition was significant different between AVHs and non-AVHs ($p = 0.002$), between AVHs and HCs was significant ($p < 0.001$), and between non-AVHs and HCs was not significant ($p = 0.072$). In either IR_{same} or IR_{diff} condition, one-way ANOVA did not reveal any significant d' difference among the three groups (for IR_{same}, $F(2,90) = 0.296$, $p = 0.744$; for IRdiff, $F(2,90) = 2.350$, $p = 0.101$). These results suggest that when considering all groups, RE long-term regularity effects are stronger than IR short-term repetition effects. In RE, irregularity effects are stronger than regularity effects.

3.3. *The relationship between clinical variables and behavior data of the speech tone recognition task*

As shown in Fig. 3A, a significant positive correlation was observed between AHRS total scores and d' of RE_{con} condition in the AVHs group ($r = 0.576$, $p < 0.001$). Further stepwise regression analysis identified the AHRS total scores as a significant predictor for d' in the RE_{con} condition (beta $= 0.113$, $t = 3.607$, $p = 0.001$) in AVHs. Furthermore, the

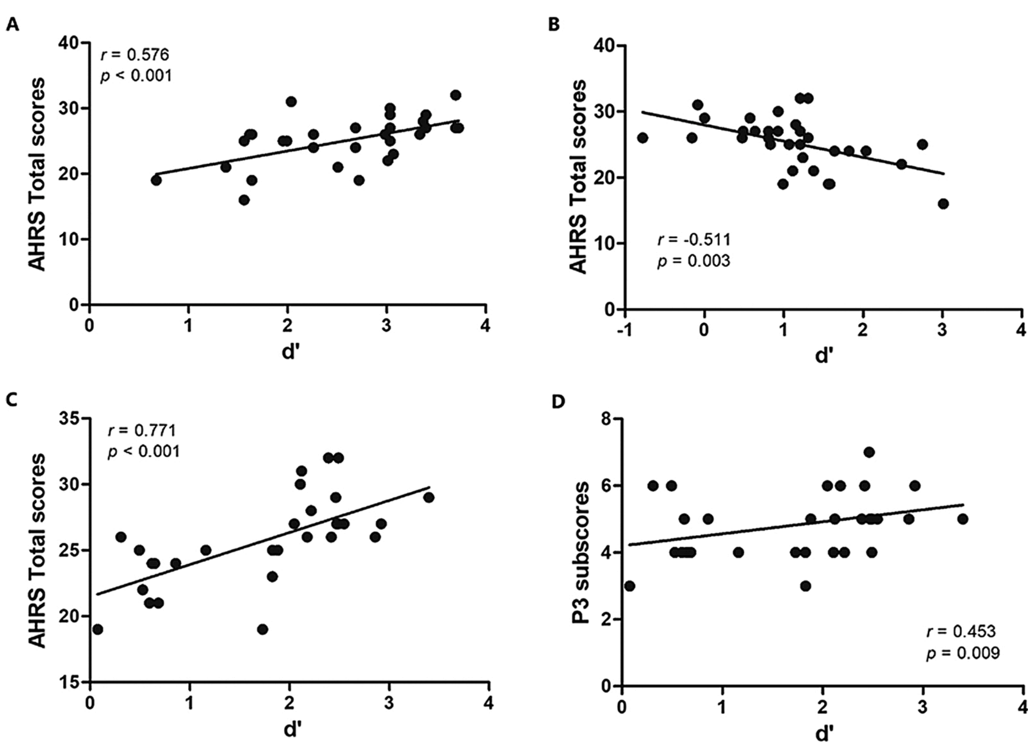

Fig. 3. Results of correlations between tone identification and indices of symptom severity.
A) The correlation between AHRS total scores and d' in the RE_{con} condition in the AVHs group.
B) The correlation between AHRS total scores and d' in the RE_{inc} condition in the AVHs group.
C) The correlation between AHRS total scores and d' difference between RE_{con} and RE_{inc} condition in the AVHs group.
D) The correlation between P3 subscores and d' difference between RE_{con} and RE_{inc} condition in the AVHs group.
AHRS, auditory hallucinations rating scale.

correlation analyses between AHRS total scores and d' in the RE_{inc} condition also revealed significant correlation (Fig. 3B, r = -0.511, p = 0.003). In contrast, other clinical variables did not correlate with d' (p > 0.05).

The difference in d' between RE_{con} and RE_{inc} was computed within the AVHs group to indicate the total influence of regularity on perceptual judgment. The following correlation analyses showed that the d' difference significantly and positively correlated with the AHRS total scores (Fig. 3C; $r = 0.771, p < 0.001$). Moreover, the PANSS P3 subscores were also significantly and positively correlated with the d' difference (Fig. 3D; r = 0.453, p = 0.009). There was no significant correlation between the d' difference and PANSS total, positive, negative and general psychopathology subscale scores (all p > 0.05). No significant correlation was found in the non-AVHs group. These results suggest that the degree of AVH severity relates to the impacts of long-term regularity on perception.

4. Discussion

We investigated the effects of top-down predictions on perceptual processing of tones in schizophrenia patients with and without AVHs. In this study, we tentatively define 'bottom up' and 'top down' in the context of the current study from an operational perspective. The 'bottom up' refers to the perceptual processes of the target tones, whereas the 'top down' refers to the processes established by the stimuli that precede the target tones (predictions established by the regularity and the repetition established by the second to last stimulus). We found that patients with AVHs identified tones better when predictions were consistent with external stimuli, whereas performance was deleterious when prediction and stimuli were inconsistent. Moreover, the modulation effects were prominent for the predictions that were derived from long-term regularities in the AVHs group. In contrast, patients without AVHs showed the effects of short-term predictions from immediate repetitions. These consistent results collaboratively revealed the featural and temporal characteristics of weighting function in source monitoring of AVHs. Impaired source monitoring in AVHs heavily weights predictions to form incorrect perception. The weighting function in source monitoring can extend to tonal features, and predictions at multiple timescales differentially modulate perception in different clinical populations.

In this study, we extended the investigation of AVHs to the basic featural level of tones. By manipulating the prediction as a function of expectancies in a trial-to-trial probabilistic fashion, we found that in the RE_{con} condition where the target tone was consistent with regularity of preceding tone sequence, patients with AVHs had higher hit rate and lower false alarm rate, so that the sensitivity indices d' became higher (Fig. 2A). Moreover, the severity of AVHs positively correlated with the modulation effects of prediction (Fig. 3A). In the RE_{inc} condition where the target tone violated the regularity, the AVHs group produced more false positives, so that the sensitivity indices d' became lower, and the severity of AVHs negatively correlated with the d' (Fig. 3B). The observations of more false positives are consistent with that verbal imagery and expectation cause more false positives of hearing speech in white noise in hallucination-prone participants [49,50]. The 'apparent' benefit of prediction in the RE_{con} condition and deleterious effect of prediction in the RE_{inc} condition are the results of confusing internal and external sources. The patients with AVHs weighted more on the internal prediction but cannot correctly perform the sensory analysis which is the relevant processing in the tone identification task. These results are consistent with our hypothesis that tonal representations induced by prediction occupy neuronal resources of the auditory cortex, making it less responsive to external stimulation. That is, AVHs may be 'parasitic' memories due to disrupted language production processes that spontaneously and erroneously activate language-based memory [25,51–53]. Moreover, these results suggest that the incorrect weighting in the source monitoring can extend to basic sound features of frequency.

The short-term repetition effects were significantly different among groups. More specifically, the non-AVHs group showed a significant difference between IR_{same} and IR_{diff} conditions, but not in the AVHs group (Fig. 2A&B), these results suggest that the monitoring of the short-term repetitions is a possible common deficit in schizophrenia patients regardless of AVH status, with that the short-term repetitions may influence perception more in the non-AVHs group.

Theoretically, in neural circuits, short-term and long-term plasticity of synaptic efficacy in sensory and motor neurons supports learning and memory [54]. Prior information can reshape synapses over different timescales by changing levels of activation, excitability and potentiation over milliseconds and minutes [55,56]. Thus, synaptic plasticity can be a neural mechanism for continuously integrating prior information into the processing of incoming information. Cognitively, the levels of information (e.g. phonemes, syllables, words, sentences) can be bases for forming predictions at multiple timescales that influence the processing of incoming information along the speech hierarchy [57]. The neural and cognitive foundations enable predictions to form in different timescales and influence perception.

The distinct modulation effects of predictions at multiple timescales suggest separate mechanisms in different clinical populations. Contrasting with the AVHs group who showed long-term prediction effects, the non-AVHs group tended to judge the target tone the same as the one immediately before. These results suggest that patients without AVHs are prone to the influences of short-term memory. Schizophrenia patients without AVHs may have a more response bias rather than perceptual sensitivity deficits. These results are consistent with the involvement of externalizing biases in schizophrenia [24,58]. Patients develop an external attribution bias to explain the confusing abnormal perceptual/cognitive experiences of psychosis. This process may be related to the conscious evaluation of the external stimuli, which is presumably caused by a lack of effective connectivity between the superior temporal gyrus (STG) and anterior cingulate cortex (ACC) – a critical component of the 'core control network'. Furthermore, the deficits in core control may link to auxiliary sensory systems and collaboratively cause the symptoms of hallucinations. For example, previous studies found that the STG volume reduced but with a hyper-activation in schizophrenia patients [59], suggesting that the auditory cortices may also be important in the process of labeling the sources and relates to the likelihood of externalizing bias in schizophrenia.

Moreover, predictive function and prediction errors that presumably involve and are generated in frontal cortices correlate with many of the features of clinical symptoms in schizophrenia. For example, previous studies found a complex relationship between clinical symptoms and predictive function in schizophrenia in an associative learning task [60]. The results showed that schizophrenic patients, compared to healthy controls, decreased activity in the right prefrontal cortex (rPFC) when predictions were violated whereas increased rPFC activation to predictable outcomes. The neural activity differences in the rPFC to unpredicted and predicted events were significantly correlated with delusional scores. This frontal-temporal neural network might be the neural mechanisms that mediate the behavioral observations in the current study. The predictions induced by the long-term regularity and short-term repetition and the monitoring signals could originate in the frontal cortices and transfer to auditory neural representations that interfere with the auditory perception. This neural hypothesis can be tested in future neuroimaging studies.

Predictions can efficiently balance the cognitive resources for processing the expected events and detecting the unexpected ones [29]. However, AVHs may be a 'by-product' of this computational advantage of prediction. The observed specific impairment in AVHs could be due to prediction-generated tone interfering with the performance on the speech tone recognition task, similar to the 'cocktail party' effects where multiple sound and speech streams compete with each other [61]. Our results are consistent with these hypotheses and suggest that prediction from long-term regularity influences perception. An interpretation of this prediction-perception interference from a neuroscience perspective is that tonal representations induced by prediction occupy neuronal resources of the auditory cortex, making it less responsive to external stimulation [62]. However, the cognitive mechanism behind this interpretation, especially the developmental dynamics of the symptoms, remains elusive. It might be explained as interference between the temporal ordering of items in memory [63]. In support, some patients reported that their experiences of hearing voices were first as intrusive and unwanted, and gradually developed from hearing their own thoughts to finally hearing verbal contents as if spoken by a third party [64]. This developmental trajectory of AVHs might arise because of incomplete encoding of memories, the vulnerability of incorrect priming, or abnormal storage that leads to weak contextual harnessing [51].

Previous studies have demonstrated that mismatch negativity (MMN) is abnormal in people who suffer from schizophrenia [65]. An alternative explanation to the observations in the current study is that AVHs may have intact ability to detect congruence but deficits in detecting incongruence. However, this alternative hypothesis is less likely because of the paradigm differences as well as the specific observed patterns in the behavioral results. Methodologically, the regularity in the MMN paradigm is usually established by repetitively presenting the standard stimuli. Moreover, MMNs are (mostly) observed in a passive paradigm (no task or irrelevant to the deviant features) that presumably involves pre-attentive 'automatic' processes. Whereas in the current paradigm the regularity is alternating tones. Moreover, participants are required to actively identify the target tones that were randomly presented in noise. These paradigm differences constrain the underlying processes and make them unlikely similar between the current experiment and MMN experiments. That is, the current study more likely involves processes that integrate top-down prediction and bottom-up perception, whereas MMN is more likely due to the release from adaptation (repetitions of standard stimuli) or violation between prediction and perception.

More important is that, empirically, the results are not consistent with the alternative hypothesis that patients with AVHs may have intact ability to detect congruence but deficits in detecting incongruence. The d' in the RE_{con} condition is significantly better in the AVHs group than that in the non-AVHs group, as well as RE_{con} significantly correlates with AHRS. This 'erroneous' boost of tone identification performance in the RE_{con} condition in AVHs supports that patients with AVHs confused prediction and perception, but the alternative hypothesis of intact ability to detect congruence is hard to explain these results.

5. Conclusion

The study revealed that schizophrenia patients with AVHs weighted predictions over sensory processing and altered the recognition of tones. Moreover, patients with and without AVHs showed distinct influences of predictions at different timescales. The impaired interaction between top-down and bottom-up processes might underlie the development of AVHs. Our results support a Bayesian cognitive account that impaired source monitoring mediates AVHs.

Funding

This study was supported by the National Natural Science Foundation of China32071099, Natural Science Foundation of Shanghai20ZR1472100, Program of Introducing Talents of Discipline to Universities, Base B16018 to XT, and East China Normal University (ECNU) Academic Innovation Promotion Program for Excellent Doctoral StudentsYBNLTS2019-026 to FY.

Author contributions

FY designed the experiment, collected and analyzed data, and drafted and edited the manuscript. HZ edited the experiment Python script and the methods. CZ, LY and WL, as psychiatrists, helped recruit schizophrenia patients. XT designed the experiment, interpreted the data, edited the manuscript, and provided critical revisions. All authors approved the final version.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] M. Stephane, S. Barton, N. Boutros, Auditory verbal hallucinations and dysfunction of the neural substrates of speech, Schizophr. Res. 50 (1-2) (2001) 61–78.
[2] H.Y. Meltzer, Treatment of the neuroleptic-nonresponsive schizophrenic patient, Schizophr. Bull. 18 (3) (1992) 515–542.
[3] F. Waters, Multidisciplinary approaches to understanding auditory hallucinations in schizophrenia and nonschizophrenia populations: the International Consortium on Hallucination Research, Schizophr. Bull. 38 (4) (2012) 693–694.
[4] I.E.C. Sommer, C.W. Slotema, Z.J. Daskalakis, E.M. Derks, J.D. Blom, M. van der Gaag, The treatment of hallucinations in schizophrenia spectrum disorders, Schizophr. Bull. 38 (4) (2012) 704–714.
[5] L. Farde, Brain imaging of schizophrenia—the dopamine hypothesis, Schizophr. Res. 28 (2-3) (1997) 157–162.
[6] S. Leucht, A. Cipriani, L. Spineli, D. Mavridis, D. Örey, F. Richter, M. Samara, C. Barbui, R.R. Engel, J.R. Geddes, W. Kissling, M.P. Stapf, B. Lässig, G. Salanti, J. M. Davis, Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis, Lancet 382 (9896) (2013) 951–962.
[7] S.S. Shergill, R.M. Murray, P.K. McGuire, Auditory hallucinations: a review of psychological treatments, Schizophr. Res. 32 (3) (1998) 137–150.
[8] R.E. Hoffman, N.N. Boutros, S. Hu, R.M. Berman, J.H. Krystal, D.S. Charney, Transcranial magnetic stimulation and auditory hallucinations in schizophrenia, Lancet 355 (9209) (2000) 1073–1075.
[9] M. Mondino, R. Jardri, M.F. Suaud-Chagny, M. Saoud, E. Poulet, J. Brunelin, Effects of fronto-temporal transcranial direct current stimulation on auditory verbal hallucinations and resting-state functional connectivity of the left temporo-parietal junction in patients with schizophrenia, Schizophr. Bull. 42 (2) (2015) 318–326.
[10] A. Priori, M. Hallett, J.C. Rothwell, Repetitive transcranial magnetic stimulation or transcranial direct current stimulation? Brain Stimul. 2 (4) (2009) 241–245.
[11] J. Brunelin, M. Mondino, L. Gassab, F. Haesebaert, L. Gaha, M.-F. Suaud-Chagny, M. Saoud, A. Mechri, E. Poulet, Examining transcranial direct-current stimulation (tDCS) as a treatment for hallucinations in schizophrenia, Am. J. Psychiatry 169 (7) (2012) 719–724.
[12] F. Yang, X. Fang, W. Tang, L. Hui, Y. Chen, C. Zhang, X. Tian, Effects and potential mechanisms of transcranial direct current stimulation (tDCS) on auditory hallucinations: a meta-analysis, Psychiatry Res. 273 (2019) 343–349.
[13] A.R. Brunoni, J. Amadera, B. Berbel, M.S. Volz, B.G. Rizzerio, F. Fregni, A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation, Int. J. Neuropsychopharmacol. 14 (8) (2011) 1133–1145.
[14] J. Kindler, P. Homan, K. Jann, A. Federspiel, R. Flury, M. Hauf, W. Strik, T. Dierks, D. Hubl, Reduced neuronal activity in language-related regions after transcranial magnetic stimulation therapy for auditory verbal hallucinations, Biol. Psychiatry 73 (6) (2013) 518–524.
[15] C.D. Frith, D.J. Done, Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action, Psychol. Med. 19 (2) (1989) 359–363.
[16] P. Moseley, C. Fernyhough, A. Ellison, Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option, Neurosci. Biobehav. Rev. 37 (10, Part 2) (2013) 2794–2805.
[17] C. Teufel, P.C. Fletcher, Forms of prediction in the nervous system, Nat. Rev. Neurosci. 21 (4) (2020) 231–242.
[18] Daniel L. Schacter, Donna R. Addis, D. Hassabis, Victoria C. Martin, R.N. Spreng, Karl K. Szpunar, The future of memory: remembering, imagining, and the brain, Neuron 76 (4) (2012) 677–694.
[19] S.T. Moulton, S.M. Kosslyn, Imagining predictions: mental imagery as mental emulation, Philos. Trans. Biol. Sci. 364 (1521) (2009) 1273–1280.
[20] K.K. Szpunar, R.N. Spreng, D.L. Schacter, A taxonomy of prospection: introducing an organizational framework for future-oriented cognition, Proc. Natl. Acad. Sci. 111 (52) (2014) 18414.
[21] M.E. Wheeler, S.E. Petersen, R.L. Buckner, Memory's echo: vivid remembering reactivates sensory-specific cortex, Proc. Natl. Acad. Sci. 97 (20) (2000) 11125.
[22] X. Tian, J.M. Zarate, D. Poeppel, Mental imagery of speech implicates two mechanisms of perceptual reactivation, Cortex 77 (2016) 1–12.
[23] X. Tian, D. Poeppel, Dynamics of self-monitoring and error detection in speech production: evidence from mental imagery and MEG, J. Cogn. Neurosci. 27 (2) (2014) 352–364.
[24] R.P. Bentall, P.D. Slade, Reality testing and auditory hallucinations: a signal detection analysis, Br. J. Clin. Psychol. 24 (3) (1985) 159–169.
[25] O. Ma, X. Tian, Distinct mechanisms of imagery differentially influence speech perception, eNeuro 6 (5) (2019). ENEURO.0261-0219.2019.
[26] P.R. Corlett, G. Horga, P.C. Fletcher, B. Alderson-Day, K. Schmack, A.R. Powers, Hallucinations and strong priors, Trends Cogn. Sci. 23 (2) (2019) 114–127.

[27] X. Tian, D.E. Huber, Playing "duck duck goose" with neurons: change detection through connectivity reduction, Psychol. Sci. 24 (6) (2013) 819–827.
[28] R.S. Keefe, M.S. Kraus, Measuring memory-prediction errors and their consequences in youth at risk for schizophrenia, J. Ann. Acad. Med. Singapore 38 (5) (2009) 414–416.
[29] A. Aleman, K.B.E. Böcker, R. Hijman, E.H.F. de Haan, R.S. Kahn, Cognitive basis of hallucinations in schizophrenia: role of top-down information processing, Schizophr. Res. 64 (2) (2003) 175–185.
[30] C.M. Cassidy, P.D. Balsam, J.J. Weinstein, R.J. Rosengard, M. Slifstein, N.D. Daw, A. Abi-Dargham, G. Horga, A perceptual inference mechanism for hallucinations linked to striatal dopamine, Curr. Biol. 28 (4) (2018) 503–514, e504.
[31] R.-P. Behrendt, Underconstrained perception: a theoretical approach to the nature and function of verbal hallucinations, Compr. Psychiatry 39 (4) (1998) 236–248.
[32] D. Dima, D.E. Dietrich, W. Dillo, H.M. Emrich, Impaired top-down processes in schizophrenia: a DCM study of ERPs, Neuroimage 52 (3) (2010) 824–832.
[33] A.R. Powers, C. Mathys, P.R. Corlett, Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors, Science 357 (6351) (2017) 596.
[34] M.H. Davis, I.S. Johnsrude, Hearing speech sounds: top-down influences on the interface between audition and speech perception, Hear. Res. 229 (1) (2007) 132–147.
[35] X. Tian, N. Ding, X. Teng, F. Bai, D. Poeppel, Imagined speech influences perceived loudness of sound, Nat. Hum. Behav. 2 (3) (2018) 225–234.
[36] A. Vercammen, E.H.F. de Haan, A. Aleman, Hearing a voice in the noise: auditory hallucinations and speech perception, Psychol. Med. 38 (8) (2008) 1177–1184.
[37] Jn.M. Fuster, Network memory, Trends Neurosci. 20 (10) (1997) 451–459.
[38] A.-L. Giraud, D. Poeppel, Cortical oscillations and speech processing: emerging computational principles and operations, Nat. Neurosci. 15 (4) (2012) 511–517.
[39] X. Teng, X. Tian, J. Rowland, D. Poeppel, Concurrent temporal channels for auditory processing: oscillatory neural entrainment reveals segregation of function at different scales, PLoS Biol. 15 (11) (2017) e2000812.
[40] G. Hickok, D. Poeppel, The cortical organization of speech processing, Nat. Rev. Neurosci. 8 (5) (2007) 393–402.
[41] X. Teng, M. Ma, J. Yang, S. Blohm, Q. Cai, X. Tian, Constrained structure of ancient Chinese poetry facilitates speech content grouping, Curr. Biol. 30 (7) (2020) 1299–1305, e1297.
[42] N. Ding, L. Melloni, H. Zhang, X. Tian, D. Poeppel, Cortical tracking of hierarchical linguistic structures in connected speech, Nat. Neurosci. 19 (1) (2016) 158–164.
[43] U. Hasson, J. Chen, C.J. Honey, Hierarchical process memory: memory as an integral component of information processing, Trends Cogn. Sci. 19 (6) (2015) 304–313.
[44] R.E. Hoffman, K.A. Hawkins, R. Gueorguieva, N.N. Boutros, F. Rachid, K. Carroll, J. H. Krystal, Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations, Arch. Gen. Psychiatry 60 (1) (2003) 49–56.
[45] NeoSpeech, NeoSpeech Text-to-Speech Voices [computer Software]. Santa Clara, CA, 2012.
[46] P. Boersma, Praat, a system for doing phonetics by computer, Glottodidactica Int. J. Appl. Linguist. 5 (9) (2002) 341–345.
[47] L.L. Kontsevich, C.W. Tyler, Bayesian adaptive estimation of psychometric slope and threshold, Vision Res. 39 (16) (1999) 2729–2737.
[48] N.A. Macmillan, C.D. Creelman, Detection Theory: A User's Guide, Psychology press, New York, 2004.
[49] A. Vercammen, A. Aleman, Semantic expectations can induce false perceptions in hallucination-prone individuals, Schizophr. Bull. 36 (1) (2008) 151–156.
[50] P. Moseley, D. Smailes, A. Ellison, C. Fernyhough, The effect of auditory verbal imagery on signal detection in hallucination-prone individuals, Cognition 146 (2016) 206–216.
[51] R.E. Hoffman, Verbal hallucinations and language production processes in schizophrenia, Behav. Brain Sci. 9 (3) (1986) 503–517.
[52] X. Tian, D. Poeppel, Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation, Front. Hum. Neurosci. 6 (2012) 314.
[53] S. Li, H. Zhu, X. Tian, Distinct neural signals in speech preparation differentially modulate auditory responses, bioRxiv (2020).
[54] U. Hasson, E. Yang, I. Vallines, D.J. Heeger, N. Rubin, A hierarchy of temporal receptive windows in human cortex, J. Neurosci. 28 (10) (2008) 2539.
[55] G. Mongillo, O. Barak, M. Tsodyks, Synaptic theory of working memory, Science 319 (5869) (2008) 1543.
[56] D.I. Perrett, D. Xiao, N.E. Barraclough, C. Keysers, M.W. Oram, Seeing the future: natural image sequences produce "anticipatory" neuronal activity and bias perceptual report, Q. J. Exp. Psychol. 62 (11) (2009) 2081–2104.
[57] U. Hasson, R. Malach, D.J. Heeger, Reliability of cortical activity during natural stimulation, Trends Cogn. Sci. 14 (1) (2010) 40–48.
[58] F. Varese, F. Smeets, M. Drukker, R. Lieverse, T. Lataster, W. Viechtbauer, J. Read, J. van Os, R.P. Bentall, Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies, Schizophr. Bull. 38 (4) (2012) 661–671.
[59] D.H. Mathalon, E.V. Sullivan, K.O. Lim, A. Pfefferbaum, Progressive brain volume changes and the clinical course of schizophrenia in men: a longitudinal magnetic resonance imaging study, Arch. Gen. Psychiatry 58 (2) (2001) 148–157.
[60] P.R. Corlett, G.K. Murray, G.D. Honey, M.R.F. Aitken, D.R. Shanks, T.W. Robbins, E.T. Bullmore, A. Dickinson, P.C. Fletcher, Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions, Brain 130 (9) (2007) 2387–2400.
[61] S.R. Jones, C. Fernyhough, Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations, Conscious. Cogn. 16 (2) (2007) 391–399.
[62] K. Kompus, L. Falkenberg, J. Bless, E. Johnsen, R.A. Kroken, B. Kråkvik, F. Larøi, E.-M. Løberg, E. Vedul-Kjelsås, R. Westerhausen, K. Hugdahl, The role of the primary auditory cortex in the neural mechanism of auditory verbal hallucinations, Front. Hum. Neurosci. 7 (2013) 144.
[63] G.E. Bruder, D.M. Alschuler, C.J. Kroppmann, S. Fekri, R.B. Gil, L.F. Jarskog, J. M. Harkavy-Friedman, R. Goetz, J. Kayser, B.E. Wexler, Heterogeneity of auditory verbal working memory in schizophrenia, J. Abnorm. Psychol. 120 (1) (2011) 88–97.
[64] P.K. McGuire, A.S. David, R.M. Murray, R.S.J. Frackowiak, C.D. Frith, I. Wright, D. A. Silbersweig, Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations, Lancet 346 (8975) (1995) 596–600.
[65] P.T. Michie, M.S. Malmierca, L. Harms, J. Todd, The neurobiology of MMN and implications for schizophrenia, Biol. Psychol. 116 (2016) 90–97.