Multi-scale predictions distinctively modulate tone perception in schizophrenia patients with auditory verbal hallucinations

Fuyin Yang[1,2], Hao Zhu[2,3], Lingfang Yu[4], Weihong Lu[4], Chen Zhang[4,**], Xing Tian[2,3,*]

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

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

3 Division of Arts and Sciences, New York University Shanghai, 1555 Century Avenue, Shanghai, 200122, China

4 Schizophrenia Program, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China

*Corresponding author: Xing Tian; New York University Shanghai, 1555 Century Avenue, Shanghai, 200122, China. Tel: (+86-21) 20595201. E-mail address: xing.tian@nyu.edu

**Corresponding author: Chen Zhang; Schizophrenia Program, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China. Tel: (+86-21) 64387250. E-mail address: zhangchen645@gmail.com

Multi-scale predictions distinctively modulate tone perception in schizophrenia patients with auditory verbal hallucinations

Fuyin Yang[1,2], Hao Zhu[2,3], Lingfang Yu[4], Weihong Lu[4], Chen Zhang[4,**], Xing Tian[2,3,*]

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

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

3 Division of Arts and Sciences, New York University Shanghai, 1555 Century Avenue, Shanghai, 200122, China

4 Schizophrenia Program, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China

Highlights

1. Malfunctioned source monitoring incorrectly weights the top-down prediction and bottom-up sensory processing underlie pathogenesis of auditory verbal hallucinations in schizophrenia.

2. The weighting function in top-down predictions and bottom-up sensory processing can extend to tonal features.

3. Predictions at multiple timescales differentially modulate perception in different clinical schizophrenia populations.

Abstracts

Auditory verbal hallucinations (AVHs) are one of the most pronounced symptoms that manifest the underlying mechanisms of deficits in schizophrenia. Cognitive models postulate that malfunctioned source monitoring incorrectly weights the top-down prediction and bottom-up sensory processing and causes hallucinations. Here, we investigate the featural-temporal characteristics of source monitoring in AVHs. Schizophrenia patients with and without AVHs, and 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, or a repetition between the target tone and the tone immediately before. The sensitivity index, d', was obtained to assess the modulation of predictions on tone identification. We found that 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 that when the targets were inconsistent with the pattern. Whereas, the short-term regularity of repetitions modulated the tone identification in patients without AVHs. Predictions did not influence tone identification in healthy controls. These findings suggest that malfunctioned source monitoring in AVHs heavily weights predictions to form incorrect perception. The weighting function in source monitoring can extend to the process of basic tonal features, and predictions at multiple timescales differentially modulate perception in different clinical populations. These collaboratively reveal the featural and temporal characteristics of weighting function in source monitoring of AVHs and suggest that the malfunctioned interaction between top-down and bottom-up processes might underlie the development of auditory hallucinations.

Introduction

Auditory verbal hallucinations (AVHs) are symptoms of hearing voices in the absence of external stimuli (Stephane *et al.*, 2001). Around 50%-70% of people who are diagnosed with schizophrenia experience AVHs (Meltzer, 1992, Waters, 2012). Antipsychotic medication such as olanzapine, risperidone and quetiapine that block D2-receptors in the brain are the most effective drugs available to treat AVHs (Farde, 1997, Sommer *et al.*, 2012). However, weight gain and sedation are serious side effects associated with antipsychotic medications (Leucht *et al.*, 2013). Medications may expose patients to metabolic complications and result in treatment non-adherence. Moreover, about 25% patients are resistant to standard antipsychotic treatment (Shergill *et al.*, 1998). Noninvasive neuro-stimulation techniques have been tested as a new treatment option for AVHs (Hoffman *et al.*, 2000). Repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) are two noninvasive techniques that are recently introduced to treat AVHs (Mondino *et al.*, 2015, Priori *et al.*, 2009). The rTMS and tDCS, as promising treatment options (Brunelin *et al.*, 2012, Sommer *et al.*, 2012), show a moderate effect size in reduction of AVHs frequency (Brunoni *et al.*, 2011, Yang *et al.*, 2019). That the treatment effects depend on the stimulation protocols and cortical targets (Kindler *et al.*, 2013, Yang *et al.*, 2019) 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 (Frith and Done, 1989). Such source monitoring account of AVHs requires an internally generated source. Prediction has been proposed as an algorithm that induce this internal source (Moseley *et*

al., 2013). Prediction refers to the set of processes that are based on the information of memory, knowledge, and belief to generate representations of future events (Moulton and Kosslyn, 2009, Schacter *et al.*, 2012, Szpunar *et al.*, 2014, Teufel and Fletcher, 2020). For example, similar representations as perception can be constructed based on retrieving memory of past experiences and regularities without external stimulations (Tian and Poeppel, 2014, Tian *et al.*, 2016, Wheeler *et al.*, 2000). These internally constructed representations from prediction could be the critical source in the cognitive monitoring account of AVHs.

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 (Bentall and Slade, 1985, Corlett *et al.*, 2019, Ma and Tian, 2019). Predictions can economically balancing the cognitive resources for processing the adapted perception and detecting the unexpected events (Tian and Huber, 2013). Breakdown in predictive function means individuals are less likely to attend to effective indicators of upcoming sensation (Keefe and Kraus, 2009). The unbalanced weighting between sources from prediction and sensory input could cause AVHs (Aleman *et al.*, 2003). For example, participants with severe hallucinations significantly increase gain over prior predictions in ambiguous perceptual situations (Cassidy *et al.*, 2018), suggesting that a relatively higher priority is assigned to top-down factors in determining the final percepts (Behrendt, 1998). To an extreme, abnormal top-down prediction processes in patients overwhelm the auditory input (Aleman *et al.*, 2003, Dima *et al.*, 2010). The malfunctioned weighting in the

source monitoring is consistent with the framework of Bayesian inferences, share the assumption that AVHs are induced when sensory predictions are activated without sensory input, or these predictions are not properly deactivated and incorrect replace the sensory analysis (Powers *et al.*, 2017).

What perceptual features and temporal characteristics the weighting function are operating on are two crucial aspects to understand the source monitory of AVHs. 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 (Davis and Johnsrude, 2007). Recently, the effects of top-down prediction on sensory analysis have extended to lower and basic sound attributes, such as pitch and loudness (Tian *et al.*, 2018, Tian and Poeppel, 2014). However, whether the weighting function in source monitoring of AVHs can extend to basic level attributes is still in debate. Some studies found that the severity of hallucination-prone was correlated with errors that were induced by semantic priming but not with phonological priming (Vercammen *et al.*, 2008). Other studies provided preliminary evidence of increased top-down influences for tonal stimuli (Aleman *et al.*, 2003). In this study, we investigated whether AVHs 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 have a ubiquitous weighting function that applies on sound attributes of all levels.

Second, prior information is available at multiple timescales and facilitate information processing across time (Fuster, 1997). For example, speech processing may operate at two distinct timescales (Giraud and Poeppel, 2012, Teng *et al.*, 2017). Multiple

levels of prior information could help comprehension of linguistic information ranging from phonemes, to words, to sentences, and to paragraphs (Ding *et al.*, 2016, Hickok and Poeppel, 2007, Teng *et al.*, 2020). 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 (Hasson *et al.*, 2015). Would the source monitoring of AVHs weight predictions at multiple timescales differently?

Together, this study investigated the featural and temporal characteristics of the weighting function in source monitoring of AVHs. Specifically, we hypothesized 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 immediate preceding one (immediate repetitions). Three groups of participants, 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 may 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.

Methods

Participants

Thirty-two (14 males) patients, who matched a DSM-V diagnosis of schizophrenia and were currently experiencing AVHs (AVHs patients) without concomitant hallucinations in other modalities, were recruited from Shanghai Mental Health Center. Furthermore, twenty-nine (12 males) patients met DSM-V diagnosis of schizophrenia who had never experienced AVHs (non-AVHs patients) were recruited from the same hospital. Two experienced psychiatrist 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 Chinese subjects (9 males) were recruited as the control group from the local communities and schools in Shanghai. 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. All participants provided signed, informed consent before their participation in the study.

Clinical measures

Demographic data were collected from patients and healthy controls. Four psychiatrists, who were blind to the study, assessed the patient's psychopathology 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, AVHs level of current AVHs patients were assessed using the 7-item Auditory Hallucinations Rating Scale-AHRS (Hoffman *et al.*, 2003). 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 rating had a correlation coefficient >0.8 on the PANSS and AHRS total scores.

Experimental design

Materials

Two different Mandarin tones of vowels /a/ (/ā/ and /á/) were synthesized via the NeoSpeech engine (NeoSpeech, 2012) with a female voice. Both Mandarin tones were 377ms in duration and scaled to 75dB SPL in intensity using Praat software.(Boersma,

2002) Additional two stimuli were created by adding white noise to the two mandarin tones. The signal-to-noise ratio was determined at individual level during a pre-test. All stimuli were digitized at 44.1kHz sampling rate and 16bit 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 experiment for all stimuli.

Threshold Test Procedures

This procedure was composed of two different steps. First, participants participated in a pre-test that measured a threshold for the detection of two Mandarin tones (/ā/ or /á/) in white noise. This pre-test session consisted of 200 trials. At the beginning of the trial, a visual cue was presented for 500ms. After the offset of the visual cue, one of the auditory stimuli in noise was presented. White noise was 1s 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 (Kontsevich and Tyler, 1999) while the intensity of the vowels fixed at 75dB. 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-staircases 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 farced-choice (2AFC) paradigm. Participants took a break of a few minutes after every 50 trials. The threshold for each tone was determined by the 75% accuracy point in the fitted the 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.

$$SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right).$$

Last, we ran a confirmation test. Participants judged twenty trials (ten for each of tone) with the SNR determined in the pre-test. When the number of correctly judgments was seven or eight in each of tone, Participants were considered pass the test. Only participants who passed the confirmation test could proceed to the next procedure.

Main Procedure

After determining an appropriate threshold of each tone, participants proceeded to the main experiment in which they heard a sequence of tones and make judgement to the last tone in noise.

At the beginning of a trial, a fixation appeared in the center of the screen for 500 msec. 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 377ms. The inter-stimulus interval (ISI) was 623ms. Therefore, the stimulus onset asynchrony (SOA) was 1s. Trails with different number of tones were randomly presented. 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 1s-long white noise and lasted for 377ms. Participants judged whether the tone in the noise was /ā/ or /á/ by pressing one of two buttons (Figure 1). The probability of each tone in the first clean tone position, in the last clean tone position, and in the noise was same.
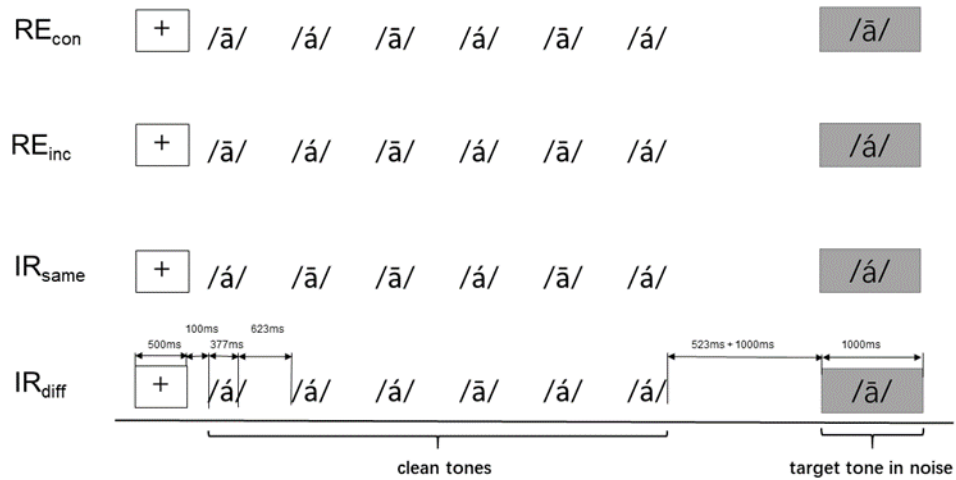
Figure 1. Schematic description of experimental procedures. At the beginning of a trial, a fixation appeared in the center of the screen for 500 msec. After the offset of the visual cue, participants passively heard four to seven clean /ā/ or /á/ Mandarin tones in a sequence. The duration of each tone was 377ms. The inter-stimulus interval (ISI) was 623ms. Trails with 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 377ms. 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.

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 an order. The sequence could be a regular pattern (RE) in which two tones are presented in an alternating manner. Or the sequence could be constructed by randomly presenting the two tones (IR). The second parameter was whether the tone in

noise was consistent with 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 tone of test was consistent with the long-term prediction formed by the 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 pseudorandom across all participants.

Statistical analysis

When computing the measures to quantify responses, we took the tone /á/ as the target tone. 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') (Macmillan and Creelman, 2004).

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. One-way analysis of variance (ANOVA) was performed with factors of group for demographic and clinical continuous variables and chi-squares (χ^2) test for categorical values. A univariate analysis of covariance (ANCOVA) was used to assess the performance of the different subject groups in four conditions. Pearson correlation

analyses were performed to determine the relationship between clinical variables and behavioral data within AVHs patients. We used stepwise multiple regression analysis

with d' scores 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.

Results

1.Demographic data

Table 1 shows the subjects' demographic data and the clinical variables. ANOVA analyses showed a significant difference in age ($p < 0.001$) among three groups, but not in education, height, and weight ($p > 0.05$). The χ^2 test showed no significant differences among three groups with regard to gender ($\chi^2 = 0.68$, $p = 0.508$). Further, P3 subscore, positive and general psychopathology subscores were significantly higher in AVHs patients than non-AVHs patients (all $p < 0.01$). Neither the age of onset, duration or the PANSS total score was significantly different between two patients' groups.

Table1 Demographics of Schizophrenia patients with and without AVHs and Healthy Controls

	AVHs Patients (n=32)	Non-AVHs Patients (n=29)	Healthy Controls (n=30)	F or χ^2	p-value
Gender (M/F)	14/18	12/17	9/21	0.68	0.508
Age (years)	27.28 (6.37)	25.79 (5.94)	21.67 (2.45)	9.41	<0.001
Education (years)	12.44 (2.47)	12.83 (2.65)	13.43 (0.77)	1.69	0.19
Height (m)	1.68 (8.31)	1.66 (9.08)	1.67 (7.07)	0.39	0.677
Weight(kg)	66.78 (15.43)	62.38 (7.10)	60.37 (8.93)	2.66	0.076
Age of onset (years)	21.09 (6.66)	19.69 (4.80)	N/A	0.82	0.368
Duration (month)	74.05 (77.64)	63.38 (52.53)	N/A	0.39	0.536
PANSS total score	82.03 (8.87)	78.55 (4.41)	N/A	1.97	0.06
P3 subscore	4.75 (0.98)	1.00 (0.00)	N/A	21.56	<0.001
P subscore	23.53 (3.82)	14.69 (2.65)	N/A	10.40	<0.001
N subscore	19.78 (3.48)	22.41 (1.43)	N/A	3.93	0.001
G subscore	38.72 (4.09)	41.45 (2.56)	N/A	3.09	0.003
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.

2.Performance on the speech tone recognition task

The sensitivity indices d' scores were first subject to a two-way mixed ANOVA with groups (AVHs, non-AVHs, and Healthy Control) as a between-subject factor and conditions (RE_{con}, RE_{inc}, IR_{same}, and IR_{diff}) as a within-subject factor. The omnibus results showed that the main effect of the groups was significant ($F_{(2,352)} = 5.089$, $p = 0.007$). The main effect of conditions was also significant ($F_{(3,352)} = 8.953$, $p = 0.001$). More importantly, the interaction was significant ($F_{(6,352)} = 5.315$, $p = 0.001$).

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 populations. Sensitivity indices d' were calculated for the four conditions in the AVHs group (Figure 2A), non-AVHs group (Figure 2B) and healthy control group (Figure 2C). In the AVHs group, as shown in Figure 2A, the d' scores in the RE_{con} condition was significantly higher than that in the RE_{inc} condition $(t_{(1,62)} = 7.45; p < 0.001)$. However, the d' was not different between the IR_{same} and IR_{diff} conditions, suggesting the tone recognition was not influenced by the immediate memory effect. This observation was consistent with our hypothesis suggesting that the long-term prediction biases perception in patients with AVHs.

In non-AVHs patients, the d' scores in the conditions with regular sequences, no significant difference was found between the Re_{con} and RE_{inc} conditions. Whereas in the IR_{same} condition was significantly higher than that in the IR_{diff} condition $(t_{(1,56)} = 2.62; p = 0.014)$. These results directly contrast with the results in the AVHs group, suggesting that the perceptual judgment in the non-AVH group was more influenced by the immediate memory effect.

In the healthy control group, 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 predictions or short-term memory effect.
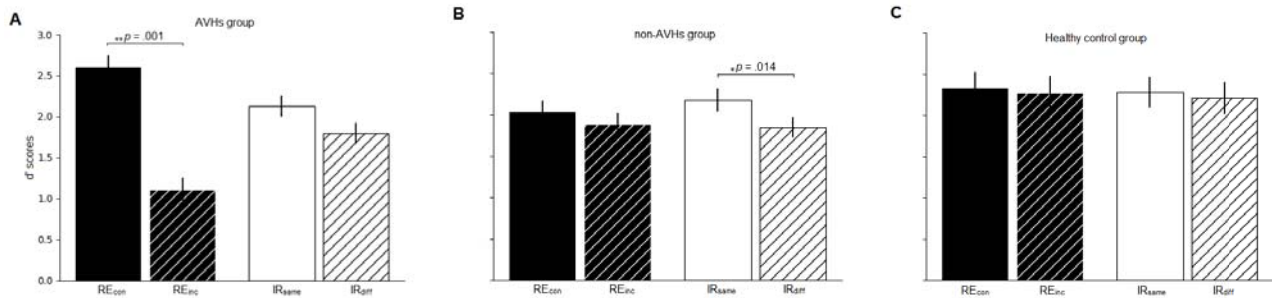
Figure 2. Tone identification results of three groups. A) Identification sensitivity index, d' results in four conditions from the group of patients with auditory verbal hallucinations. B) Results from the group of patients without auditory verbal hallucinations. C) Results from the goup of health participants. Error bars indicate standard errors. **significant at the level of p < 0.01; *significant at the level of p < 0.05.

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

As shown in Figure 3A, a significantly positive correlation was found between AHRS total scores and d' scores of RE_{con} condition in AVHs patients group ($r = 0.576$, $p < 0.001$). Further stepwise multiple regression analysis identified the AHRS total scores as a significant predictor for d' scores in the RE_{con} condition (beta = 0.113, $t = 3.607$, $p = 0.001$) in AVHs patients. Whereas the other variables showed no effects ($p > 0.05$). The difference in d' scores between RE_{con} and RE_{inc} was computed within AVHs groups to indicate the total influence of regularity on perceptual judgment. The following correlation analyses showed a significantly positive correlation between the d' difference and the AHRS total scores (Figure 3B; $r = 0.771$, $p < 0.001$). Moreover, the PANSS P3 subscores were also significantly positively correlated with the d' difference (Figure 3C; $r = 0.453$, $p = 0.009$). 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 non-AVHs patients. These results suggest that the degree of severity in AVH relates to the impacts of memory on perception.
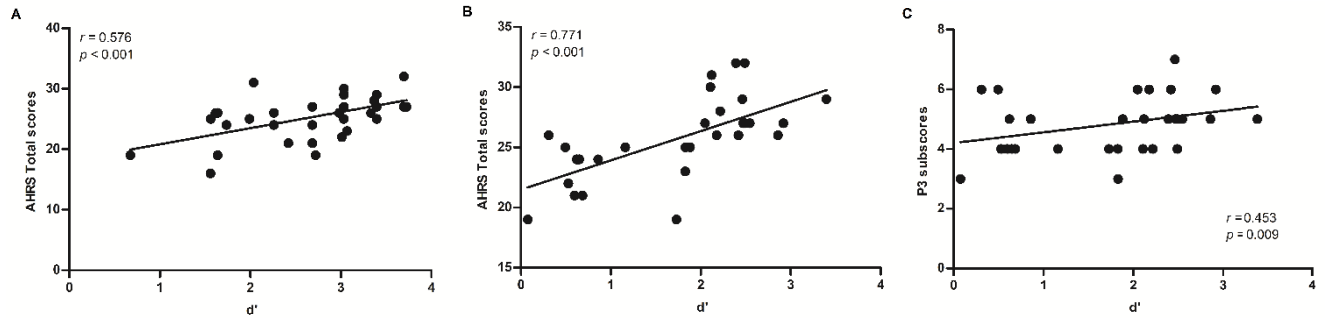

Figure 3. Results of correlations between tone identification and indices of symptom severity. A) The correlation between AHRS total scores and d' in the RE_{con} condition in AVHs patients. B) The correlation between AHRS total scores and d' difference between RE_{con} and RE_{inc} condition in AVHs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

Discussion

We investigated the effects of top-down predictions on perceptual processing of tones in schizophrenia patients with and without AVHs. We found that AVHs patients identified tones better when predictions met 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 AVHs patients. 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. Malfunctioned 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 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, AVHs patients had higher hit rate and lower false alarm rate, so that the sensitivity indices d' became higher (Figure 2A). Moreover, the severity of AVHs positively correlated with the modulation effects of prediction (Figure 3A). In the RE_{inc} condition where the target tone violated the regularity, the AVHs patients produced more false positives, so that the sensitivity indices d' became lower. 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 (Moseley *et al.*, 2016, Vercammen and Aleman, 2008). The 'apparent' benefit of prediction in the RE_{con} condition and deleterious effect of prediction in the RE_{inc} condition are actually the results of confusing internal and external sources. The AVHs patients 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 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 (Hoffman, 1986, Li *et al.*, 2020, Ma and

Tian, 2019, Tian and Poeppel, 2012).Moreover, these results suggest that the incorrect weighting in the source monitoring can extend to basic sound features of frequency.

Interestingly, a double dissociation was found between AVHs and non-AVHs groups regarding the effects of long-term and short-term predictions (Figure 2A.2B). Theoretically, in neural circuits, short-term and long-term plasticity of synaptic efficacy in sensory and motor neurons support learning and memory (Hasson *et al.*, 2008). Prior information can reshape synapses over different timescales by changing levels of activation, excitability and potentiation over milliseconds and minutes (Mongillo *et al.*, 2008, Perrett *et al.*, 2009). Thus, the synaptic plasticity can be a neural mechanism for continuously integrating prior information into 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 processing of incoming information along the speech hierarchy (Hasson *et al.*, 2010). The neural and cognitive foundations enable predictions 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 AVHs patients who showed long-term prediction effects, the non-AVHs patients tend to judge the target tone the same as the one immediately before. These results suggest that non-AVHs patients are prone to the influences of short-term memory. Schizophrenia patients without AVHs may have a greater response bias, rather than perceptual sensitivity deficits. These results are consistent with the involvement of externalizing biases in schizophrenia (Bentall and Slade, 1985, Varese *et al.*, 2012). Patients develop an external attribution

bias as an explanation for 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 lacks of effective connectivity between superior temporal gyrus(STG) and anterior cingulate cortex(ACC) – a critical component of the 'core control network'.

In summary, we found that schizophrenia patients with AVHs weighted predictions more over sensory processing and altered the recognition of tones. Moreover, patients with and without AVHs showed distinct influences of predictions at different timescales. Our results support a Bayesian cognitive account that malfunctioned source monitoring mediate AVHs.

Acknowledgments

Conflicts of Interest

None

Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

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.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

Aleman, A., Böcker, K. B. E., Hijman, R., de Haan, E. H. F. & Kahn, R. S. (2003). Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research* **64**, 175-185.

Behrendt, R.-P. (1998). Underconstrained perception: A theoretical approach to the nature and function of verbal hallucinations. *Comprehensive Psychiatry* **39**, 236-248.

Bentall, R. P. & Slade, P. D. (1985). Reality testing and auditory hallucinations: A signal detection analysis. *British Journal of Clinical Psychology* **24**, 159-169.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot. Int.* **5**, 341-345.

Brunelin, J., Mondino, M., Gassab, L., Haesebaert, F., Gaha, L., Suaud-Chagny, M.-F., Saoud, M., Mechri, A. & Poulet, E. (2012). Examining Transcranial Direct-Current Stimulation (tDCS) as a Treatment for Hallucinations in Schizophrenia. *American Journal of Psychiatry* **169**, 719-724.

Brunoni, A. R., Amadera, J., Berbel, B., Volz, M. S., Rizzerio, B. G. & Fregni, F. (2011). A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. *International Journal of Neuropsychopharmacology* **14**, 1133-1145.

Cassidy, C. M., Balsam, P. D., Weinstein, J. J., Rosengard, R. J., Slifstein, M., Daw, N. D., Abi-Dargham, A. & Horga, G. (2018). A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. *Current Biology* **28**, 503-514.e4.

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K. & Powers, A. R. (2019). Hallucinations and Strong Priors. *Trends in Cognitive Sciences* **23**, 114-127.

Davis, M. H. & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research* **229**, 132-147.

Dima, D., Dietrich, D. E., Dillo, W. & Emrich, H. M. (2010). Impaired top-down processes in schizophrenia: A DCM study of ERPs. *NeuroImage* **52**, 824-832.

Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* **19**, 158-164.

Farde, L. (1997). Brain imaging of schizophrenia—the dopamine hypothesis. *Schizophrenia Research* **28**, 157-162.

Frith, C. D. & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine* **19**, 359-363.

Fuster, J. n. M. (1997). Network memory. *Trends in Neurosciences* **20**, 451-459.

Giraud, A.-L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience* **15**, 511-517.

Hasson, U., Chen, J. & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences* **19**, 304-313.

Hasson, U., Malach, R. & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences* **14**, 40-48.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J. & Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. *The Journal of Neuroscience* **28**, 2539.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience* **8**, 393-402.

Hoffman, R. E. (1986). Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences* **9**, 503-517.

Hoffman, R. E., Boutros, N. N., Hu, S., Berman, R. M., Krystal, J. H. & Charney, D. S. (2000). Transcranial magnetic stimulation and auditory hallucinations in schizophrenia. *The Lancet* **355**, 1073-1075.

Hoffman, R. E., Hawkins, K. A., Gueorguieva, R., Boutros, N. N., Rachid, F., Carroll, K. & Krystal, J. H. (2003). Transcranial Magnetic Stimulation of Left Temporoparietal Cortex and Medication-Resistant Auditory Hallucinations. *Archives of General Psychiatry* **60**, 49-56.

Keefe, R. S. & Kraus, M. S. (2009). Measuring memory-prediction errors and their consequences in youth at risk for schizophrenia. *J Ann Acad Med Singapore* **38**, 414-416.

Kindler, J., Homan, P., Jann, K., Federspiel, A., Flury, R., Hauf, M., Strik, W., Dierks, T. & Hubl, D. (2013). Reduced Neuronal Activity in Language-Related Regions After Transcranial Magnetic Stimulation Therapy for Auditory Verbal Hallucinations. *Biological Psychiatry* **73**, 518-524.

Kontsevich, L. L. & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research* **39**, 2729-2737.

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Örey, D., Richter, F., Samara, M., Barbui, C., Engel, R. R., Geddes, J. R., Kissling, W., Stapf, M. P., Lässig, B., Salanti, G. & Davis, J. M. (2013). Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *The Lancet* **382**, 951-962.

Li, S., Zhu, H. & Tian, X. (2020). Distinct neural signals in speech preparation differentially modulate auditory responses. *bioRxiv*.

Ma, O. & Tian, X. (2019). Distinct Mechanisms of Imagery Differentially Influence Speech Perception. *eNeuro* **6**, ENEURO.0261-19.2019.

Macmillan, N. A. & Creelman, C. D. (2004). *Detection theory: A user's guide.* New York: Psychology press.

Meltzer, H. Y. (1992). Treatment of the neuroleptic-nonresponsive schizophrenic patient. *Schizophrenia bulletin* **18**, 515-542.

Mondino, M., Jardri, R., Suaud-Chagny, M.-F., Saoud, M., Poulet, E. & Brunelin, J. (2015). 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. *Schizophrenia Bulletin* **42**, 318-326.

Mongillo, G., Barak, O. & Tsodyks, M. (2008). Synaptic Theory of Working Memory. *Science* **319**, 1543.

Moseley, P., Fernyhough, C. & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience & Biobehavioral Reviews* **37**, 2794-2805.

Moseley, P., Smailes, D., Ellison, A. & Fernyhough, C. (2016). The effect of auditory verbal imagery on signal detection in hallucination-prone individuals. *Cognition* **146**, 206-216.

Moulton, S. T. & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 1273-1280.

NeoSpeech (2012). NeoSpeech Text-to-Speech voices [computer software]. *Santa Clara, CA.*

Perrett, D. I., Xiao, D., Barraclough, N. E., Keysers, C. & Oram, M. W. (2009). Seeing the future: Natural image sequences produce "anticipatory" neuronal activity and bias perceptual report. *Quarterly Journal of Experimental Psychology* **62**, 2081-2104.

Powers, A. R., Mathys, C. & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science* **357**, 596.

Priori, A., Hallett, M. & Rothwell, J. C. (2009). Repetitive transcranial magnetic stimulation or transcranial direct current stimulation? *Brain Stimulation* **2**, 241-245.

Schacter, Daniel L., Addis, Donna R., Hassabis, D., Martin, Victoria C., Spreng, R. N. & Szpunar, Karl K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron* **76**, 677-694.

Shergill, S. S., Murray, R. M. & McGuire, P. K. (1998). Auditory hallucinations: a review of psychological treatments. *Schizophrenia Research* **32**, 137-150.

Sommer, I. E. C., Slotema, C. W., Daskalakis, Z. J., Derks, E. M., Blom, J. D. & van der Gaag, M. (2012). The Treatment of Hallucinations in Schizophrenia Spectrum Disorders. *Schizophrenia Bulletin* **38**, 704-714.

Stephane, M., Barton, S. & Boutros, N. (2001). Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophrenia research* **50**, 61-78.

Szpunar, K. K., Spreng, R. N. & Schacter, D. L. (2014). A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences* **111**, 18414.

Teng, X., Ma, M., Yang, J., Blohm, S., Cai, Q. & Tian, X. (2020). Constrained Structure of Ancient Chinese Poetry Facilitates Speech Content Grouping. *Current Biology* **30**, 1299-1305.e7.

Teng, X., Tian, X., Rowland, J. & Poeppel, D. (2017). Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLoS biology* **15**, e2000812-e2000812.

Teufel, C. & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nat Rev Neurosci* **21**, 231-242.

Tian, X., Ding, N., Teng, X., Bai, F. & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour* **2**, 225-234.

Tian, X. & Huber, D. E. (2013). Playing "Duck Duck Goose" With Neurons: Change Detection Through Connectivity Reduction. *Psychological Science* **24**, 819-827.

Tian, X. & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience* **6**, 314.

Tian, X. & Poeppel, D. (2014). Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG. *Journal of Cognitive Neuroscience* **27**, 352-364.

Tian, X., Zarate, J. M. & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex* **77**, 1-12.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., Read, J., van Os, J. & Bentall, R. P. (2012). Childhood Adversities Increase the Risk of Psychosis: A Meta-analysis of Patient-Control, Prospective- and Cross-sectional Cohort Studies. *Schizophrenia Bulletin* **38**, 661-671.

Vercammen, A. & Aleman, A. (2008). Semantic Expectations Can Induce False Perceptions in Hallucination-Prone Individuals. *Schizophrenia Bulletin* **36**, 151-156.

Vercammen, A., de Haan, E. H. F. & Aleman, A. (2008). Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychological Medicine* **38**, 1177-1184.

Waters, F. (2012). Multidisciplinary approaches to understanding auditory hallucinations in schizophrenia and nonschizophrenia populations: the International Consortium on Hallucination Research. *Schizophrenia bulletin* **38**, 693-694.

Wheeler, M. E., Petersen, S. E. & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences* **97**, 11125.

Yang, F., Fang, X., Tang, W., Hui, L., Chen, Y., Zhang, C. & Tian, X. (2019). Effects and potential mechanisms of transcranial direct current stimulation (tDCS) on auditory hallucinations: A meta-analysis. *Psychiatry Research* **273**, 343-349.

Figure Legends

Figure 1. Schematic description of experimental procedures.

At the beginning of a trial, a fixation appeared in the center of the screen for 500 msec. 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 377ms. The inter-stimulus interval (ISI) was 623ms. Trails with 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 377ms. Participants judged the target tone by pressing one of two buttons. The target tone in the RE conditions was either consistent (REcon) or inconsistent (REinc) with the alternating pattern in the preceding clean tones. Whereas the target tone in the IR conditions was either the same (IRsame) or different (IRdiff) from the tone immediately before.

Table1 Demographics of Schizophrenia patients with and without AVHs and Healthy Controls

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.

Figure 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. B) Results from the group of patients without auditory verbal hallucinations. C) Results from the goup of health participants. Error bars indicate standard errors. **significant at the level of $p < 0.01$; *significant at the level of $p < 0.05$.

Figure 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 AVHs patients. B) The correlation between AHRS total scores and d' difference between RE_{con} and RE_{inc} condition in AVHs patients. C) The correlation between P3 subscores and d' difference between RE_{con} and RE_{inc} condition in AVHs patients. AHRS, auditory hallucinations rating scale.

	AVHs Patients (n=32)	Non-AVHs Patients (n=29)	Healthy Controls (n=30)	F or χ^2	*p*-value
Gender (M/F)	14/18	12/17	9/21	0.68	0.508
Age (years)	27.28 (6.37)	25.79 (5.94)	21.67 (2.45)	9.41	<0.001
Education (years)	12.44 (2.47)	12.83 (2.65)	13.43 (0.77)	1.69	0.19
Height (m)	1.68 (8.31)	1.66 (9.08)	1.67 (7.07)	0.39	0.677
Weight(kg)	66.78 (15.43)	62.38 (7.10)	60.37 (8.93)	2.66	0.076
Age of onset (years)	21.09 (6.66)	19.69 (4.80)	N/A	0.82	0.368
Duration (month)	74.05 (77.64)	63.38 (52.53)	N/A	0.39	0.536
PANSS total score	82.03 (8.87)	78.55 (4.41)	N/A	1.97	0.06
P3 subscore	4.75 (0.98)	1.00 (0.00)	N/A	21.56	<0.001
P subscore	23.53 (3.82)	14.69 (2.65)	N/A	10.40	<0.001
N subscore	19.78 (3.48)	22.41 (1.43)	N/A	3.93	0.001
G subscore	38.72 (4.09)	41.45 (2.56)	N/A	3.09	0.003
AVHs total score	25.25 (3.85)	N/A	N/A	N/A	N/A

Table1 Demographics of Schizophrenia patients with and without AVHs and Healthy Controls

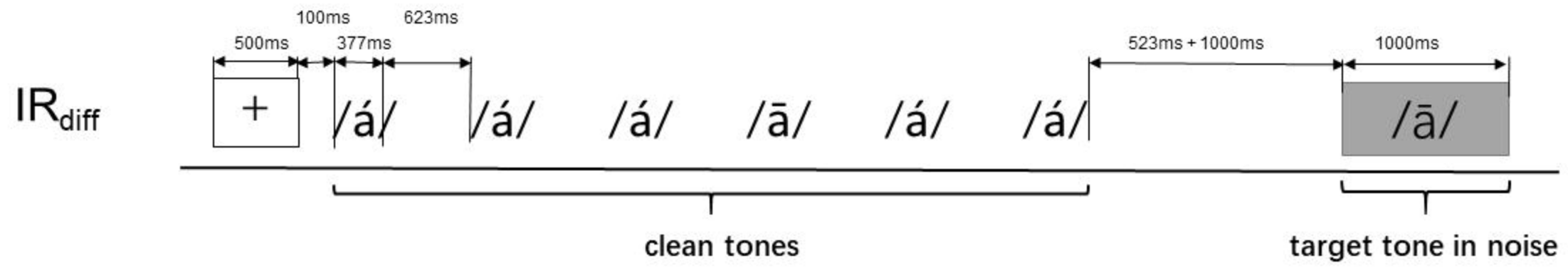Notes: PANSS: Positive and Negative Syndrome Scale; P subscore: positive symptom subscore;

N subscore: negative symptom subscore; G subscore: general psychopathology subscore;

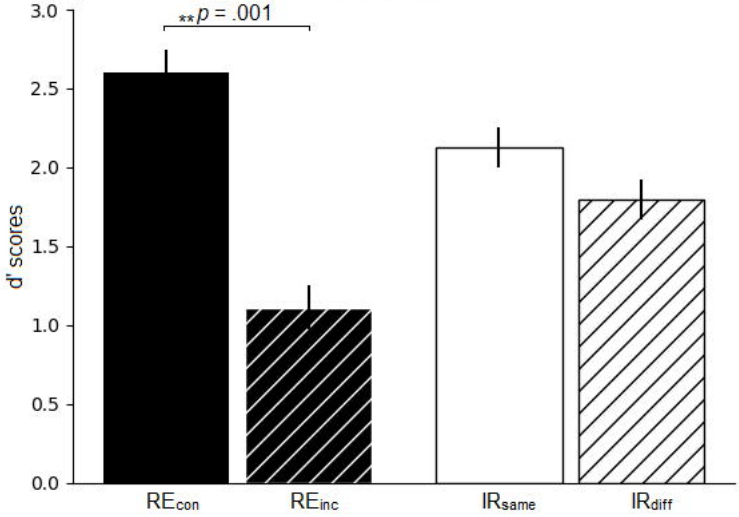N/A: not applicable.

RE_con   + /ā/ /á/ /ā/ /á/ /ā/ /á/   /ā/

RE_inc   + /ā/ /á/ /ā/ /á/ /ā/ /á/   /á/
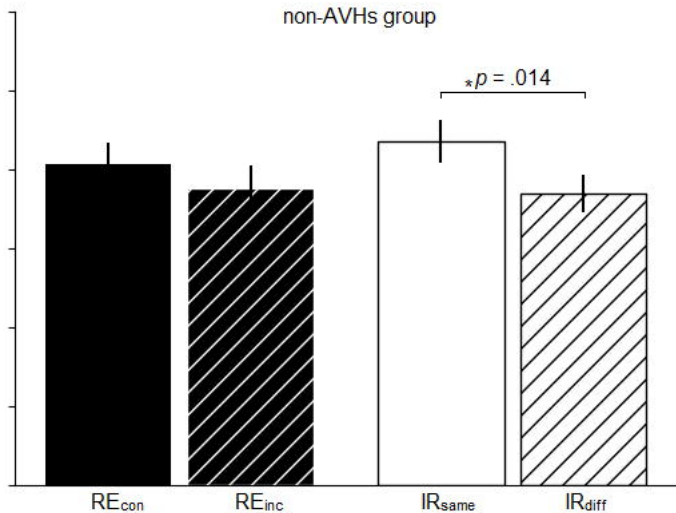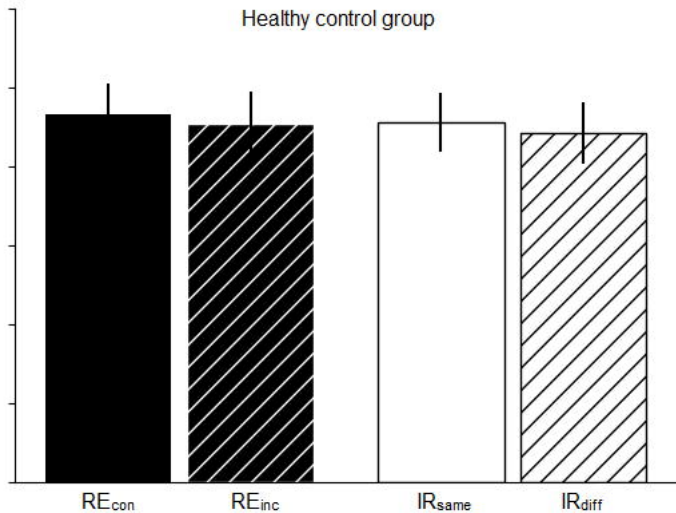
IR_same   + /á/ /ā/ /ā/ /á/ /ā/ /á/   /á/
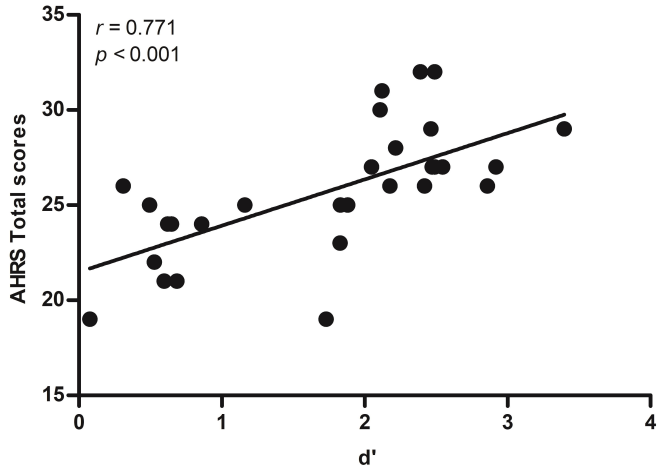
500ms 100ms 623ms
       377ms

IR_diff   + /á/ /á/ /á/ /ā/ /á/ /á/   523ms + 1000ms   1000ms   /ā/

clean tones   target tone in noise

A AVHs group

B

non-AVHs group

c

Healthy control group

RE$_{con}$ RE$_{inc}$ IR$_{same}$ IR$_{diff}$

A

B

$r = 0.771$
$p < 0.001$

C

P3 subscores (y-axis)

d' (x-axis)

$r = 0.453$
$p = 0.009$