Psychiatry Research 189 (2011) 149–152

Contents lists available at ScienceDirect

Psychiatry Research

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

Brief report

Categorization and timbre perception of environmental sounds in schizophrenia

Jean-Arthur Micoulaud-Franchi [a,b,c,*], Mitsuko Aramaki [b,c], Adrien Merer [d], Michel Cermolacce [a,b,c], Solvi Ystad [d], Richard Kronland-Martinet [d], Jean Vion-Dury [a,b,c]

[a] Pôle de Psychiatrie "Solaris", Service Pr. Naudin, Centre Hospitalier Universitaire de Sainte-Marguerite, 270 Bd de Sainte-Marguerite, 13009 Marseille, France
[b] Institut de Neurosciences Cognitives de la Méditerranée, INCM-CNRS UMR 6193, 31 Chemin Joseph Aiguier 13402 Marseille cedex 20, France
[c] Université d'Aix-Marseille, 58 Bd Charles Livon, 13284 Marseille cedex 07, France
[d] Laboratoire de Mécanique et d'Acoustique, LMA-CNRS UPR 7051, 31 Chemin Joseph Aiguier 13402 Marseille cedex 20, France

ARTICLE INFO

Article history:
Received 17 May 2010
Received in revised form 22 December 2010
Accepted 2 March 2011

Keywords:
Schizophrenia
Auditory perception
Psychophysics

ABSTRACT

Perception of environmental sounds from impacted materials (Wood, Metal and Glass) was examined by conducting a categorization experiment. Stimuli consisted of sound continua evoking progressive transitions between material categories. Results highlighted shallower response curves in subjects with schizophrenia than healthy participants, and are discussed in the framework of Signal Detection Theory and in terms of impaired perception of specific timbre features in schizophrenia.

1. Introduction

Schizophrenia is associated with complex cognitive (Frith, 1987) and perceptual (Holzman, 1972) impairments. An extensive literature analyzed perceptual impairments that were mainly interpreted as consequences of sensory processing dysfunctions (Javitt, 2009a,2009b). In the auditory modality (Rabinowicz et al., 2000), anomalies were found in tone matching (Javitt et al., 1997), pitch identification (Holcomb et al., 1995), or sound duration discrimination (Todd et al., 2003). Leitman et al. (2008) evaluated the contribution of these underlying dysfunctions, especially tone matching, to emotion recognition from vocal expressions in schizophrenia. Further on, Bach et al. (2009) suggested that perceptual impairments can be related to a more general dysfunction in cortical brain networks according to the Signal Detection Theory (SDT) that allowed the authors to explain the shallower response curves in Schizophrenia found by Kee et al. (2006) in an identification task on emotional facial continua (i.e., two facial expressions morphed into each other).

However, perceptual impairments in schizophrenia have not been thoroughly investigated for the perception of environmental sounds, which is an essential way to perceive the surrounding world (Gaver, 1993). We here aim at examining disturbances of sound categorization in schizophrenia by using impact sounds from different material categories (Wood, Glass, and Metal) which present the following advantages: i) they afford relevant information to perceive surrounding objects in a very concise way, which is important in everyday situations, ii) the parametric manipulation of these sounds was previously explored by creating sound continua (impacted sounds morphed between two material categories) with analysis–synthesis and interpolation methods and iii) the relevancy of specific acoustic features for material perception, i.e., damping and spectral complexity (roughness), was previously highlighted (Aramaki et al., 2011). With a reduced sound corpus from this previous study, we used the same experimental protocol based on a forced-choice categorization task of the perceived material and measured the response curve for these transitions. We compared group data between patients with schizophrenia (SCZ) and control participants (CTL).

Analogous to results from Kee et al.'s (2006) study based on emotional facial continua, we expect disturbances of the sound categorization reflected by shallower response curves in SCZ than in CTL. Going further, based on the knowledge of acoustic features responsible for the material perception, we expect to identify the effect of specific timbre-related features on the possible disturbances in SCZ.

2. Methods

2.1. Participants

We included 20 chronic inpatients and outpatients from the Department of Psychiatry, Marseille University Hospital, France on the Structured Clinical Interview for DSM-IV Axis I Disorders: Clinical Version (SCID-CV) confirmed the diagnosis of schizophrenia (First et al., 1997). The Positive and Negative Syndrome Scale (Kay et al., 1987) assessed severity of

Abbreviations: SCZ, patients with schizophrenia; CTL, control participants.

* Corresponding author at: Pôle de Psychiatrie "Solaris", Service Pr. Naudin, Centre Hospitalier Universitaire de Sainte-Marguerite, 270 Bd de Sainte-Marguerite, 13009 Marseille, France. Tel.: +33 622 364 019; fax: +33 491 745 578.
E-mail address: jarthur.micoulaud@gmail.com (J.-A. Micoulaud-Franchi).

0165-1781/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.psychres.2011.03.001

Table 1
Demographic and clinical characteristics, behavioral data and values of parameter *d* (slope) and *c* (shift point) obtained from bootstrap procedure. Statistical differences were tested with Mann–Whitney *U*-test.

	SCZ ($n=20$)	CTL ($n=20$)	Differences (SCZ-CTL)	Statistics
Demographic				
Age, mean (S.D.)	31.90 (6.95)	29.65 (10.87)	+2.25	0.440
Gender (male)	43%	43%	0	1.000
Education, mean (S.D.)	6.40 (3.12)	7.80 (2.31)	−1.40	0.124
Hearing threshold (dB)				
Right, mean (S.D.)	14.88 (4.42)	13.08 (3.85)	+1.80	0.347
Left, mean (S.D.)	14.94 (4.43)	12.89 (4.25)	+2.05	0.231
Clinical				
PANSS				
Positive, mean (S.D.)	17.45 (6.47)	–	–	–
Negative, mean (S.D.)	20.4 (7.62)	–	–	–
Age 1st hospitalization, mean (S.D.)	20.80 (2.67)	–	–	–
Lifetime hospitalizations, mean (S.D.)	11.05 (6.74)	–	–	–
Chlorpromazine equivalents, mean (S.D.)	368.58 (218.54)	–	–	–
Behavioral data				
RT (s), mean (S.D.)	4.49 (1.34)	4.26 (1.25)	+0.23	0.310
Unexpected responses (%), mean (S.D.)				
Glass (in Wood–Metal),	5.75 (10.5)	1.25 (3.2)	+4.50	0.357
Wood (in Glass–Metal)	3 (5.5)	0.75 (2.45)	+2.25	0.296
Metal (in Glass–Wood)	25.5 (18.2)	29.75 (15.75)	−4.25	0.670
Bootstrap data				
Slope, median (Quartile 25, Quartile 75)				
Wood–Metal	0.53 (0.45, 0.66)	0.93 (0.81, 1.04)	−0.40	≤0.001
Glass–Metal	0.70 (0.58, 0.86)	0.88 (0.75, 1.07)	−0.18	≤0.001
Glass–Wood	0.72 (0.59, 0.82)	0.95 (0.69, 2.99)	−0.23	≤0.001
Shift point, median (Quartile 25, Quartile 75)				
Wood–Metal	9.91 (9.46, 10.19)	9.49 (9.13, 9.73)	+0.42	≤0.001
Glass–Metal	9.25 (8.74, 9.71)	8.78 (8.48, 9.09)	+0.53	≤0.001
Glass–Wood	8.75 (8.33, 9.14)	8.94 (8.59, 9.44)	−0.19	0.157

symptoms. All patients were medicated. CTL constituted 20 healthy subjects screened for any current or lifetime history of a DSM-IV axis I disorder based on MINI interviews (Sheehan et al., 1998) (Table 1).

The exclusion criteria were presence of neurological illness, brain injury or other significant medical illness, current or past substance abuse or dependence, and any kind of auditory impairment. In particular, all participants presented a normal screening audiogram. They gave written informed consent, according to the guidelines established by the institutional review board.

2.2. Stimuli

Stimuli were selected from the sound data bank used in the previous study (Aramaki et al., 2011). To design these sounds, 15 impact sounds from everyday life objects (wood beams, glass bowls and metallic plates) that were 5 sounds per material, were recorded, analyzed and re-synthesized with an additive synthesis model. The synthetic sounds were then tuned to the same chroma and equalized by gain adjustments to minimize influence of pitch and loudness variations in the categorization judgments. From these 15 reference sounds, 15 continua were constructed to simulate progressive transitions between material categories (i.e., 5 Wood–Metal, 5 Glass–Metal, and 5 Glass–Wood continua). Each continuum was composed of 20 hybrid sounds that were obtained by interpolating the synthesis parameters of the 2 reference sounds located at the extreme positions. A total of 300 hybrid sounds (without reference sounds) were created.

To shorten the duration of the experiment for SCZ, we reduced the previous sound corpus to 3 continua (one per transition): based on the response curves obtained in the previous study, we selected the continua presenting curves that were most centered on the middle sound position. We also verified that all reference sounds differed from each other (e.g. the reference sound of Metal in Wood–Metal and Glass–Metal continua differed). Thus, 60 impact sounds,[1] i.e. 3 continua of 20 hybrid sounds, were kept for this current study (Fig. 1A). The mean sound duration was 875 ms for Wood–Metal, 1044 ms for Glass–Metal, and 653 ms for Glass–Wood continuum.

2.3. Procedure

Impact sounds were presented through the internal sound card of a PC computer in a quiet room, using open headphones (HD650 Sennheiser) amplified with Samsom (s-type amp). Subjects were free to adjust the intensity level of the sounds, once at the beginning of the test. The experimenter controlled that this level was high enough to provide a comfortable listening condition for each subject and that this level was approximately set to the same value across participants in both groups.

The experimenter was present with the participant during the entire test and followed the process through second headphones connected on the same Samsom amplifier. The experiment began with a 4-trials training session to familiarize participants with the task. Then, all sounds were randomly presented in a single session. Participants could listen to the sound as many times as they wanted, and were asked to categorize each sound once as Wood, Glass or Metal, by selecting with a mouse the material label displayed on the computer interface specifically designed[2] at LMA-CNRS on Max/MSP software.[3] Positions of labels displayed on the screen were randomly balanced across participants. Responses (one selected material category per sound) and Reaction Times (RT, recorded from sound onset the first time it was played) were collected for each subject. Although no time constraint was imposed, the experimenter verified that subjects did not dwell too much in each trial.

3. Results

3.1. Behavioral data

Individual responses were averaged so that each sound was associated with a percentage of categorization in Wood, Metal and Glass category for each group. Then, for each material continuum, a response curve was constructed from the percentage of categorization in the second category of material (e.g., Metal for Wood–Metal) as a function of sound position along the continuum (Fig. 1B). Percentages of unexpected responses (e.g. Glass in Wood–Metal continuum) and

[1] http://www.lma.cnrs-mrs.fr/~kronland/Categorization/sounds.html.

[2] A pretest was conducted to evaluate the handiness of the interface with 14 healthy subjects (5 women and 9 men; mean age: 34 years old, S.D. = 13.75) and 3 schizophrenia patients. They did not belong to the groups that were tested in the formal experiment. We controlled: i) that they well understood the task, and that they easily answered with the interface, and ii) that data were recorded, correctly stored, and easy to export toward statistic software.

[3] http://www.cycling74.com/.

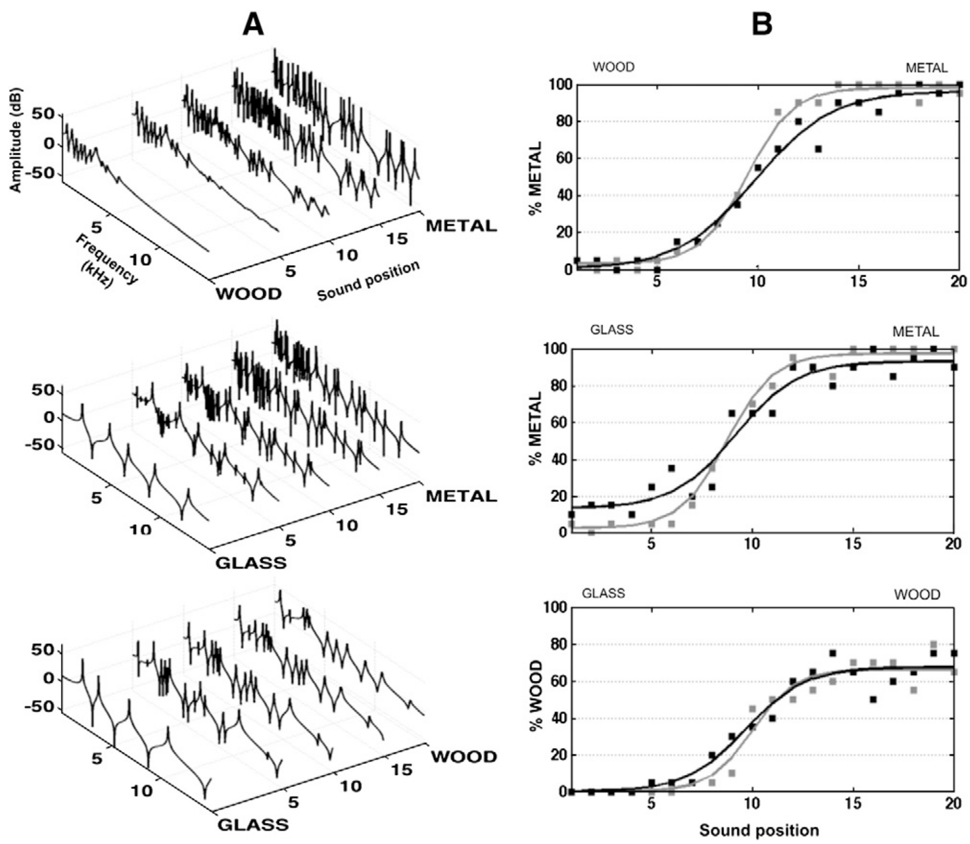

Fig. 1. (A) Spectra of the hybrid sounds at position 5, 10 and 15 along the Wood–Metal, Glass–Metal and Glass–Wood continua. The spectra of the reference sounds are represented at the extremes of each continuum. (B) Performance on the categorization task for SCZ (black) and CTL (grey): percentage of categorization in the second category of material (e.g. Metal in Wood–Metal continuum) as a function of sound position (20 points) in each continuum (dots), and the corresponding logistic function (solid line).

RT values are shown in Table 1. No significant differences in unexpected responses and RT were found between SCZ and CTL. Note that RTs were quite short (about 4 s) in both groups, suggesting that subjects in SCZ did not experience the test as too difficult.

3.2. Data modeling

Since the response curves presented a sigmoid shape as previously found in Aramaki et al. (2011), they were modeled by the following logistic function:

$$y = a + b / \left(1 + e^{-[(x-c).d]}\right) \quad (1)$$

where *y* is the probability to categorize a sound in the second category of material in the continuum, *x* the sound position, *a* the lower asymptote, *b* the difference between upper and lower asymptotes, *c* the sound position at midpoint (shift point) and *d* the slope. The logistic parameters were estimated by fitting data at best in a least-squares sense (Fig. 1B). Note that Kee et al. (2006) and Pollak and Kistler (2002) used this modeling in studies based on an emotional continuum paradigm.

Since participants categorized each sound only once, we obtained one single response curve for each continuum and for each group. Therefore, a bootstrap randomization method (Efron, 1979; Kee et al., 2006) was used to generate empirical standard error estimates and to evaluate statistical group differences in shift point and slope. In practice, sets of 20 records (i.e. individual responses) were chosen randomly (with replacement) from all available records for each continuum and for each group. The logistic parameters *c* and *d* were calculated from the resulting response curve and the procedure was repeated 100 times. Finally, a Mann–Whitney *U*-test was conducted on the two series of 100 values to evaluate differences between groups.

In SCZ, the slopes were significantly smaller for all continua and the shift points moved towards the Metal category in Wood–Metal and Glass–Metal continua (Table 1). The same shift towards Metal was obtained in SCZ by considering the percentages of categorization in the first category of material (i.e., Wood in Wood–Metal and Glass in Glass–Metal continuum).

4. Discussion

As expected, data analysis highlighted shallower curves for all transitions with significantly smaller slopes in SCZ than in CTL. The slope indicates how abrupt the categorization between materials is. Thus, this suggests that SCZ perceive category transitions in a more smooth way than CTL, similar to studies investigating facial emotional continua in visual modality (Kee et al., 2006) and *Ba–Da* continuum with phonetic sounds (Cienfuegos et al., 1999).

Since damping and roughness were previously shown to be relevant timbre features for the material perception (Aramaki et al., 2011), this result lead us to assume that SCZ are less able than CTL to perceive these specific acoustic features. This assumption is in line with the literature that considers perceptual impairments as consequences of sensory processing dysfunctions (Cienfuegos et al., 1999; Javitt, 2009a,2009b), such as Leitman et al. (2008) who related the prosodic emotion recognition impairment in schizophrenia to pitch processing dysfunction.

However, based on computational models of schizophrenia (Rolls et al., 2008), we cannot discard a more general deficit in the sensory information processing across modalities. In line with this consideration, Bach et al. (2009) hypothesized that their findings in emotional prosody recognition impairment in schizophrenia can be related to global increased fluctuations in cerebral networks. The authors discussed this hypothesis in the SDT framework, as a broader distribution of noise in the internal representation of the signal (i.e., a reduced signal-to-noise ratio) leading to shallower response curves observable in different modalities, for example in visual modality in Kee et al.'s (2006) study. Hence, for our concern, we can interpret our findings as the fact that the internal representation of material categories might be less distinctive in SCZ.

Going further, data analysis highlighted additional results specifically related to the Metal perception. Indeed, we found that the shift point was moved towards the Metal endpoint category in Wood–Metal and Glass–Metal continua in SCZ. The shift point indicates the perceptual boundary between categories. Thus, in the SDT framework, we can conclude on a biased response criterion for the Metal category in SCZ. For the Glass–Wood continuum, there was no significant shift point difference, but interestingly, the number of Metal responses (unexpected responses) was high in both groups, but tended to be lower in SCZ than in CTL (neither significant). From an acoustic point of view, Metal sounds are characterized by a longer duration (related to damping) and an increased spectral complexity (related to roughness) compared to Glass and Wood sounds (Aramaki et al., 2011). The increase in spectral complexity of sounds located at intermediate positions of the Glass–Wood continuum (Fig. 1A; higher number of spectral components) might explain this high number of unexpected Metal responses.

As well as these acoustic considerations, the specific perception of Metal sounds was previously highlighted by electrophysiological data (collected on healthy subjects) showing early differences of the auditory ERPs to Metal (from 150 ms from the sound onset) compared with both Glass and Wood sounds that most likely reflect timbre processing (Aramaki et al., 2009). Based on the auditory ERP anomalies found in 50–200 ms latency in schizophrenia (Buchsbaum, 1977; Turetsky et al., 2009), we can hypothesize that the biased response criterion for Metal in SCZ is due to sensory bias in perception of acoustic features specific to Metal timbre, i.e., longer duration and/or high spectral complexity.

To document our hypotheses, we need, i) to explore the concept of internal representation of material categories from a theoretical point of view as in Smits et al. (2006), and by an experimental approach by measuring, for instance, the synchrony EEG activity during impact sound listening (Lachaux et al., 1999; Tallon-Baudry and Bertrand, 1999); then the possible decrease of this synchrony especially for sounds at the intermediate positions of the continua might be associated with broader noise distribution of the material representation, ii) to evaluate auditory deficits relatively to specific timbre-related features by psychophysical investigations and ERP measurements, and iii) to include effect of therapeutic drugs, psychiatric symptoms, and cognitive disorders into multivariate analyses.

In conclusion, this is the first study to our knowledge that used calibrated and parameterized sounds to assess environmental sound perception in schizophrenia. Results highlighted disturbances of sound categorization that we supposed to be induced by impairments in the perception of specific timbre features and in the internal representation of material category in the SDT framework. This study confirmed the importance of investigating timbre perception in schizophrenia and constitutes a new step to better understand distortions related to this disorder, which induce deep handicap in the everyday situation.

Acknowledgments

The authors thank Dr. Laurent Boyer for providing help with statistical analysis.

References

Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S., 2009. Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. In: Ystad, S., Kronland-Martinet, R., Jensen, K. (Eds.), Computer Music Modeling and Retrieval — Genesis of Meaning of Sound and Music. Springer-Verlag, Berlin-Heidelberg, pp. 1–17.

Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S., 2011. Controlling the perceived material in an impact sound synthesizer. IEEE Transactions on Audio, Speech and Language Processing 19, 301–314.

Bach, D.R., Buxtorf, K., Grandjean, D., Strik, W.K., 2009. The influence of emotion clarity on emotional prosody identification in paranoid schizophrenia. Psychological Medicine 39, 927–938.

Buchsbaum, M.S., 1977. The middle evoked response components and schizophrenia. Schizophrenia Bulletin 3, 93–104.

Cienfuegos, A., March, L., Shelley, A.M., Javitt, D.C., 1999. Impaired categorical perception of synthetic speech sounds in schizophrenia. Biological Psychiatry 45, 82–88.

Efron, B., 1979. Bootsrap methods: another look at the jackknife. Annals of statistics 7, 26.

First, R., Gibbon, M., Williams, J., 1997. Structured Clinical Interview for DSM-IV Axis I Disorders: Clinical Version (SCID-CV). American Psychiatric Press, Washington, DC.

Frith, C.D., 1987. The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. Psychological Medicine 17, 631–648.

Gaver, W.W., 1993. What in the world do we hear? An ecological approach to auditory source perception. Ecological Psychology 5, 29.

Holcomb, H.H., Ritzl, E.K., Medoff, D.R., Nevitt, J., Gordon, B., Tamminga, C.A., 1995. Tone discrimination performance in schizophrenic patients and normal volunteers: impact of stimulus presentation levels and frequency differences. Psychiatry Research 57, 75–82.

Holzman, P.S., 1972. Assessment of perceptual functioning in schizophrenia. Psychopharmacologia 24, 29–41.

Javitt, D.C., 2009a. Sensory processing in schizophrenia: neither simple nor intact. Schizophrenia Bulletin 35, 1059–1064.

Javitt, D.C., 2009b. When doors of perception close: bottom-up models of disrupted cognition in schizophrenia. Annual Review of Clinical Psychology 5, 249–275.

Javitt, D.C., Strous, R.D., Grochowski, S., Ritter, W., Cowan, N., 1997. Impaired precision, but normal retention, of auditory sensory ("echoic") memory information in schizophrenia. Journal of Abnormal Psychology 106, 315–324.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin 13, 261–276.

Kee, K.S., Horan, W.P., Wynn, J.K., Mintz, J., Green, M.F., 2006. An analysis of categorical perception of facial emotion in schizophrenia. Schizophrenia Research 87, 228–237.

Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. Human Brain Mapping 8, 194–208.

Leitman, D.I., Laukka, P., Juslin, P.N., Saccente, E., Butler, P., Javitt, D.C., 2008. Getting the Cue: Sensory Contributions to Auditory Emotion Recognition Impairments in Schizophrenia. Schizophrenia Bulletin.

Pollak, S.D., Kistler, D.J., 2002. Early experience is associated with the development of categorical representations for facial expressions of emotion. Proceedings of the National Academy of Sciences 99, 9072–9076.

Rabinowicz, E.F., Silipo, G., Goldman, R., Javitt, D.C., 2000. Auditory sensory dysfunction in schizophrenia: imprecision or distractibility? Archives of General Psychiatry 57, 1149–1155.

Rolls, E.T., Loh, M., Deco, G., Winterer, G., 2008. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. Nature Reviews Neuroscience 9, 696–709.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry 59 (Suppl 20), 22–33 quiz 34–57.

Smits, R., Sereno, J., Jongman, A., 2006. Categorization of sounds. Journal of Experimental Psychology: Human Perception and Performance 32, 733–754.

Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. Trends in Cognitive Sciences 3, 151–162.

Todd, J., Michie, P.T., Jablensky, A.V., 2003. Association between reduced duration mismatch negativity (MMN) and raised temporal discrimination thresholds in schizophrenia. Clinical Neurophysiology 114, 2061–2070.

Turetsky, B.I., Bilker, W.B., Siegel, S.J., Kohler, C.G., Gur, R.E., 2009. Profile of auditory information-processing deficits in schizophrenia. Psychiatry Research 165, 27–37.