



Test-retest reliability of time-frequency measures of auditory steady-state responses in patients with schizophrenia and healthy controls

Brian J. Roach^a, Deepak Cyril D'Souza^{b,c,d}, Judith M. Ford^{a,e}, Daniel H. Mathalon^{a,e,*}

^a Psychiatry Service, San Francisco VA, San Francisco, CA, USA

^b Psychiatry Service, VA Connecticut Healthcare System, West Haven, CT, USA

^c Abraham Ribicoff Research Facilities, Connecticut Mental Health Center, New Haven, CT, USA

^d Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

^e Department of Psychiatry, UCSF, San Francisco, CA, USA

ARTICLE INFO

Keywords:

Auditory steady-state response (ASSR)

Gamma

Electroencephalography (EEG)

Schizophrenia

Reliability

Generalizability

Oscillations

Power

Phase angle

ABSTRACT

Background: Auditory steady-state response (ASSR) paradigms have consistently demonstrated gamma band abnormalities in schizophrenia at a 40-Hz driving frequency with both electroencephalography (EEG) and magnetoencephalography (MEG). Various time-frequency measures have been used to assess the 40-Hz ASSR, including evoked power, single trial total power, phase-locking factor (PLF), and phase-locking angle (PLA). While both EEG and MEG studies have shown power and PLF ASSR measures to exhibit excellent test-retest reliability in healthy adults, the reliability of these measures in patients with schizophrenia has not been determined.

Methods: ASSRs were obtained by recording EEG data during presentation of repeated 20-Hz, 30-Hz and 40-Hz auditory click trains from nine schizophrenia patients (SZ) and nine healthy controls (HC) tested on two occasions. Similar ASSR data were collected from a separate group of 30 HC on two to three test occasions. A subset of these HC subjects had EEG recordings during two tasks, passively listening and actively attending to click train stimuli. Evoked power, total power, PLF, and PLA were calculated following Morlet wavelet time-frequency decomposition of EEG data and test-retest generalizability (G) coefficients were calculated for each ASSR condition, time-frequency measure, and subject group.

Results: G-coefficients ranged from good to excellent (> 0.6) for most 40-Hz time-frequency measures and participant groups, whereas 20-Hz G-coefficients were much more variable. Importantly, test-retest reliability was excellent for the various 40-Hz ASSR measures in SZ, similar to reliabilities in HC. Active attention to click train stimuli modestly reduced G-coefficients in HC relative to the passive listening condition.

Discussion: The excellent test-retest reliability of 40-Hz ASSR measures replicates previous EEG and MEG studies. PLA, a relatively new time-frequency measure, was shown for the first time to have excellent reliability, comparable to power and PLF measures. Excellent reliability of 40 Hz ASSR measures in SZ supports their use in clinical trials and longitudinal observational studies.

1. Introduction

The auditory steady-state response (ASSR) is a stimulus-evoked oscillation that can be measured with either electroencephalography (EEG) or magnetoencephalography (MEG) in response to a sound that is repeated at a fixed rate or frequency. Such steady-state stimulation drives the ASSR in EEG at the same rate (Galambos et al., 1981), and classic MEG source localization analyses revealed generator contributions from primary auditory cortex (Pantev et al., 1993). The ASSR has been shown to reach its maximum at 40-Hz (i.e., 40 repetitions per

second), its “resonant” frequency (Galambos et al., 1981), and this result has been widely replicated (Brenner et al., 2003; Hamm et al., 2011; Krishnan et al., 2009; Light et al., 2006; Roach et al., 2013; Spencer et al., 2008). The 40-Hz ASSR has been shown to be reduced in schizophrenia patients (see (Thune et al., 2016) for a review), their first-degree relatives (Hong et al., 2004), other psychosis spectrum disorders including schizoaffective disorder and bipolar disorder (Zhou et al., 2018), and in one study of individuals at clinical high risk for psychosis (Tada et al., 2016). Because of the known role of fast-spiking, parvalbumin-expressing inhibitory γ -aminobutyric acid (GABA)

* Corresponding author at: San Francisco VA Healthcare System, Psychiatry Service (116D), 4150 Clement Street, San Francisco, CA 94121, USA.

E-mail address: daniel.mathalon@ucsf.edu (D.H. Mathalon).

<https://doi.org/10.1016/j.nicl.2019.101878>

Received 11 February 2019; Received in revised form 18 April 2019; Accepted 25 May 2019

Available online 05 June 2019

2213-1582/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

interneurons and *N*-methyl-D-aspartate (NMDA) receptor glutamatergic signaling in the interneuron-pyramidal neuron microcircuits (Mathalon and Sohal, 2015) that generate normal gamma band (e.g., 30–80 Hz) oscillations (Cardin et al., 2009; Carlen et al., 2012; Sohal et al., 2009), the 40-Hz ASSR is increasingly regarded as a biomarker of impaired GABAergic and/or glutamatergic neural activity, or impairment in the excitatory-inhibitory balance achieved through their interplay (Gonzalez-Burgos and Lewis, 2008; Light et al., 2017; Roopun et al., 2008), in schizophrenia.

ASSR oscillations can be isolated and quantified through time-frequency decomposition of EEG/MEG single trial or averaged data. Traditional ASSR quantification involved applying a Fast-Fourier Transformation (FFT) to the time-domain event-related potential (ERP) average to obtain an evoked power spectrum. Time-frequency decomposition of the ERP (e.g., by sliding window FFT) is another evoked power measure, in this case retaining time-domain information. However, time-frequency decomposition of single trials allows the ASSR to be represented by event-related change in single trial signal magnitude, referred to as total power, and by the consistency in signal phase across single trials, referred to as phase-locking factor (PLF (Tallon-Baudry et al., 1997)). While no time-frequency measure dominates the psychosis ASSR literature, these three time-frequency measures (i.e., evoked power, total power, PLF) are interrelated (Roach et al., 2018) and have all been used to quantify selective 40-Hz ASSR reductions in schizophrenia (Thune et al., 2016). In addition to these measures, a new time-frequency measure, the phase-locking angle (PLA), has recently been introduced (Roach et al., 2018). PLA quantifies the degree to which a subject's oscillatory phase leads or lags the average phase angles of time-locked oscillations taken from a reference sample (e.g., a healthy control group), and it has been shown to be more sensitive to patient-control group differences than power or PLF measures (Roach et al., 2018). Unlike other time-frequency measures, which quantify aspects of signal variance, the PLA quantifies a mean. This mean phase angle provides a frequency-specific estimate of oscillation latency, relative to stimulus onset.

While the 40-Hz ASSR paradigm has shown great promise as a probe of gamma oscillation abnormalities in psychosis-spectrum disorders, it is important to determine the test-retest reliability of each of these four time-frequency measures to better evaluate the potential of each to serve as an outcome measure in longitudinal clinical trials or observational studies. In healthy subjects, both EEG (McFadden et al., 2014; van Deursen et al., 2009) and MEG (Legget et al., 2017; Tan et al., 2015) reliability studies have shown strong, positive correlations between 40-Hz ASSR measures assessed on two separate occasions. However, only one study (Tan et al., 2015) used the intra-class correlation (ICC) coefficient (Shrout and Fleiss, 1979) to quantify reliability, while the others used Pearson's correlation coefficient. ICC quantifies the extent to which responses from one test occasion to the next have similar values and covary, while Pearson's *r* only quantifies covariance and ignores agreement of values. Notably, Tan et al. showed excellent 40-Hz ASSR test-retest reliability for PLF and evoked power measures in both sensor and source space using amplitude modulated tones in an ASSR paradigm that included an instruction to attend to an infrequent, target tone within the sequence of 40-Hz amplitude modulated tones (Tan et al., 2015). This manipulation may be important as attention has been shown to enhance the 40-Hz ASSR (Skosnik et al., 2007), but attention is rarely monitored or controlled with a behavioral task in most ASSR studies.

The goals of the current study were to (1) establish test-retest reliability of ASSR time-frequency measures, including evoked power, total power, PLF, and PLA, in patients with schizophrenia and healthy controls; (2) determine if test-retest reliability of these time-frequency measures is similar between unattended (passive listening) compared to attended (auditory oddball) task conditions; and (3) assess 40-Hz ASSR PLA test-retest reliability for the first time.

Table 1
Study characteristics.

Subject group	N participants	N occasions	Paradigms
Study 1 - All Days	25	3	Auditory Oddball
Study 1 - All Subjects	30	2	Auditory Oddball
Study 1 - Active/ Passive	14	2	Auditory Oddball, Passive Listening
Study 2 - HCs	9	2	Auditory Oddball
Study 2 - SZs	9	2	Auditory Oddball

2. Methods

2.1. Studies

EEG data for reliability analyses came from two different studies, referred to throughout as Study 1 and Study 2. The participants in these studies contribute to 5 sets of reliability analyses, which are outlined in Table 1. All participants completed an auditory oddball task on each EEG test occasion, which has been used previously (D'Souza et al., 2012; Gunduz-Bruce et al., 2012; Hamilton et al., 2018) to efficiently obtain both ASSR and oddball ERP components with one brief task while also helping to control attention. Only a small subset ($n = 14$) of Study 1 participants also completed a passive ASSR task, which allowed for the comparison of the impact of attention on ASSR reliability. These paradigms are described in greater detail below. Subjects from Study 1 and Study 2 were never combined for any analysis. Additional details of the design and execution of these two studies are provided in Supplementary Material, with limited details relevant to the reliability studies described below.

2.2. Subjects

All subjects provided written informed consent to participate in these IRB-approved studies. Study 1 subjects ($n = 30$; 18–35 years old) have been described in a previous report (D'Souza et al., 2012), while clinical and demographic characteristics of Study 2 subjects are presented in Table 2.

2.3. EEG oddball/passive paradigms and acquisition

In a three-stimulus auditory oddball task, a random series of infrequent (8.33%) “target” tones (1000 Hz, 500 ms), frequent (83.33%) “standard” click trains (20, 30, or 40 Hz), and infrequent task-irrelevant novel distractor sounds (8.33%) (Friedman et al., 1993), were presented at approximately 80 dB SPL (C weighting) with a 1.25-s stimulus onset asynchrony in three separate blocks. Each block had 15 targets, 15 distractor stimuli, and 150 standards (20 Hz click trains for Block 1, 30 Hz click trains for Block 2, and 40 Hz click trains for Block 3). The order of blocks was randomized for each test day, and test days were separated by approximately 1 week (Study 1 median: 7 days, Study 2 median: 8 days). All subjects completed this oddball task. On the same test occasions, a subset of the Study 1 HC subjects ($n = 14$) also had a passive 20 and 40-Hz ASSR recording with separate blocks of 150 click trains presented every 1.25-s. All click trains comprised 1 ms rarefaction clicks presented at the specified frequency for a duration of 500 ms.

Subjects sat in an acoustically shielded booth in front of a computer monitor and wore insert earphones (Etymotic Research, Inc., Elk Grove Village, IL, USA). EEG data were digitized at 1000-Hz from 22 (Study 1) or 28 (Study 2) scalp electrodes, bandpass filtered between 0.05 and 200-Hz, and referenced to linked ears (Study 1) or the right mastoid (Study 2) using Neuroscan SynAmps amplifiers (Compumedics Neuroscan, Charlotte, NC, USA). Additional electrodes were placed at the outer canthi of both eyes and above and below the left eye to record eye movements and blinks (vertical and horizontal electro-oculogram [EOG]; VEOG, HEOG). All electrode impedances were maintained at or

Table 2
Demographic and clinical data for patients and controls.

	Patients (N = 9)				Controls (N = 9)			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Age (years)	36.56	12.95	22	54	35.56	11.81	20	53
Average Parental SES [‡]	43.28	19.04	9.5	60	35.44	14.92	11	51.5
Education (years) [‡]	13.11	1.36	12	16	15.33	2.12	12	19
Race	5 Caucasian, 3 African American, 1 Hispanic/Latino				7 Caucasian, 1 African American			
Handedness	9 Right				9 Right			
Gender	5 Male, 4 Female				5 Male, 4 Female			
Diagnosis	Paranoid Schizophrenia (5) Undifferentiated Schizophrenia (2) Schizoaffective Depressed Type (1) Schizoaffective Bipolar Type (1)							
Antipsychotic Medication	9 atypical							

[‡] Socioeconomic status (SES) based on the Two-Factor (1958) Hollingshead Scale; Higher scores indicate lower socioeconomic status.

* $p < .05$ with independent samples *t*-test.

below 10 kΩ. Because Study 1 subjects were participating in an intravenous drug challenge study, the EEG data used for reliability analysis came from pre-infusion EEG recordings on the morning of each test day. Due to Study 1 time constraints and lower priority of these pre-infusion EEG recordings relative to other test day procedures, not all 30 Study 1 subjects provided useable 20, 30, and 40-Hz oddball task runs on all three test occasions. Specifically, two subjects dropped out and did not complete a third test occasion, and there were acquisition errors with at least one task run on one occasion for three additional subjects. Therefore, two, parallel sets of Study 1 reliability analyses were conducted to account for this: either (1) using the first two test occasions for all ($n = 30$) subjects (“all subject” analyses) or (2) using all three test occasions for the subset of subjects ($n = 25$) who completed all test days (“all day” analyses).

2.4. EEG processing and time-frequency analysis

Study 2 data were re-referenced to an average mastoid reference. Using Brain Vision Analyzer software, all continuous EEG data from both studies were initially 1-Hz high-pass filtered, segmented into 3 s (−1 to 2 s around click train onset) single trial epochs for each driving frequency and, if applicable, oddball versus passive attention conditions, subjected to regression-based ocular correction (Gratton et al., 1983), and baseline corrected using the 100 ms baseline period preceding click train onset. Single trial data were then exported to MATLAB for additional processing steps described in greater detail in Supplementary Material. Briefly, canonical correlation analysis and automated single trial cleaning were applied to all data, using the same approaches described previously (Kort et al., 2017). Following these artifact reduction and rejection steps, Morlet wavelet time-frequency transformation was implemented using FieldTrip (Oostenveld et al., 2011) separately for all driving frequencies and oddball/passive tasks. ERPs were calculated prior to wavelet decomposition to allow calculation of evoked power, but all other time-frequency measures were derived from the single trial data. PLA time-frequency data for each subject and condition were re-expressed as the difference from the expected angle at each point in the time-frequency matrix. The expected angle at each time and frequency data point was defined separately for Study 1, using the circular mean across all Study 1 subjects and test days, and Study 2, using the circular mean across both test days of the HC only. The difference between an individual subject phase angle and the expected (circular mean) phase angle defines the PLA, in units of radians (see Supplementary Material). For the PLA measure only, a *z*-score transformation using the same group means used to define expected angles was applied prior to conducting reliability analyses. While such standardization does not change the underlying distribution of linear variables (e.g., evoked power, total power, or PLF), it can be

considered a non-linear transformation of PLA data because it changes periodic values, in radians, to units of standard deviation from the reference group (all subjects for Study 1 and HC only for Study 2). Because of its frequent use in previous studies of schizophrenia, results from electrode Fz are presented in the main text, with parallel results from other electrodes presented in Supplementary Materials. The majority of ASSR studies in schizophrenia have focused on either the entire steady-state stimulation period (e.g., 0 to 500 ms) or a time window after the transient onset auditory evoked potential components (e.g., 250–500 ms) in statistical analyses (see Supplementary Material). However, most ASSR *reliability* studies have focused on later (i.e., after 200 ms) windows. Separate averages were calculated from each time-frequency measure at each of the three steady-state frequencies for early (0 to 250 ms), late (250 to 500 ms), and complete (0 to 500 ms) steady-state stimulation time windows.

2.5. Reliability analysis

Test-retest reliability was determined with random effects analysis using a single facet of observation (test occasion) generalizability study (“G-study”) design to estimate variance components (Shavelson, 1991). Such a design allows estimation of three variance components for any time-frequency measure. The variance components for Person (σ_p^2), Occasion (σ_o^2), and Person $\times$ Occasion (plus Error, σ_{po+e}^2) are estimated separately for each study and group (e.g., HC or SZ). Once variance components are estimated, the G-coefficient, which provides a measure of generalizability or reliability of the measured score, can be calculated as in Eq. (1):

$$G = \frac{\sigma_p^2}{\left(\sigma_p^2 + \frac{\sigma_{po+e}^2}{n_o}\right)} \quad (1)$$

The majority of ASSR studies in schizophrenia are cross-sectional. If multiple time points are used to assess longitudinal or treatment effects, time-frequency measures from each session would be treated separately, indicating the best choice for the number of test occasions on which the measurement is based (n_o) is 1. Therefore, the G-coefficient in Eq. (1) is equal to the ICC defined by Shrout and Fleiss (e.g., ICC(3,1) in (Shrout and Fleiss, 1979)) when $n_o = 1$. As all G-coefficients in this study were calculated with $n_o = 1$, they could be considered interchangeable with ICCs. Variance components were estimated using a restricted maximum likelihood approach in MATLAB (Witkovský, 2012). Components were estimated separately and saved for the steady-state frequencies (20-Hz, 30-Hz, or 40-Hz), two tasks (Study 1 only), and four time-frequency measurements (evoked power, total power, PLF, and PLA).

The goal of a G-study is not to test a specific hypothesis. Thus, there

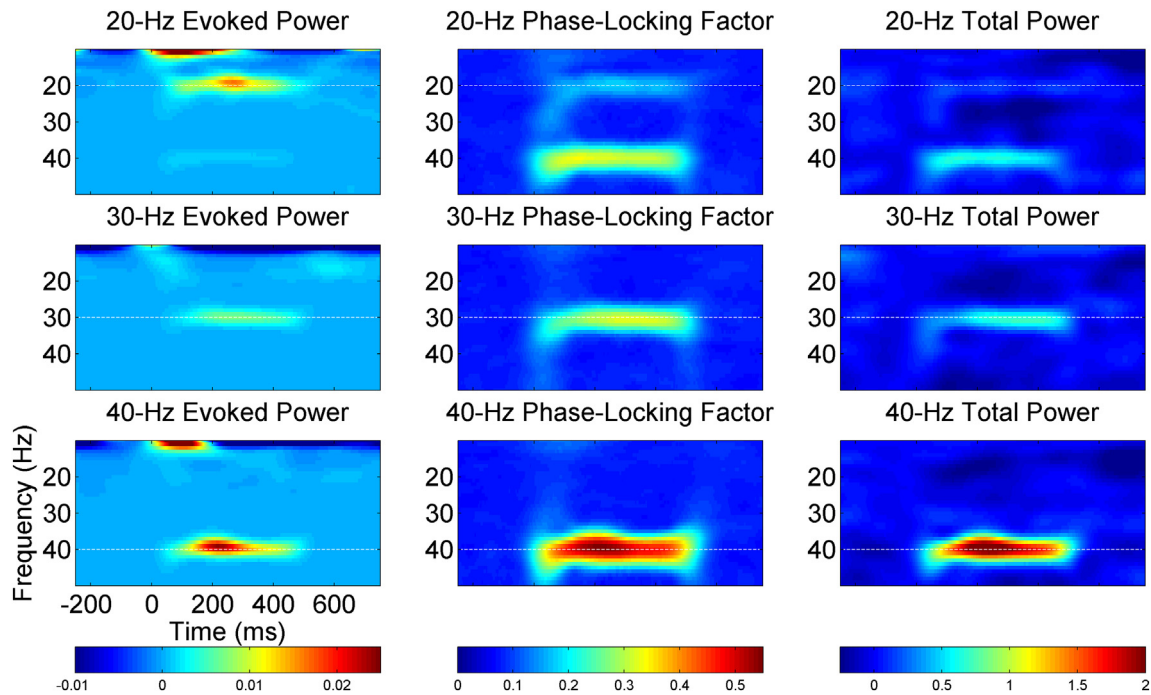


Fig. 1. Grand average time-frequency maps summarize 25 Study 1 subjects and three, separate test occasions at electrode Fz. Evoked power (left column), phase-locking factor (middle column), and total power (right column) were estimated from EEG data collected during 20-Hz (top row), 30-Hz (middle row), and 40-Hz auditory steady state stimulation. EEG frequency, in Hertz (Hz), is shown on the y-axis and time, in milliseconds (ms), is shown on the x-axis, with horizontal dashed white lines indicating the frequency of the steady-state stimulation. Darker red colors indicate greater auditory steady-state responses.

are no *p*-values associated with estimated variance components or G-coefficients. However, existing guidelines for determining clinical significance of ICCs suggest that the reliability coefficient can be qualitatively categorized as follows: $ICC < 0.4$ is poor, $0.4 \leq ICC < 0.6$ is fair, $0.6 \leq ICC < 0.75$ is good, and $0.75 \leq ICC < 1$ is excellent (Cicchetti and Sparrow, 1981). Therefore, G-coefficients were categorized using these four labels for descriptive purposes, but alternative categories have been proposed (e.g., (Shrout, 1998)), and could similarly be applied to characterize reliability.

3. Results

Grand average time-frequency maps for each measure and steady-state stimulation condition are presented in Fig. 1 for the Study 1 completers.

Across each of the 3 test days for Study 1 subjects, the largest ASSR total power and PLF were observed for 40-Hz click trains, followed by 30-Hz trains, and finally 20-Hz trains. While 40-Hz evoked power is also greatest on two of the three test days, the 20-Hz evoked power is greatest on the first test occasion and almost equal to 30-Hz evoked power on the other test occasions (Fig. 2).

Estimated G-coefficients for various combinations of time window (0–250, 250–500, and 0–500 ms), ASSR frequency (20-Hz, 30-Hz, 40-Hz), task (oddball vs. passive), time-frequency measure, and study group are presented in Table 3 for electrode Fz. To summarize, there was tendency for reliability to be greatest for the 40-Hz ASSR, followed next by 30-Hz ASSR, and finally 20-Hz ASSR across study groups and time-frequency measures. The Active (auditory oddball) versus Passive steady-state paradigm comparisons revealed that the Passive version of the paradigm was associated with greater 40-Hz ASSR reliability across Evoked/Total Power and PLA, but not PLF, measures. In both paradigms, the 20-Hz ASSR reliability was poor for all measures but PLF. One unexpected result was poor reliability in the Study 2 HC sample in the late (250–500 ms) time window for the 40-Hz ASSR PLF measure. To better understand the underlying source of this poor reliability value, variance components as well as fixed steady-state frequency effects

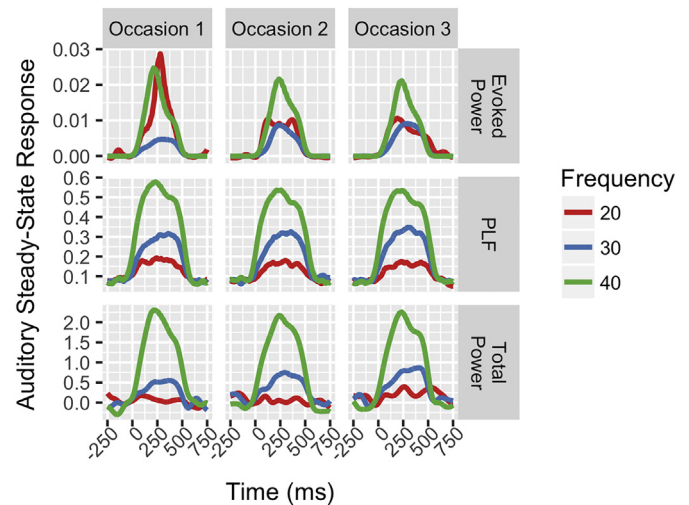


Fig. 2. Grand average waveforms summarize all 25 Study 1 subjects' auditory steady-state responses at each of the three, separate test occasions at electrode Fz. Waveforms represent the time-frequency measure (either Evoked Power, top row; Phase-Locking Factor (PLF), middle row; or Total Power, bottom row) at the steady-state stimulation frequency, with the 20-Hz activity plotted from 20-Hz stimulation trials (Red), the 30-Hz activity plotted from the 30-Hz stimulation trials (Blue) and the 40-Hz activity plotted from the 40-Hz stimulation trials (Green). Time-frequency measures are plotted and labeled on the y-axis with time, in milliseconds (ms), depicted on the x-axis in sub-plots for Day or Occasion 1 (left column), Occasion 2 (middle column), and Occasion 3 (right column).

(e.g., 40-Hz > 30-Hz > 20-Hz) were qualitatively compared between these HC, the SZ, and the Study 1 all subjects sample. This Study 1 all subjects sample was selected because it contained the most subjects ($n = 30$) but only 2 test occasions, which matched the Study 2 design. Fig. 3 demonstrates that both HC and SZ exhibit similar fixed SSR frequency effects on PLF (e.g., 20-Hz < 30-Hz < 40-Hz) to those

Table 3
Generalizability coefficients for electrode Fz.

	Early (0–250 ms)			Late (250–500 ms)			Complete (0–500 ms)		
	20 Hz	30 Hz	40 Hz	20 Hz	30 Hz	40 Hz	20 Hz	30 Hz	40 Hz
Evoked power									
Study 1 - All Days	0.15	0.54	0.70	0.22	0.52	0.72	0.20	0.55	0.73
Study 1 - All Subjects	0.11	0.29	0.32	0.29	0.42	0.52	0.23	0.35	0.41
Study 1 - Active	0.09		0.34	0.25		0.63	0.20		0.49
Study 1 - Passive	0.22		0.87	0.38		0.86	0.36		0.88
Study 2 - HCs	0.76	0.83	0.92	0.28	0.76	0.97	0.61	0.93	0.96
Study 2 - SZs	0.18	0.96	0.50	0.91	0.71	0.43	0.53	0.89	0.46
Total power									
Study 1 - All Days	0.12	0.42	0.71	0.00	0.43	0.77	0.00	0.47	0.75
Study 1 - All Subjects	0.13	0.30	0.55	0.00	0.44	0.63	0.00	0.43	0.61
Study 1 - Active	0.11		0.50	0.00		0.73	0.00		0.65
Study 1 - Passive	0.15		0.63	0.10		0.81	0.11		0.74
Study 2 - HCs	0.11	0.64	0.76	0.00	0.79	0.33	0.00	0.77	0.57
Study 2 - SZs	0.61	0.68	0.85	0.28	0.81	0.88	0.31	0.76	0.88
PLF									
Study 1 - All Days	0.48	0.69	0.74	0.50	0.70	0.81	0.55	0.71	0.79
Study 1 - All Subjects	0.55	0.71	0.70	0.64	0.73	0.73	0.67	0.75	0.73
Study 1 - Active	0.71		0.67	0.77		0.75	0.81		0.73
Study 1 - Passive	0.47		0.69	0.53		0.74	0.61		0.73
Study 2 - HCs	0.61	0.73	0.65	0.60	0.82	0.16	0.73	0.82	0.41
Study 2 - SZs	0.22	0.53	0.73	0.79	0.65	0.76	0.57	0.55	0.77
PLA									
Study 1 - All Days	0.30	0.73	0.70	0.44	0.79	0.52	0.33	0.80	0.63
Study 1 - All Subjects	0.13	0.60	0.67	0.31	0.72	0.62	0.31	0.71	0.65
Study 1 - Active	0.17		0.68	0.76		0.68	0.54		0.69
Study 1 - Passive	0.22		0.82	0.86		0.72	0.00		0.84
Study 2 - HCs	0.00	0.56	0.77	0.08	0.73	0.78	0.38	0.63	0.80
Study 2 - SZs	0.97	0.77	0.69	0.58	0.89	0.86	0.97	0.86	0.81

G-coefficients greater than 0.75 are in bold font.

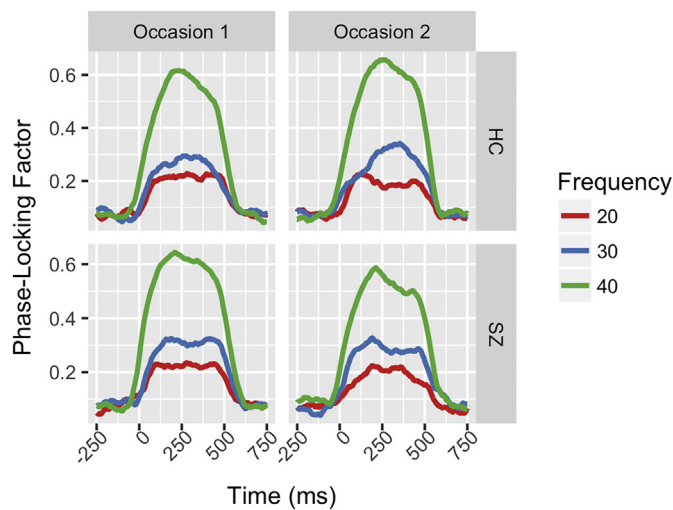


Fig. 3. Grand average waveforms summarize Study 1 subjects' auditory steady-state responses from two, separate test occasions at electrode Fz. Waveforms represent the phase-locking factor at the steady-state stimulation frequency, with the 20-Hz activity plotted from 20-Hz stimulation trials (Red), the 30-Hz activity plotted from the 30-Hz stimulation trials (Blue) and the 40-Hz activity plotted from the 40-Hz stimulation trials (Green). The phase-locking factor values are plotted on the y-axis separately for Healthy Controls (HC, top row) and Schizophrenia Patients (SZ, bottom row) with time, in milliseconds (ms), depicted on the x-axis in sub-plots for Occasion 1 (left column) and Occasion 2 (right column).

observed in Study 1 subjects (Fig. 2, middle row).

However, stacked bar plots (Fig. 4) demonstrate that relative to both Study 2 SZ subjects and Study 1 subjects, the Study 2 HC sample has relatively small person variance. The error variance components, which include both Person X Day interaction plus error variance, are similar

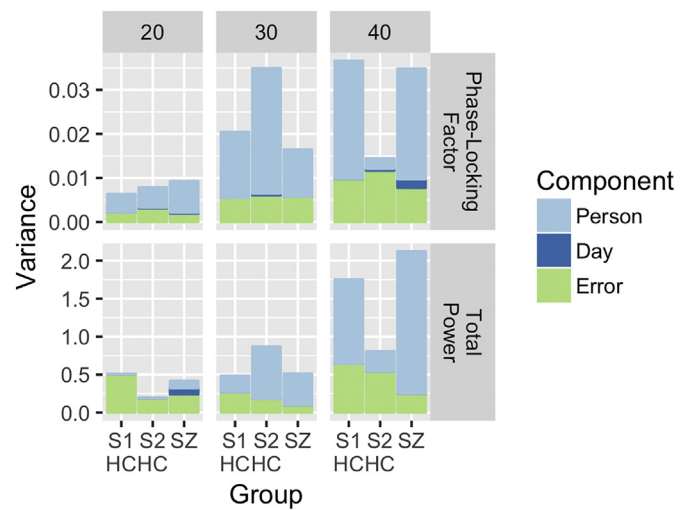


Fig. 4. Stacked bar plots demonstrate color-coded contributions of the three different variance components (Person, Day, and Person X Day plus Error) to the total variance (y-axis) for the phase-locking factor (top row) and total power (bottom row) from electrode Fz in the 250 to 500 millisecond time-window. Separate bars depict the 30 Study 1 subjects (S1 HC), Study 2 healthy controls (S2 HC), and Study 2 schizophrenia patients (SZ) with separate sub-plots for the 20-Hz stimulation trials (left column), 30-Hz stimulation trials (middle column), and 40-Hz stimulation trials (right column).

across these subject groups, as are error variance components in the 20-Hz and 30-Hz frequency conditions. Thus, the reduction in reliability reflects reduced true variance of the 40-Hz PLF measure in the Study 2 HC sample during the 250–500 ms 40-Hz click-train window. 40-Hz total power, which also showed reduced reliability in this same time window in Study 2 HCs, appeared to have a similar reduction in 40-Hz ASSR true-score variance across participants.

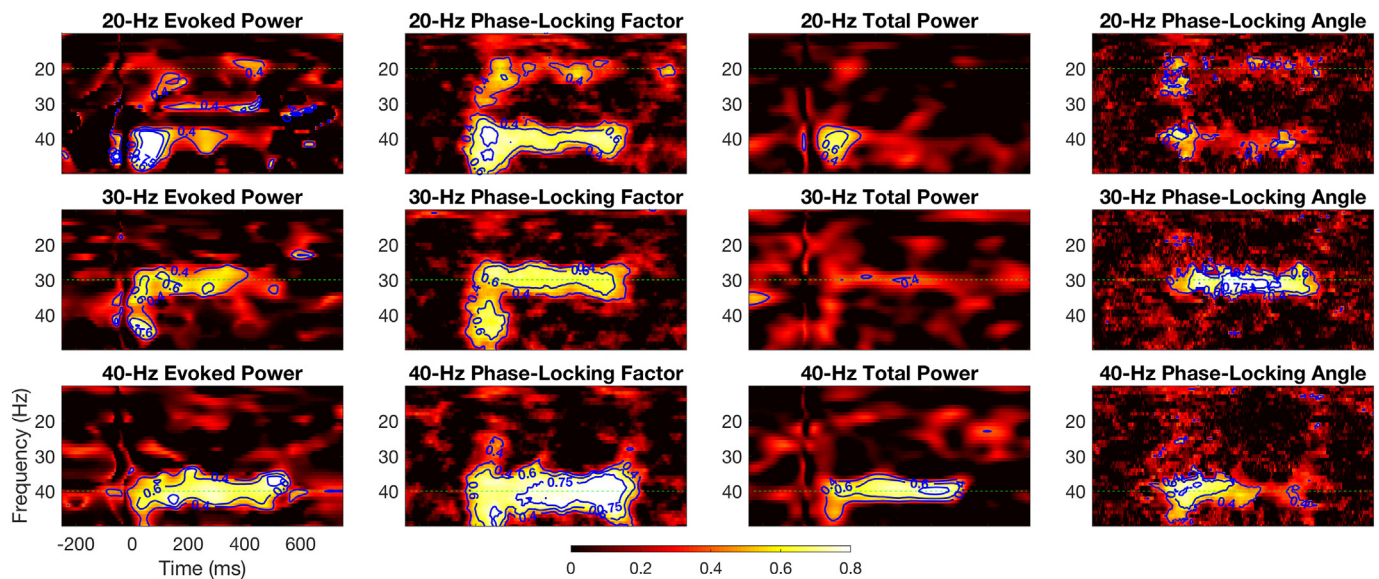


Fig. 5. Generalizability- (or G-) coefficient time-frequency maps summarize the test-retest reliability of time-frequency measures using the 25 Study 1 subjects and three test occasions at electrode Fz. Evoked power (left column), phase-locking factor (inner-left column), total power (inner-right column), and phase-locking angle (right column) were estimated from EEG data collected during 20-Hz (top row), 30-Hz (middle row), and 40-Hz auditory steady state stimulation. EEG frequency, in Hertz (Hz), is shown on the y-axis and time, in milliseconds (ms), is shown on the x-axis, with horizontal dashed green lines indicating the frequency of the steady-state stimulation. Brighter yellow to white colors indicate greater auditory steady-state response test-retest reliability, and labeled blue contour lines indicate boundaries between fair (G-coefficient ≥ 0.4), good (G-coefficient ≥ 0.65) and excellent (G-coefficient ≥ 0.75) reliability.

Fig. 5 shows time-frequency G-coefficient maps for the Study 1 completers. These maps clearly indicate that the 40-Hz ASSR has good reliability (G-coefficients > 0.6) for most of the steady-state stimulation time period across time-frequency measures. The PLF reliability on a millisecond basis is excellent (G-coefficients > 0.75) in the 200–500 ms time window. Notably, the early evoked gamma band response (40–50 Hz, 0–100 ms) exhibits good reliability across time-frequency measures and steady-state driving frequencies.

4. Discussion

This study assessed the test-retest reliability of multiple time-frequency measures of 20-Hz, 30-Hz, and 40-Hz ASSR in two samples, including HC and SZ groups. Reliability analyses established that the test-retest reliability of the ASSR in patients with schizophrenia is as good and in some instances better than ASSR reliability in healthy controls. While the reliability of ASSR measures in both passive listening and oddball task conditions was high, consistent with prior ASSR reliability studies (Legget et al., 2017; McFadden et al., 2014; Tan et al., 2015; van Deursen et al., 2009) controlling for attention with a behavioral (i.e., auditory oddball) task manipulation appears to slightly reduce ASSR reliability. Finally, comprehensive assessment of evoked power, total power, PLF, and PLA time-frequency measures of the ASSR revealed a pattern of effects that was in line with prior ASSR reliability studies for power and PLF measures and highlighted a consistent pattern of reliability across all time-frequency measures: 40-Hz ASSR measures exhibit excellent reliability that are often greater than those observed for 20-Hz and 30-Hz ASSR measures.

While several studies have demonstrated a strong, positive correlation between 40-Hz ASSR measures from two separate occasions (Legget et al., 2017; McFadden et al., 2014; van Deursen et al., 2009), this study is only the second to assess test-retest reliability using ICCs, which depend on agreement and not just shared variance between measurements from separate test occasions. Tan et al. (Tan et al., 2015) used ICCs to demonstrate high test-retest reliability of the 40-Hz ASSR in MEG data. In addition to replicating these previous reports of high 40-Hz ASSR reliability, this study demonstrates that various ASSR time-frequency measures are equally reliable in SZ and HC. This is the first

study to assess ASSR reliability in SZ. Importantly, patients showed good (ICCs > 0.6) or excellent (ICCs > 0.75) 40-Hz ASSR reliability for all time-frequency measures and time windows assessed except evoked power. The 40-Hz ASSR, relative to 30-Hz or 20-Hz ASSR, is most consistently found to be deficient in SZ, making it a reasonable target for intervention studies that aim to improve gamma oscillations. That the 40-Hz ASSR has excellent reliability for total power, PLF, and PLA measures provides quantitative support for the use of these three 40-Hz ASSR measures in intervention studies with multiple assessment occasions. The data also support tracking changes in 40-Hz ASSR measures over time in longitudinal observational studies of schizophrenia patients and healthy controls.

This is the first study to assess test-retest reliability of ASSRs in the same subjects during different task conditions. Increasing attention to auditory steady-state stimulation has been shown to enhance the 40-Hz ASSR (Ross et al., 2004; Skosnik et al., 2007) and increased arousal level has been shown to diminish the 40-Hz ASSR (Griskova et al., 2007). While it is unclear to what extent an auditory oddball task would increase arousal relative to passive listening, the task itself along with the behavioral readout provided by participant responses to target tones at the very least indicates subjects are listening to the sounds being presented. Given that several previous studies have shown ASSR enhancement with increased attention (Bidet-Caulet et al., 2007; Ross et al., 2004; Saupe et al., 2009a; Saupe et al., 2009b), perhaps the slight reductions in G-coefficients observed in the auditory oddball task condition were a function of decreased person variance in this task, relative to passive tasks, due to increased attention. This can be seen in Fig. 6, where overall variance and person variance in particular in all 40-Hz ASSR time-frequency measures are reduced in the active condition compared to the passive condition.

A subject who was not actively attending to the click train stimuli during passive listening might increase his or her ASSR through increased attention during the oddball task. A subject who was very vigilant in his or her auditory attention during passive listening might decrease attention to the click trains in the oddball task, as these were non-target sounds, leading to a decrease in the ASSR. In both cases, these trait-like differences between individuals would be associated with larger person variance (and G-coefficients) in the passive listening

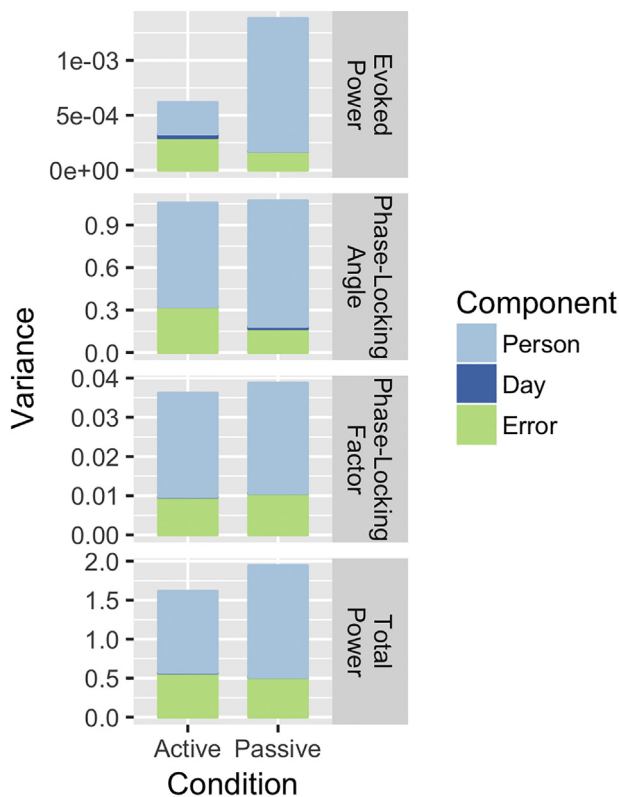


Fig. 6. Stacked bar plots demonstrate color-coded contributions of the three different variance components (Person, Day, and Person X Day plus Error) to the total variance (y-axis) for the evoked power (top row), phase-locking angle (second row), phase-locking factor (third row) and total power (bottom row) at electrode Fz in the 0 to 500 millisecond time-window for 40-Hz auditory steady-state stimulation. Separate bars depict the active auditory oddball (left) and passive listening task conditions from 14 Study 1 subjects.

task and smaller person variance in the auditory oddball task. It is also important to note that P300 ERP measures obtained from the infrequent sounds in the oddball task have previously been shown to have good to excellent reliability (ICCs ≥ 0.7), indicating this one task condition yields multiple EEG/ERP measures that are reliable (D'Souza et al., 2012).

The PLA, a novel measure of latency- and frequency-specific phase lag relative to a group mean, exhibited 40-Hz ASSR test-retest reliability that was good to excellent for most of the study samples and time windows evaluated. The PLA G-coefficients for electrode Fz were comparable to other time-frequency measures in general, but PLF, total power, and evoked power G-coefficients were all greater than PLA G-coefficients for the Study 1 completers (Table 3 “all days” sample), which represents the largest sample ($n = 75$ total EEG sessions) in this study. Analysis of electrode Cz (Supplementary Material) demonstrated that this pattern (i.e., PLA G < PLF, evoked and total power Gs) did not hold across sites, and PLA had more G-coefficients in the “excellent” range (i.e., ≥ 0.75) than any other measure for Cz. It is interesting that the PLA G-coefficients were greater than all other time-frequency measures for the 30-Hz ASSR in the relatively large Study 1 sample. In the one prior report comparing SZ to HC, 30-Hz ASSR PLA had greater circular phase variability than 40-Hz PLA, but the group comparison was only marginally significant at 30-Hz (Roach et al., 2018). Given the excellent test-retest reliability of 30-Hz PLA, it should continue to be explored along with the 40-Hz ASSR in subsequent studies of PLA that attempt to determine what clinical, demographic, or cognitive variables might be associated with phase delay.

In addition to PLA, assessment of the reliability of power and PLF measures revealed that across all 4 measures, 40-Hz ASSR had greater

reliability than 20-Hz and 30-Hz ASSR in two-thirds (48/72) of the samples and time windows tested at Fz. When considering only the power measures, 40-Hz ASSR had the greatest G-coefficients in $\sim 86\%$ (31/36) of the reliability analyses, which is also much greater than what would be expected by chance (i.e., 1/3 or 33.3%). While no single measure was uniformly most reliable among the samples and time windows assessed, the phase measures tended to outperform power measures on average. Consistent with previous reliability studies of 40-Hz ASSR (Legget et al., 2017; McFadden et al., 2014; Tan et al., 2015), power measures had a wider range of reliability coefficients, including poor G-coefficients as low as zero for 20-Hz ASSR. Given the previously observed strong correlations between evoked power, total power, and PLF in both HC and SZ (Roach et al., 2018) and the fact that no single measure was dominant in terms of reliability in these analyses, continued use of multiple time-frequency measures and multivariate statistical tests in the assessment of 40-Hz ASSR is well justified.

4.1. Limitations

As was the case in almost every study reporting abnormal 40-Hz ASSR in SZ, the patients in study 2 were all prescribed antipsychotic medications at the time of EEG assessments. Both the SZ and HC groups in study 2 had small sample sizes, which could lead to less stable variance component estimates, and subsequently, less accurate G-coefficient estimates. Subjects in Study 1 included cannabis users, and these participants were run in the morning, fasted on the test day, and had intravenous lines placed for the drug challenge study to follow. The data pre-processing methods and choice of reference electrodes could impact reliability coefficients and should be carefully considered in future studies. Any one of these factors as well as any additional factors (e.g., poor quality control) that could lead to increased error variance might make the reliability results from this sample less generalizable. However, the G-coefficients observed were both consistent with Study 2 samples and the ICCs reported by Tan et al. (Tan et al., 2015), providing at least some indication that the unique aspects of the Study 1 design did not greatly influence the reliability results.

5. Conclusions

The previously observed excellent test-retest reliability of the 40-Hz ASSR using either MEG (Legget et al., 2017; Tan et al., 2015) or EEG (McFadden et al., 2014) in healthy controls was replicated in multiple samples. Furthermore, similar test-retest reliability was observed in SZ patients for the first time. This is important because it supports the use of the 40-Hz ASSR paradigm as a probe of gamma band oscillatory activity in clinical trials examining within-subjects treatment effects aimed at ameliorating abnormal gamma oscillations in schizophrenia. Minimal reduction of test-retest reliability during an auditory oddball task indicates that it could be used when some behavioral measure of attention is required to monitor task compliance, and when additional measures of ERP P300 components are of interest without sufficient time to run separate auditory oddball and ASSR paradigms. Finally, excellent test-retest reliability was also observed for the PLA, which should encourage additional use of this novel measure of phase delay in future studies.

Acknowledgements

We thank Angelina Genovese, RNC, MBA; Michelle San Pedro, RN; Elizabeth O'Donnell, RN; Brenda Breault, RN, BSN; Sonah Yoo, RPh; Rachel Galvan, RPh; and Willie Ford of the Neurobiological Studies Unit at the VA Connecticut Healthcare System, West Haven Campus for their central contributions to the success of this project.

Funding

This work was supported by the National Institute on Drug Abuse (R21 DA020750 to DCD), an investigator-initiated AstraZeneca study (to DHM), and the Department of Veterans Affairs (Senior Research Career Scientist award to JMF; I01 CX0004971 to JMF).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2019.101878>.

References

- Bidet-Caulet, A., Fischer, C., Besle, J., Agüera, P.E., Giard, M.H., Bertrand, O., 2007. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.* 27, 9252–9261.
- Brenner, C.A., Sporns, O., Lysaker, P.H., O'Donnell, B.F., 2003. EEG synchronization to modulated auditory tones in schizophrenia, schizoaffective disorder, and schizotypal personality disorder. *Am. J. Psychiatry* 160, 2238–2240.
- Cardin, J.A., Carlen, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.H., Moore, C.I., 2009. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459, 663–667.
- Carlen, M., Meletis, K., Siegle, J.H., Cardin, J.A., Futai, K., Vierling-Claassen, D., Ruhlmann, C., Jones, S.R., Deisseroth, K., Sheng, M., Moore, C.I., Tsai, L.H., 2012. A critical role for NMDA receptors in parvalbumin interneurons for gamma rhythm induction and behavior. *Mol. Psychiatry* 17, 537–548.
- Cicchetti, D.V., Sparrow, S.A., 1981. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am. J. Ment. Defic.* 86, 127–137.
- van Deursen, J.A., Vuurman, E.F., van Kranen-Mastenbroek, V.H., Verhey, F.R., Riedel, W.J., 2009. 40-Hz steady state response in Alzheimer's disease and mild cognitive impairment. *Neurobiol. Aging* 32 (1), 24–30.
- D'Souza, D.C., Fridberg, D.J., Skosnik, P.D., Williams, A., Roach, B., Singh, N., Carbutto, M., Elander, J., Schnakenberg, A., Pittman, B., Sewell, R.A., Ranganathan, M., Mathalon, D., 2012. Dose-related modulation of event-related potentials to novel and target stimuli by intravenous Delta(9)-THC in humans. *Neuropsychopharmacology* 37, 1632–1646.
- Friedman, D., Simpson, G., Hamberger, M., 1993. Age-related changes in scalp topography to novel and target stimuli. *Psychophysiology* 30, 383–396.
- Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2643–2647.
- Gonzalez-Burgos, G., Lewis, D.A., 2008. GABA neurons and the mechanisms of network oscillations: implications for understanding cortical dysfunction in schizophrenia. *Schizophr. Bull.* 34, 944–961.
- Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484.
- Griskova, I., Morup, M., Parnas, J., Ruksenas, O., Arnfred, S.M., 2007. The amplitude and phase precision of 40 Hz auditory steady-state response depend on the level of arousal. *Exp. Brain Res.* 183, 133–138.
- Gunduz-Bruce, H., Reinhart, R.M., Roach, B.J., Gueorguieva, R., Oliver, S., D'Souza, D.C., Ford, J.M., Krystal, J.H., Mathalon, D.H., 2012. Glutamatergic modulation of auditory information processing in the human brain. *Biol. Psychiatry* 71, 969–977.
- Hamilton, H.K., Perez, V.B., Ford, J.M., Roach, B.J., Jaeger, J., Mathalon, D.H., 2018. Mismatch negativity but not P300 is associated with functional disability in schizophrenia. *Schizophr. Bull.* 44, 492–504.
- Hamm, J.P., Gilmore, C.S., Picchetti, N.A., Sponheim, S.R., Clementz, B.A., 2011. Abnormalities of neuronal oscillations and temporal integration to low- and high-frequency auditory stimulation in schizophrenia. *Biol. Psychiatry* 69 (10), 989–996.
- Hong, L.E., Summerfelt, A., McMahon, R., Adami, H., Francis, G., Elliott, A., Buchanan, R.W., Thaker, G.K., 2004. Evoked gamma band synchronization and the liability for schizophrenia. *Schizophr. Res.* 70, 293–302.
- Kort, N.S., Ford, J.M., Roach, B.J., Gunduz-Bruce, H., Krystal, J.H., Jaeger, J., Reinhart, R.M., Mathalon, D.H., 2017. Role of N-methyl-D-aspartate receptors in action-based predictive coding deficits in schizophrenia. *Biol. Psychiatry* 81, 514–524.
- Krishnan, G.P., Hetrick, W.P., Brenner, C.A., Shekhar, A., Steffen, A.N., O'Donnell, B.F., 2009. Steady state and induced auditory gamma deficits in schizophrenia. *Neuroimage* 47, 1711–1719.
- Legget, K.T., Hild, A.K., Steinmetz, S.E., Simon, S.T., Rojas, D.C., 2017. MEG and EEG demonstrate similar test-retest reliability of the 40Hz auditory steady-state response. *Int. J. Psychophysiol.* 114, 16–23.
- Light, G.A., Hsu, J.L., Hsieh, M.H., Meyer-Gomes, K., Sprock, J., Swerdlow, N.R., Braff, D.L., 2006. Gamma band oscillations reveal neural network cortical coherence dysfunction in schizophrenia patients. *Biol. Psychiatry* 60, 1231–1240.
- Light, G.A., Zhang, W., Joshi, Y.B., Bhakta, S., Talledo, J.A., Swerdlow, N.R., 2017. Single-dose memantine improves cortical oscillatory response dynamics in patients with schizophrenia. *Neuropsychopharmacology* 42, 2633–2639.
- Mathalon, D.H., Sohal, V.S., 2015. Neural oscillations and synchrony in brain dysfunction and neuropsychiatric disorders: It's about time. *JAMA Psychiatry* 72, 840–844.
- McFadden, K.L., Steinmetz, S.E., Carroll, A.M., Simon, S.T., Wallace, A., Rojas, D.C., 2014. Test-retest reliability of the 40 Hz EEG auditory steady-state response. *PLoS One* 9, e85748.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intelligence Neurosci.* 2011, 156869.
- Pantev, C., Elbert, T., Makeig, S., Hampson, S., Eulitz, C., Hoke, M., 1993. Relationship of transient and steady-state auditory evoked fields. *Electroencephalogr. Clin. Neurophysiol.* 88, 389–396.
- Roach, B.J., Ford, J.M., Hoffman, R.E., Mathalon, D.H., 2013. Converging evidence for gamma synchrony deficits in schizophrenia. *Suppl. Clin. Neurophysiol.* 62, 163–180.
- Roach, B.J., Ford, J.M., Mathalon, D.H., 2018. Gamma band phase delay in schizophrenia. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4 (2), 131–139.
- Roopun, A.K., Cunningham, M.O., Racca, C., Alter, K., Traub, R.D., Whittington, M.A., 2008. Region-specific changes in gamma and beta2 rhythms in NMDA receptor dysfunction models of schizophrenia. *Schizophr. Bull.* 34, 962–973.
- Ross, B., Picton, T.W., Herdman, A.T., Pantev, C., 2004. The effect of attention on the auditory steady-state response. *Neurol. Clin. Neurophysiol.* 2004, 22.
- Saupe, K., Schroger, E., Andersen, S.K., Müller, M.M., 2009a. Neural mechanisms of intermodal sustained selective attention with concurrently presented auditory and visual stimuli. *Front. Hum. Neurosci.* 3, 58.
- Saupe, K., Widmann, A., Bendixen, A., Müller, M.M., Schroger, E., 2009b. Effects of intermodal attention on the auditory steady-state response and the event-related potential. *Psychophysiology* 46, 321–327.
- Shavelson, R.J., 1991. In: Shavelson, Richard J., Webb, Noreen M. (Eds.), *Generalizability Theory: A Primer*. Sage Publications, Newbury Park, Calif.
- Shrout, P.E., 1998. Measurement reliability and agreement in psychiatry. *Stat. Methods Med. Res.* 7, 301–317.
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428.
- Skosnik, P.D., Krishnan, G.P., O'Donnell, B.F., 2007. The effect of selective attention on the gamma-band auditory steady-state response. *Neurosci. Lett.* 420, 223–228.
- Sohal, V.S., Zhang, F., Yizhar, O., Deisseroth, K., 2009. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* 459, 698–702.
- Spencer, K.M., Salisbury, D.F., Shenton, M.E., McCarley, R.W., 2008. Gamma-band auditory steady-state responses are impaired in first episode psychosis. *Biol. Psychiatry* 64 (5), 369–375.
- Tada, M., Nagai, T., Kirihaara, K., Koike, S., Suga, M., Araki, T., Kobayashi, T., Kasai, K., 2016. Differential alterations of auditory gamma oscillatory responses between pre-onset high-risk individuals and first-episode schizophrenia. *Cereb. Cortex* 26, 1027–1035.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1997. Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *J. Neurosci.* 17, 722–734.
- Tan, H.R., Gross, J., Uhlhaas, P.J., 2015. MEG-measured auditory steady-state oscillations show high test-retest reliability: a sensor and source-space analysis. *NeuroImage* 122, 417–426.
- Thune, H., Recasens, M., Uhlhaas, P.J., 2016. The 40-Hz auditory steady-state response in patients with schizophrenia: a meta-analysis. *JAMA Psychiatry* 73, 1145–1153.
- Witkovský, V., 2012. Estimation, testing, and prediction regions of the fixed and random effects by solving the Henderson's mixed model equations. *Measure. Sci. Rev.* 234.
- Zhou, T.H., Mueller, N.E., Spencer, K.M., Mallya, S.G., Lewandowski, K.E., Norris, L.A., Levy, D.L., Cohen, B.M., Ongur, D., Hall, M.H., 2018. Auditory steady state response deficits are associated with symptom severity and poor functioning in patients with psychotic disorder. *Schizophr. Res.* 201, 278–286.