IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 21, 2021, accepted July 19, 2021, date of publication July 26, 2021, date of current version August 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100157

Towards Real-Time Detection of Auditory Steady-State Responses: A Comparative Study

LEI WANG, ELISABETH NOORDANUS, AND ADRIANUS JOHANNES VAN OPSTAL
Donders Centre for Neuroscience, Department of Biophysics, Radboud University, 6525 AJ Nijmegen, The Netherlands

Corresponding author: Lei Wang (l.wang@donders.ru.nl)

The work of Lei Wang and Elisabeth Noordanus was supported by the Project Neuromodulation by Continuous and Integrated Monitoring and Treatment (NeuroCIMT)-Otocontrol under Grant 14689. The work of Adrianus Johannes van Opstal was supported by European Union (EU) Horizon 2020 European Research Council (ERC) Advanced Grant 2016 Orient under Grant 693400.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee from the Faculty of Social Sciences at Radboud University under Protocol No. ECSW2016-2208-41 and performed in line with the human experiment guidelines and regulations of Radboud University.

ABSTRACT To objectively evaluate weak neural responses such as auditory steady-state responses (ASSRs) often involves repetitive measurements, after which the accumulated trials are averaged or analyzed by statistical methods to detect a significant response. Such detection methods are often performed off-line on all measured trials. However, the number of required trials, and therefore the measurement time, is often heuristically chosen, and might be suboptimal. In this work, we compared three real-time signal-detection algorithms that could yield improved detection performance, at reduced measurement time. The classical Neyman-Pearson (NP) detector was evaluated by quantifying the signal-to-noise ratio (SNR), and detection probability of ASSRs, as a function of accumulated trial number. We also analyzed the performance of the Bayes factor (BF) to detect ASSRs at different thresholds. Finally, we modified the sequential probability ratio test (SPRT), with a cropped maximum likelihood (ML) estimator, such that it can detect ASSRs (with unknown SNRs) sequentially. We compared the three real-time detectors by using Monte Carlo simulation, and evaluated their performance on detecting ASSRs, generated from the superposition of a pair of amplitude-modulated (AM) tones near 40 Hz in the EEG of nine subjects with normal hearing. The low-order ASSRs (i.e., envelope frequency-following responses) have been sufficiently evaluated in the literature. However, the higher-order ASSRs might reflect further nonlinear mechanisms in the upper ascending auditory pathway. Results show that the real-time detectors can detect not only all low-order ASSRs but also higher-order ASSRs at frequencies with lower SNR. The NP detector yielded the best simulation and actual detection performance. For all subjects, the second-order ASSRs could already be detected with the NP and BF detectors within five trials, but more trials were needed for the modified SPRT detector. In general, higher-order ASSRs require more trials to detect, with the exception of ASSRs near 40 Hz and 80 Hz. In conclusion, compared with traditional off-line detectors, both (real-time) NP and BF detectors showed improved detection performance, and their application in EEG experiments can save valuable measurement time.

INDEX TERMS Auditory steady-state response, amplitude modulation stimuli, Bayes factor, EEG, high-order monaural beats, Neyman-Pearson detector, spectral F test, sequential probability ratio test.

I. INTRODUCTION

The auditory steady-state response (ASSR) is an envelope-following response [1] to periodic sound complexes, which manifests as stable brain oscillations that are locked to the frequencies present in the sound envelopes [2]. ASSRs are due to nonlinear signal processes that may be generated at different levels within the auditory pathway, ranging from as early as the cochlear nerve and subcortical sources, to the neocortex [3], [4]. Because of their reproducibility and involuntary nature, ASSRs have been considered a valid objective biomarker to assess hearing thresholds for clinical use and auditory system disorders [3], [5].

This characteristic of ASSRs thus allows objective response detection methods to extract the relatively weak ASSRs from noisy EEG measurements, by analyzing

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li.

a number of repetitive trials. The commonly used detection methods include the t-test and the spectral F test [6]–[8], the phase synchrony measure (PSM) [9] and the component synchrony measure (CSM) [10], the magnitude squared coherence (MSC) [11], the Hotelling T^2 (T square) test [12] and the circular T^2 test [13]. Comparative studies [6], [14], [15] have evaluated the reliability of these different methods, although no consensus has been achieved regarding the most efficient detection method.

Note that the aforementioned detection methods are mainly used off-line, after having measured a sufficiently large number of trials, judged necessary (a priori) to detect significant responses. However, it is so far not known: (i) how many experimental trials are required for detecting a significant response, and (ii) whether it is possible to determine the minimum number of required trials in real-time while performing an EEG recording experiment.

To answer these questions, we investigated the potential use of real-time signal detectors for analyzing ASSRs. In this study, we use the spectrum F test [7] combined with several real-time detectors, as it allows to easily estimate the signal-to-noise ratio (SNR) of a target response. This facilitates the evaluation of the relationship between detection performance and accumulated number of trials. Compared with off-line detectors, real-time detectors have the potential advantage to save valuable measurement time by not having to record more trials than needed. Thus, one could terminate the experiment as soon as the detected target signal reaches significance, or one may decide early on that the number of trials required will exceed the maximally allowed duration for the experiment. Furthermore, using more trials than needed is not a guarantee for optimal results, because of potential confounding factors that could interfere with the measurements, such as subjects' fatigue [16] or other nonstationarities in the EEG signals.

Existing off-line detection methods are mainly based on hypothesis testing, in which the null hypothesis (H_0) refers to the absence of a response to a stimulus vs. the alternative (H_1), which indicates the presence of a response [17]. A classical off-line method is the Neyman-Pearson (NP) detector [18], which can be readily applied in real-time [7]. The NP detector maximizes the probability of detection (P_D) of a target response with the constraint of a pre-defined Type-I error (e.g., $\alpha = 0.05$) for false detection. According to a frequentist approach, the NP detector is optimal for maximizing P_D, given a fixed threshold for the probability of a false alarm (P_{FA}). A fixed threshold for P_{FA} will also set a fixed threshold on the estimated SNR of ASSRs.

Alternatively, a Bayesian approach shifts the emphasis from hypothesis testing to an estimation based on confidence [19]. In Bayesian hypothesis testing, a common decision rule is based on the Bayes factor (BF), which is computed from the marginal likelihood ratio. The BF is the degree to which the model estimation shifts from prior to posterior probability. In the decision procedure, the BF is compared against a decision threshold, which is set by a criterion regarding the evidence. For example, the BF detector with the threshold of 3 has been employed as the evidence of the ASSR at 40 Hz [20]. In general, a BF between 3 and 10 is indicative for 'moderate' evidence for the alternative hypothesis, a BF between 10 and 30 indicates 'strong' evidence, and BF $>$ 30 indicates 'very strong' evidence [21], [22]. Here, we studied the reliability of the Bayes detector for ASSR identification by varying the BF. In addition, we evaluated the relationship between a BF detector and the decision rule of minimizing the Bayes risk.

For continuous sequential analyses, Wald [23] proposed a sequential probability ratio test (SPRT), which decides for the presence (or not) of a signal as soon as the likelihood ratio exceeds a certain upper bound (or falls below a certain lower bound). Otherwise, it continues to collect more data samples. SPRT is optimal when comparing two alternatives [24]. To extend the classical SPRT from single hypothesis testing to composite alternative hypotheses, the maxSPRT algorithm [25] employs a maximum likelihood (ML) estimate. However, so far the maxSPRT method has only been applied to simple Poisson and binomial distributions. To extend the use of SPRT to the F distributions (of SNRs) and to complete hypothesis tests (as SNR is unknown in this study), we modified the classical SPRT with a 'cropped' ML estimator, which can estimate SNR values and enable an early stop for data collecting when the estimated SNR falls below a predetermined lower threshold.

Figure 1 illustrates how we applied these three real-time detectors in this study. The real-time detectors will stop collecting more data as soon as a (positive or negative) decision is made. Each detector employed a different threshold strategy for the estimated SNRs of ASSRs from the spectral F test. Specifically, the NP detector uses a fixed threshold, the so-called NP critical value; the Bayes factor (BF) uses dynamic thresholds, which are updated with the arrival of each new trial; the modified SPRT method uses two (upper and lower) bounds to divide the estimated likelihood ratio into three decision zones (Z_0, Z_1, Z_u), which correspond to 'non-significant', 'significant', and 'unknown yet' (i.e., need more data). To validate these detectors, we applied the three real-time detectors to the EEG data from nine normal-hearing human subjects. From the results of our earlier study [5], we imposed a measurement time of 12.3 s for each EEG trial to ensure a sufficient frequency resolution. The maximum trial number (i.e., 100) was determined from one pilot measurement (on one subject) to ensure the detection for most of target ASSRs. Then, we set this maximum trial number for all experimental participants. To mimic real-time detection, we evaluated each detector on a trial-based manner, computed the detection probabilities, and reported the minimum number of trials needed to reach a significance decision for each ASSR frequency.

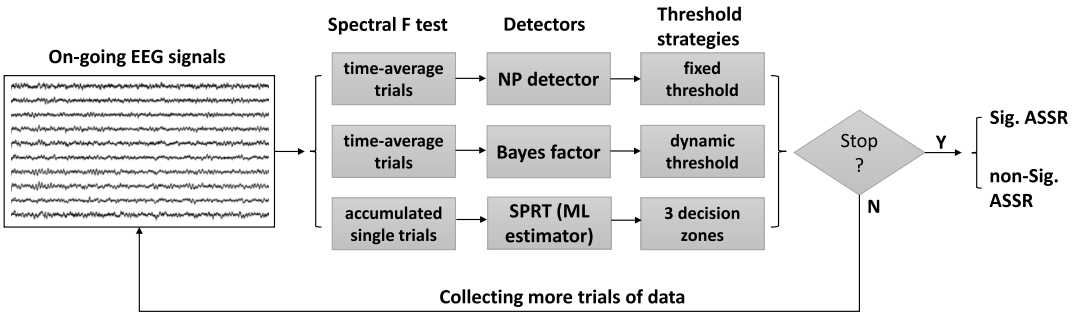

FIGURE 1. Application of three real-time ASSR detectors on EEG signals, using their specific thresholding strategies as described in the text. The NP and BF detectors apply the spectral F test on time-averaged EEG trials, whereas the modified SPRT detector applied it to accumulated single EEG trials.

II. METHODS

A. PROBLEM DESCRIPTION

A typical application of objective response detection is to detect the steady-state responses that are phase-locked to periodic stimuli, e.g., extracting sinusoids from noisy backgrounds. As the steady-state responses are often too weak to detect from one measurement trial, it is necessary to repeat many trials to determine the presence of a response.

There are mainly two types of approaches for detecting a steady-state response. First, statistical methods are used on accumulated trials at only the target frequency component. For example, phase coherence [26] measures the degree to which the phases in K measurement trials are clustered (i.e., signal is present) or randomly dispersed (i.e., noise) at the target frequency. Similarly, the Hotelling T^2 test and the circular T^2 test (CT2) have been proposed to quantify the amplitude and phase information. The latter is related to the magnitude squared coherence (MSC) measure [6]. In addition to consider only the target frequency components, the second type of approaches considers also neighboring frequency components around the target frequency. One commonly used method was the spectral F test [7], [27]. The performance of the spectral F test (see below) has been reported to be identical to the MSC and CT2 test [6], [14].

We here used the spectral F test, as it allows to readily estimate the SNR of the signal at the target frequency, which facilitates real-time detection of ASSRs. The goals of this study were to quantify the detection probability of an ASSR (second-order and higher orders) during the accumulation of EEG trials for different real-time signal detectors, and to assess to what extent valuable measurement time can be optimized.

Objective response detection is often performed in the frequency domain [14]. The null hypothesis H_0 is that the observed random frequency component, $x_0\,(k)$ (extracted from the Fourier transform at the target frequency of the k^{th} trial) contains only noise, with real part, R_k, and imaginary part, I_k. The alternative hypothesis, H_1, holds that the observed random variable, $x_1\,(k)$ contains both signal and noise, with the signal's real part, A, and imaginary part, 0, and the noise real part, R_k, and imaginary part, I_k. Note that the signal is considered to be a phase-locked ASSR component (cosine) with amplitude A and initial phase zero.

$$\begin{aligned} H_0 &: x_0(k) = R_k + jI_k, \quad k = 1, 2, \ldots, K \\ H_1 &: x_1(k) = (A + R_k) + jI_k, \quad k = 1, 2, \ldots, K, \end{aligned} \tag{1}$$

where $R_k, I_k \sim N(0, \sigma^2)$, and K is the maximum number of observation trials. The signal-to-noise ratio (SNR) is a function of amplitude A ($SNR = A^2/2\sigma^2$).

To estimate the unknown SNR from the EEG, a common method is to repeat the measurement with a certain number of trials. Then, the improved SNR can be estimated from time averaged signals across trials (see section 'Time-averaged trials'). Alternatively, the single-trial SNR can be directly estimated from accumulated single trials (as in the SPRT method). The problem is to minimize the number of trials needed for accepting H_1 (when SNR > 0) with desired detection performance, e.g., detection probability (P_D), acceptance confidence (AC), and constraints on false detections (see below sections).

B. TIME-AVERAGED TRIALS

Temporal averaging of sequential EEG trials can improve the SNR of a phase-locked weak response. The averaged signal across K trials is

$$\bar{x}_1(K) = \frac{1}{K} \sum_{k=1}^{K} x_1(k). \tag{2}$$

Given that $SNR^{(1)}$ is the SNR of a single-trial EEG signal $x_1\,(k)$, the SNR of K averaged trials, $\overline{SNR}^{(K)}$, is K times of $SNR^{(1)}$, or equivalently (see also Fig. 4),

$$\overline{SNR}_{dB}^{(K)} = 10\log_{10}(K) + SNR_{dB}^{(1)}, \tag{3}$$

where $SNR_{dB}^{(1)} = 10\log_{10}(SNR^{(1)})$. Whereas the SNR of time averaging noisy variables, $x_0\,(k)$, will not change, as SNR $= 0$ ($-\infty$ dB).

C. SPECTRAL F TEST

The spectral F-test is used to estimate the SNR of the time-averaged signal $\bar{x}_1(K)$. The spectral F-score on a target

frequency bin, $F(f_t)$, is calculated by the relative neighboring SNR around the target frequency of the ASSR:

$$F(f_t) = \frac{|X(f_t)|^2}{\frac{1}{M}\sum_{i=1}^{M}|X(f_i)|^2}, \tag{4}$$

where $X(f_t)$ is the Fourier transform of a time-averaged signal at the target frequency bin, f_t, obtained from a single trial, or from the time-averaged trials, $\bar{x}_1(K)$. M is the number of neighboring frequency bins used for estimating the SNR. Note that the F score is not an unbiased estimate of SNR (see Appendix) [6]. The relationship between the expected F score and SNR is:

$$E(F) = SNR + 1. \tag{5}$$

Thus, the single-trial SNR is estimated by $SNR^{(1)} = \frac{1}{K}(F-1)$, where F (>1) is computed from the time-averaged signal with trial number K. In practice, F is often reported in F (dB), i.e., $F_{dB} = 10\log_{10}(F)$. Thus, the single-trial $SNR^{(1)}_{dB}$ is estimated by

$$SNR^{(1)}_{dB} = 10\log_{10}(\frac{10^{(F_{dB}/10)}-1}{K}). \tag{6}$$

Given H_0 (no signal present), $F(f_t)$ will have an F distribution with (2, 2M) degrees of freedom. Given H_1 (signal present), $F(f_t)$ will have a noncentral F distribution with (2, 2M) degrees of freedom, and the non-centrality parameter, 2SNR [7], as below.

$$F(f_t) \sim \left\{ \begin{array}{ll} F_{2,2M}, & under\ H_0 \\ F_{2,2M}(2SNR), & under\ H_1 \end{array} \right\}, \tag{7}$$

where SNR can be estimated from $F(f_t)$ of the time-averaged signal. See examples of probability density functions (PDF)s of F distributions (H_0) and noncentral F distributions (H_1) in Fig. 11 in Appendix.

D. NEYMAN-PEARSON DETECTOR

The Neyman-Pearson (NP) detector [18] maximizes the probability of detecting an ASSR component, P_D, under a fixed constraint for the probability of a false alarm, P_{FA}, a so-called type I error (α). A type I error is the rejection of a true null hypothesis (also known as a false positive) [28]. The NP detector will accept H_1 (signal present), if the likelihood ratio

$$\Lambda(x) = \frac{p(x;H_1)}{p(x;H_0)} > \gamma. \tag{8}$$

Otherwise, the NP detector will decide H_0. Note that γ is determined from $P_{FA} = \int_{\{x:\Lambda(x)>\gamma\}} p(x;H_0)dx = \alpha$. Accordingly, a fixed threshold of x, so-called NP critical value (NP_{crit}), is determined from the PDF of H_0, such that,

$$\int_{NP_{crit}}^{\infty} p(x;H_0)dx = \alpha. \tag{9}$$

The detection rate, P_D, hence is computed as,

$$P_D = \int_{NP_{crit}}^{\infty} p(x;H_1)dx. \tag{10}$$

E. BAYES-FACTOR DETECTOR

The classical Bayes risk (R) is defined by [18]:

$$R = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}\pi_j P(H_i|H_j), \tag{11}$$

where C_{ij} represents the cost of selecting hypothesis H_i when hypothesis H_j is true; $P(H_i|H_j)$ is the probability of selecting hypothesis H_i conditional on that H_j is true. π_j is the prior probability of the hypothesis H_j ($\pi_0 + \pi_1 = 1$). Minimizing the Bayes risk (R) will lead to the decision rule accepting hypothesis H_1 if:

$$\Lambda(x) = \frac{p(x|H_1)}{p(x|H_0)} \geq \frac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})} = \eta, \tag{12}$$

where $\Lambda(x)$ is the marginal likelihood ratio, referred to as the Bayes factor (BF) [21] in Bayesian hypothesis testing, and η is the BF threshold, which is a lower bound for accepting H_1 and an upper bound for accepting H_0.

1) RELATIONSHIP BETWEEN THE BAYES RISK AND TWO TYPES OF ERRORS

When assigning a null cost to correct decisions (i.e., $C_{00} = C_{11} = 0$), and a unit cost to wrong decisions (i.e., $C_{01} = C_{10} = 1$), minimizing the Bayes risk is identical to minimize the sum of probabilities of two types of errors, as shown below.

$$R = \pi_0 P(H_1|H_0) + \pi_1 P(H_0|H_1) \triangleq PE. \tag{13}$$

That is, $PE = \pi_0 P_{FA} + (1-\pi_0)(1-P_D)$, where $P_D = \int_{\{x:\Lambda(x)>\eta\}} p(x|H_1)dx$ and $P_{FA} = \int_{\{x:\Lambda(x)>\eta\}} p(x|H_0)dx$.

With equal prior ($\pi_0 = \pi_1 = 0.5$),

$$PE = 0.5(P_{FA} + (1-P_D)). \tag{14}$$

The probability of correct decision is

$$PC = 1 - PE. \tag{15}$$

Note that a predefined BF threshold (e.g., $\eta = 1$) will determine different thresholds of x (denoted by θ_F later) for different SNRs that change when averaging more trials. Hence, we call θ_F a dynamic threshold on F. See an illustration ($\eta = 1$) in Fig. 11.

2) RELATIONSHIP BETWEEN BF AND ACCEPTANCE CONFIDENCE

Besides, according to the Bayes' theorem, the posterior probability can be written as follows,

$$\frac{P(H_1|x)}{P(H_0|x)} = \frac{P(x|H_1)\pi_1}{P(x|H_0)\pi_0} = \Lambda(x)\frac{\pi_1}{\pi_0}. \tag{16}$$

For a binary hypothesis test, $P(H_1|x)+P(H_0|x) = 1$. It then follows that

$$P(H_1|x) = \frac{\pi_1 \Lambda(x)}{\pi_0 + \pi_1 \Lambda(x)}. \quad (17)$$

Assuming equal prior ($\pi_0 = \pi_1 = 0.5$) in the absence of prior knowledge of either hypothesis before testing, the posterior probability $P(H_1|x)$, also referred to as the acceptance confidence (AC) [29], is then defined by:

$$AC = \frac{\Lambda(x)}{1 + \Lambda(x)}. \quad (18)$$

Therefore, $\Lambda(x) = 0$ indicates zero confidence for accepting H_1, and $\Lambda(x) \to \infty$ indicates 100% confidence. In contrast to PC, AC is independent of the BF threshold, η, and it thus can objectively assess the confidence (between 0 and 1) to accept the alternative hypothesis. The chosen BF threshold η will determine a fixed threshold θ_{AC} for AC:

$$\theta_{AC} = \frac{\eta}{1 + \eta}. \quad (19)$$

Thus, θ_{AC} is between 0.5 and 1 with $\eta \geq 1$, and in practice η is often chosen as an integer larger than 1.

F. MODIFIED SEQUENTIAL PROBABILITY RATIO TEST (SPRT)

1) STANDARD SPRT

According to the definition of the standard SPRT [23], a sequence of independent and identically distributed (iid) F-score observations for each trial, under hypotheses H_0 and H_1, is described as:

$$H_i : F_1, F_2, \ldots, F_k \overset{iid}{\sim} p_i, \quad i = 0, 1. \quad (20)$$

Rather than fixing n observations, SPRT provides a sequential approach to test whether it is necessary to continue gathering samples until a confident decision regarding the presence (H_1) or absence (H_0) of a signal can be made. The sequential probability ratio is defined as

$$\Lambda(k) = \ln\left(\prod_{i=1}^{k} \frac{p_1(F_i)}{p_0(F_i)}\right), \quad k = 1, 2, \ldots \quad (21)$$

To implement SPRT, two thresholds, γ_1 and γ_0, ($\gamma_1 > \gamma_0$), are preset. if $\Lambda(k) > \gamma_1$, one accepts H_1 and stops collecting more data. Alternatively, if $\Lambda(k) < \gamma_0$, one accepts H_0 and stops collecting data; else continue.

By specifying two types of error probabilities, $0 \leq \alpha, \beta \leq 1$, where α and β are the probabilities for a false positive (i.e., $P(H_1|H_0)$) and false negative (i.e., $P(H_0|H_1)$) decision, respectively, SPRT will accept H_1 if $\Lambda(k) \geq \ln \frac{1-\beta}{\alpha}$, and will accept H_0 if $\Lambda(k) \leq \ln \frac{\beta}{1-\alpha}$. Note that SPRT is conservative, the actual detection rate P_D and false alarm rate P_{FA} need to be computed from simulations [24].

2) MODIFIED SPRT WITH A CROPPED ML ESTIMATOR

The standard SPRT applies only when the PDFs under both hypotheses (SNRs) are known. We extended the standard SPRT to allow for composite hypotheses testing (where SNR is unknown) by using a cropped maximum likelihood (ML) estimate, which estimates the PDF parameter (i.e., SNR) from currently available observations. Thus, the log-likelihood ratio (LLR) reads

$$\Lambda(k) = \ln\left(\prod_{i=1}^{k} \frac{p_1(F_i|SNR = SNR_{ML}^{k})}{p_0(F_i)}\right), \quad k = 1, 2, \ldots \quad (22)$$

where the estimated SNR from the cropped ML is defined as,

$$SNR_{ML}^{k} = \max\{SNR_\theta, \arg\max_{SNR}\left(\prod_{i=1}^{k} p_1(F_i(SNR))\right)\}, \quad (23)$$

where SNR_θ is the predetermined low band of the estimated SNR (e.g., -13 dB for a P_D of 0.8). The expected stop time (trial numbers) of the SPRT, and its relationship to P_D, is provided in the Appendix D.

G. SIMULATION METHODS

Monte Carlo simulations are often used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables (in this case, the spectral F distribution). Monte Carlo simulation is a technique that uses repeated random sampling to obtain the likelihood for the occurrence of target results [30]. Instead of using traditional statistical methods (e.g., t-test) that compare the obtained p-value against a threshold (e.g., 0.05) to make a binary decision (i.e., significant or not), we here used Monte Carlo simulation to compute the probability of detection, P_D (i.e., the likelihood of exceeding a threshold) for the three detectors, and also to compare results with the theoretical prediction of P_D based on the PDFs of the spectral F distributions for the NP detector (see Fig. 3).

The signals (H_1, SNR > 0) and noise (H_0, SNR $= 0$) for simulation were generated by using (1). For signals, the amplitude A and standard deviation ($\sigma = 1$) were chosen such that SNRs varied between -30 dB and $+20$ dB, which will cover the potential range of actual ASSRs in EEG signals. For noise (no signal present), the amplitude A $= 0$ and standard deviation $\sigma = 1$. The number of neighboring frequency bins, M (in (4)), was affected by the frequency resolution and the measurement duration of each trial, which can vary from seconds to several minutes in different studies [6]. To simulate the spectral F test, we evaluated different values for M (4, 6, 12, 24). For the ASSR detection in EEG, we used M $= 12$ and the width of each frequency bin of 1/12 Hz. For each SNR, we generated 100,000 (n = 1e5) samples of signals (H_1) and M*n samples of neighboring frequency bins (H_0) for the Monte Carlo simulation.

For both NP and BF detectors, each SNR (every 1 dB between -30 and $+20$ dB) was specified to analyze the

detection performance. The PDF of H_1 was thus determined for each SNR. To assess the detection performance, we performed both theoretical analysis and numerical simulation to compute P_D and PC. Specifically, after computing the fixed NP critical value (NP_{crit}) for the NP detector and the threshold θ_F (determined by SNR) for the BF detector, we used (9-10) and (13-15) to compute the theoretical values of P_D and PC. The corresponding numerical results of P_D were computed by counting the number of cases (n = 1e5) in which F exceeded the thresholds (NP_{crit} or θ_F for the NP or BF detectors, respectively). The modified SPRT detector, instead of specifying an SNR, estimated a (single-trial) SNR by using a cropped ML estimator, P_D was obtained from the numerical simulation (n = 1e4 for the Monte Carlo simulation). Therefore, the modified SPRT could be readily applied to actual EEG signals, for which the SNR is unknown. In contrast, the NP and BF detectors need to estimate the SNR from time-averaged trials by using (5), or from a single-trial SNR by using (6). The expected F can be obtained from time-averaged trials by using (4).

Matlab (R2018a) was used for the simulations and for the ASSR detection in recorded EEG. We used Matlab's 'Statistics and Machine Learning Toolbox', and built-in functions including normrnd.m (for random number generation) for Monte Carlo simulations. For the NP detector, given a type-I error (α), the NP critical value is a function of the number of neighboring frequency bins M, i.e., $NP_{crit} = F^{-1}(1-\alpha, 2, 2M)$ (with $F^{-1}(x)$ the inverse of the cumulative F-distribution, calculated by using the function 'finv.m'). For a predefined BF threshold η (e.g., η = 1, 3, and 6 in this study), thresholds of F (i.e., θ_F) were computed for each SNR from (12) and (7). To implement (7), Matlab's functions fpdf.m and ncfpdf.m were used to quantify PDFs of both F distributions (H_0) and noncentral F distributions (H_1). To help reproduce this study, we have made the Matlab scripts for the simulations available online: https://github.com/ieeeWang/ASSR-Realtime-Detectors

H. EEG EXPERIMENTS

Amplitude modulation (AM) is the elementary feature of natural stimuli, which could reveal dynamic properties of the auditory system not addressed with simpler, static stimuli [3]. The contour of an AM sound is the envelope, which shows a single sinusoidal component corresponding the modulation frequency (MF). The low-order ASSRs (i.e., brain responses found at the MFs) generated from AM tones with a range of carrier frequencies are used to obtain objective audiometry [31], [32]. Here, we used an ensemble of a pair of AM tones to represent more complex stimuli in the real world, and to generate a number of higher-order ASSRs, which include harmonics and intermodulation of the MFs [33]. Studies have shown that ASSRs tend to reach maximum responses near 40 Hz [32], which may reflect a superposition of auditory brainstem responses and middle latency responses [34]. We therefore selected a pair of MFs at 37 and 43 Hz to ensure well-detectable ASSRs that would also include a series of unique higher-order ASSR frequencies (See Table 3).

We evaluated performance of the three real-time detectors on the detection of ASSRs, generated from AM sounds in the EEG of nine normal-hearing subjects (<30 dB SPL HL) in a passive listening task as shown in Fig. 2A. The AM sound stimulus used in this study was given as:

$$x(t) = [1 + m\cos(2\pi f_m t)]\cos(2\pi f_c t) \tag{24}$$

where m = 1 (i.e., 100% amplitude modulation), and the initial phase set to zero. The MF and carrier frequency (CF) are denoted by f_m and f_c, respectively. For each trial, t = 12.3 sec with the first 0.3 sec (evoked potentials, that included non-ASSRs, which was called the pre-steady-state phasic response phase [35]) excluded for analysis. We presented the superposition of two AM tones to the left ear with the same CF of 500 Hz, and two MFs of 37 and 43 Hz (see Fig. 2B). Stimuli were generated by TDT (Tucker-Davis Technologies) System-3 hardware, and presented through ER3C insert earphones (Etymotic Research), which were connected to the listeners' ears via 30 cm-long plastic tubes and foam earplugs. The stimuli were calibrated by using the Bruel & Kjaer sound-level calibration meter (model 2260), such that the level of each AM tone was 60 ± 1 dB SPL. We confined this study to the detection of monaural beats from the left ear only. Thus, the stimuli in the right ear and potential binaural beat responses were not included and analyzed here. We repeated 100 trials on each subject. The experiments were approved by the ethics committee at Radboud University and performed in accordance with the human experiment guidelines and regulations of Radboud university.

We applied a three-stage preprocessing protocol on the raw EEG recordings as described in our previous study [5]. First, EEG signals were re-referenced to a common average reference (CAR) after excluding the EEG electrodes above the eyes due to eye-blink and eye-movement artifacts. Second, each EEG channel was filtered with a zero-phase shift filter that consisted of a 10^{th}-order Butterworth high-pass filter with a cutoff frequency of 1 Hz, and two notch filters at 50 Hz and 100 Hz to remove the line noise. Third, to exclude severely contaminated EEG trials by electromyographic artifacts from further analysis, we computed the mean and the variance of the peak-to-peak amplitude range as indicators for each EEG trial. The EEG trials with the two indicators as positive outliers were excluded (up to 5% of 100 trials) from the analysis.

It was found that ASSRs show a general strong activation in the frontal-central brain [5], [36]. Therefore, in this study, to illustrate the real-time detectors, we performed the ASSR detection for all subjects on the same frontal-central brain region (near EEG channel FCz). The five EEG channels (Fz, Cz, FCz, FC1, FC2) around FCz were averaged as one (denoted as AVG-FCz) to improve the robustness of EEG signals over one signal channel.

Note that ASSRs are generated because of the nonlinear mechanisms in the auditory system [3], [37].

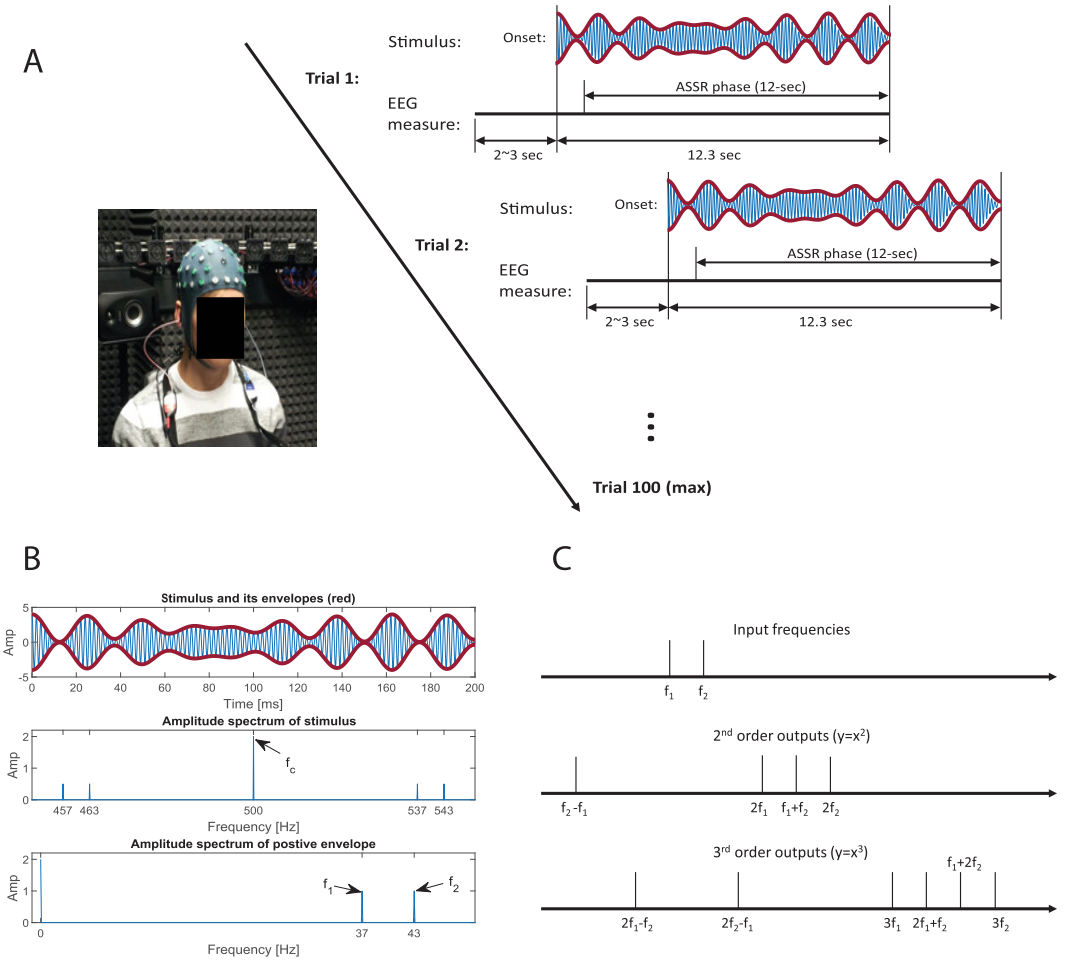

FIGURE 2. Experimental design, i.e., a passive listening task (A), sound stimulus (B), and expected nonlinear ASSRs (C). (A) The sound stimuli (through ER3C insert earphones) were present (trial by trial) to both ears of a normal-hearing subject while the EEG signals were continuously measured (64-channel cap) from the listener sitting in an anechoic chamber, watching a silent video. Each trial started with a silent break with a random duration between 2 and 3 sec, and lasted 12.3 seconds. The first 0.3-sec EEG of each trial was excluded from the analysis to avoid the influence of event-related potentials (i.e., non-ASSRs), and the remaining 12-sec EEG segment (i.e., ASSR phase) was used to detect significant ASSRs. We repeated maximum 100 trials for each subject. Note the stimulus illustrated here were present to the left ear, and we showed only 200 ms at the beginning of the stimulus for clear visualization. (B) The stimulus (at the left ear) for each trial was the superposition of two AM tones (each with 60 dB SPL) with the same CF of 500 Hz, and two MFs of 37 and 43 Hz. The waveform (blue) and its envelopes (red) were shown in the top panel. The amplitude spectrum of the stimulus and the positive envelop (in red) were shown in the middle panel and bottom panel, respectively. (C) The nonlinear system outputs (i.e., ASSRs) were demonstrated with a 2^{nd}-order system and a 3^{rd}-order system. Note that the two MFs (37 and 43 Hz) are already 2^{nd}-order outputs of the stimulus frequencies (i.e., the difference between two stimulus frequencies as shown in the middle pane of B). Therefore, the 2^{nd} and 3^{rd} order outputs of the MFs correspond to 4^{th} and 6^{th} outputs (ASSRs) of the stimulus, respectively. See these ASSR frequencies in Table 3.

The 2^{nd}-order nonlinear ASSRs (also known as binaural and monaural beats) were sufficiently evaluated in previous studies [3], [36], and they were thought to already arise at the auditory nerve [38]. However, in the ascending auditory pathway, the nonlinear mechanisms might also evoke higher-order nonlinear components, which in turn could give rise to additional ASSRs. The higher-order ASSRs often cover a larger frequency range and have been used to estimate multiple frequency-related apparent latencies of the auditory system [5]. For a nonlinear system with system order R, given two input frequencies (f_1 and f_2), a series of combination frequencies is produced at the output, characterized by $nf_1 \pm mf_2$ (> 0), with n and m positive integers, such that $n + m \leq R$ [5], [39]. See Fig. 2C for an example of $R = 2$ and 3. In this study, by using a pair of MFs as stimuli, we expect significant ASSRs at not only the MFs (also known as the envelope frequencies, which are the 2^{nd} order ASSRs), but also higher-order ASSRs (4^{th} and 6^{th}), which correspond to the 2^{nd} and 3^{rd} order outputs of the MFs, respectively. These higher-order ASSRs often show lower SNRs, and are therefore harder to detect [5]. Table 3 presents the relevant ASSR frequencies. So far, few studies have

systematically evaluated the detection task for such higher-order ASSRs in the EEG responses of the human auditory system.

I. ASSR DETECTION PERFORMANCE CRITERIA

All potential ASSR frequencies were considered as positive class while non-ASSR frequencies as negative class. The detection performance was assessed in two ways. First, receiver operating characteristic (ROC) curves were plotted and the area under curve (AUC) was used to quantify the overall performance [6]. Second, for unbalanced class distributions, the precision and recall (P-R) curves and its AUC have been cited as an alternative to ROC curves [40], [41]. Also, algorithms that optimize the AUC of the ROC curve are not guaranteed to optimize the AUC of the P-R curve [42]. Therefore, P-R curves were also used here because we employed a large non-ASSR frequency set (n = 88) as the negative class (control group), which significantly outnumbered the positive class (n = 12). Each point in ROC and P-R curves represents a specific classifier with a specified threshold that determines a positive sample if this sample is larger than the threshold. We used a full set of threshold values from all available samples to ensure a maximum 'resolution' plot. Subsequently, the interpolation method [42] was used to estimate the AUC of both curves. Furthermore, to show the effect of parameters of these detectors (e.g., values of α), we reported the commonly used performance indicators including sensitivity (i.e., TP/#positive class), specificity (i.e., TN/#negative class) and precision (i.e., TP/#positive detection; 1-precision is known as false detection rate), where TP and TN denote the numbers of true positive samples and true negative samples, respectively.

III. RESULTS

A. SIMULATIONS

1) NEYMAN-PEARSON (NP) DETECTOR

Figure 3 shows the detection performance of the NP detector on each SNR value. The lower panel shows that the detection probability, P_D, of the signal increases with SNR. In addition, P_D slightly increases with the number of neighboring frequency bins, M (see Fig. 12 in Appendix).

Table 1 shows that the critical value of the NP detector is affected by both M and the type-I error α (i.e., P_{FA}).

Temporal averaging of a number of EEG trials leads to an improved SNR for the averaged trial. Figure 4 (A, top) shows the improved SNR (and corresponding P_D, bottom) as a function of the number trials used for the averaging, for signals with different (single-trial) SNRs. For example, when $SNR_{dB}^{(1)} = 0$ dB, after averaging 10 trials, the P_D is very close to 1.

Table 2 shows a slice of information extracted from Fig. 4 (B). It shows, for example, that the minimum single-trial SNRs (dB), needed for achieving $P_D = 0.8$, are -3, -13, and -23 dB when time-averaging over $N_{trials} = 10$, 100, and 1000, respectively.

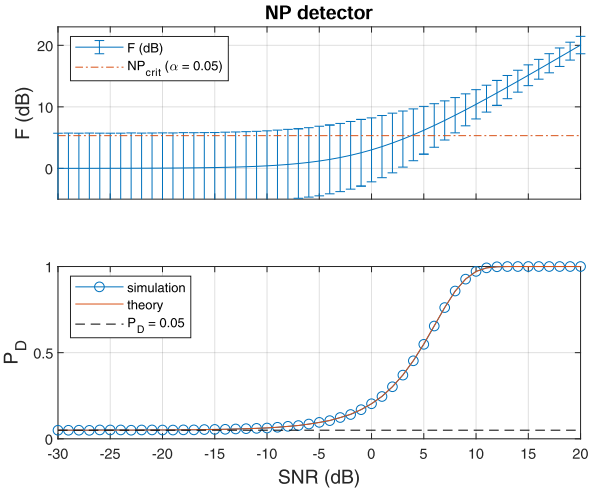

FIGURE 3. Performance of the NP detector as a function of the SNR (with M = 12). (top) Expected value of the F (dB) (i.e., 10log10(F)) with the 95% confidence interval (vertical bars); The critical value of the NP detector for a probability of a false alarm of 5% (red dashed line at NP_{crit} = 5.32 dB); (bottom) The probability of detection (P_D) obtained in the simulations (denoted by circles, n = 1e5 for Monte Carlo simulations) overlaid on the theoretical line (red line; see methods).

TABLE 1. Critical values of F for the NP detector.

M (# freq bins)	4	6	12	24
NP_{crit} ($\alpha = 0.05$)	4.46 (6.49 dB)	3.89 (5.89 dB)	3.40 (5.32 dB)	3.19 (5.04 dB)
NP_{crit} ($\alpha = 0.01$)	8.65 (9.37 dB)	6.93 (8.41 dB)	5.61 (7.49 dB)	5.08 (7.06 dB)

TABLE 2. Minimum single-trial SNR (dB) needed for each P_D.

P_D	0.5	0.6	0.7	0.8	0.9	1
N_{trials} =10	-6	-5	-4	-3	-1	9
N_{trials} =100	-16	-15	-14	-13	-11	-1
N_{trials} =1000	-26	-25	-24	-23	-21	-11

P_D was computed with $\alpha = 0.05$,
N_{trials} is the number of accumulated EEG trials.

2) BAYES-FACTOR (BF) DETECTOR

In our simulation, the PDFs of H_1 and H_0 were both known because the SNR value was specified. Given a specified BF threshold η, we computed the dynamic thresholds of F (i.e., θ_F) by searching for the F that satisfied the relationship of the two PDFs, $p(F|H_1) = \eta p(F|H_0)$, according to (12). See Fig. 11 (in Appendix) for an illustration of θ_F that was computed for a number of different SNR values when $\eta = 1$. When the BF threshold η increases, the θ_F will move to the right side of the according intersection of both PDFs (H_1 and H_0), yielding a larger θ_F. However, since it is difficult to obtain an analytical solution of θ_F, we performed a numerical approximation with an SNR resolution of 0.1 dB (more details were given in our open-access code), and assessed the BF detector with three typical thresholds of $\eta = 1, 3$ and 6. For each value of η, the obtained dynamic thresholds (θ_F) were a function of E(F) (as shown in Fig. 5A) and this function

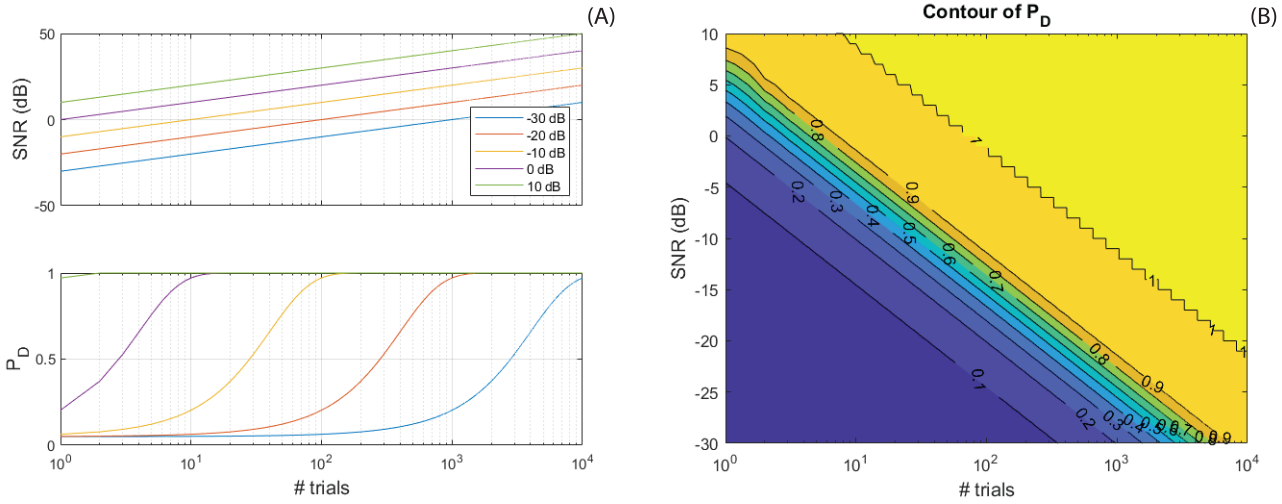

FIGURE 4. NP detector: Effect of time-averaging. (A) The SNR (top) and probability of detection (P_D) (bottom) both increase with the number of time-averaged trials, which is demonstrated for different values of single-trial SNRs, $SNR_{dB}^{(1)} = -30$ dB, -20 dB, -10 dB, 0 dB and $+10$ dB, respectively. (B). Contour plot of P_D as a function of SNR and the accumulated trial numbers (SNR resolution is 1 dB for this simulation).

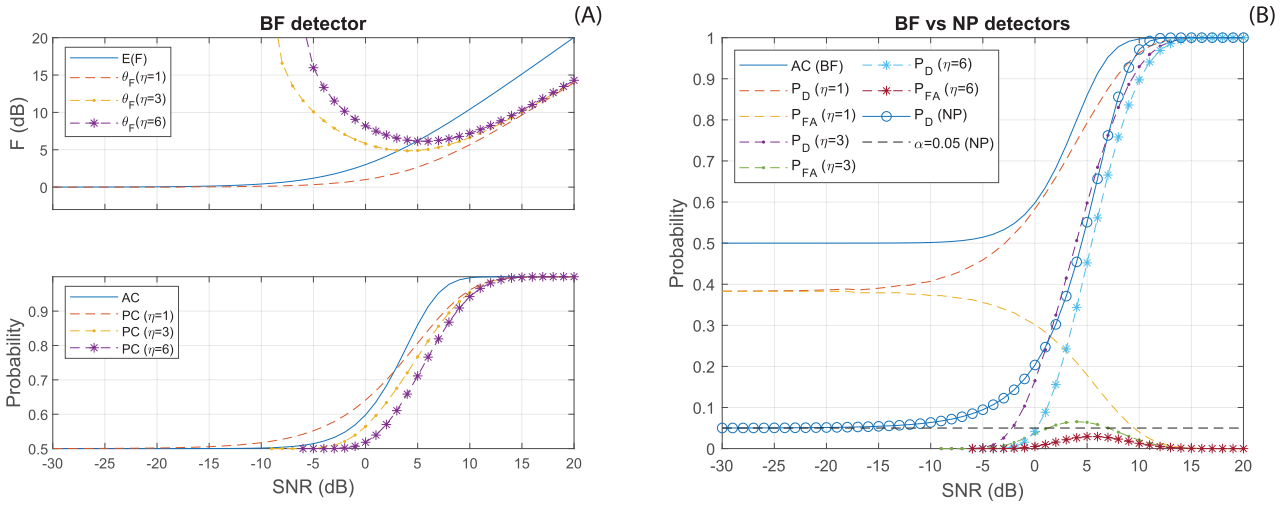

FIGURE 5. (A) Performance of the BF detector as a function of the SNR: (top) The expected F scores, E(F), and the thresholds of F (θ_F) corresponding to $\eta = 1$, 3, and 6. (bottom) the AC determined from E(F). The values of PC were computed by 1-PE for $\eta = 1$, 3, and 6. (B) Performance comparison between the BF and NP detectors. The BF detector with three thresholds ($\eta = 1$, 3, and 6) yields three pairs of P_D and P_{FA} values. When $\eta = 1$, P_D and P_{FA} converge to 0.38 at -30 dB, respectively, and when $\eta = 3$ and 6, P_D and P_{FA} were computed for SNRs larger than -9 dB and -6 dB, respectively. AC does not depend on η. The P_D of the NP detector was computed for $\alpha = 0.05$.

was also applied to the ASSR detection in the EEG, with the E(F) estimated from (4).

As described in Methods, the BF detector with $\eta = 1$ is equivalent to the decision rule that minimizes the Bayes risk by assigning a unit cost to wrong decisions (zero cost to correct decisions), with equal prior probabilities. Figure 5A shows the dynamic thresholds of F (θ_F) determined by η (top panel), and corresponding PC values (bottom panel). The minimum value of PC is 0.5 because it considers the probability of correct decisions for both hypotheses. In contrast, P_D (with the minimum value of zero) of the NP detector considers only correct detection for hypothesis H_1. The AC was computed from the estimated SNR from F scores and thus did not depend on η. Note that both AC and PC show the same range (between 0.5 and 1). However, it is not to be expected that PC and AC are equal because the former quantifies the probability of correct decisions, while the latter quantifies the posterior probability of H_1.

The comparative analysis in Fig. 5B compares both P_D and P_{FA} between the NP and BF detectors. The BF($\eta = 1$) detector yields a high false detection rate (approaching 38%) when the SNR is low, which is thus not feasible in practice. Compared with the NP detector ($\alpha = 0.05$), BF($\eta = 3$) yields a slightly

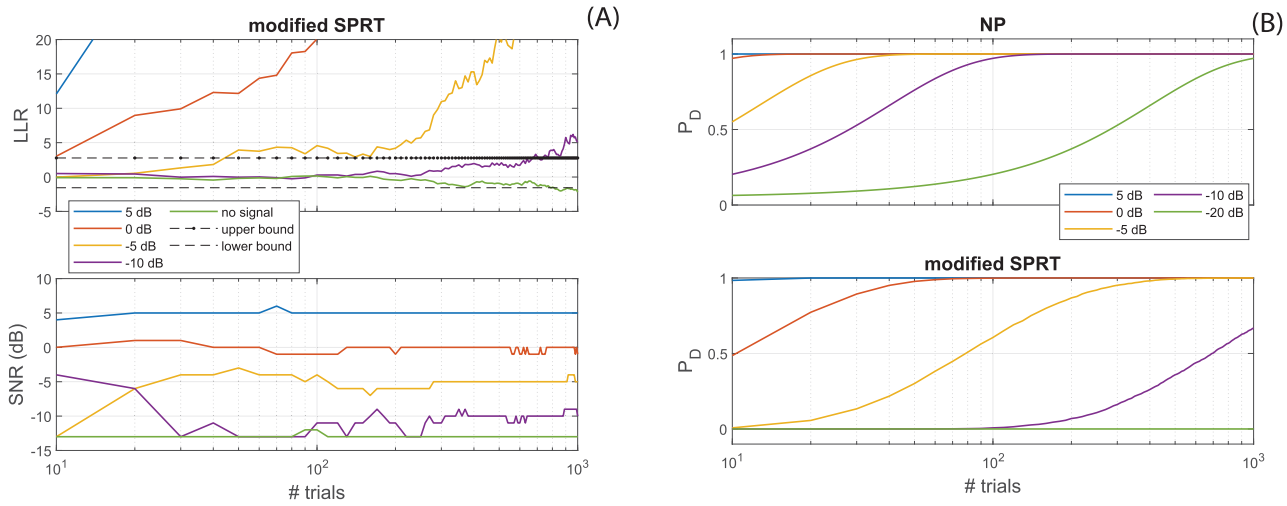

FIGURE 6. (A) A simulation of the modified SPRT with a cropped ML estimator ($SNR_\theta = -13$ dB). (top) LLR of each signal on accumulated trials. The upper and lower bounds of SPRT are 2.77 and −1.56, respectively (with $\alpha = 0.05$, $\beta = 0.2$). (bottom) The estimated SNR from the cropped ML estimator, which updates every ten samples (trials). (B) The detection probability P_D of several signals with single-trial SNRs (5, 0, −5, −10, −20 dB) as a function of accumulated trials by using NP and the modified SPRT detectors (n = 1e4 for the Monte Carlo simulations).

higher P_D, but also a higher P_{FA} (between 1 and 7 dB), while BF($\eta = 6$) yields an overall lower P_D and P_{FA}. Furthermore, the BF($\eta = 6$) will not generate a positive signal detection for lower SNR (e.g., <-6 dB) to ensure an extremely low P_{FA}. Therefore, the BF($\eta = 6$) can be viewed as a more 'conservative' detector than the NP detector. The BF detector yields a dynamic P_{FA} that is affected by the SNR. In contrast, the P_{FA} for the NP detector remains constant (e.g., 5% with $\alpha = 0.05$). In general, the false detection rate of the BF detector declines with increasing η.

3) MODIFIED SEQUENTIAL PROBABILITY RATIO TEST (SPRT)

To estimate the unknown SNRs and to enable an early stop for accepting the null hypothesis (no signals), we modified the standard SPRT with the cropped ML estimator, which has a lower bound for estimating the SNR of weak signals, i.e., lower SNRs will be replaced by a predetermined lower bound. Here, we set the lower bound at −13 dB, which was chosen on the basis of Table 2, from which we determined that the minimum single-trial SNR was −13 dB to reach a $P_D = 0.8$ within 100 trials.

Figure 6A illustrates the results for a simulation, in which several signals were generated from the non-central F distribution with the non-centrality parameter (see Methods). Signals with SNRs exceeding the lower bound were detected (i.e., LLR beyond the upper bound), while the LLR of noise (i.e., F distribution with M = 12) fell below the lower bound, and hence was indicated as non-significant. We repeated this simulation 10,000 times to calculate the detection rate (P_D). As SPRT has been reported to be a 'conservative' detector, Fig. 6B shows that the P_D of several signals by the modified SPRT were much lower than obtained for the NP detector. Note that the P_D provided by the NP detector represents the chance exceeding NP_{crit} (here, set at $\alpha = 0.05$). In contrast, the P_D of SPRT represents the probability exceeding the LLR upper bound, here at 2.77, (i.e., $\alpha = 0.05$, $\beta = 0.2$). The lower value of P_D for SPRT results from the extra constraint (i.e., $\beta = 0.2$), which means that only signals for which $P_D > 0.8$ will surpass the LLR upper bound, and will be counted as a significant detection.

B. ASSR DETECTION IN EEG DATA

1) ESTIMATION OF NUMBER OF REQUIRED TRIALS

We performed a pilot experiment ($N_{trials} = 100$) with subj #1 to determine the required number of experimental trials for detecting ASSRs. We evaluated the monaural beats up to 6^{th} order (see Table 3) generated from MFs at 37 Hz and 43 Hz presented to the left ear, and computed the spectral F of the ASSRs from the time-averaged EEG (100 trials of 12 s). Then, the single-trial SNR was estimated from the F-score (by (6)). From the results in Table 2, we estimated the required number scales of trials, which are given in the bottom row of Table 3. These results show that most ASSRs could be detected within 100 trials with $P_D > 0.8$. Only the 6^{th} order ASSR at 111 Hz could not be detected, and the two ASSRs at 6 Hz (4^{th} order) and 129 Hz (6^{th} order) required more trials (10^3) to detect. Therefore, we set the maximum number of trials to 100 for all subjects to perform the remaining experiments.

2) ASSR DETECTION RESULTS

Figure 7 illustrates the detection process of the three real-time detectors on the averaged EEG channel (AVG-FCz) of subject S1 for three example ASSRs, indicated as MF (43 Hz), 2MF (80 Hz) and 3MF (117 Hz), respectively. For the NP (Fig. 7A) and BF detectors (Fig. 7C, D), the spectral F was computed on time-averaged accumulated trials up to the maximum of 100 trials. The difference between the

TABLE 3. Estimated SNR and required trial numbers for ASSR detection (subj#1).

Order	2^{nd}		4^{th}				6^{th}					
ASSR (Hz)	37	43	6	74	80	86	31	49	111	117	123	129
F (dB)	25	26	1.9	13.9	21.4	12	8.6	12.9	-10.8	10.3	6.8	1.6
$SNR_{dB}^{(1)}$	5	6	-23	-6	1	-8	-12	-7	$-\infty$	-10	-14	-24
# Trials*	10	10	10^3	10^2	10	10^2	10^2	10^2	∞	10^2	10^2	10^3

$*$ estimated number of trials for $P_D > 0.8$.
∞ denotes not detectable.

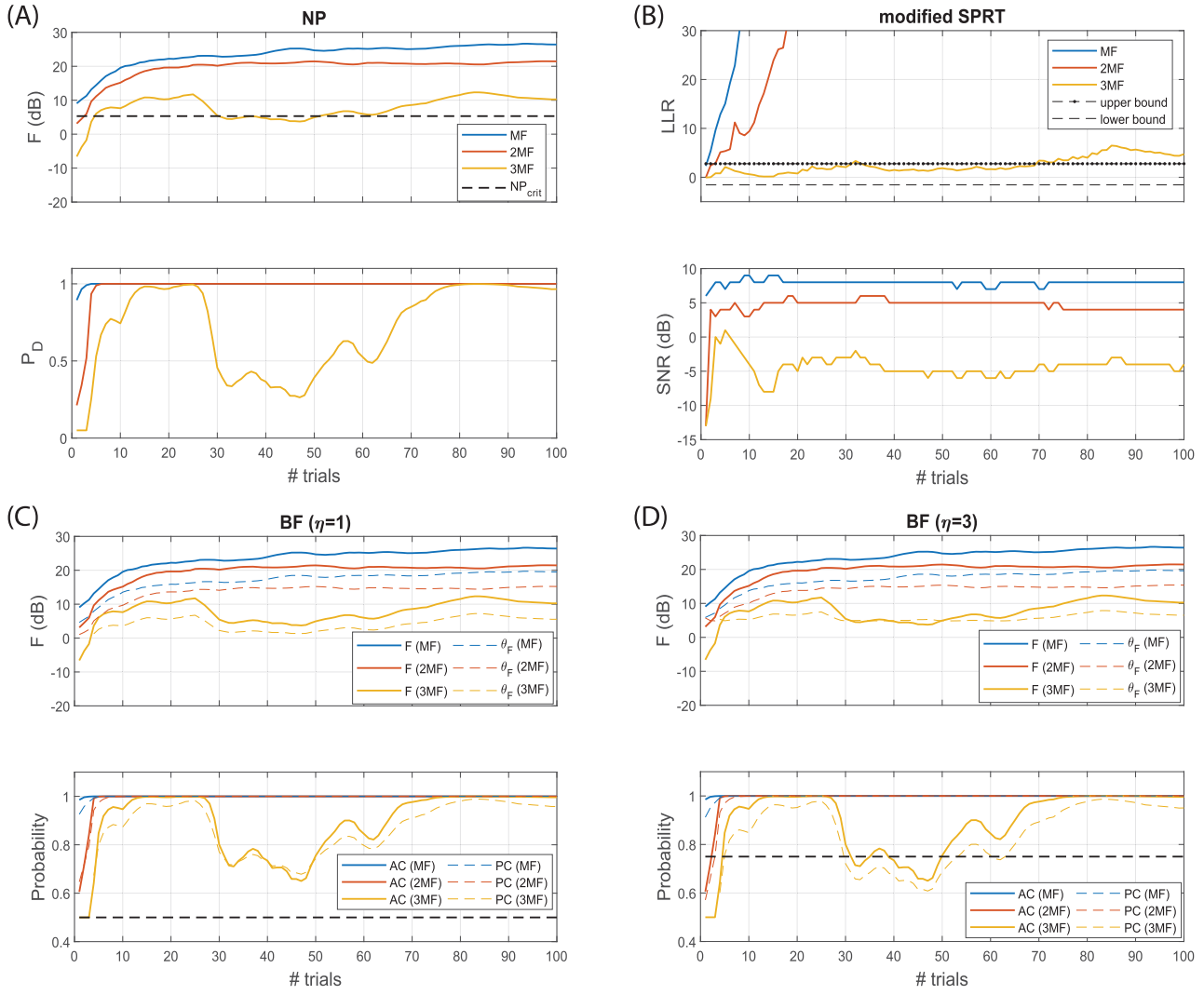

FIGURE 7. Performance of the three real-time detectors for the EEG data of subj#1 (on AVG-FCz) over the course of 100 trials. Three ASSR example frequencies: 43 Hz (MF; 2^{nd} order) 80 Hz (2MF; 4^{th} order), and 117 Hz (3MF; 6^{th} order). (A) F (dB) of the three ASSRs on accumulated trials, and the critical value (with $\alpha = 0.05$) shown by the dashed line, and corresponding P_D of the three ASSRs (lower panel). (B) LLR of the three ASSRs on accumulated trials (top panel) and the estimated single-trial SNR from accumulated trials (bottom panel). (C) and (D) (top panels) F of the three ASSRs on accumulated trials, and the corresponding thresholds (θ_F) determined by the BF threshold η; (bottom panels) corresponding AC and PC of the three ASSRs. The black dashed lines correspond to θ_{AC} shown in (19) for the BF thresholds $\eta = 1$ in (C) and 3 in (D).

two detectors is that NP adopts a fixed threshold (NP_{crit}) on F (black dashed line), the same for all ASSR frequencies, while BF adopts a dynamic (estimated) SNR-dependent threshold (θ_F) on F, resulting in frequency-dependent thresholds. The latter was determined by η, and θ_F corresponded to a fixed threshold for AC (i.e., θ_{AC}), as shown by the dashed lines in bottom panels of Fig. 7C and D ($\theta_{AC} = 0.5$ and 0.75 for $\eta = 1$ and 3, respectively). Note that AC was fully specified by the data (i.e., F), whereas PC was also affected by the threshold, η (see Methods). Hence, the AC values are the same in Fig. 7 C and D, but the values of PC differ.

We evaluated the detection performance of the three detectors on the EEG data from all subjects, and compared the results with the commonly used off-line Hotelling T^2 test, which computes the value of T^2 from the distribution of bivariate values of the real and imaginary parts of a frequency component [12]. To make a fair comparison among detectors, we used the same number (i.e., 100) of accumulated trials. In addition to the target ASSR frequencies (i.e., 2^{nd} and higher-order monaural beats generated from the two modulation frequencies of 37 and 43 Hz in the left ear), we also evaluated the detection on a large group of non-ASSR frequencies as the control group. These non-ASSRs corresponded to 88 integer frequencies (<150 Hz) at ≥1 Hz distance from potential ASSR frequencies (including monaural and binaural ASSRs caused by stimuli from both ears). Assuming that all theoretical ASSRs (n = 12) encompassed the positive class, and all non-ASSRs the negative class, we evaluated the binary classification performance by using ROC and P-R curves.

Figure 8 shows the detection performance of three real-time detectors and the Hotelling T^2 off-line detector over 100 accumulated trials. For easy visualization we also plotted the grand-average indicators (i.e., F, AC and LLR) on each integer frequency. The ROC and P-R curves were obtained for each detector from pooling the samples (of ASSRs and non-ASSRs) from all subjects. The higher AUC values indicate that the NP detector yielded the best overall performance (with the BF detector achieving a similar performance), while the modified SPRT performed worst, but similar to T^2.

In practice, use of the real-time detectors may take fewer than 100 trials for a reliable detection. Figure 9 shows the minimum number of accumulated trials required for detecting ASSRs of different order. The 2^{nd} order ASSRs (at 37 and 43 Hz), were detected as significant by the NP and BF detectors already after the first available trial for most subjects. In contrast, the modified SPRT requires at least 20 trials. In general, the higher-order ASSRs (4^{th} and 6^{th} orders) required more trials to detect than the lower order (2^{nd} and 4^{th} order) ASSRs. However, we also found some exceptions to this rule for the ASSRs at 80 Hz (4^{th} order) and 49 Hz (6^{th} order), which were detected for all subjects within 100 trials.

Figure 10 summarizes the probabilities (maximum values obtained within 100 trials) of the NP and BF detectors. The detection probability P_D of the NP detector was affected by the threshold (NP_{crit}) determined by the α values. In contrast, the acceptance confidence, AC, of the BF detector did not depend on the BF threshold (η). Both detectors showed similar trends: the detection probability (or confidence) tended to drop with increasing ASSR order, with some notable exceptions, such as 80 Hz (=37 + 43) and 49 Hz (=43 ∗ 2 − 37). Note that the minimum value for AC is 0.5, which means that for signals with extremely low SNR, the confidence that signals are present is 50% (given equal priors for signals present, or not). A lower α value (e.g., 0.01) for the NP detector will yield fewer false positives on the non-ASSR

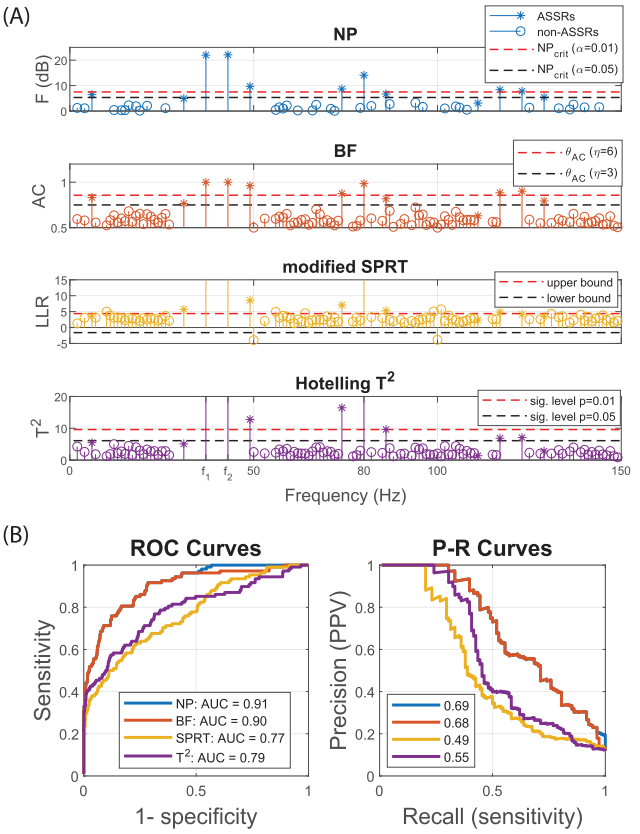

FIGURE 8. Detection performance of the three real-time detectors over 100 accumulated trials. (A) The grand-average values (average across subjects) of F, AC, LLR and Hotelling T^2 on integer frequencies including ASSRs and non-ASSRs. For each detector, typical thresholds were denoted by horizontal dash lines. Two thresholds (θ_{AC}) plotted for BF are 0.75 and 0.86 corresponding to the BF thresholds $\eta = 3$ and 6, respectively. Two MFs were denoted by f_1 and f_2 (37 and 43 Hz). The three cut-off values for LLR are 126, 137 and 30 at 37, 43 and 80 Hz, respectively. The three cut-off values for T^2 are 374, 400 and 103 at 37, 43 and 80 Hz, respectively. (B) The ROC curves and P-R curves of four detectors were plotted by varying the thresholds from the minimum to maximum on the pooling data from all subjects.

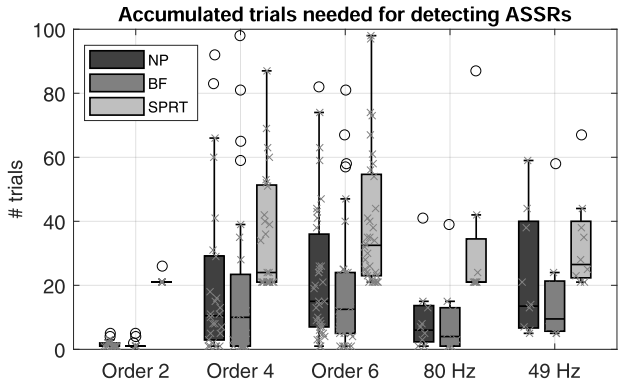

FIGURE 9. Boxplot of the number of accumulated trials needed to detect ASSRs of specific order (see Table 3), and for the two higher-order ASSRs at 80 and 49 Hz, mentioned in the text. The parameters are: $\alpha = 0.05$ (NP); $\eta = 3$ ($\theta_{AC} = 0.75$) for the BF detector; $\alpha = 0.05$, $\beta = 0.2$ for the modified SPRT.

frequencies, but also overall lower detection probabilities for ASSRs. It is also worth noting that (1-P_D) is the

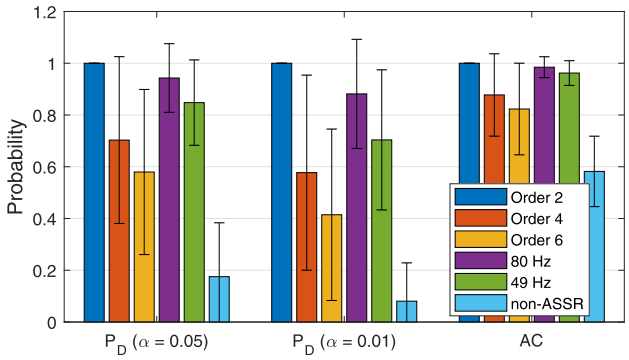

FIGURE 10. Mean and standard deviation (SD) of the probability measures for the performance of the NP (P_D) and BF (AC) detectors. The SDs on 'Order 2' are close to zero. The sample sizes were 2*n (Order 2), 4*n (Order 4), 6*n (Order 6), n (80 and 49 Hz), and 88*n (non-ASSR), respectively, where n = 9 subjects.

TABLE 4. The performance of detectors.

Detectors	Sensitivity	Specificity	Precision
NP (α=0.05)	0.73 ($\pm$0.08)	0.88 ($\pm$0.04)	0.47 ($\pm$0.09)
NP (α=0.01)	0.53 ($\pm$0.13)	0.96 ($\pm$0.02)	0.66 ($\pm$0.16)
BF (η=3)	0.76 ($\pm$0.09)	0.85 ($\pm$0.04)	0.42 ($\pm$0.07)
BF (η=6)	0.69 ($\pm$0.12)	0.92 ($\pm$0.03)	0.55 ($\pm$0.13)
SPRT* (α=0.05)	0.70 ($\pm$0.13)	0.62 ($\pm$0.06)	0.20 ($\pm$0.03)
SPRT* (α=0.01)	0.58 ($\pm$0.13)	0.81 ($\pm$0.02)	0.29 ($\pm$0.05)
T^2 (p=0.05)	0.49 ($\pm$0.15)	0.90 ($\pm$0.03)	0.41 ($\pm$0.09)
T^2 (p=0.01)	0.42 ($\pm$0.19)	0.97 ($\pm$0.02)	0.66 ($\pm$0.17)

* $\beta = 0.2$ for the modified SPRT,
T^2 denotes the Hotelling T^2 test (an off-line detector).

probability of false negatives, and the P_D on the category of non-ASSR represents the probability of false positives, as all non-ASSR frequencies belong to the negative class.

The AUC of the ROC (or P-R) curves is an indicator for the overall detection performance (i.e., discriminative power between classes) as it can assess classification performance for all possible threshold values. In practice, we need to specify a threshold (by choosing proper parameters) that reaches an adequate trade-off between sensitivity and specificity (or precision). Sensitivity (also known as recall on the P-R curve) indicates the percentage of how many theoretical ASSRs (n = 12) can be detected as significant. Specificity considers the percentage of correct negative detection on the negative (control) group, i.e., the non-ASSR frequencies (n = 88). The precision indicates the percentage of correct detection among all positive detections (1− precision = false detection rate), and it is often used for imbalanced classes, i.e., the negative sample size is much larger [43]. We evaluated the real-time detectors with the thresholds determined by commonly used parameters (e.g., α, η and p values) and compared the results with an off-line detector. The obtained performance metrics are shown in Table 4.

Table 4 clearly illustrates the trade-off between sensitivity and specificity (or precision) for each detector, i.e., a lower sensitivity will correspond to higher specificity (or precision). It is worth noting that not all theoretical ASSRs of all subjects can be detected within 100 trials, as higher-order ASSRs are generally harder to detect due to lower SNRs. As a result, the maximum sensitivity (and AUC of both ROC and P-R curves) should be lower than 100%. The baseline performance was provided by the off-line Hotelling T^2 test, for which 90% specificity was achieved with the p value of 0.05. Thus, we considered only detectors that yield larger specificity than 0.9, which means that less than 10% non-ASSR frequencies were falsely detected as significant. In practice, only the NP detector with α of 0.01 (or smaller) and BF($\eta \geq 6$) are suitable. (See for a more detailed comparison for the NP detector between α of 0.01 and 0.05, Fig. 14 of the Appendix.) The NP($\alpha = 0.01$) detector with yields similar performance as the BF($\eta = 6$) detector, achieving specificities of 96% and 92%, respectively. The low precision values in Table 4 were partly due to the unbalanced sample sizes of the ASSR and non-ASSR frequencies. For imbalanced datasets, the precision tends to fall with the increasing size of the negative class [42], [43]. Thus, precision values may differ across different studies, due to different negative sample sizes.

IV. DISCUSSION

A. SUMMARY

In this study, we evaluated the real-time performance of the classical NP detector, by quantifying the probability of detecting an auditory steady-state response in the EEG as function of the accumulated number of recorded trials and the signal-to-noise ratio. The NP detector was used to estimate the required number of trials from recorded data obtained in a pilot study, either off-line, or in real-time measurements (see section 'ASSR detection in EEG data'). It is relatively straightforward to apply the NP detector on EEG signals to detect ASSRs, as it uses a fixed threshold of F(dB). By contrast, the BF detector uses a dynamic threshold of F, and is therefore more tedious to implement. To simplify its application, we proposed to use the acceptance confidence, AC, which does not depend on the empirically-chosen BF threshold (η). Furthermore, we demonstrated that using a dynamic threshold for F was equivalent to using a fixed threshold on AC (Fig. 7). Therefore, AC is more suitable for application in the BF detector as an objective indicator. We also analysed the performance of the BF detector on ASSRs, and compared the results (i.e., P_D and false detection rate) with the NP detector for typical threshold values.

The NP and BF detectors can be readily used in both off-line and real-time applications, whereas the SPRT method was particularly designed for real-time detection. Here, we modified the standard SPRT detector by a cropped ML estimator, such that it can sequentially detect ASSRs with unknown SNRs. Setting a lower bound for the ML estimator yielded two advantages. First, it resulted in faster ML estimation because only a limited range of candidate SNR values needs to be evaluated. Second, given a limited number of trials (e.g., N = 100), an ASSR with an

extremely low SNR value technically cannot be detected. After cropping estimated SNR values at a lower bound, signals with SNRs below the lower bound will be determined as non-significant, rather than remaining in an undetermined status during the remainder of the experiment. However, the modified SPRT yielded only a similar performance to the off-line detection, which is due to two reasons. First, SPRT is a 'conservative' method because it generates a positive detection under the two preset constraints: a detection rate (i.e., $1-\beta$) and a false detection rate (α). Second, in this study we used a limited sample size (maximum 100 trials), which limited an accurate estimation for the unknown SNR by point estimation with the maximum likelihood algorithm. We expect improved performance of the modified SPRT detector when a larger number of trials (>1000) is available.

Finally, the NP detector yielded the best overall detection performance for the simulations as well as the EEG signal analysis, with the BF detector yielding a similar performance. The detection results on EEG signals show that the 2^{nd} order ASSRs (i.e., monaural beats) were detected within five trials by the NP and BF detectors, and that more trials were required for the modified SPRT detector (Fig. 9). In general, the higher-order ASSRs took more trials to detect, with exception of the higher-order ASSRs at 49 and 80 Hz. These two particular ASSRs could be detected within 100 trials for all subjects. These frequencies have been argued to arise mainly from brainstem (around 80 Hz) and cortical processes (around 45 Hz) in the auditory pathways [4], [37], and have been shown to be quite robust in many studies [44], [45].

B. RELATION TO OTHER STUDIES

The Hotelling T^2 test and the spectral F test (which was implemented in both NP and BF detectors) represent the two main approaches for detecting a steady-state response. The former quantifies the amplitude and phase information only on the target frequency, the latter also considers neighboring frequencies around the target frequency. The spectral F test has been reported to be equivalent to many classical off-line methods, like the magnitude squared coherence (MSC) measure and the circular T^2 test (CT2) (see Methods). Here, we evaluated the spectral F test for real-time application, i.e., making a decision for each newly obtained trial and existing trials.

Compared with off-line detectors that rely on averaging or statistically analyze all available trials, the proposed real-time (NP and BF) detectors showed improved performance because the maximum SNRs and corresponding P_D were often achieved well before reaching the maximum number of trials. In addition, we also observed that the SNRs of weak ASSRs often fluctuated during the measurement (as shown e.g. in Fig. 7), which might be caused by factors such as subjects' fatigue [16], or other non-stationarities in the EEG signals. In practice, during real-time detection, if the current accumulated set of trials is sufficient to determine the presence of a particular target ASSR, no more trials are needed. Furthermore, we showed that one can estimate an approximate estimate for the required number of trials for target ASSRs from the calculated single-trial SNR (Fig. 4 and Table 2). If the estimated number of trials is much larger than the maximum experimental time allowed, one could immediately decide to make an early stop. This will save valuable measurement time and unnecessary inconvenience for hearing impaired subjects, whose hearing thresholds are higher than normal-hearing listeners, and perhaps may lack the healthy nonlinearities in their early auditory processing stages, which may therefore prevent them from generating the corresponding ASSRs.

Many previous studies did not evaluate their methods on the ensemble of non-ASSR frequencies as a control group. Consequently, the reported high detection rates (or sensitivities) in the literature may come at a potential cost: a high false-detection rate on non-ASSR frequencies (as shown by Table 4). Therefore, it is important to choose proper parameters for the detectors, in order to make a 'reliable' detection with an acceptable low false-detection rate. We recommend to use the NP and SPRT detectors with $\alpha = 0.01$ (a stricter criterion than 0.05, see an example in Appendix Fig. 14), and the BF threshold $\eta \geq 6$. These settings will reduce the false detection rates for both ASSRs and non-ASSRs, especially when the EEG channels with higher SNRs are selected for detection [5].

C. POSSIBLE APPLICATIONS IN OTHER DOMAINS

In this study, the online detection was performed to detect both low and high order auditory steady-state responses. However, it is worth noting that these online detection methods could also be applied for the detection tasks of event-related potentials, such as P300-based brain-computer interfaces, for which multiple trials are required to improve SNRs [46]. Similar online detection strategies have been used to reduce the calibration time [47], and for visual stimulus studies to reduce the intensity of stimuli, yielding reduced visual fatigue [48]. In addition, a similar (online) Neyman-Pearson (NP) detection rule has been used in other signal detection tasks, such as photoplethysmogram (PPG) signals endowed with motion artifacts [49].

D. FUTURE WORK

A limitation of the real-time detectors is that their performance still relies on empirically chosen parameters, which determine a trade-off between correct detections and false alarms. Future studies could consider two options to fundamentally improve detection performance: (i) increasing the SNRs of target ASSRs, and (ii) properly setting the prior of ASSRs. First, ASSRs can be detected at increased SNRs by combining multiple electrode EEG with source-localization techniques [50]. However, the improved SNR of ASSRs will be affected by the proper estimation of multiple ASSR sources. Second, for the BF detector in this study we did not vary the prior probabilities of potential ASSRs (we used equal priors for signals present or not). However, another study [7]

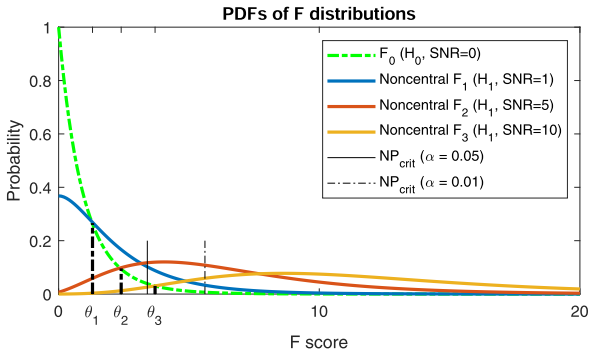

FIGURE 11. PDFs of F distributions (H_0; green)) and non-central F distributions (H_{1-3}) under different SNRs with the neighboring frequency bin number, M = 12. The critical value (NP_{crit}) of the NP detector was determined by $\alpha = 0.05$ or 0.01). The thresholds of F (θ_{1-3}, denoted by the thick dashed lines) for the BF($\eta = 1$) detector are determined by intersections of the PDFs for H_0 and H_1 with SNR values of 1, 5 and 10.

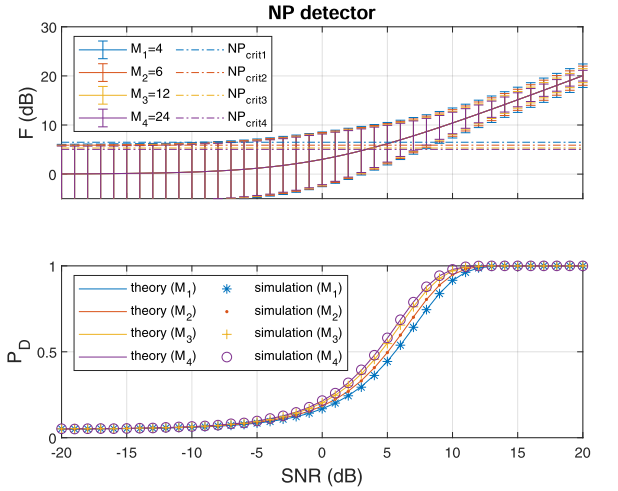

FIGURE 12. Performance of the NP detector as a function of SNR for different numbers of neighboring frequency bins (M = 4, 6, 12, and 24). Top panel shows the values of the expected F and its 95% confidence intervals (solid lines) and the corresponding critical values (dashed lines) of the NP detector ($\alpha = 0.05$). For each M, the expected F values completely overlap, but their confidence intervals (and corresponding critical values) differ. The bottom panel shows that the detection probability, P_D, computed by theory (solid lines) and simulation (symbols) overlap (n = 1e5 in the Monte Carlo simulations). The P_D for the lowest SNRs converges at 0.05 (i.e., the preset value of α).

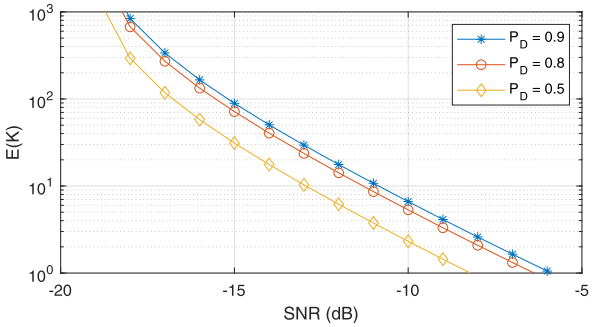

FIGURE 13. The expected stop time, i.e., trial number K (in (14)) for each SNR using SPRT with specified P_D values of 0.9, 0.8 and 0.5 (with $\alpha = 0.05$).

has suggested that the use of prior probabilities of ASSRs potentially further reduces measurement time.

V. CONCLUSION

This study evaluated three real-time ASSR detectors: the classical NP detector, the BF detector and the modified SPRT method, by simulations and by detecting higher-order ASSRs in the EEG, generated by AM sound stimuli. Compared with traditional off-line detectors, the NP and BF (real-time) detectors showed improved performance and can potentially save measurement time. The detection results show that the second order ASSRs can be detected already within five trials for both NP and BF detectors. In general, higher-order ASSRs take more trials to detect, with the exception of those near 40 Hz and 80 Hz. To guide the implementation of the real-time detectors, we evaluated the influence of commonly-used parameters to reach an acceptable trade-off for detection performance against false positives.

APPENDIX A
RELATIONSHIP BETWEEN SNR AND F SCORE

The following formulas show the relationship between the expected F score (by the spectral F test) and SNR.

$$SNR = \frac{p(s)}{p(n)} = \frac{E(A^2)}{E(R_k^2 + I_k^2)} = \frac{A^2}{2\sigma^2} \tag{A1}$$

$$E(F) = \frac{p(s+n)}{p(n)} = \frac{E((A+R_k)^2 + I_k^2)}{E(R_k^2 + I_k^2)} = \frac{A^2 + 2\sigma^2}{2\sigma^2} \tag{A2}$$

Hence, $E(F) = SNR + 1$. See (1) in the main text for an explanation of each symbol.

APPENDIX B
ILLUSTRATION OF F DISTRIBUTIONS

Figure 11 illustrates how the PDFs of the null hypothesis (H_0) and the alternative hypothesis (H_1, with different SNR values) affect the thresholds of F for the NP and BF detectors. The F threshold for the NP detector is known as critical values (NP_{crit}), which is independent of the SNR. The F threshold for the BF detector is denoted by θ, which is affected by the SNR.

APPENDIX C
INFLUENCE OF FREQUENCY BINS

Figure 12 illustrates that P_D slightly increases with the number of neighboring frequency bins, M.

APPENDIX D
EXPECTED STOPPING TIME OF SPRT

The expected stopping time ($E_0(K)$ and $E_1(K)$) of SPRT can be estimated when either H_0 or H_1 is true, respectively [51]:

$$E_0(K) = \frac{(1-\alpha)\ln(\frac{1-\alpha}{\beta}) - \alpha\ln(\frac{1-\beta}{\alpha})}{D(p_0||p_1)} \tag{A3}$$

$$E_1(K) = \frac{(1-\beta)\ln(\frac{1-\beta}{\alpha}) - \beta\ln(\frac{1-\alpha}{\beta})}{D(p_1||p_0)} \tag{A4}$$

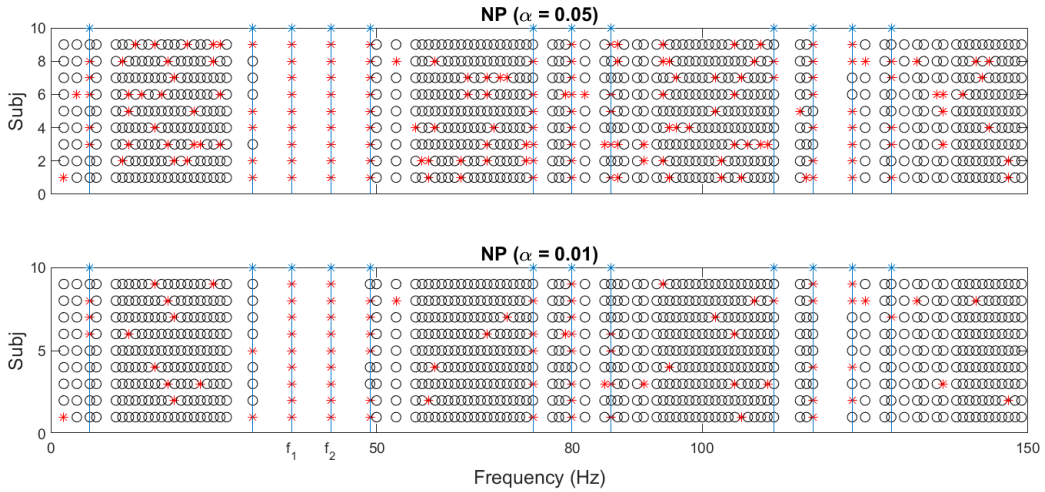

FIGURE 14. A comparison of ASSR detection at integer frequencies with the NP detector for $\alpha = 0.05$ (top) and 0.01 (bottom) for all subjects. Positive detections are denoted by stars (red), and negative detections by circles. The theoretical ASSR frequencies (n = 12) are denoted by vertical bars; the other frequencies are non-ASSR controls (n = 88).

where $D(p_0||p_1)$ and $D(p_1||p_0)$ are the Kullback-Leibler (KL) divergences between $p_1(F_i)$ and $p_0(F_i)$, and $D(p_i||p_j) = \int_0^\infty p_i(F)\log(\frac{p_i(F)}{p_j(F)})dF$. Figure 13 illustrates the expected stopping time (as trial numbers) for different detection probabilities P_D.

APPENDIX E
INFLUENCE OF α VALUE

Figure 14 illustrates how α value affects the detection of ASSRs for the NP detector.

REFERENCES

[1] E. B. J. Coffey, S. C. Herholz, A. M. P. Chepesiuk, S. Baillet, and R. J. Zatorre, "Cortical contributions to the auditory frequency-following response revealed by MEG," *Nature Commun.*, vol. 7, no. 1, p. 11070, Apr. 2016.

[2] T. W. Picton, M. S. John, A. Dimitrijevic, and D. Purcell, "Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos," *Int. J. Audiol.*, vol. 42, no. 4, pp. 177–219, Jan. 2003.

[3] P. X. Joris, C. E. Schreiner, and A. Rees, "Neural processing of amplitude-modulated sounds," *Physiol. Rev.*, vol. 84, no. 2, pp. 541–577, Apr. 2004.

[4] T. W. Picton, *Human Auditory Evoked Potentials*. San Diego, CA, USA: Plural Publishing, 2010.

[5] L. Wang, E. Noordanus, and A. J. van Opstal, "Estimating multiple latencies in the auditory system from auditory steady-state responses on a single EEG channel," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, Art. no. 2150.

[6] R. A. Dobie and M. J. Wilson, "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," *J. Acoust. Soc. Amer.*, vol. 100, no. 4, pp. 2236–2246, Oct. 1996.

[7] M. Romao and C. J. Tierra-Criollo, "A Bayesian approach to the spectral F-test: Application to auditory steady-state responses," *Comput. Methods Programs Biomed.*, vol. 183, Jan. 2020, Art. no. 105100.

[8] L. B. Felix, P. F. F. Rocha, E. M. A. M. Mendes, and A. M. F. L. M. de Sá, "Multivariate approach for estimating the local spectral F-test and its application to the EEG during photic stimulation," *Comput. Methods Programs Biomed.*, vol. 162, pp. 87–91, Aug. 2018.

[9] M. Aoyagi, T. Fuse, T. Suzuki, Y. Kim, and Y. Koike, "An application of phase spectral analysis to amplitude—Modulation following response," *Acta Oto-Laryngol.*, vol. 113, no. 504, pp. 82–88, Jan. 1993.

[10] R. Schoonhoven, C. J. R. Boden, J. P. A. Verbunt, and J. C. de Munck, "A whole head MEG study of the amplitude-modulation-following response: Phase coherence, group delay and dipole source analysis," *Clin. Neurophysiol.*, vol. 114, no. 11, pp. 2096–2106, Nov. 2003.

[11] G. D. M. Silva, F. Antunes, C. S. Henrique, and L. B. Felix, "Assessment of auditory threshold using multiple magnitude-squared coherence and amplitude modulated tones monaural stimulation around 40 Hz," *Comput. Methods Programs Biomed.*, vol. 159, pp. 71–76, Jun. 2018.

[12] F. J. Vanheusden, S. L. Bell, M. A. Chesnaye, and D. M. Simpson, "Improved detection of vowel envelope frequency following responses using Hotelling's T2 analysis," *Ear Hearing*, vol. 40, no. 1, pp. 116–127, Jan. 2019.

[13] J. D. Victor and J. Mast, "A new statistic for steady-state evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, no. 5, pp. 378–388, May 1991.

[14] R. A. Dobie and M. J. Wilson, "Objective response detection in the frequency domain," *Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sect.*, vol. 88, no. 6, pp. 516–524, Nov. 1993.

[15] D. B. Melges, A. F. C. Infantosi, and A. M. F. L. M. de Sá, "Using objective response detection techniques for detecting the tibial somatosensory evoked response with different stimulation rates," *J. Neurosci. Methods*, vol. 195, no. 2, pp. 255–260, Feb. 2011.

[16] J. H. Grose, E. Buss, and J. W. Hall, "Binaural beat salience," *Hearing Res.*, vol. 285, nos. 1–2, pp. 40–45, Mar. 2012.

[17] A. F. C. Infantosi, "Frequency-domain objective response detection techniques applied to evoked potentials: A review," in *Applied Biological Engineering—Principles and Practice*, G. R. Naik, Ed. IntechOpen, 2012.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[19] J. K. Kruschke and T. M. Liddell, "The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective," *Psychonomic Bull. Rev.*, vol. 25, no. 1, pp. 178–206, Feb. 2018.

[20] M. Szychowska and S. Wiens, "Visual load does not decrease the auditory steady-state response to 40-Hz amplitude-modulated tones," *Psychophysiology*, vol. 57, no. 12, Dec. 2020, Art. no. e13689.

[21] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Assoc.*, vol. 90, no. 430, pp. 773–795, 1995.

[22] Z. Dienes, "How Bayes factors change scientific practice," *J. Math. Psychol.*, vol. 72, pp. 78–89, Jun. 2016.

[23] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.

[24] J. P. C. Kleijnen and W. Shi, "Sequential probability ratio tests: Conservative and robust," *Simulation*, vol. 97, no. 1, pp. 33–43, Jan. 2021.

[25] M. Kulldorff, R. L. Davis, M. Kolczak, E. Lewis, T. Lieu, and R. Platt, "A maximized sequential probability ratio test for drug and vaccine safety surveillance," *Sequential Anal.*, vol. 30, no. 1, pp. 58–78, Jan. 2011.

[26] T. W. Picton, A. Dimitrijevic, M. S. John, and P. Van Roon, "The use of phase in the detection of auditory steady-state responses," *Clin. Neurophysiol.*, vol. 112, no. 9, pp. 1698–1711, Sep. 2001.
[27] E. D. Farahani, J. Wouters, and A. van Wieringen, "Contributions of non-primary cortical sources to auditory temporal processing," *NeuroImage*, vol. 191, pp. 303–314, May 2019.
[28] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL, USA: CRC Press, 2020.
[29] X. Jiang and S. Mahadevan, "Bayesian risk-based decision method for model validation under uncertainty," *Rel. Eng. Syst. Saf.*, vol. 92, no. 6, pp. 707–718, Jun. 2007.
[30] S. Raychaudhuri, "Introduction to Monte Carlo simulation," in *Proc. Winter Simulation Conf.*, Dec. 2008, pp. 91–100.
[31] T. W. Picton, A. Dimitrijevic, M.-C. Perez-Abalo, and P. Van Roon, "Estimating audiometric thresholds using auditory steady-state responses," *J. Amer. Acad. Audiol.*, vol. 16, no. 3, pp. 140–156, Mar. 2005.
[32] C. S. van der Reijden, L. H. M. Mens, and A. F. M. Snik, "Frequency-specific objective audiometry: Tone-evoked brainstem responses and steady-state responses to 40 Hz and 90 Hz amplitude modulated stimuli," *Int. J. Audiol.*, vol. 45, no. 1, pp. 40–45, Jan. 2006.
[33] Y. Yang, T. Solis-Escalante, J. Yao, A. Daffertshofer, A. C. Schouten, and F. C. T. van der Helm, "A general approach for quantifying nonlinear connectivity in the nervous system based on phase coupling," *Int. J. Neural Syst.*, vol. 26, no. 1, Feb. 2016, Art. no. 1550031.
[34] J. Bohórquez and Ö. Özdamar, "Generation of the 40-Hz auditory steady-state response (ASSR) explained using convolution," *Clin. Neurophysiol.*, vol. 119, no. 11, pp. 2598–2607, Nov. 2008.
[35] B. Roß, T. W. Picton, and C. Pantev, "Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field," *Hearing Res.*, vol. 165, nos. 1–2, pp. 68–84, Mar. 2002.
[36] D. W. Schwarz and P. Taylor, "Human auditory steady state responses to binaural and monaural beats," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 658–668, 2005.
[37] J. V. Opstal, *The Auditory System and Human Sound-Localization Behavior*. New York, NY, USA: Academic, 2016.
[38] M. van der Heijden and P. X. Joris, "Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies," *J. Neurosci.*, vol. 23, no. 27, pp. 9194–9198, Oct. 2003.
[39] Y. Yang, T. Solis-Escalante, F. C. T. van der Helm, and A. C. Schouten, "A generalized coherence framework for detecting and characterizing nonlinear interactions in the nervous system," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 12, pp. 2629–2637, Dec. 2016.
[40] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[41] L. Wang, J. B. A. M. Arends, X. Long, P. J. M. Cluitmans, and J. P. van Dijk, "Seizure pattern-specific epileptic epoch detection in patients with intellectual disability," *Biomed. Signal Process. Control*, vol. 35, pp. 38–49, May 2017.
[42] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.
[43] L. Wang, X. Long, J. B. A. M. Arends, and R. M. Aarts, "EEG analysis of seizure patterns using visibility graphs for detection of generalized seizures," *J. Neurosci. Methods*, vol. 290, pp. 85–94, Oct. 2017.
[44] E. D. Farahani, T. Goossens, J. Wouters, and A. van Wieringen, "Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway," *NeuroImage*, vol. 148, pp. 240–253, Mar. 2017.
[45] B. Ross, T. Miyazaki, J. Thompson, S. Jamali, and T. Fujioka, "Human cortical responses to slow and fast binaural beats reveal multiple mechanisms of binaural hearing," *J. Neurophysiol.*, vol. 112, no. 8, pp. 1871–1884, 2014.
[46] J. Jin, Z. Chen, R. Xu, Y. Miao, X. Wang, and T.-P. Jung, "Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, pp. 2585–2593, Sep. 2020.
[47] J. Jin, S. Li, I. Daly, Y. Miao, C. Liu, X. Wang, and A. Cichocki, "The study of generic model set for reducing calibration time in P300-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 3–12, Jan. 2020.
[48] M. Xu, X. Xiao, Y. Wang, H. Qi, T.-P. Jung, and D. Ming, "A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1166–1175, May 2018.
[49] R. Krishnan, B. Natarajan, and S. Warren, "Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1867–1876, Aug. 2010.
[50] M. X. Cohen and R. Gulbinaite, "Rhythmic entrainment source separation: Optimizing analyses of neural responses to rhythmic sensory stimulation," *NeuroImage*, vol. 147, pp. 43–56, Feb. 2017.
[51] Z. Govindarajulu, *Sequential Statistics*. Singapore: World Scientific, 2004.

LEI WANG was born in Henan, China, in 1987. He received the B.S. degree from China Jiliang University, in 2009, the M.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2013, and the Ph.D. degree in the signal processing systems, electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 2018.

Since 2017, he has been working as a Postdoctoral Researcher with Radboud University, and the Donders Institute for Brain, Cognition and Behavior, The Netherlands. His research interests include biomedical signal processing and machine learning, neuromodulation, human auditory system modeling, and EEG-based epilepsy and seizure detection.

ELISABETH NOORDANUS was born in Huizen, The Netherlands, in 1961. She received the M.Sc. degree in physics from Utrecht University, in 1987, and the master's degree in neuroscience from Radboud University, Nijmegen, in 2014.

After working in product development and technical management with Philips and ASML, she started her own company specialized in internet-based applications and databases. She currently conducts Ph.D. research with the Department of Biophysics, focusing on EEG-based and psychophysical test methods of the auditory processing of cochlear implant users.

ADRIANUS JOHANNES VAN OPSTAL was born in Zevenbergen, The Netherlands, in 1957. He received the M.Sc. degree in physics and the Ph.D. degree in systems neuroscience from the Department of Biophysics, Radboud University, Nijmegen, in 1983 and 1989, respectively.

After his postdoctoral work in Zurich, Switzerland, he became an Assistant Professor, in 1990, and an Associate Professor of systems neuroscience with the Biophysics Department, in 1995. He became a Full Professor, in 2005, and the Head of the Department of Biophysics, in 2010. Since 2019, he has been the Scientific Director of the Donders Centre for Neuroscience, Science Faculty. He is the author of more than 150 articles and 1 book. His research interests include human auditory systems, multisensory integration, gaze control, and neurocomputational modeling.