FASTER AUTOMATIC ASSR DETECTION USING SEQUENTIAL TESTS

Tiago Zanotellia, b\*,Felipe Antunesc, David Martin Simpsond, Eduardo Mazoni Andrade Marçal Mendesa and Leonardo Bonato Felixd,e

*a Graduate Program in Electrical Engineering, Federal University of Minas Gerais, Belo Horizonte – MG, Brazil.*

b Federal Institute of Education Science and Technology of Espírito Santo-São Mateus, São Mateus – ES, Brazil.

*c Federal Institute of Education Science and Technology of Minas Gerais-Ipatinga Advanced Campus, Ipatinga – MG, Brazil.*

d Institute of Sound and Vibration and Research, University of Southampton, Southampton, United Kingdom.

*e Department of Electrical Engineering, Federal University of Viçosa, Viçosa – MG, Brazil.*

\*Corresponding author: tiagoz@ifes.edu.br

FASTER AUTOMATIC ASSR DETECTION USING SEQUENTIAL TESTS

**Abstract.** *Objective:* Objective Response Detection (ORD) can be used for auditory steady-state response (ASSR) detection. In conventional ORD methods, the statistical tests are applied at the end of data collection (‘single-shot tests’). In sequential ORD methods, statistical tests are applied repeatedly, while data is being collected. However, repeated testing can increase False Positive (FP) rates. One solution is to infer that response is present only after the test remains significant for a predefined number of consecutive detections (NCD). Thus, this paper describes a new method for finding the required NCD that control the FP rate for ASSR detection. *Design:* NCD values are estimated using Monte Carlo simulations. *Study sample:* ASSR signals were recorded from 8 normal-hearing subjects. *Results:* The exam time was reduced by up to 38.9% compared to the single-shot test with loss of approximately 5% in detection rate. Alternatively, lower gains in time were achieved for a smaller (non-significant) loss in detection rate. The FP rates at the end of the test were kept at the nominal level expected (1%). *Conclusion:* The sequential test strategy with NCD as the stopping criterion can improve the speed of ASSR detection and prevent higher than expected FP rates.

**Keywords:** Auditory Steady-State Response; Objective Response Detection; Sequential Testing; Stopping Criteria.

**Abbreviations:**

AEP Auditory evoked potential

AM Amplitude Modulated

ASSR Auditory Steady-State Response

CSM Component Synchrony Measure

DFT Discrete Fourier Transform

EEG Electroencephalogram

FP False Positive

H0 Null Hypothesis

MSC Magnitude-Squared Coherence

NCD Number of Consecutive Detections

NTMAX Maximum Number of ORD tests

MMAX Maximum Number of epochs

MMIN  Minimum Number of epochs

MSTEP Step Width of epochs

ORD Objective Response Detection

# Introduction

Auditory evoked potentials (AEPs) can be defined as a change in neural activity in response to acoustic stimuli (Picton, 2013). Transient responses are evoked by auditory stimuli administered at a slow repetition rate, such that the response to one stimulus ends before the next stimulus starts (Stapells, 2011). When auditory stimuli are administered at a sufficiently high repetition rate, these responses are superimposed (Rance, 2008) and are then known as auditory steady-state responses (ASSRs). The ASSRs are usually obtained by electroencephalography (EEG) and can be used for the non-invasive assessment of hearing impairments (Picton et al, 2003; Korczak et al, 2012; Seidel et al, 2015; Resende et al, 2015; Israelsson et al, 2015; Sininger et al, 2018). This is particularly important for patients who cannot (or will not) provide the behavioural responses used in conventional hearing tests (e.g. pure-tone audiometry), e.g. infants and the cognitively impaired.

The ASSR is usually evoked by Amplitude-Modulated (AM) tones, and results in an increase in amplitude of the EEG at the modulation frequency, compared to no stimulation or adjacent frequencies. In addition, the stimulation leads to phase locking of the EEG at the modulation frequency. The amplitude and phase of the modulation frequency can be obtained by applying the Fast Fourier Transform (FFT) to the EEG data, after which the presence or absence of an ASSR can be inferred by using techniques known as objective response detection (ORD) methods, applied at the modulation frequency (and/or its harmonics). These ORD techniques utilize statistical methods (hypothesis tests) to test for the presence of a response. There are many different ORD techniques, such as Hotelling’s T2 (Mijares et al, 2013; Vanheusden et al, 2019), spectral F-test (Fisher, 1929), q-sample (Cebulla et al, 2006), TCIRC (Victor & Mast, 1991), component synchrony measure (CSM) (Fridman et al, 1984), or magnitude-squared coherence (MSC) (Dobie & Wilson, 1989). The ASSR is considered to be present when the value of the estimated ORD parameter exceeds a critical value (or the p-value drops below the significance level), otherwise, the ASSR is considered to be either absent or too weak to be detected. The critical value is derived from the detector’s distribution under the null hypothesis (H0) (i.e. absence of a response) for a desired significance level α. False Positives (FPs, or type I error) arise when H0 is true, but a response is ‘detected’. By definition, α is the probability of obtaining a FP (if the test is performing according to theory, when all assumptions are justified). It is thus desirable to keep α low, but this increases the probability of failing to detect a response when present (False Negatives or type II error), i.e. it decreases the sensitivity of the test (Kay, 1998). These statistical ORD tests are ‘objective’ in allowing the automatic assessment of an ASSR test without subjective (e.g. visual) input from an assessor (John et al, 2002; Lins & Picton, 1995; Van der Reijden et al, 2004).

In clinical applications, ASSRs have been used for hearing threshold estimation. However, both the accuracy of threshold prediction and exam time need to be improved (Picciotti et al, 2013; Hatton & Stapells, 2013; Sininger et al, 2018). Studies have reported ASSR thresholds to be 10-20 dB higher than those from behavioural tests (Picton et al, 2003). While it typically takes 30 minutes for conventional pure tone audiometry (behavioural test) (Bachmann & Arvedson, 2007), it has been reported that ASSR threshold assessment (at eight frequencies) can take around 19.93 min (Sininger et al, 2018) with the support of automated detection methods and state of the art stimulation paradigms.

Improvements in the time required for ASSR detection using a sequential test strategy have been reported (Stürzebecher et al, 2005; D’haenens et al, 2010; Luts et al, 2008; Torres Fortuny et al, 2011). Sequential strategies consist of applying the ORD test repeatedly while stimulation continues and EEG data is accumulated. When the ASSR is detected, the exam can stop. Otherwise it continues until a predefined maximum exam time has been reached. However, as with all sequentially applied statistical tests, the repeated application of the ORD test leads to an increased FP rate (above the specified significance level), unless appropriate measures are taken (Stürzebecher et al, 2005; D’haenens et al, 2010; Luts et al, 2008).

In order to prevent an increased FP rate, traditional correction methods for independent tests could be used, e.g. Bonferroni or Dunn-Šidák (Hochberg & Tamhane, 1987), but these are known to be conservative and furthermore, the assumption of independence is not justified when the statistical tests are applied repeatedly to accumulating data. These corrections thus tend to reduce test sensitivity and/or increase exam time (Stürzebecher et al, 2005). To avoid the conservative correction associated with these methods, two different approaches have been proposed in the literature on AEP detection: 1) An adjusted critical value for rejecting the null hypothesis is obtained for the sequentially applied ORD (Stürzebecher et al, 2005; Chesnaye et al, 2019a), and 2) The presence of response is only inferred when the ORD test remains significant for a predefined number of consecutive detections (NCD) (Luts et al, 2008).

With respect to the first method, Stürzebecher et al. (2005) used extensive Monte Carlo simulations to find a critical value (in their case for the Rayleigh test) to be used throughout the sequential detection strategy, such that the overall FP rate would equal α at the end of the sequential test session. In (Chesnaye et al, 2019a), critical decision boundaries are calculated by combining the p-values obtained from applying the ORD methods to sequentially acquired groups of epochs, which makes this method applicable to any ORD method that can provide an accurate p-value, and thus more flexible.

The second approach, which defines the presence of response when a predefined NCD occurs in the sequential test strategy, was proposed in Luts et al. (2008) and later evaluated for ASSR detection (D’haenens et al, 2010; Wilding et al, 2012). The results in D’haenens et al (2010) showed that the sequential test strategy using NCD as the stopping criteria achieved a mean reduction in exam time of 44.73% without reducing the sensitivity of ASSR detection when compared to the single-shot test. In the latter, the statistical test is applied to the data just once (after all data has been collected). In D’haenens et al (2010), the ORD was applied to 0 dB HL ASSR data from 31 subjects using a range of NCD values (1 to 17, in steps of 2) as criteria for response detection. The final NCD value for response detection was then chosen such that the FP rate was controlled. In Wilding et al (2012), using a minimum time of 2.4 minutes to start the sequential test strategy and using NCD equal to 1 as the stopping criterion in 7 sequentially applied tests (i.e. tests at 9, 10 ... 15), reported a reduction in the examination time of 38.8% when compared to the single-shot test. Additionally, these studies showed that the FP rate was inversely related to the NCD value e.g. in D’haenens et al (2010) the FP rate found was 36.4%, 16.4% 8.2% and 5.0% for the NCD equal to 1, 3, 5 and 7 respectively in 64 sequentially applied tests (each test with a 5% significance level). However, to the best of our knowledge, to date no general method has been proposed to automatically obtain the required NCD for an arbitrary FP rate. Moreover, D’haenens et al (2010) suggested future research to clarify the effect on exam time and the detection of ASSR as a function of protocol parameters such as the length of the intervals (number of stimuli) between the tests and the minimum time until the first test.

The present work thus addresses this challenge, and proposes a new method, using Monte-Carlo simulations, for obtaining the required NCD values for a sequentially applied ORD. The impact of minimum time before the first test (i.e. the buffer) and the length of the interval between consecutive tests on the NCD were assessed. ASSR detection rates were evaluated in a sample of recordings from normal hearing subjects. The ORD methods chosen were MSC and CSM (Dobie & Wilson, 1994) with α=0.01, though the method could readily be applied to other ORD techniques and different levels of α, or other modalities of evoked responses, e.g. during visual or somatosensory stimulation.

# Materials and Methods

## Objective Response Detection (ORD) techniques

### Magnitude-squared coherence (MSC)

The MSC between a periodic signal (the stimulus) and a digital signal $y[k]$ (EEG signal) can be estimated by dividing $y[k]$ into M non-overlapping segments (aligned with the stimuli) and applying the expression: (Miranda de Sá et al, 2002):

 $M\hat{S}C(f)= \frac{|\sum\_{i=1}^{M}Y\_{i}(f)|^{2}}{M\sum\_{i=1}^{M}\left|Y\_{i}\left(f\right)\right|^{2}} $ (1)

where *Yi(f)* is the Discrete Fourier Transform (DFT) of the *i-th* epoch of *y*[*n*] and ‘^’ indicates estimation. Under the null hypothesis H0, defined as“no response present at frequency *f*”, the $M\hat{S}C\left(f\right)$ follows a central beta distribution with 1 and M-1 degrees of freedom. This further assumes that the epochs are uncorrelated and the data follow a Gaussian distribution and are stationary. In this case, the critical value for $M\hat{S}C(f)$ at a significance level of α is given by the equation (Miranda de Sá & Infantosi, 2007):

 $ MSC\_{crit}=1-α^{\frac{1}{M-1}} $ (2)

H0 is rejected if $M\hat{S}C(f)$≥$MSC\_{crit}$. Thus, if the MSC at a particular modulation frequency is higher than the associated critical value, then an ASSR is considered to be present.

### Component Synchrony Measure (CSM)

The Component Synchrony Measure (CSM) of the discrete-time signal, $y[k]$, divided into M non-overlapping epochs, is estimated by the expression (Fridman et al, 1984)

 $CSM\left(f\right)=\left[\frac{1}{M}\sum\_{i=1}^{M}\cos(\left(θ\_{i}\left(f\right)\right))\right]^{2}+\left[\frac{1}{M}\sum\_{i=1}^{M}\sin(\left(θ\_{i}\left(f\right)\right))\right]^{2}$ (3)

where $θ\_{i}(f)$ is the phase angle at frequency $f$ of $Y\_{i}\left(f\right), $the DFT of the *i-th* epoch. This is equivalent to the Rayleigh test (Lütkenhöner, 1991). The critical value for the CSM, beyond which H0 is rejected, is obtained with the following equation (Miranda de Sá & Felix, 2003):

 $CSM\_{crit}\left(f\right)=\frac{χ\_{2,α}^{2}}{2M}$ (4)

where $χ\_{2,α}^{2}$ is the critical value of the chi-squared distribution with 2 degrees of freedom and a significance level α.


## Detection protocol

The detection protocol is based on the sequential test strategy (Stürzebecher et al, 2005; Luts et al, 2008; D’haenens et al, 2010), which consists of applying the ORD test (e.g. MSC and CSM) as soon as a predefined minimum number of epochs (MMIN) is available. If the stopping criterion (MSC or CSM beyond the critical value, or more generally a significant p-value in the ORD method) is not achieved, an additional new predefined number of epochs (denoted by “step width” or MSTEP) are collected, which are pooled with the previously collected epochs, and the ORD test is performed again (now on the pooled ensemble of epochs). This process continues until either 1) the number of consecutive detections is equal to the pre-determined NCD, or the 2) a predefined maximum number of epochs, MMAX, is reached. In the former case the ASSR is deemed present, in the latter, absent (or more precisely, that there is insufficient evidence to conclude that a response is present).

The challenge is thus to determine the NCD value such that the FP rate equals the desired overall significance level α. Monte Carlo simulations were performed in order to find the NCD with predefined set of values for parameters MMIN, MSTEP, and MMAX, as discussed in the next section.

## Obtaining NCD

Simulated data was generated as white Gaussian noise, with 75 epochs and epoch lengths of 1024 samples (corresponding to those obtained with the recorded EEG). This emulates data under the null hypothesis, and 1,000,000 such signals were generated. MSC and CSM ORD methods were then applied in a sequential manner according to the selected MMIN, MSTEP and MMAX. Then the FP rate was determined for a range of NCD in the range from 1 to NTMAX, where NTMAX is the maximum number of possible ORD tests in the detection protocol, calculated using NTMAX=(MMAX-MMIN)/MSTEP+1. This was repeated for different choices of MMIN, MSTEP, and MMAX (further explained below). The FP rate was estimated using:

 $ FP=\frac{NP\left(M\_{MIN},M\_{step},M\_{MAX},NCD\right)}{1,000,000}$ (5)

where NP is the number of FP, i.e. the number of recordings where H0 was rejected when using parameters MMIN, MSTEP ,MMAX, and a chosen NCD.. Finally, as illustrated in Figure 1, the optimal NCD value was estimated by finding the lowest NCD value that obtained a FP rate equal to or lower than the chosen significance level α. In this study, α=0.01 was used, in accordance with previous related works (Stürzebecher et al, 2005; Cebulla et al, 2006). The simulation was performed for different combinations of MMIN, MSTEP, and MMAX so as to obtain the corresponding NCD for the chosen α. The method was then tested on recorded EEG signals, to assess performance in terms of test time and overall sensitivity.



**Figure 1**. False Positive (FP) rate as a function of the NCD. The horizontal line indicates the chosen significance level α (α=0.01). The vertical arrow shows the NCD value at the intersection between the FP curve and the level of significance. In this example the NCD value chosen is 15, as the first integer value giving a FP rate below 0.01.

## EEG Data

The detection protocol proposed in this work was applied to a database of EEG recordings (Felix et al, 2018) in order to evaluate its performance. The EEG data were recorded from the electrode Fz with Oz as a reference and Fpz as ground. Eight volunteers with normal hearing (thresholds ≤ 20 dB HL at frequencies between 500 and 4000 Hz), aged 20-43 years old (mean age of 26.4 and standard deviation of 7.8 years, 2 female and 6 male), participated in the study (Felix et al, 2018). The study was approved by the Local Ethics Committee (CEP/UFV No. 2.105.334). The data recordings were performed in an acoustically isolated cabin in the Interdisciplinary Center for Signal Analysis (NIAS), Federal University of Viçosa (UFV), Brazil. As the (nominally) 40-Hz ASSR has been shown to be sensitive to the state of arousal of the subject (Korczak et al, 2012), volunteers were instructed to sit comfortably and keep their eyes closed without sleeping. The biological signal amplifier (Brainnet BNT-36, EMSA, Brazil) was used for EEG acquisition with the parameters: 0.1 Hz high pass, 70 Hz low pass, 60 Hz notch filter and a sampling frequency of 600 Hz.

The data acquisition for each volunteer was performed in twelve recordings of 2 minutes and 8 seconds (2.13 minutes) each. During the recordings, the volunteers were stimulated by AM tones delivered diotically (John et al, 2004), with 100% modulation depth and sound intensity of 70, 50 and 30 dB SPL. The stimuli were presented by means of an insert ear phone model 5A (Aearo Technologies, Indianapolis, IN, USA). The modulation frequencies were 35.24 Hz for the left ear and 37.01 Hz for the right ear. Four distinct carrier frequencies (0.5, 1, 2 and 4 kHz) were used in different recordings (leading thus to 4 carrier x 3 intensity = 12 recordings in each subject). Further details can be found in Felix et al. (2018). After the recordings, the EEGs signals were divided into epochs of 1024 points (approximately 1.7s time intervals) resulting in 75 epochs (2.13 min) per recording that were stored on a hard disk for offline analysis. In order to minimize spectral leakage, the duration of the epochs was chosen so that the modulation frequency has an integer number of cycles in each epoch (John et al, 1998). All offline processing was performed using Matlab R2016a (MathWorks, Natick, MA, USA).

## Evaluation of the Detection protocol

In order to evaluate the performance of the detection protocol as a function of the test parameters, different parameters MMIN and MSTEP and and MMAX = 75 epochs were applied to the EEG data, with

 $M\_{MIN}+M\_{STEP}∙(NT\_{MAX}-1)=75$ (6)

where $2\leq M\_{MIN}\leq 75,$ $1\leq M\_{STEP}\leq 73$ and NTMAX can be any integer value between 1 and 74. As a result, there were 328 different parameter combinations to assess. The maximum possible duration of all analysed protocols did not exceed 2.13 min (the duration of the 75 epochs). In addition, the MSC and CSM single-shot test (using all 75 epochs at once) were also evaluated in order to provide a comparison with the sequential method.

The FP rate, detection rate and the mean exam time was estimated for all 328 sequential test protocols. The FP rate was estimated using the number of detections at the 42 frequencies between 23Hz and 48 Hz where no response is expected, i.e. away from the stimulus frequency and its harmonics. A total of 4032 tests were thus performed (8 volunteers x 42 frequencies x 12 recordings = 4032). The detection rate is the ratio of the number of detections at the modulation frequencies and the total number of possible detections (8 volunteers x 2 modulation frequencies x 12 recordings = 192). The mean exam time is the average time, taken all across recordings, until a stopping criterion (detection of a response, or the maximum number of epochs, MMAX, without detection) was reached.

McNemar´s test (Siegel & Castellan, 1988) and Wilcoxon signed-rank test were used to compare – with a confidence level of 95% - the detection rate and exam time, respectively among different options for sequential tests and also with results from the single-shot tests.

# Results and Discussion

Figure 2 shows some “optimal” NCD values estimated by Monte Carlo simulation as a function of MMIN, MSTEP, and the choice for ORD technique (MSC is shown in plot a) and CSM is shown in plot b), for α=0.01 and MMAX = 75 epochs. The NCD for MSC and CSM are similar but not identical. This suggests that for other ORD methods the NCD would also need to be recalculated. Note that the NCD value increases as MMIN or MSTEP decreases, i.e. as the maximum number of sequentially applied statistical tests increases, the NCD also increases.



**Figure 2.** Number of consecutives detections (NCD) as a function of the minimum number of epochs (MMIN) and step widths (MSTEP) for a) MSC and b) CSM, using a significance level α = 0.01 and a maximum number of epochs (MMAX) equal to 75.

The detection protocol was applied to the EEG database for all 328 sets of parameters and ORD techniques (MSC and CSM, both with α=0.01). In Figure 3, the FP rates are shown, considering only frequencies away from the stimulus rate or its harmonics. As desired, the FP rates were at 0.01 or lower. However, many of the parameter combinations lead to conservative results (FP<0.01), which would also reduce the sensitivity when responses are present.

It was found, for example, that the combinations with NTMAX=2and MSTEP>20, such as {MMIN= 10, MSTEP = 65, MMAX= 75}, were particularly poor, forming the band of results at the bottom of Fig. 3. An additional analysis confirmed that, as expected, the FP rates increase, when NCD values lower than the “optimal” NCD values estimated (see Figure 2) were used (Luts et al, 2008; D’haenens et al, 2010). It was also found that the FP determined from the Monte Carlo simulation corresponded closely to those found with the EEG (for the same NCD), with Spearman correlations of 0.981 and 0.982 for MSC and CSM (both, *p*≤0.001), respectively. The small discrepancy between the results from simulation and those from the EEG can probably be explained by small violations of the assumptions underlying the simulation, including stationarity, Gaussianity and zero correlation between successive epochs.



**Figure 3.** False Positive (FP) rate calculated from the EEG signals for all 328 sets of parameters for the MSC (plot a) and the CSM (plot b). Only the sets with FP rates between 0.8 (horizontal black line) and 1% were included in the subsequent analysis.

Figure 4 shows the detection rates and the mean exam times for all possible sets of parameters and ORD techniques applied to the recorded ASSR signals. As expected, the exam time is lower than that for the single-shot test (unless all 75 epochs were used) and detection rates are also generally lower. Note that, as expected, the performance of the detection protocol for different sets of parameters show a trade-off between the mean exam time and detection rate, as also noted by others (Picton et al, 2005; D’haenens et al, 2010), i.e. as expected, test sensitivity increases when more data is used. The greater sensitivity of the MSC compared to the CSM agrees with previous observations (Dobie & Wilson, 1993; Simpson et al, 2000; Cebulla et al, 2001).



**Figure 4.** The detection rates and mean exam times for the 328 sets of parameters for (a) the MSC and (b) the CSM. Vertical and horizontal dash-doted lines represent the detection rate and mean exam time for the single-shot test ORD, respectively. The solid line and the circles (‘o’) represent the choices that gave the shortest exam time for a given detection rate in these experiments, with their parameters shown in the format {MMIN, MSTEP, MMAX}. In all cases MMAX = 75 and α=0.01.

Across all 328 cases examined, the average test time were 1.58 min for MSC and 1.60 min for CSM with an average detection rate of 74.14 for MSC and 68.38 for CSM. Some combinations of MMIN, MSTEP lead to minimal gain in exam time, with a notable loss of detection rate (top-left corner of the plot). On the other hand, the test along the right-hand and bottom edge of the plot (as marked in the figure) can be deemed as optimal; for all other points there is either a higher detection rate for the same exam time, or a shorter exam time for the same detection rate than all other tests.

In this dataset, the shortest exam time was given by parameters {MMIN=2, MSTEP=1, MMAX=75} for the MSC and {MMIN=13, MSTEP=1, MMAX=75} for the CSM, and showed, respectively, a 38.9% and 35.7% reduction in the average exam time compared to the single-shot test (both, *p* ≤ 0.001). However, this was associated with a 5.57% (*p*=0.02) and 1.98% (*p*=0.42) reduction in the detection rate for MSC and CSM respectively, compared to the single shot test.

In some cases (4 for the MSC and 29 for the CSM, out of the total of 328 tests), the detection rate exceeded the single shot test, but the differences in detection rates is not statistically significant. Sequential tests are not expected to exceed the sensitivity of the single shot tests, given the general principle that with a reduced (average) sample size, sensitivity decreases (Bauer & Kohne, 1994; Chesnaye et al, 2019b). However, due to random variations and some non-stationarity of background activity or time-variation of ASSR responses with accompanying changes in signal-to-noise ratios, repeated testing can in some cases allow sequential tests to find a response, when the single shot test does not. This might occur for example if the data quality degrades towards the end of the recording, degrading the MSC and CSM for the single-shot test, but the sequential test may already have detected a response.

A limitation of the study is the relatively short duration of the recordings (MMAX = 75) such that the detection rates never reached 100%. It is thus not possible to compare the number of epochs required to detect all responses between the single-shot and sequential tests (and varying values of {MMIN, MSTEP and MMAX}). Longer recordings (MMAX > 75), which allow 100% detection rate might also have been desirable, but on the other hand, this would lead to a ceiling effect in the detection rate. Ceiling effects with a very large MMAX would allow impressive reductions in exam time when using sequential methods, but such results would be deceptive.

Other limitation of the study is the small dataset that does not allow clear recommendations for the optimal choice of {MMIN, MSTEP and MMAX}. The current results show some trends (e.g. larger MMIN leads to improved detection rates at the cost of exam time), but no optimal value can be given. Such optimization should be carried out in the future with a larger dataset and with specific applications in mind (e.g. which patient-group, stimulus protocol and desired compromise between sensitivity and test time).

# Conclusion

In this work, a new method to estimate NCD values as a stopping criterion for sequential tests with ASSRs is proposed, based on extensive Monte Carlo simulations. Results showed that this approach limits the FP rate to below the desired (nominal) α value and allows exam time to be reduced compared to a single-shot test, however this is at the expense of sensitivity, which generally reduces for a sequential test compared to a single shot test.

While the current work focused on the use of MSC and CSM as the ORD method, the method could be applied to alternative techniques also, as well as other forms of evoked responses. The current results demonstrate the potential of these methods in improving the efficiency of the tests for the clinic or in research.

**Acknowledgements:** This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, CNPq and FAPEMIG, Brazil.

**Conflicts of interest:** none

**Ethical approval:** Work approved by the Local Ethics Committee. CEP/UFV No. 2.105.334

# References

Bachmann, K.R., Arvedson, J.C., 2007. Early Identification and Intervention for Children Who Are Hearing Impaired. *Pediatr. Rev.*, 19(5), p.155–165.

Bauer, P., Kohne, K., 1994. Evaluation of Experiments with Adaptive Interim Analyses. *Biometrics*, 50(4), p.1029.

Cebulla, M., Stürzebecher, E., 2015. Automated auditory response detection: Further improvement of the statistical test strategy by using progressive test steps of iteration. *Int. J. Audiol.*, 2027(January), p.1–5. Available at: http://informahealthcare.com/doi/abs/10.3109/14992027.2015.1017659.

Cebulla, M., Stürzebecher, E., Elberling, C., 2006. Objective detection of auditory steady-state responses: comparison of one-sample and q-sample tests. *J. Am. Acad. Audiol.*, 17(2), p.93–103.

Cebulla, M., Stürzebecher, E., Wernecke, K., 2001. Objective detection of the amplitude modulation following response (AMFR). *Audiology*.

Chesnaye, M.A., Bell, S., Harte, J.M., Simpson, D.M., 2019a. The Convolutional Group Sequential Test: reducing test time for evoked potentials. *IEEE Trans. Biomed. Eng.*, 9294(c), p.1–1.

Chesnaye, M.A., Bell, S.L., Harte, J.M., Simpson, D.M., 2019b. A group sequential test for ABR detection. *Int. J. Audiol.*, 0(0), p.1–10. Available at: https://doi.org/10.1080/14992027.2019.1625486.

D’haenens, W., Vinck, B.M., Maes, L., Bockstael, A., Keppler, H., et al, 2010. Determination and evaluation of clinically efficient stopping criteria for the multiple auditory steady-state response technique. *Clin. Neurophysiol.*, 121(8), p.1267–1278. Available at: http://dx.doi.org/10.1016/j.clinph.2010.03.008.

Dobie, R.A., Wilson, M.J., 1989. Analysis of auditory evoked potentials by magnitude-squared coherence. *Ear Hear.*, 10(1), p.2–13. Available at: https://insights.ovid.com/crossref?an=00003446-198902000-00002.

Dobie, R.A., Wilson, M.J., 1994. Objective detection of 40 Hz auditory evoked potentials: phase coherence vs. magnitude-squared coherence. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials*, 92(5), p.405–413.

Dobie, R.A., Wilson, M.J., 1993. Objective response detection in the frequency domain. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials*.

Felix, L.B., Antunes, F., Carvalho, J.A. da S., Barroso, M.F. dos S., Miranda de Sá, A.M.F.L., 2018. Comparison of univariate and multivariate magnitude-squared coherences in the detection of human 40-Hz auditory steady-state evoked responses. *Biomed. Signal Process. Control*, 40, p.234–239. Available at: https://www.sciencedirect.com/science/article/pii/S1746809417302331.

Fisher, R.A., 1929. Tests of Significance in Harmonic Analysis. *Proc. R. Soc. London, Ser. A*, 125(1929), p.54–59.

Fridman, J., Zappulla, R., Bergelson, M., Greenblatt, E., Mails, L., et al, 1984. Application of phase spectral analysis for brain stem auditory evoked potential detection in normal subjects and patients with posterior fossa tumors. *Int. J. Audiol.*

Hatton, J.L., Stapells, D.R., 2013. Monotic versus dichotic multiple-stimulus auditory steady state responses in young children. *Ear Hear.*, 34(5), p.680–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23985978.

Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. *Wiley Ser. Probab. Stat.*, 312(5), p.2014–2015.

Israelsson, K.-E., Bogo, R., Berninger, E., 2015. Reliability in Hearing Threshold Prediction in Normal-Hearing and Hearing-Impaired Participants Using Mixed Multiple ASSR. *J. Am. Acad. Audiol.*, 26(3), p.299–310. Available at: http://openurl.ingenta.com/content/xref?genre=article&issn=1050-0545&volume=26&issue=3&spage=299.

John, M.S., Brown, D.K., Muir, P.J., Picton, T.W., 2004. Recording auditory steady-state responses in young infants. *Ear Hear.*, 25(6), p.539–553.

John, M.S., Lins, O.G., Boucher, B.L., Picton, T.W., 1998. Multiple Auditory Steady-state Responses (MASTER): Stimulus and Recording Parameters. *Int. J. Audiol.*

John, M.S., Purcell, D.W., Dimitrijevic, A., Picton, T.W., 2002. Advantages and caveats when recording steady-state responses to multiple simultaneous stimuli. *J. Am. Acad. Audiol.*, 13(5), p.246–259.

Kay, S.M., 1998. Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory. *Signal Processing*, II, p.672. Available at: http://www.amazon.com/Fundamentals-Statistical-Signal-Processing-Detection/dp/013504135X.

Korczak, P., Smart, J., Delgado, R., M. Strobel, T., Bradford, C., 2012. Auditory Steady-State Responses. *J. Am. Acad. Audiol.*

Lins, O., Picton, T., 1995. Auditory steady-state responses to multiple simultaneous stimuli. *Electroencephalogr. Clin. Neurophysiol.*

Lütkenhöner, B., 1991. Theoretical considerations on the detection of evoked responses by means of the Rayleigh test. *Acta Otolaryngol.*, 111(S491), p.52–60.

Luts, H., Van Dun, B., Alaerts, J., Wouters, J., 2008. The influence of the detection paradigm in recording auditory steady-state responses. *Ear Hear.*, 29(4), p.638–650.

Mijares, E., Pérez Abalo, M.C., Herrera, D., Lage, A., Vega, M., 2013. Comparing statistics for objective detection of transient and steady-state evoked responses in newborns. *Int. J. Audiol.*, 52(1), p.44–49.

Miranda de Sá, A.M.F.L., Felix, L.B., 2003. Multi-channel evoked response detection using only phase information. *J. Neurosci. Methods*, 129(1), p.1–10. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0165027003001523.

Miranda de Sá, A.M.F.L., Infantosi, A.F.C., 2007. Evaluating the relationship of non-phase locked activities in the electroencephalogram during intermittent stimulation: A partial coherence-based approach. *Med. Biol. Eng. Comput.*, 45(7), p.635–642.

Miranda de Sá, A.M.F.L., Infantosi, A.F.C., Simpson, D.M., 2002. Coherence between one random and one periodic signal for measuring the strength of responses in the electro-encephalogram during sensory stimulation. *Med. Biol. Eng. Comput.*, 40, p.99–104.

Picciotti, P.M., Giannantonio, S., Paludetti, G., Conti, G., 2013. Steady state auditory evoked potentials in normal hearing subjects: Evaluation of threshold and testing time. *Orl*, 74(6), p.310–314.

Picton, T.W., 2013. Hearing in Time: Evoked potential studies of temporal processing. *Ear Hear.*, 34(4), p.385–401.

Picton, T.W., Dimitrijevic, A., Perez-Abalo, M.C., Van Roon, P., 2005. Estimating audiometric thresholds using auditory steady-state responses. *J. Am. Acad. Audiol.*, 16(3), p.140–156.

Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D., 2003. Human auditory steady-state responses. *Int. J. Audiol.*, 42(4), p.177–219.

Rance, G. [Ed], 2008. Auditory steady-state response: Generation, recording, and clinical applications. *Audit. steady-state response Gener. Rec. Clin. Appl.* Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc6&NEWS=N&AN=2009-06805-000.

Van der Reijden, C.S., Mens, L.H.M., Snik, A.F.M., 2004. Signal-to-noise ratios of the auditory steady-state response from fifty-five EEG derivations in adults. *J. Am. Acad. Audiol.*, 15(2004), p.692–701.

Resende, L., Carvalho, S., dos Santos, T., Abdo, F., Romão, M., et al, 2015. Auditory steady-state responses in school-aged children: a pilot study. *J. Neuroeng. Rehabil.*, 12(1), p.13.

Seidel, D.U., Flemming, T.A., Park, J.J.H., Remmert, S., 2015. Hearing threshold estimation by auditory steady-state responses with narrow-band chirps and adaptive stimulus patterns: implementation in clinical routine. *Eur. Arch. Oto-Rhino-Laryngology*, 272(1), p.51–59.

Siegel, S., Castellan, N.J., 1988. *Nonparametric statistics for the behavioral sciences*, Available at: http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1988-97307-000&site=ehost-live.

Simpson, D.M., Tierra-Criollo, C.J., Leite, R.T., Zayen, E.J.B., Infantosi, A.F.C., 2000. Objective Response Detection in an Electroencephalogram During Somatosensory Stimulation. *Ann. Biomed. Eng.*, 28(6), p.691–698. Available at: http://link.springer.com/10.1114/1.1305530.

Sininger, Y.S., Hunter, L.L., Hayes, D., Roush, P.A., Uhler, K.M., 2018. Evaluation of Speed and Accuracy of Next-Generation Auditory Steady State Response and Auditory Brainstem Response Audiometry in Children With Normal Hearing and Hearing Loss. *Ear Hear.*, p.1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29624540%0Ahttp://Insights.ovid.com/crossref?an=00003446-900000000-98953.

Stapells, D.R., 2011. Frequency-Specific ABR and ASSR Threshold Assessment in Young Infants. , p.409–448.

Stürzebecher, E., Cebulla, M., 2013. Automated auditory response detection: Improvement of the statistical test strategy. *Int. J. Audiol.*, 52(12), p.861–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24219121.

Stürzebecher, E., Cebulla, M., Elberling, C., 2005. Automated auditory response detection: Statistical problems with repeated testing. *Int. J. Audiol.*, 44(2), p.110–117.

Torres Fortuny, A., Pérez Abalo, M.C., Sotero Díaz, R.C., Rioja Rodríguez, L., Rodríguez Dávila, E., et al, 2011. Stopping criteria for averaging the multiple auditory steady-state response. *Acta Otorrinolaringol. Esp.*, 62(3), p.173–180. Available at: http://dx.doi.org/10.1016/j.otoeng.2010.10.001.

Vanheusden, F.J., Bell, S.L., Chesnaye, M.A., Simpson, D.M., 2019. Improved Detection of Vowel Envelope Frequency Following Responses Using Hotelling’s T 2 Analysis. *Ear Hear.*

Victor, J.D., Mast, J., 1991. A new statistic for steady-state evoked potentials [published erratum appears in Electroencephalogr Clin Neurophysiol 1992 Oct;83(4):270]. *Electroencephalogr. Clin. Neurophysiol.*

Wilding, T.S., McKay, C.M., Baker, R.J., Kluk, K., 2012. Auditory steady state responses in normal-hearing and hearing-impaired adults: an analysis of between-session amplitude and latency repeatability, test time, and F ratio detection paradigms. *Ear Hear.*, 33(2), p.267–78. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3378473&tool=pmcentrez&rendertype=abstract.



**Figure 1**. False Positive (FP) rate as a function of the NCD. The horizontal line indicates the chosen significance level α (α=0.01). The vertical arrow shows the NCD value at the intersection between the FP curve and the level of significance. In this example the NCD value chosen is 15, as the first integer value giving a FP rate below 0.01.



**Figure 2.** Number of consecutives detections (NCD) as a function of the minimum number of epochs (MMIN) and step widths (MSTEP) for a) MSC and b) CSM, using a significance level α = 0.01 and a maximum number of epochs (MMAX) equal to 75.



**Figure 3.** False Positive (FP) rate calculated from the EEG signals for all 328 sets of parameters for the MSC (plot a) and the CSM (plot b). Only the sets with FP rates between 0.8 (horizontal black line) and 1% were included in the subsequent analysis.



**Figure 4.** The detection rates and mean exam times for the 328 sets of parameters for (a) the MSC and (b) the CSM. Vertical and horizontal dash-doted lines represent the detection rate and mean exam time for the single-shot test ORD, respectively. The solid line and the circles (‘o’) represent the choices that gave the shortest exam time for a given detection rate in these experiments, with their parameters shown in the format {MMIN, MSTEP, MMAX}. In all cases MMAX = 75 and α=0.01.