Stopting Criteria for Averaging the Multiple Auditory Steady-State Response

Alejandro Torres Fortuny*, María C. Pérez Abalo, Roberto C. Sotero Díaz, Lilliam Rioja Rodríguez, Ernesto Rodríguez Dávila, Lidice Galán García, Eduardo Emil Suarez

Departamento de Audición y Lenguaje, Centro de Neurociencias de Cuba, La Habana, Cuba

Received 23 August 2010; accepted 17 October 2010

Abstract

Introduction and objectives: The aim of this study was to examine the efficiency of the averaging technique for estimating multiple auditory steady state responses in normal hearing subjects and to provide quantifiable stopping criteria at near-threshold intensities.

Methods: Multiple amplitude-modulated (89-115 Hz) tones (500, 1000, 2000, and 4000 Hz) were simultaneously presented to both ears at a fixed intensity of 40 dB HL. A total of 128 data epochs were averaged (23.9 min).

Results: The results showed that “classic” ensemble averaging, although accurate and time-efficient in most cases, could not extract all near-threshold MSSR from noise, even after recording a considerable number of sweeps. The present study also proposed a different approach to evaluate the background noise based on evaluating the mean of the variance close to the signal.

Conclusions: The study proposed quantitative parameters to establish stopping criteria during auditory steady-state recordings.

© 2010 Elsevier España, S.L. All rights reserved.

PALABRAS CLAVE
Potencial evocado auditivo estado estable; Promediación; Criterio de parada

Resumen

Introducción y objetivos: El presente estudio pretende examinar la eficiencia de la técnica de promediación para obtener potenciales evocados auditivos de estado estable en sujetos normoyentes y proveer de un criterio de parada cuantitativo a intensidades cercanas al umbral de audibilidad.


* Corresponding author.
E-mail address: fortuny@cneuro.edu.cu (A. Torres Fortuny).

2173-5735/$ - see front matter © 2010 Elsevier España, S.L. All rights reserved.
Introduction

The reliability of steady-state auditory evoked potentials (SSAEPs) as an electroaudiometric tool has been widely demonstrated in both healthy subjects and patients with hypoacusis.\(^1\)\(^-\)\(^7\) It has also been shown that the multiple stimulation SSAEP technique makes an evaluation possible with a considerable saving of time, without an appreciable loss of reliability of the test.\(^5\)

However, there are certain disadvantages in SSAEP recording that can reduce the efficiency of the test in clinical practice.\(^3\) One is the difficulty in ensuring an adequate signal-to-noise ratio (SNR), especially at intensities close to the threshold. This is particularly true in children and elderly patients, in whom the noise levels are higher than those reported in young adults.\(^8\) Additionally, responses in children have smaller amplitudes.\(^1\)\(^,\)\(^9\)\(^,\)\(^10\) In such conditions, SSAEP recording may be impractical due to the excessive number of averages necessary, with a consequent increase in the recording time.

For these reasons, many studies have focused on increasing the efficiency in SSAEP recording time. A valid approach to this problem has been the development of acoustic stimuli that produce responses of greater amplitude, which can be easily removed from the residual noise.\(^11\)\(^-\)\(^13\) Digital filtering techniques and the use of weighted averaging have also been proposed to improve SSAEP estimation.\(^14\)

However, despite the possibility of using any of the above methods to improve SSAEP estimates, “classical” averaging is still the established method to extract a signal from the background noise in which it is immersed.

In practice, the idea that averaging for a sufficiently long period of time will enable all the SSAEP signals to be extracted is often accepted.\(^8\)\(^,\)\(^15\) However, the time-effectiveness of the averaging procedure can be a limiting factor in clinical practice. Consequently, it is necessary to know how long the averaging process should last to obtain a response.

However, in SSAEP recording, we still do not have a quantifiable criterion that can be used to stop the averaging process in the absence of a significant response or in situations where the SNR is low (intensities close to the threshold). The latter condition usually requires a longer recording period to obtain a signal, or perhaps it may be more efficient to stop the recording, improve any conditions that can be improved and start again.

Despite the number of devices available for a reliable detection of steady state auditory signals,\(^2\)\(^,\)\(^16\)\(^,\)\(^17\) there are no criteria for deciding when to stop averaging. A lack of response could indicate that the threshold was reached or it could indicate the need to continue the averaging process to increase the SNR. In this case, the question would be: how long should averaging last and how effective will this be?

Consequently, the objective of this study was to evaluate the effectiveness of the averaging process in extracting SSAEP signals, by examining the behaviour and characteristics of both signal and noise during the process of averaging as a method to obtain quantitative criteria to help decide when to stop averaging in adults with normal hearing.

Methods

Subjects

We selected a sample of 8 subjects (16 ears), who were audiologically normal, aged between 20 and 28 years (24.1 \(\pm\) 3.7). They underwent a liminal tone audiometric test prior to the electrical recording. This was done using a MADSEN Orbiter 922 Version 2 clinical audiometer. All studies were performed in a soundproof environment, with a global acoustic noise level of 54 dB SPL (sound pressure level). The spectral composition of ambient acoustic noise was 31, 32, 30, and 31 dB SPL for the frequencies of 500, 1000, 2000, and 4000 Hz, respectively. The bandwidth of the measurement was between 250 and 8000 Hz and measurements were made with a Brüel & Kjaer sound level meter (Investigator 2260 model) and a 4165-type microphone. All subjects had behavioural thresholds below 15 dB HL (hearing level) for the audiometric frequencies in the range between 500 and 4000 Hz. All participants gave their informed consent prior to the start of the study.

SSAEP Recordings

The subjects were instructed to remain calm, relaxed and, if possible, to fall asleep during the recording period.
The recordings were made with the AUDIX electroaudiome-
ter (NEURONIC SA). We used disk electrodes (Ag/AgCl) attached
to the scalp with conductive paste, placing the active electrode on the vertex (Cz), the reference 1.5 cm below the inion and the ground electrode at Fpz. The impedance of the electrodes was maintained below $5 \, \text{k}\Omega$ to 10 Hz in all records. Bioelectrical activity was digitized with a 16-bit analogue-digital converter, amplified with a gain of 12 000 and analogically filtered between 10 and 300 Hz. All subjects underwent a recording until reaching 128 averages, regardless of the moment when they reached statistical significance. Each analysis window contained 8192 digitized samples scanned with a period of 1.08 ms, so the total recording time for each subject was 23.9 min. Artefact rejection was fixed at 50 $\mu$V.

**Stimulus**

We used a mixture of continuous tones of 500, 1000, 2000, and 4000 Hz modulated in amplitude at frequencies of 89, 97, 104, and 111 Hz and 93, 100, 107, and 115 Hz for the left and right ears, respectively, with a modulation depth of 95%. This stimulus was presented to both ears simultaneously through TDH49 supra-auricular headphones at an intensity of 40 dB HL. The spectral composition of the stimulus was analysed using a Brüel & Kjaer sound level meter (Investigator 2260 model), an artificial ear (4152-type) and a microphone (4144-type). The global energy of the stimulus was 57 dB SPL, with a spectral composition of 52, 47, 50, and 50 dB SPL for frequencies of 500, 1000, 2000, and 4000 Hz, respectively.

**Efficiency of Averaging**

To assess how quickly the SSAEP could be differentiated from residual noise, we analysed: 1) the number of responses detected as statistically significant; 2) the recording time required to judge the response as statistically significant and 3) the number of responses that were not differentiated from residual noise at the end of registration.

**Signal–Noise Ratio**

In order to evaluate the accuracy of SSAEP estimation, we characterised the magnitude and its fluctuations for each subject, for both the signal and for the residual noise based on the number of averages. The quantification of these parameters would help in deciding when to stop averaging in the absence of statistically significant responses.

**Signal**

In each subject (ear×frequency), the signal was characterised by measuring the following parameters in each recording segment: 1) amplitude of the response; 2) amplitude difference between 2 consecutive tests (first order derivative) to evaluate amplitude fluctuations and 3) Z index of the first order derivative, with the intention of statistically evaluating amplitude fluctuations of the signal (standardised distance from the mean). The values of this standardised vector ($\delta i$) would approach zero when (and if) the signal amplitude became stabilised. Empirically, the $Z$ index between $\pm \sigma/4$ ($\sigma=1$) was defined as the limit of a constant signal.

**Noise**

The residual noise was characterised throughout the averaging using the following index numbers: 1) residual noise level (RNL), calculated as the root mean square of the spectral components between 70 and 115 Hz, without excluding those components corresponding to the response (represented at the different modulation frequencies of each carrier tone)$^7$; and 2) variance of the noise in the vicinity of each response frequency, calculated from $\pm 60$ spectral components.

**Results**

**Efficiency of the Estimation of Steady-State Auditory Evoked Potential**

The efficiency of the averaging technique to estimate SSAEP was assessed by measuring each subject×ear×frequency, the minimum number of averages needed to recognise the signal as being significantly different from background noise. Table 1 shows the distribution of these recording times throughout the entire sample.

Note that most (89%) of the responses (57 of 64 estimates [4 frequencies×8 subjects×2 ears]) could be

<table>
<thead>
<tr>
<th>Time of Recording, min</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt;0 \times \leq 3$</td>
<td>9 (56.2%)</td>
<td>10 (62.5%)</td>
<td>6 (37.5%)</td>
<td>5 (31.2%)</td>
<td>30 (46.8%)</td>
</tr>
<tr>
<td>$\geq 3 \times \leq 6$</td>
<td>2 (12.5%)</td>
<td>3 (18.7%)</td>
<td>5 (31.2%)</td>
<td>5 (31.2%)</td>
<td>15 (23.4%)</td>
</tr>
<tr>
<td>$&gt;6 \times \leq 9$</td>
<td>2 (12.5%)</td>
<td>1 (6.2%)</td>
<td>3 (18.7%)</td>
<td>2 (12.5%)</td>
<td>8 (12.5%)</td>
</tr>
<tr>
<td>$&gt;9 \times \leq 17$</td>
<td>1 (6.2%)</td>
<td>0 (0%)</td>
<td>1 (6.2%)</td>
<td>2 (12.5%)</td>
<td>4 (6.2%)</td>
</tr>
<tr>
<td>Undetectable</td>
<td>2 (12.5%)</td>
<td>2 (12.5%)</td>
<td>1 (6.2%)</td>
<td>2 (12.5%)</td>
<td>7 (11%)</td>
</tr>
</tbody>
</table>

The values of each cell represent the total count and percentage (in brackets) of the signals detected as statistically significant in each time interval.
Figure 1 (a–c) In a superimposed manner, the behaviour of amplitude and noise values (Y axis) versus the number of averages (X axis). (a) A typical efficient estimation, where the signal was detected as significantly different from the noise before 6 min recording; (b) inefficient estimation, where it took a longer recording time (between 6 and 17 min) to detect the signal as noise; and (c) an example of undetectable signal even after averaging for 24 min. The data points of each curve represent the test-by-test update of the amplitude values and residual noise level.

Table 2 Mean Amplitude of the SSAEP Responses in Different Recording Situations.

<table>
<thead>
<tr>
<th>Frequencies, Hz</th>
<th>Efficient</th>
<th>Inefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>63±28</td>
<td>23±14</td>
</tr>
<tr>
<td>1000</td>
<td>51±24</td>
<td>21</td>
</tr>
<tr>
<td>2000</td>
<td>37±18</td>
<td>24±14</td>
</tr>
<tr>
<td>4000</td>
<td>32±11</td>
<td>20±10</td>
</tr>
</tbody>
</table>

The values of each cell represent the mean and standard deviation for each frequency response calculated in: 1) time-efficient estimates (the signal reached statistical significance before 30 averages); 2) inefficient estimates (more than 30 averages were required to detect the signal as statistically different from noise, 6–17 min recording). This analysis included only signals significantly different from baseline noise (T2H P<.05). The amplitudes are expressed in nanovolts.

(b) those requiring between 6 and 17 min recording (time inefficient estimates) and (c) responses that never reached statistical significance (undetectable). Each graph shows the fluctuations of the signal amplitude and noise based on the number of averages.

Note that in the first case (efficient), the signal amplitude reached a criterion value after averaging 17 (3.1 min) and then remained almost constant until the end of the recording.

In Fig. 1b (inefficient estimates), the signal amplitude varied greatly during the first averages and was then attenuated considerably, stabilising at values lower than the RNL, to then increase slowly and reach a criterion value only after averaging 50–60.

In Fig. 1c, in contrast, the signal amplitude remained close to and even below the values of noise and never reached a statistically significant criterion.

Three representative cases are exemplified in Fig. 1a–c: (a) responses that required less than 6 min recording to reach the criterion of significance (time efficient estimates); (b) those requiring between 6 and 17 min recording (time inefficient estimates) and (c) responses that never reached statistical significance (undetectable). Each graph shows the fluctuations of the signal amplitude and noise based on the number of averages.

Note that in the 3 graphs, regardless of the efficiency of the averaging process, the residual noise decreased (as expected) following an inverse square relationship with the number of averages. It stabilised at slightly higher values in the case of the undetectable estimates compared with efficient or inefficient estimates.

Characterisation of Signals from Steady-State Auditory Evoked Potential by Multiple Frequencies

The changes in magnitude of the signal amplitude were analysed in two different situations (efficient versus inefficient estimates). Table 2 shows a comparison between the means and standard deviations of the signal amplitude for each frequency response. Note that for all frequencies explored, the signal amplitude is lower for the case of inefficient estimates compared with efficient estimates.

To investigate possible fluctuations in the magnitude of the signal throughout the averaging process, the Z index was calculated from differences in amplitude. Fig. 2 (a and b) shows a comparison (between efficient and inefficient estimates) of variations in Z index, calculated for the individual records throughout the entire sample.

Note that in the case of efficient estimates, the curve of the Z index is close to zero (within the acceptable range...
Stopping Criteria for Averaging the Multiple Auditory Steady-State Response

1. **Characterisation of Baseline Activity of Steady-State Auditory Evoked Potential by Multiple Frequencies**

   **Residual Noise Level**
   
   **Fig. 3** shows the mean and standard deviation of the RNL during averaging throughout the entire sample. Note that in all the subjects explored, the baseline noise decreased following an inverse quadratic relationship with respect to the number of averages, reaching a relatively stable value (between 20 and 10 nanovolts [nV]) after averaging for about 5 min, regardless of the statistical significance of the signal. The magnitude of the RNL decreased to values even below 10 nV at the end of the recording (128 data epochs averaged).

   **Mean Noise Variance in the Vicinity of the Signal**
   
   This time we used the same criteria as before, where estimates were divided into 3 subgroups: 1) efficient estimates (<6 min), 2) inefficient estimates (6–17 min recording), and 3) undetectable (did not reach a statistically significant criterion).

   In this study, the mean noise variance was also used to examine the behaviour of the baseline noise (see Methods).

   **Table 3** shows a comparison between the 3 subgroups. Note that the mean noise variance was increased by the same measure that the efficiency of the averaging process decreased. An ANOVA test (one factor) showed that the group effect reached statistical significance \[F(1.55)=5.82; P<.005\].

   **Discussion**
   
   This study provides a quantitative description of steady-state evoked response and background noise over a long recording period, providing us with the appropriate framework for evaluating the efficacy of the averaging technique. Likewise, this framework is appropriate for examining whether averaging over long periods of time is sufficient to remove the steady state signal from background noise at intensities of stimulation near the auditory threshold in subjects with normal hearing.

   Our results show that classical averaging, although time efficient in most cases, is not always capable of extracting the steady state signal, even after averaging over a considerable period of time. In our opinion, this relative inefficiency of the averaging technique can be explained by some interactions between the properties of background noise, such as the high degree of variance of noise in the vicinity of the signal and the variable behaviour of the signal throughout the averaging process at some response frequencies.

   Our results show that regardless of the explored frequency, most signals (70%) were efficiently estimated by using the standard averaging technique during a relatively...
short period of time (less than 6 min of recording). Multiple studies have demonstrated the usefulness of steady state auditory evoked potentials by multiple frequencies (SSAEP-MF) in the assessment of subjects with normal hearing.1-7 Our results are consistent with these previous results and also show the feasibility of using the averaging technique to efficiently extract most of the steady-state auditory responses at near-threshold intensities (at least under the experimental conditions used in our study).

However, in apparent contradiction with the above interpretation, our results also showed that the averaging technique was not always effective in estimating the SSAEP-MF at near-threshold intensities. There were some signals that needed between 6 and 17 min of recording (32–86 averages) to be extracted from the background noise, and a small number of them (7/64, 11%) that were not recognized even after averaging for 24 min (128 averages). These results contradict the hypothesis that sufficiently long averaging periods are sufficient to estimate the hearing threshold using SSAEP-MF.8

Consequently, we analysed the behaviour of the signal and background noise in SSAEP-MF during the averaging process as a possible cause of this phenomenon.

Firstly, we analysed the behaviour of the steady-state auditory potential signal by multiple stimulations. The amplitude of the signals efficiently estimated in time reached a plateau (showing only very slight fluctuations in their mean value) after averaging 22±7. These findings may indicate that the different frequency responses contained in the SSAEP became stabilized during the averaging process and could be assumed to be constant from one stimulus repetition to the next.

Moreover, the amplitude of the signals inefficiently extracted from the background noise (those requiring longer recording times) showed large fluctuations around their mean value during the entire averaging process. This implies that the responses did not remain constant in amplitude and phase between one stimulus repetition and the next as has been proposed.18

Efficiency in the estimation of SSAEP-MF may also be affected by the characteristics of background noise. Using statistical methods, we analysed the RNL and mean variance in the vicinity of the spectral components corresponding to the signals. The results obtained and their possible implications for estimating steady-state auditory signals are discussed below.

There is a previous study that provides a quantitative description of SSAEP baseline activity.5 The authors measured the residual noise by calculating the root mean square of the Fourier components within the region between 70 and 1110 Hz of the response spectrum and showed that the RNL decreased (as assumed theoretically) following an inverse quadratic relationship with the number of averages. Our results confirm and expand, over a long recording period, the description of noise variations previously reported.5 In the present study, we demonstrate how RNL decreases similarly in all subjects studied. However, our results also showed that from the time when the RNL showed a similar pattern in all subjects, it was insufficient to explain the behaviour of the estimates that took more time or never reached statistical significance.

In addition to RNL, this study aimed to analyse the mean variance in the vicinity of each response in question (the same spectral components used in the automatic detection of the response) as a method to examine background noise during SSAEP-MF recording. Our results showed that the average behaviour of the variance was more consistent with the actual efficiency of the averaging process than the RNL, commonly used during SSAEP recording as an indicator of background noise. Moreover, unlike the RNL, the mean noise variance can be calculated independently for each
frequency response, even when using the multiple frequency technique.

Criterion for Stopping of Averaging

The results of this study may be helpful in deciding when to stop the averaging procedure in the absence of meaningful responses, a common problem in clinical practice.

The results show that when recording SSAEP-MF (at least in similar recording conditions), responses can be distinguished from background noise by using the classical averaging technique (in adults with normal hearing) when the background noise amplitude decreases to values below $20\pm10\text{nV}$ or less and the magnitude of the signal is between 40 and $60\text{nV}$. The results also show that the signal amplitude remains almost constant after averaging $22\pm7$.

Consequently, based on our results, SNR cannot be improved by the averaging technique or at least the process would require long periods of time, which is impractical in routine clinical practice.

In any case, if the subject is adequately relaxed and there are no changes in the experimental conditions, we would be able to detect 89% of the signals as significantly different from background noise (using the T2H device) in the first 17 min of recording.

These criteria could represent a practical (approximate) guide to determining, in the absence of statistically significant responses, if the SNR is inadequate to identify the steady-state auditory response (and we should continue the averaging process) or if we can stop the process because the response will never be achieved.

Due to the phenomenon of recruitment, the hearing threshold is easier to obtain in patients with hearing losses of cochlear origin than in other types of auditory losses, so the above criteria must be verified in a sample of patients with hearing loss. Moreover, according to data from Savio et al., these criteria may also be different in children.

Conclusions

This study shows a verification of the efficiency of averaging as a standard technique to extract the steady-state auditory responses at intensities near the audibility threshold. The averaging technique, despite being efficient in extracting most of the responses from background noise in a relatively short recording time, showed some deviations in the behaviour of the signal and the noise, which could interfere with its efficiency to some extent.

Our study contributes some measurable parameters that can help to establish a criterion to stop averaging during SSAEP-MF recording, proposing the minimum number of required averaging and the magnitude of the signal and noise required for deciding (in the absence of significant responses) if the averaging process should be stopped or whether it should continue.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References