Patient-adaptive Objective Response Detection using Reinforcement Learning

Alexandre Gomes Caldeira[1] and Leonardo Bonato Felix[2]

[1] Graduate Program in Electrical Engineering, Universidade Federal de Minas Gerais, alexandrecaldeira@ufmg.br
[2] Electrical Engineering Department, Universidade Federal de Viçosa, leobonato@ufv.br

Abstract— **In objective response detection (ORD) based on electroencephalograms (EEGs), different methods have been developed to detect brain responses to stimuli. Many rely on statistical hypothesis testing to objectively detect stimuli-induced responses, such as in Auditory Steady State Response (ASSR) detection. However, existing methods often require long test durations or assumptions about signal-to-noise ratios (SNR) and sample sizes, which can limit their practicality. To address these limitations, this paper introduces a novel approach to ASSR ORD using Reinforcement Learning (RL) methods, and a model is presented and tested. In both simulated and experimental data, the RL model is shown to enable online optimization of false positive rates while maintaining detection rates similar to other published methods, thereby custom-tailoring detectors to patients during exams or Brain-Computer Interface use. The potential of RL for enhancing ORD methods for any evoked response detection is highlighted and avenues for further research in this area are suggested. All the source code for replicating or improving upon these findings is made openly available online, and experimental data is accessible upon request to the authors.**

Keywords— **Personalized medicine, objective response detection, auditory steady state response, reinforcement learning, machine learning applications.**

In order to detect electrical brain responses to stimuli, Objective Response Detection (ORD) methods have been developed based on properties of steady state evoked potentials (SSEP) both in the time and frequency domain. Essentially, using specific visual, auditory or other sources of stimulation, measurable electrical responses can be found in electroencephalograms (EEGs), which can be then modeled and detected by comparing samples collected during and without sources of stimulation [1, 2].

Noninvasive EEGs are extensively used in Brain-Computer Interface (BCI) and ORD research for being relatively easier, cheaper and faster than intracranial (invasive) EEG or other methods - in terms of research approval and experiment design. The invisible cost of using noninvasive EEG is that it effectively measures the SSEP of stimuli on broader, farther anatomical regions and therefore measurements on the scalp are noisier than intracranial samples [3, 4], thus requiring preprocessing and more complex detection methods.

Despite this inherent noise in noninvasive EEGs, seminal works [5, 6] in Auditory Steady State Response (ASSR) detection have shown that statistical hypothesis testing can be used to objectively reject the hypothesis that no stimulus is present and that, therefore, the subject is able to hear a given frequency at some sound pressure level stimulation. This approach has the advantage of controlled false positive rates (FPRs) during testing, based on the known theoretical confidence interval of the hypothesis testing method that is used - which has been extensively proven true in experiments.

However, these methods rely on assumptions of signal-to-noise ratio (SNR) levels, sample size, as well as the probability distribution of the spontaneous and stimulated brain response. In practice, this results in either long test duration - in order to guarantee adequate sample size - or different sequential hypothesis testing approaches that seek to maintain detection power and false positive rates while reducing the test duration.

This leads to multi-objective optimization problem with constraints to sample size and statistical power: how short can a experiment be and still detect statistically relevant responses within a given confidence interval? In addition, SNR levels are not necessarily constant - the problem is not stationary - and this means larger sample sets may have different levels of response when experiments are too long.

Our aim is to contribute a Reinforcement Learning (RL) agent model that is able to detect objective responses in EEG and optimize false positive rates even during the exam. In this sense, a patient-adaptive ORD method is developed with applications to ASRD ORD, as well as discussions of its advantages and limitations.

A. Literature Review and Motivation

A traditional (narrative) review was conducted in order to establish the historical and latest methods used for ORD, especially for ASSR and ABR, but including detection of other evoked potentials as well as BCI results based on noninvasive EEG. Historically, statistical hypothesis testing has been used, improved and extended in time [7], space [8] and frequency [6] domains for ORD. Currently, research is focused in finding novel stimulation techniques that have faster but

statistically significant and detectable responses - e.g. chirp or click stimulation for Auditory Brainstem Responses (ABR) detection [9] - and sequential hypothesis testing methods that can reduce test time whilst controlling FPR [10, 11, 12]. Our approach is to combine frequency measures of synchrony, coherence and relative energy in the stimulated and noise frequencies.

Similar works in the past have attempted to combine these ORDs using linear combinations, neural networks, evolutionary algorithms, naive bayes and other algorithms [13, 14]. This often resulted in either faster detection or higher detection rates (DR, true positive rates) than the individual ORDs for a given SNR level. But in terms of Machine Learning, most other research is focused towards BCI, not specifically the ORD problem [3, 2]. There are indeed recent results using stacked ensemble models for ABR detection [15], and Reinforcement Learning for robotic BCI [16, 17].

However, to the best of the authors' knowledge, there are no previous works regarding the problem of Objective Response Detection using Reinforcement Learning. Therefore, we model the ASSR ORD problem for the RL framework, conduct tests in simulated environments as well as experimental data, and discussed the advantages and limitations of the method, proposing further improvements.

B. Goals, Contributions and Impact

This work proposes a novel method for modeling the Objective Response Detection problem for Reinforcement Learning, aiming to achieve online optimization for each patient, towards faster experiments with lower false positive rates. Our contributions are in four domain areas: (i) patient-adaptive technology, (ii) evoked potential detection models, (iii) experimental validation for ML agents in ORD and (iv) avenues for improvement. Respectively, these contributions are: (i) a novel method for online optimization during exams based in objective response detection; (ii) a pioneer model of evoked potential detection using Reinforcement Learning; (iii) a proof of concept that the first two contributions are valid in experimental ASSR data; and (iv) discussion of the advantages and limitations of this approach, offering suggestions and further work for improving our findings.

I. Theoretical Background

In order to model the ASSR ORD problem in RL, there are key concepts and mathematical definitions that describe the experiments. This section presents a summary of the necessary background in ORD measures in the frequency domain that have been successfully applied to statistical hypothesis testing for ASSR. Next, the Temporal Difference Tabular Q-Learning framework for RL is described and contextualized.

A. Objective Response Detection

The Component Synchrony Measurement (CSM), Global F-test (GFT), and Magnitude-Square Coherence (MSC) are three popular and effective techniques used for Objective Response Detection in Auditory Steady-State Responses (ASSRs). These methods are commonly employed to create a statistical distribution for hypothesis testing, and are used in this work as representations of patient response to stimulus at a frequency, in other words, the current state.

In terms of patient state representation and measurement, the Component Synchrony Measurement is defined as:

$$CSM(f_i) = \left[\frac{1}{M}\sum_{n=1}^{M}\cos\phi_n(f_i)\right]^2 + \left[\frac{1}{M}\sum_{n=1}^{M}\sin\phi_n(f_i)\right]^2 \quad (1)$$

where f_i is the current frequency of interest, M is the total number of samples available, and ϕ is the phase computed based on the Fast Fourier Transform, in our case. Similarly, the Global F-test is given by:

$$GFT(f_i) = \frac{\sum_{i=1}^{M}\left|Y_n(f_i)\right|^2}{\sum_{n=1}^{M}\left|Y_n(f_i)\right|^2 + \sum_{n=1}^{M}\left|X_n(f_i)\right|^2} \quad (2)$$

where Y_i and X_i are the Fourier Transforms of EEG measurements during and without stimulation respectively. Finally, the magnitude-square coherence is defined as:

$$MSC(f_i) = \frac{\left|\sum_{n=1}^{M}Y_n(f_i)\right|^2}{M\sum_{n=1}^{M}\left|Y_n(f_i)\right|^2} \quad (3)$$

using the same definitions of the other variables. Further examples of usage and demonstrations of validity of these measures can be found in several works in the literature [8, 6, 11].

B. Temporal Difference Tabular Q-Learning

Reinforcement Learning seeks to compute the optimal policy in Markov Decision Processes in order to maximize expected returns. In order words, this framework defines an agent that interacts with an environment or process and receives rewards based on its actions at any given state. After taking an action, the agent will reach a new state and receive a reward. By interacting with the environment, the agent accumulates rewards and by selecting different actions, an agent

can navigate the environment and collect lower or higher total rewards. One approach for implementing this is Tabular Q-Learning, which uses a matrix for pairing every possible state and action with a value. This is the value of each action a at each state s, based on the expected return (proportional to the sum of current and future expected rewards) and is named the Q-value. Temporal Difference (TD) error in RL is the deviation between the estimation of Q(s,a) before and after interacting with the environment.

Mathematically, this means that the Q-value table is updated by the agent using the law:

$$Q(s,a) = Q(s,a) + \alpha[R + \gamma \max_a Q(s',a) - Q(s,a)] \qquad (4)$$

where s is the current state; s' is the next state; a is an action; $Q(s,a)$ is the estimated Q-value of a state-action pair; α is a scalar hyperparameter known as learning rate; R is the reward received after taking the action; γ is another hyperparameter known as discount rate, used for balancing current and future expected rewards. The difference $\max_a Q(s',a) - Q(s,a)$ is a measure of TD error computed at every step of an episode within the environment and used for refining Q-values, resulting in learning.

II. Materials

Having defined the usual response detection metrics and selected a RL approach to the ASSR ORD problem, this section discusses data to be used in training and testing agents. Computational simulations similar to published literature [14, 7, 11] are conducted sampling noise from normal distributions for training and testing the agent for proof of concept, and a experimental database of EEG collected during auditory stimulation is used as reference for results.

A. Monte Carlo Simulations

Synthetic data for training the agent was produced using a common Monte Carlo approach, assuming that noise has Gaussian distribution of a certain measurable and approximately constant power (SNR). In Figure 1, the simulated signal is shown in blue, where a pure tone is the theoretical stimulation response. This response has a noise added to it by the sampling hardware, other brain responses and mechanisms, and this noise is commonly assumed to be normally distributed. Usually, in experimental data, the sampled time series is normalized to have zero mean and unit variance, in order to attend conditions to some ORD methods, which is also applied here.

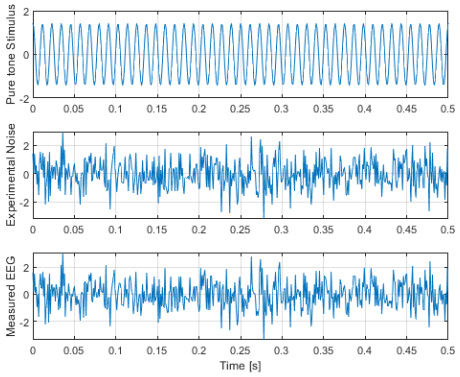


Fig. 1: A pure tone stimulus is the theoretical detectable signal, which has noise added to it at 5 SNR, resulting in a simulated EEG measurement.

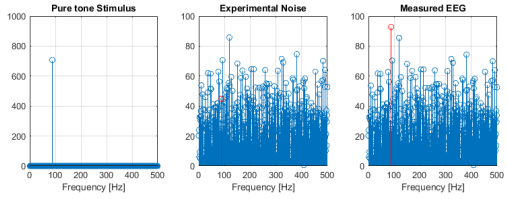
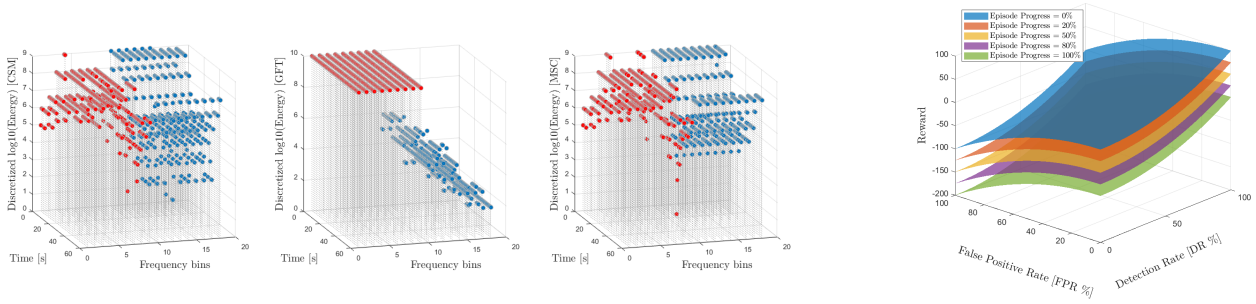


Fig. 2: The simulated signals in Fig. 1 are shown in the frequency domain, and the stimulation frequency is shown in red for the noise and measurement spectra.

As demonstration, Figure 2 displays the frequency spectra of each simulated signal, and the stimulation is shown in red for the noisy spectra. Notice that, for a relatively high SNR response is evident based on the energy at the stimulation frequency, even after adding noise. This is practical in experiments for noninvasive EEG, where SNR levels are often ranging -25 to -5. Also, SNR may change because of patient attention, sleep, other sources of stimulation in the environment and even interference in the sampling equipment. Therefore, training must be conducted at a similar or identical SNR to experiments, and it is important to analyze the impact of different SNR levels to the performance of the detector.

B. Experimental ASSR Database

Model validation was conducted on a previously published database [11], approved by the Local Ethics Committee (CEP/UFV No.2.105.334). It is comprised by EEG from 11 voluntary, normal hearing adults, sampled at 1000 or 1750 Hz in 16 electrodes during different lengths and sound pressure levels of auditory AM^2 stimulation in 8 different frequencies, within a soundproof cabin. All preprocessing methods in [11] were applied for reproducibility, and data is available upon request to the original corresponding author.



(a) (b)

Fig. 3: (a) The discretized states, represented in time for different frequencies - colored blue if stimulation is absent and red if present for the given frequency. (b) A smooth reward function is proposed based on the False Positive Rate, Detection Rate (True Positive) and Episode Progress. Notice that faster, correct detection tends to yield positive rewards whereas longer episodes and detection at incorrect states return more negative rewards, reinforcing true positives.

III. Methods

Based on the theoretical background and the data available from simulations and experiments, an agent was modeled, implemented[1], trained and tested in MATLAB using the methodology described in this section.

In the context of ORD for ASSR, an episode within the environment can be defined as each window where the Fourier Transform will be computed in order to measure response. Therefore, the states are the resulting measures of response, where the ORDs in Equations 1, 2 and 3 will be discretized to define the current state. In addition, only two actions are available for the agent: to wait and see another window (action a_0), or reject the null hypothesis and detect response at the current window (action a_1). Notice that states selected here are based on frequency metrics and other methods are available in the literature, including for modeling evoked potentials from different experiment and stimulus features.

A. State Representation and Discretization

Since the TD Tabular Q-Learning algorithm requires updating a discrete table at each step, the continuous states proposed in Eqs. 1, 2 and 3 were discretized into 10 levels. The Fourier transform was normalized for ease of discretization (the mean was subtracted, and the resulting distribution divided by its standard deviation). Figure 3(a) displays the resulting representation, where each time step is a window with 1000ms of sampled data where the Fourier Transform is computed. In order to increase data separation, $\log_{10}(\cdot)$ was applied to the measures. Moreover, the M number of samples were fixed, such that only up to the 50 last samples were used to compute the states at any given step in simulated or experimental data. In Figure 3(a), notice that the energy in red samples - frequencies with stimulation - may vary between states and bins. Moreover, notice that despite being regarded as mostly noise, the blue samples collected from non-stimulated frequencies have changing intensities in time, as other stimulus may have been present to the patient.

[1] All the source code for producing figures, results and other interesting complementary analysis for this paper are available at the code repository: https://github.com/Alexandre-Caldeira/CBEB24-RL-ORD

B. Reward Shaping

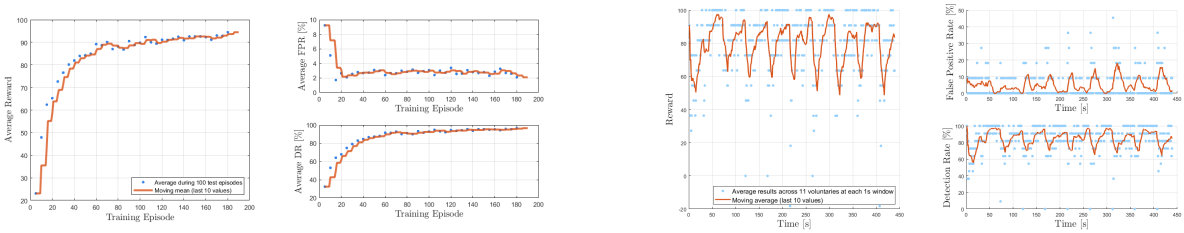
To reinforce low false positive values, short exam duration and high detection probability, a possible reward function is:

$$R(DR, FPR, EP) = \frac{DR^2}{100} - \frac{FPR^2}{100} - EP \tag{5}$$

where DR is the detection rate percentage measured as the mean detection on stimulated frequencies; FPR is the false positive rate percentage as calculated as the rate of detection in unstimulated frequencies (noise, in simulation); and EP is the episode progress percentage. This reward function is shown in Figure 3, demonstrating that higher DR, lower FP and lower EP yields higher and more positive rewards.

C. Training and Testing Strategy

During training, a Q-table was initialized with zero value for all states and spaces, sized $(10, 10, 10, 10, 2)$. Next, the frequency where signal and noise are expected are defined, as well the SNR, maximum number of episodes and the maximum number of states to be visited in each episode. Then, the learning hyperparameters α, γ are defined for training and testing. Finally, these values are used for creating simulated data and, for each training episode, for each window, the agent is presented with states regarding different frequencies and uses the $Q(s,a)$ to detect response or skip the state. When



(a) (b)

Fig. 4: (a) After each 5 training episodes, 100 tests are conducted and the average results are shown in blue. In orange, the moving mean demonstrates the agent learning, resulting in better detection rate (DR) and false positive rate (FPR). (b) Mean results for experimental ORD in ASSR using RL to optimize detection and false positive rates for 11 patients are show. On the first 80 seconds, agent's adaptation to the patient is observed as FPR and DR optimization.

response is detected, a count is updated for Detection Rate (DR) and False Positive Rate (FPR) respectively for states in each of the stimulus and non-stimulus frequencies. In each step, for every episode, the reward is calculated based on Eq. 5 for the current Episode Progress percentage and the mean DR and FPR. Then, these rewards are applied to Eq. 4 for updating each of the states visited based on the actions that were selected. This is done iteractively until the last window of the last episode is reached.

To achieve patient-wise false positive rate optimization, the agent updates its Q-values at each window during testing, using the same reward and update laws as in training. The convergence of the algorithm is analyzed by conducting 100 tests at every fifth training episode. The average performance is then collected in terms of reward, detection rate and false positive rate. At each window the action with the highest $Q(s,a)$ is selected to possibly detect responses.

IV. Results and Discussion

Multiple instances of Monte Carlo simulations were conducted to tune the hyperparameters by trial and error. After preliminary tests, convergence was first achieved by using $\alpha = 5 \cdot 10^{-3}$, $\gamma = 0.5$, 185 episodes with 40 windows each during training. Epsilon-greedy action selection was used based on a decaying epsilon value, given by:

$$\varepsilon(EpisodeNumber) = 0.2e^{-0.5\log_{10}(EpisodeNumber)} \quad (6)$$

which was updated at each episode. This results in a probability of selecting a random action at each state in training that ranges from 20% to $\approx 5\%$. In terms of Q-table size, the 4 states proposed in 10 levels were sufficient. This resulted in 96.62% detection rate and 2.10% false positive rate during tests at the final episodes at SNR 5, as seen in Figure 4(a).

Note in Figure 4(a), that the left plot displays the average test reward being optimized as the main objective function of the agent. Simultaneously, the bottom right plot shows the DR progressively improving, the top right plot shows FPR minimization in order to maximize rewards, and therefore the agent learns to detect. Similar results to Figure 4(a) were obtained for lower SNR (-8, -6, -2) but yielded higher FPR (10 to 20% approximately in average), despite maintaining DR at approximately 80% in average.

A. Experimental results

In terms of ASSR experimental data, EEG from the FC_z electrode was selected during 40 dB SPL stimulation for 440 seconds, stimulation frequency at 88 Hz was used as detection target, while 373 Hz was selected as noise for false positive rate estimation and learning. A model trained in 5 SNR was applied, and Figure 4(b) shows the resulting perfomance with 84.24% mean DR, 4.87% mean FPR that improves over time for the selected patient, as designed. In comparison with results using the same database and preprocessing techniques ([11] reaches 67% mean DR and 6.4% mean FPR and [7] reaches 85% mean DR and 1% mean FPR), the results from this technique show promising improvements using RL but not necessarily the best yet.

Despite occasionally surpassing 5% FPR, the same convergence behaviour designed and seen in simulated tests was shown during the exams. This suggests[2] that improvements to this work can preserve the patient-wise optimization behaviour and refine FPR, as windows 60-72 reached 95.8% mean DR, 0.7% mean FPR. Regarding the orange average on Figure 4(b) as a performance approximation for the agent during the exam, it is relevant to highlight that FPR reaches 5% for the first time at 12 seconds (faster than [7]) and this threshold is achieved in 284 of the 440 windows, representing a total of 64.7% of relevant detections at different times.

[2]Note that performance variation may be rooted in SNR changing during stimulation, as also seen in Fig. 3(a), see more in the code repository.

V. Final Remarks

Objective response detectors of evoked potentials are the engine of BCIs and clinical exams. Current research focuses in sequential testing and combination of existing methods to achieve faster, more accurate detection and have shown promising results. In a novel direction, this work proposes using RL for combining information from different tests and achieve detectors capable of adapting to different patients during the exam or BCI use.

This paper describes the features of RL applied to the ASSR ORD problem, proposing a detection-focused model with potential variations and improvements. One such variation is the single-shot detector, limiting the agent to one detection per episode, common in classic hypothesis tests, or updating the method to be comparable with sequential tests. Future research can also explore other state representations using correlation, entropy, SNR estimation, among others.

While this work focuses on simulating signals akin to EEG in ASSR detection, the framework can be easily adapted to detect other evoked responses. With this paper, we hope to direct the attention of researchers to the advantages of online optimization with RL, which may be an important breakthrough for improving solutions in personalized medicine.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) through the Academic Excellence Program (PROEX). Authors also extend their grateful regards to Prof. Tiago Zanotelli, Prof. Armando Neto, and colleagues Isabela Santos Silveira and Victor Hugo de Souza Singulani Ragazzi for the insightful discussions and their support during the conceptual stage of this work.

References

1. Alamanda Monik, Hohman Marc H. Auditory Steady-State Response 2023.
2. Aggarwal Swati, Chugh Nupur. Review of machine learning techniques for EEG based brain computer interface *Archives of Computational Methods in Engineering.* 2022;29:3001–3020.
3. Karikari Evelyn, Koshechkin Konstantin A. Review on brain-computer interface technologies in healthcare *Biophysical Reviews.* 2023;15:1351–1358.
4. Ahmadian Pouya, Cagnoni Stefano, Ascari Luca. How capable is non-invasive EEG data of predicting the next movement? A mini review *Frontiers in human neuroscience.* 2013;7:124.
5. Picton Terence W, John M Sasha, Dimitrijevic Andrew, Purcell David. Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos *International journal of audiology.* 2003;42:177–219.
6. Infantosi Antônio Fernando Catelli, Melges DB, Tierra-Criollo Carlos Julio. Use of magnitude-squared coherence to identify the maximum driving response band of the somatosensory evoked potential *Brazilian Journal of Medical and Biological Research.* 2006;39:1593–1603.
7. Zanotelli Tiago, Antunes Felipe, Simpson David Martin, Mazoni Andrade Marcal Mendes Eduardo, Felix Leonardo Bonato. Faster automatic ASSR detection using sequential tests *International Journal of Audiology.* 2020;59:631–639.
8. Souza Ana Paula, Soares Quenaz B, Felix Leonardo B, Mendes Eduardo MAM. Classification of auditory selective attention using spatial coherence and modular attention index *Computer Methods and Programs in Biomedicine.* 2018;166:107–113.
9. Chesnaye MA, Bell SL, Harte JM, Simpson DM. A group sequential test for ABR detection *International Journal of Audiology.* 2019;58:618–627.
10. Vaz Patrícia Nogueira, Antunes Felipe, Mendes Eduardo Mazoni Andrade Marçal, Felix Leonardo Bonato. Automated detection of auditory response: non-detection stopping criterion and repeatability studies for multichannel EEG *Computer Methods in Biomechanics and Biomedical Engineering.* 2023:1–11.
11. Zanotelli Tiago, Soares Quenaz Bezerra, Simpson David Martin, Mendes Eduardo Mazoni Andrade Marçal, Felix Leonardo Bonato, others . Choosing multichannel objective response detectors for multichannel auditory steady-state responses *Biomedical Signal Processing and Control.* 2021;68:102599.
12. Shojaeemend Hassan, Ayatollahi Haleh. Automated audiometry: a review of the implementation and evaluation methods *Healthcare informatics research.* 2018;24:263.
13. De Souza Ana Paula. Detecção de Respostas Evocadas no EEG usando técnicas de inteligência computacional 2007.
14. Soares Quenaz Bezerra. Proposta de algoritmos evolutivos para construção de detectores objetivos de resposta por meio da recombinação de elementos de detectores existentes. 2019.
15. McKearney Richard M, Bell Steven L, Chesnaye Michael A, Simpson David M. Auditory brainstem response detection using machine learning: a comparison with statistical detection methods *Ear and Hearing.* 2022;43:949–960.
16. Luo Tian-jian, Fan Ya-chao, Lv Ji-tu, others . Deep reinforcement learning from error-related potentials via an EEG-based brain-computer interface in *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*:697–701IEEE 2018.
17. Boubchir Larbi, Touati Youcef, Daachi Boubaker, Chérif Arab Ali. EEG error potentials detection and classification using time-frequency features for robot reinforcement learning in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*:1761–1764IEEE 2015.

Enter the information of the corresponding author:

Author: Alexandre Gomes Caldeira
Institute: Universidade Federal de Minas Gerais
Street: Av. Antônio Carlos 6627, 31270-901
City: Belo Horizonte
Country: Brazil
Email: alexandrecaldeira@ufmg.br