AUTHOR(S)	da Silva Lima, G., Cota, V.R., Bessa, W.M.
TITLE	In Silico Application of the Epsilon-Greedy Algorithm for Frequency Optimization of Electrical Neurostimulation for Hypersynchronous Disorders.
YEAR	2024
DOI	10.1007/978-3-031-63848-0_5
CITATION	da Silva Lima, G., Cota, V.R., Bessa, W.M. (2024). In Silico Application of the Epsilon-Greedy Algorithm for Frequency Optimization of Electrical Neurostimulation for Hypersynchronous Disorders. In: Riascos Salas, J.A., Cota, V.R., Villota, H., Betancur Vasquez, D. (eds) Computational Neuroscience. LAWCN 2023, pp. 57-68. Communications in Computer and Information Science, vol 2108. Springer, Cham. https://doi.org/10.1007/978-3-031-63848-0_5
VERSION	Accepted Manuscript
LICENSE	In Copyright © 2024 The Author(s), under exclusive license to Springer Nature Switzerland AG

In silico application of the epsilon-greedy algorithm for frequency optimization of electrical neurostimulation for hypersynchronous disorders

Gabriel da Silva Lima[1,3][0000−0001−6615−078X], Vinícius Rosa Cota[2][0000−0002−2338−5949], and Wallace Moreira Bessa[3][0000−0002−0935−7730]

[1] RoboTeAM - Robotics & Machine Learning, Universidade Federal do Rio Grande do Norte, Natal, Brazil
[2] Rehab Technologies Lab, Istituto Italiano di Tecnologia, Genova, Italy
vinicius.rosacota@iit.it
[3] Faculty of Technology, University of Turku, Turku, Finland
{gdasil,wmobes}@utu.fi

Abstract. One of most promising alternatives to suppress epileptic seizures in drug-resistant and neurosurgery-refractory patients is using electro-electronic devices. By applying an appropriate pulsatile electrical stimulation, the process of ictogenesis can be quickly suppressed. However, in designing such stimulation devices, a common problem is defining suitable parameters such as pulse amplitude, duration, and frequency. In this work, we propose a machine learning technique based on the epsilon-greedy algorithm to optimize the pulse frequency which could prevent abnormal neuronal activity without exceeding energy usage for the stimulation. Five different simulations were carried out in order to evaluate the contribution of the energy consumption in determining the minimum frequency. The results show the efficacy of the proposed algorithm to search the minimum pulse frequency necessary to suppress epileptic seizures.

Keywords: Epilepsy · Seizure suppression · Machine Learning · Epsilon-greedy · Optimization.

1 Introduction

Neuropathologies such as epilepsy, motor disorders, neuropsychiatric dysfunctions, and many others still represent a major cause of adult-onset disability worldwide [5]. This has a profoundly negative impact on the lives of patients, families, communities, and in the society as whole, displaying a muti-faceted impairment of aspects that range from the quality of life of individuals to the economic health of nations (particularly those of aging populations). Furthermore, first-choice treatments such as pharmacotherapy, neurosurgery, or physical therapy often fall short of delivering full relief of symptoms or robust recovery of neural function. For instance, in epilepsy, circa 30% of the patients do not obtain full control of seizures with anti-epileptic drugs, and, among these, half can not be submitted to curative ablative neurosurgery [9, 17]. Considering the great prevalence of the disease worldwide ($1-2\%$ of the world population), these numbers represent millions of individuals suffering from uncontrollable seizures [7, 16].

A promising alternative is the use of technology to interface the brain with electro-electronic devices in the pursuit of controlling and/or rectifying aberrant brain activity which may provide treatment of symptoms or even full recovery and cure. Broadly termed neurotechnologies, such an approach consists both of reading descriptive signals from the brain and of controlling the activity of neuronal cells by means of directly applying stimuli of different physical forms, bypassing somatosensory neural functions. Among many possibilities, electrophysiological interfaces are the most well developed and broadly used in both experimental and clinical scenarios. In fact, electrophysiological recordings (such as the electroencephalogram, electrocorticogram, local field potentials, and multi-unit recordings) and electrical stimulation of the nervous systems (Deep Brain Stimulation (DBS), Vagus Nerve Stimulation (VNS), transcranial Electrical Stimulation (tES), etc.) have served not only as powerful diagnostic tools, but also as effective alternatives to pharmacoresistant neurological scenarios [12, 2, 10, 8, 1].

From the start of its modern age in the sixties, neurotechnology in general and electrical stimulation in particular have seen great progress due to important advances in neuroscientific knowledge, neurosurgery, and digital technology [6]. More recently, unprecedented technological breakthrough such as disruptive neuronal sensor technology, very-large scale integration of electronic circuits, machine learning techniques, neuromorphism, and other breakthroughs are spurring yet another major paradigm shift in the field: close-loop approaches [4, 3]. In reality, the vast majority of clinical or experimental applications, neurotechnology is still used in an open-loop fashion only, in which neural stimulation is delivered with a fixed set of parameters, changing only in the case of the intervention of the expert. In this case, electrophysiological recordings are used only prior to stimulation as a means of diagnosis or after treatment to assess therapeutic efficacy of neuromodulation. Conversely, closed-loop neuroengineering systems, in which stimulation is directly controlled in real time by features of electrophysiological recordings, have been shown to be superior as it can dynamically respond to the highly non-stationary nature of brain activity to find privileged windows for delivering stimulus and to fine-tune its parameters. This, in turn, provides increased efficacy (stronger therapeutic effect) and efficiency (stimulation delivered only when needed).

Naturally, to be able to close the loop, one must design computational/mathematical strategies capable of connecting in real time features of the electrophysiological recordings to the parameters of electrical stimulation. Preferentially, this should be done in a way to optimize hardware resources while decreasing the risk of the therapy. For instance, in a closed-loop system to treat epilepsy, once detected, seizures must be suppressed as fast as possible by using the minimum amount of energy in the stimulation protocol, which can be translated to reduced pulse frequency, amplitude, and duration in a conventional pulsatile electrical stimulation.

In this work, we used a network of Izhikevich neurons which undergoes gradual increases of synaptic weights to simulate an ictogenesis process (development of a seizure) in a computational testbed. Local field potentials were computed to detect aberrant patterns of neural activity and electrical stimulation was applied to control it. Closing the loop, an ε-greedy algorithm was applied to quickly suppress seizures as quickly as possible, while searching for the minimum frequency of stimulation.

2 Neuronal model

The nonlinear integrate-and-fire model introduced by Izhikevich captures the diverse firing patterns observed in real biological neurons [14, 11] efficiently. This renders the model biophysically plausible, making it highly suitable for conducting large-scale simulations. The model can be represented by the following set of equations:

$$\dot{\boldsymbol{v}} = 0.04 D_v \boldsymbol{v} + 5\boldsymbol{v} + 140 - \boldsymbol{u} + \boldsymbol{I} + \boldsymbol{J} \tag{1}$$

$$\dot{\boldsymbol{u}} = D_a (D_b \boldsymbol{v} - \boldsymbol{u}) \tag{2}$$

$$\text{if } v_k \geq 30 \text{ mV, then } \begin{cases} v_k \leftarrow c_k \\ u_k \leftarrow u_k + d_k \end{cases} \tag{3}$$

which v represents the membrane potential of the neuron, u stands for a membrane recovery variable, D_x is the diagonal matrix of x, $\boldsymbol{I}$ denotes the synaptic currents, $\boldsymbol{J}$ the inject current, and a, b, c, and d being dimensionless parameters.

Considering that the model (1) represents a network of N_e excitatory neurons and N_i inhibitory, the Euler method with time steps of 0.02 ms was employed to simulate a neuronal population with $N_e = 800$ and $N_i = 200$. The network is full-connected by the matrix of synaptic weights $\boldsymbol{S} \in \mathbb{R}^{N \times N}$, where each s_{mn} weight is obtained from a continuous uniform distribution with amplitude 0.1 for the excitatory neurons and -0.1 for the inhibitory.

This work considered the initial states of v and u as being $\boldsymbol{v}_0 = -70 + 10\boldsymbol{r}$ mV and $\boldsymbol{u}_0 = -15 + 3\boldsymbol{r}$ mV, with $\boldsymbol{r}$ standing for a N-dimensional vector of random numbers extracted from a continuous uniform distribution between 0 and 1, with $N = N_e + N_i$. The dimensionless parameters are defined as follows:

$$\begin{aligned} D_{a,k} &= 0.02 \\ D_{b,k} &= 0.3 \\ c_k &= -65 - 13r_e \\ d_k &= 8 + 0.8r_e \end{aligned} \qquad 1 \leq k \leq N_e$$

$$\begin{aligned} D_{a,k} &= 0.02 + 4 \times 10^{-4} r_i \\ D_{b,k} &= 0.3 - 0.06r_i \\ c_k &= -65 \\ d_k &= 8 \end{aligned} \qquad N_e + 1 \leq k \leq N$$

with r_e and r_i being random numbers extracted from a continuous uniform distribution between 0 and 1. The synaptic current $\boldsymbol{I}$ can be modeled by the sum of the fired neuron currents with the background current which represents the unmodeled electrical activity of the brain:

$$\boldsymbol{I} = \boldsymbol{I}_f + \boldsymbol{I}_b \tag{4}$$

where $I_{f,k} = \sum_m s_{k,m}$, which $s_{k,m}$ represents the synaptic weight from the f-fired neuron connected to the k-neuron. The background current $\boldsymbol{I}_b$ is given by a normal distribution with null mean and variance of 4 for the excitatory neurons and 36 for the inhibitory.

The figure 1 shows the raster plot and the local field potential (LFP) of the neuronal model simulated as described above. The inject current was set to be null. The LFP was calculated by the weighted average of $\boldsymbol{v}$ with weights obtained from a continuous uniform distribution.

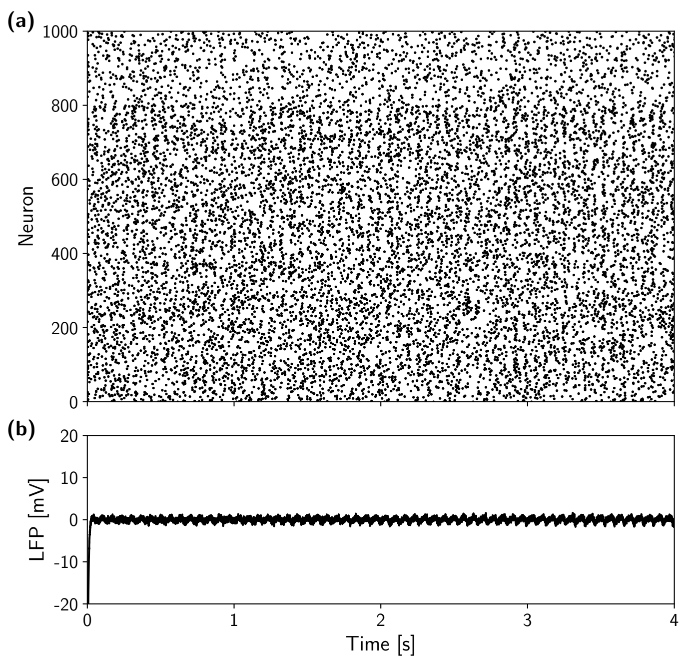

Fig. 1. Simulation with normal network activity: (a) raster plot and (b) LFP signal.

2.1 Induction of Ictogenesis

The process of ictogenesis was simulated by a gradual increment of the synaptic weights from 0.1 and -0.1 to 2 and -4 corresponding to the excitatory and inhibitory neurons, respectively. The results are shown in the figure 2.

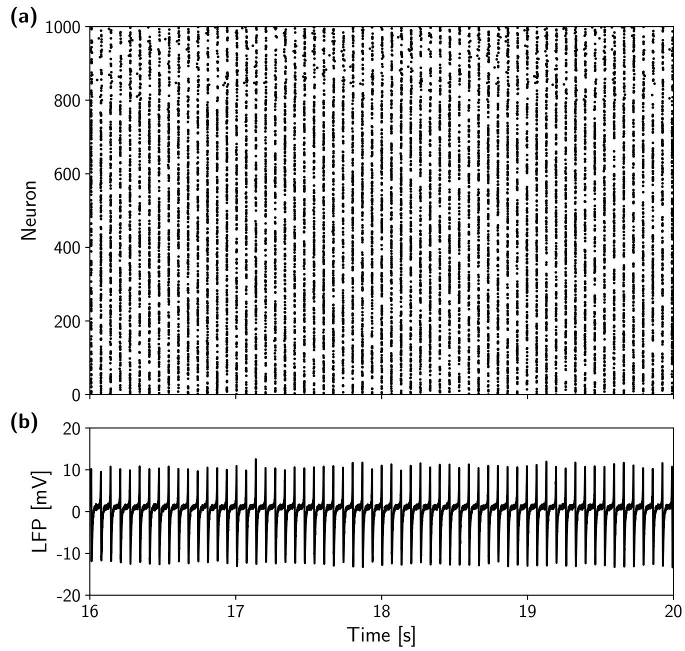

Fig. 2. Network synchronization by modification in synaptic weights: (a) raster plot and (b) LFP signal.

In order to suppress this aberrant activity, a train of pulses of electrical current J_k should be applied as follows:

$$J_k = \nu_k \bar{J} \max\{0, \operatorname{sgn}[\cos(2\pi t f) - \cos(\pi d f)]\} \tag{5}$$

where ν_k is the weight, obtained by an uniform distribution, that relates the train of amplitude $\bar{J}$ to the k-neuron, d is the pulse duration, f the frequency of pulses, and $\operatorname{sgn}(x)$ is the standard relay function. The objective here is to search the optimum frequency f given $\bar{J} = 200$ μA and $d = 0.1$ ms.

3 The ε-greedy algorithm

The ε-greedy is a reinforcement learning algorithm designed to solve the multi-armed bandit (MAB) problem [18]. The problem of MAB was introduced by [15], and it consists of an agent who is presented with a range of actions, commonly known as "arms", where each action generates a specific reward upon selection. The primary goal of the agent is to optimize its total accumulated reward throughout a sequence of engagements. It achieves this by acquiring an understanding of which actions are more advantageous and leveraging this knowledge to enhance its decision-making abilities as time progresses.

Basically, the ε-greedy algorithm is based on the control of the degree of exploration by changing the parameter ε. If the value of ε increases (more exploratory), the average reward will be initially low but will become higher as the agent discovers and explores the best actions. However, in the long term, due to its exploratory nature, the agent might not stick to the optimal action as much as desired. On the other hand, if the value of ε decreases (more exploitative), the average reward will increase more slowly, but this is the right strategy to maximize the expected reward in the next step. Unlike the exploratory case, in the long run, it will more frequently choose the optimal case.

Our objective here is the desynchronization of the neuron firing pattern by the injection of electrical current in the form of a train of pulses. Given that both the amplitude and pulse duration remain fixed, the variable we will manipulate is the frequency of these pulses, adjusting it through an algorithm. To determine the appropriate action—whether to increase, decrease, or maintain the pulse frequency—we must establish a criterion for discerning aberrant behavior within the neuronal population. To achieve this, our approach involves utilizing the frequency with the highest magnitude, denoted as f_a, obtained through the application of a Fast Fourier Transform (FFT) on the LFP signal. In Figure 3, you can observe the frequency spectrum of the LFP signal depicted in Figures 1b and 2b, considering a 300 ms time window. Notably, the frequency associated with the highest amplitude in the desynchronized signal, as shown in Figure 1, exhibits a greater magnitude than that of the synchronized LFP signal.

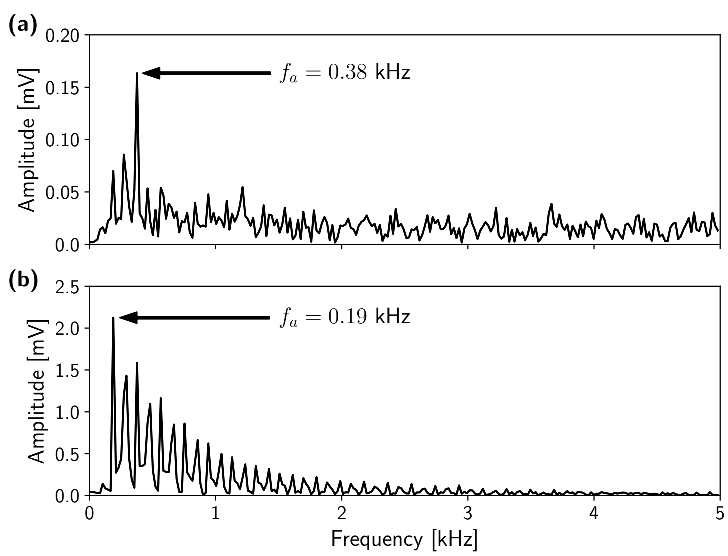

Fig. 3. Frequency spectrum for the LFP signal: (a) presented at the figure 1b and (b) presented at the figure 2b.

The algorithm 1 provides a concise summary of the steps involved in the ε-greedy scheme. It is worth noting that the actions taken within this scheme are directly linked to the adjustment of pulse frequency, either increasing or decreasing it by 10 Hz. This decision carries significance as higher frequencies not only induce desynchronization in neuron firing but also result in an increase in energy consumption associated with the external stimulus. To address this dual concern, the reward function is designed to take into account both the current state variable and the pulse frequency itself, aiming to identify a frequency that not only promotes desynchronization but also minimizes the overall energy expenditure of the stimulus. The reward function is defined as follows:

$$r = 0.01[10\exp(5|\vartheta - \vartheta_{max}|/\vartheta_{max}) + p_s \exp(5f/f_{max}) - 40] \tag{6}$$

with $\vartheta_{max} = 4$ representing the maximum attainable state variable, while f_{max} denotes the expected maximum frequency. Additionally, the parameter p_s plays a pivotal role, representing the weight of the pulse frequency in determining its optimal value without exceeding the total energy of the external stimulus.

Algorithm 1 ε-greedy algorithm.

```
Set exploration probability ε
Initialize main reward R̄ ← 0
Initialize specific reward R(a) ← 0, for a from 1 to k
loop
    Pick a random number p between 0 and 1
    if p < ε then
        Choose an action randomly
        A ← random(a)
    else
        Choose the action that minimizes R(a)
        A ← arg min_a R(a)
    end if
    Apply A into the model and receive the reward r
    According to the previous R̄, define the specific reward R_a
    Calculate the main reward: R̄ ← R̄ + α_1(r − R̄)
    Calculate the specific reward: R(a) ← R(a) + α_2(R_a − R(a))
end loop
```

As mentioned previously, we have established three distinct actions: action 0, which maintains the pulse frequency unchanged; action 1, designed to increase the frequency; and action 2, intended to decrease it. To further incentivize the action that leads to a significant enhancement in the main reward, denoted as $\bar{R}$, and to discourage the other actions, we have defined a set of specific rewards as follows:

$$\boldsymbol{R}_a = \begin{array}{ccc} a=0 & a=1 & a=2 \\ \begin{bmatrix} -1 & 0.5 & 0.5 \\ 0.5 & -1 & -1 \end{bmatrix} & \begin{bmatrix} -0.1 & -1 & 0.5 \\ -1 & 0.5 & -1 \end{bmatrix} & \begin{bmatrix} -0.1 & 0.5 & -1 \\ -1 & -1 & 0.5 \end{bmatrix} \end{array} \tag{7}$$

For instance, consider a scenario where the preceding action involved an increase in the parameter f, resulting in a decrease in the primary reward $\bar{R}$. In such cases, we opt for the first row of specific rewards within the central matrix defined in equation (7). This entails assigning values of $R_0 = -0.1$ to action 0, $R_1 = -1$ to action 1, and $R_2 = 0.5$ to action 2. Conversely, if the prior action led to an improvement in $\bar{R}$, the second row of the same matrix is utilized to determine the corresponding rewards. It should be pointed out that Gaussian noise with a variance of 0.1 was added to the reward matrix (7) in order to emulate a stochastic environment and increase the robustness of the algorithm.

4 Simulation results

The ε-greedy algorithm was employed with time steps set at 400 ms, utilizing parameter values of $\varepsilon = 0.2$, $\alpha_1 = 0.9$, and $\alpha_2 = 0.2$. To comprehensively assess the impact of varying p_s, we conducted 10 distinct simulations for each p_s. As illustrated in Figure 4, this figure displays the average outcomes alongside their corresponding variances for each p_s. As anticipated, due to the inclusion of pulse frequency within the reward function, in a task aimed at minimization, an increase in the value of p_s corresponds to a reduction in the final frequency. This outcome signifies a lower energy consumption by the electrode.

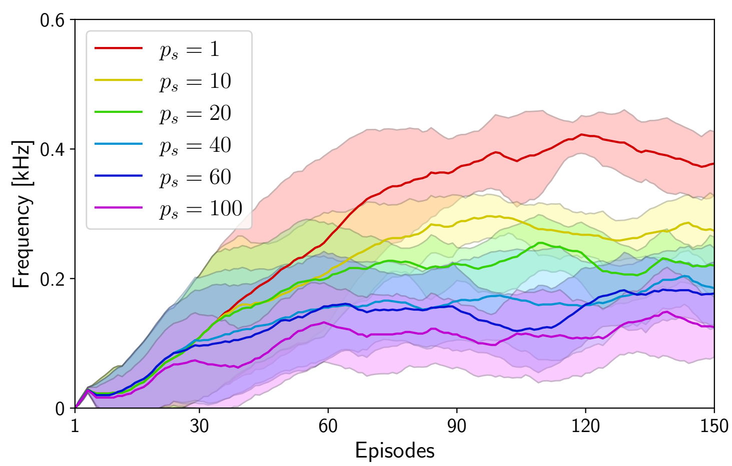

Fig. 4. Simulation results for the ε-greedy algorithm changing the p_s parameter.

Concerning the impact on the frequency of highest amplitude in the Fourier spectrum, denoted as f_a, it becomes evident that these effects diminish as the values of p_s increase, as depicted in Figure 5. However, even with p_s set to 60, we observe a f_a value of 996 Hz, which exceeds the requirement, especially when comparing it to the frequency of highest amplitude in a desynchronous signal, as shown in Figure 3b, where f_a is at 380 Hz. For the scenario with $p_s = 60$, the mean pulse frequency stabilizes at approximately $f = 159$ Hz in the steady state, as illustrated in Figure 4. This particular frequency translates to a more efficient energy consumption profile when compared to the higher values of p_s.

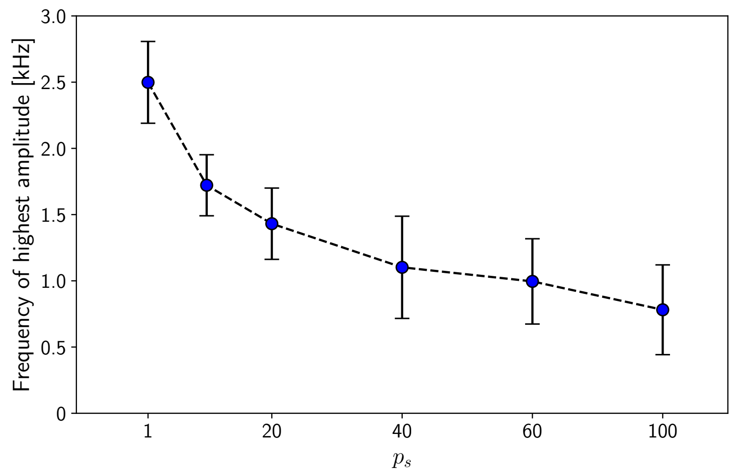

Fig. 5. Simulation results for the ε-greedy algorithm changing the p_s parameter.

The results depicted in Figure 6 provide a visual representation of the raster plot and LFP generated when the pulse frequency is set to $f = 159$ Hz, as determined by the ε-greedy algorithm. As observed, this choice of pulse frequency achieved the desired desynchronization of the raster plot without necessitating excessive control effort.

4.1 Results remarks

Some points of the obtained results must be highlighted. The parameter ε is directly related to the algorithm's convergence. While increasing ε, the convergence cannot be guaranteed due to the fact that the agent will choose more random actions; decreasing it will restrict the exploration of other scenarios (frequencies). The chosen value of 0.2 was the best value found to balance the trade-off.

Other important aspects are the limitations of the reductionist computational approach used here. Beginning with the neural tissue model, as an all-to-all connected network with few neurons (1000) of no particular type (dimensionless parameters varied randomly), it does not faithfully reproduce cytoarchitectonics of any circuitry involved in ictogenic phenomena. Furthermore, the simplistic method for the calculation of LFP does not consider specific aspects of neurophysics. Instead, given its tremendous complexity in real life, here we merely used a uniformly randomized weight vector to encompass all possible factors

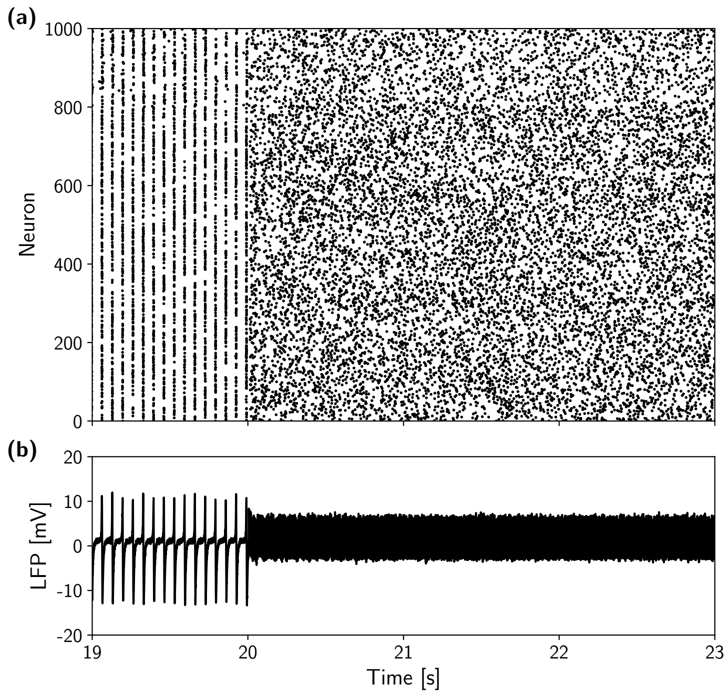

Fig. 6. Simulation results for $f = 159$ Hz: (a) raster plot and (b) LFP signal.

(e.g. cell dimensions, geometry, distance, field conduction pathway, etc.). On the other hand, we do not believe that these issues represent a significant impact on the conclusions here. First, although simple, the network can produce dynamical features of interest, particularly stable states related to different levels of synchronization, the transition of which can be induced or modulated (such as in epilepsy) [13]. By its turn, LFP signals were used here only as a tool to detect the state of the network by means of an indirect measurement of biophysical factors, mimicking how it would be done by electrophysiological recordings in real-life closed-loop neurotechnologies. In any case and of course, frequency and time values found in this study must not be taken by their face values. Yet, this was never the intention. Conversely, the main purpose of this work was to assess the feasibility of the investigation of automatic intelligent control strategies using computational neuroscience tools. Experiments with more sophisticated and realistic approaches must be carried out if one wishes to implement such technologies in lab setups or even clinical scenarios.

5 Conclusions

This paper introduces the application of a machine learning algorithm to discover the optimal frequency of a pulse train used to mitigate ictogenesis in a network of neurons. To model the neuronal mass, a mathematical framework based on the work of Izhikevich [11], capable of representing both normal and pathological firing patterns, is employed. To counteract these abnormal neuronal activities, a pulse train is administered. The ε-greedy algorithm is leveraged to determine the pulse frequency required to disrupt synchronous behavior—a hallmark of undesired signals observed in epileptic seizures—while minimizing excessive control effort. The algorithm's effectiveness is assessed through computational simulations, revealing that the ε-greedy approach successfully identifies the essential pulse frequency to suppress the disorder. In any case, given the very straightforward nature of computational methods applied to simulate neural tissue activity here, it is crucial to perform additional experiments with more sophisticated approaches to better understand the feasibility of using automatic intelligent control methods in closed-loop neurotechnologies.

Acknowledgments. This work was supported by the Marie Skłodowska-Curie Individual Fellowship MoRPHEUS granted to VC, Grant Agreement no. 101032054, funded by the European Union under the framework programme H2020-EU.1.3. – EXCELLENT SCIENCE, and by the Brazilian research agencies CNPq and CAPES.

References

1. Alamri, A., Pereira, E.A.: Deep brain stimulation for chronic pain. Neurosurgery Clinics **33**(3), 311–321 (2022)
2. Bao, S.c., Khan, A., Song, R., Tong, R.K.y.: Rewiring the lesioned brain: electrical stimulation for post-stroke motor restoration. Journal of stroke **22**(1), 47 (2020)
3. Carè, M., Chiappalone, M., Cota, V.R.: Personalized strategies of neurostimulation: from static biomarkers to dynamic closed-loop assessment of neural function. Frontiers in Neuroscience **18**, 1363128 (2024)
4. Chiappalone, M., Cota, V.R., Carè, M., Di Florio, M., Beaubois, R., Buccelli, S., Barban, F., Brofiga, M., Averna, A., Bonacini, F., et al.: Neuromorphic-based neuroprostheses for brain rewiring: state-of-the-art and perspectives in neuroengineering. Brain Sciences **12**(11), 1578 (2022)
5. Chin, J.H., Vora, N.: The global burden of neurologic diseases. Neurology **83**(4), 349–351 (2014)
6. Cota, V.R., Cançado, S.A.V., Moraes, M.F.D.: On temporal scale-free non-periodic stimulation and its mechanisms as an infinite improbability drive of the brain's functional connectogram. Frontiers in Neuroinformatics **17**, 1173597 (2023)
7. Engel, J.: Epilepsy: global issues for the practicing neurologist. Demos Medical Publishing. **2** (2005)
8. Figee, M., Riva-Posse, P., Choi, K.S., Bederson, L., Mayberg, H.S., Kopell, B.H.: Deep brain stimulation for depression. Neurotherapeutics **19**(4), 1229–1245 (2022)
9. French, J.A.: Refractory epilepsy: clinical overview. Epilepsia **48**, 3–7 (2007)
10. Hariz, M., Blomstedt, P.: Deep brain stimulation for parkinson's disease. Journal of internal medicine **292**(5), 764–778 (2022)
11. Izhikevich, E.M.: Simple model of spiking neurons. IEEE Transactions on neural networks **14**(6), 1569–1572 (2003)
12. Laxpati, N.G., Kasoff, W.S., Gross, R.E.: Deep brain stimulation for the treatment of epilepsy: circuits, targets, and trials. Neurotherapeutics **11**, 508–526 (2014)

13. Oliveira, J.P.S.e., Discacciati, V.R.P., Medeiros, D.d.C., Moraes, M.F.D., Pereira, G.S., França, K.L.d.A., Cota, V.R.: In silico investigation of the effects of distinct temporal patterns of electrical stimulation to the amygdala using a network of izhikevich neurons. In: Latin American Workshop on Computational Neuroscience. pp. 132–152. Springer (2021)
14. Pena, R.F., Zaks, M.A., Roque, A.C.: Dynamics of spontaneous activity in random networks with multiple neuron subtypes and synaptic noise: Spontaneous activity in networks with synaptic noise. Journal of computational neuroscience **45**, 1–28 (2018)
15. Robbins, H.: Some aspects of the sequential design of experiments. Bulletin of the American Mathematical Society **58**(5), 527–535 (1952)
16. Savage, N.: Epidemiology: The complexities of epilepsy. Nature **511**(7508) (2014)
17. Spencer, S.S.: When should temporal-lobe epilepsy be treated surgically? The Lancet Neurology **1**(6), 375–382 (2002)
18. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press. (2018)