Biomedical Signal Processing and Control 63 (2021) 102196

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Parameter investigation of topological data analysis for EEG signals

Fatih Altındiş [a,c,*], Bülent Yılmaz [a,b,c], Sergey Borisenok [a,e], Kutay İçöz [a,b,d]

[a] Electrical and Electronics Engineering Department, Abdullah Gül University, Kayseri, 38080, Turkey
[b] Bioengineering Department, Abdullah Gül University, Kayseri, 38080, Turkey
[c] BISA (Biomedical Instrumentation and Signal Analysis) Laboratory, Abdullah Gül University, Kayseri, 38080, Turkey
[d] BioMINDS (Bio Micro/Nano Devices and Sensors) Laboratory, Abdullah Gül University, Kayseri, 38080, Turkey
[e] Feza Gürsey Center for Physics and Mathematics, Boğaziçi University, Istanbul, 34684, Turkey

ARTICLE INFO

Keywords:
Topological data analysis
EEG
Brain-Computer interface
Persistent homology
False nearest neighbors
Motor intention waves

ABSTRACT

Topological data analysis (TDA) methods have become appealing in EEG signal processing, because they may help the scientists explore new features of complex and large amount of data by simplifying the process from a geometrical perspective. Time delay embedding is a common approach to embed EEG signals into the state space. Parameters of this embedding method are variable and the structure of the state space can be entirely different depending on their selection. Additionally, extracted persistent homologies of the state spaces depend on filtration level and the number of points used. In this study, we showed how to adapt false nearest neighbor (FNN) test to find out the suitable/optimal time embedding parameters (*i.e.*, time delay and embedding dimension) for EEG signals, and compared their effects on different types of artefacts and motor intention waves that are commonly used in brain-computer interfaces. We extracted and compared persistent homologies of state spaces that were reconstructed with four different sets of parameters. Later, the effect of filtration level on extracted persistent homologies was compared, and statistical significance levels were computed between left- and right-hand movement imaginations. Finally, computational cost of the discussed methods was found, and the adaptability of this method to a real-time application was evaluated. We demonstrated that the discussed parameters of the TDA approach were highly crucial to extract true topological features of the EEG signals, and the adapted testing approaches depicted the applicability of this approach on real-time analysis of EEG signals.

1. Introduction

As the acquisition of high resolution multichannel EEG signals got cheaper and easier, its usage in brain-computer interfaces (BCIs) has become widespread. Sampling frequency and the number of channels used have been increased in order to collect more information from subjects leading to more data to process and analyze, which requires more time and processing power. On the other hand, BCIs need to respond quickly to patient's requests in real life. For that matter, topological data analysis (TDA) methods have become appealing to be utilized in BCI systems, because they may help the scientists explore new features of complex and large amount of data by simplifying the process from a geometrical perspective [1,2]. Especially, when one considers that TDA is more convenient to handle complex and large amount of data, it is a good idea to focus on TDA of EEG signals [3]. Despite in many studies TDA was used on EEG signals [4–6], many questions remain to be answered. EEG signals are acquired from multiple channels and frequently contain artefacts and noise. This study aimed at tackling the following questions from the perspective of persistent homology analysis method of topology: What is the performance of TDA under noisy EEG signals? Is it critical to have clear EEG signals for TDA to perform well? At what extend do noise and artefacts spoil TDA performance? What are the key parameters of TDA for an effective EEG analysis? How can we optimize those parameters? Can this method operate in real-time if implemented in BCI systems?

EEG signals can be treated as scalar time series that are collected from multiple channels, and there are two main ways of mapping such time series into multidimensional Euclidean space, which are time delay embedding and spatial embedding. Time delay embedding is based on Taken's embedding theorem [7], and it is very useful to reconstruct state space from single signal source [8]. Even if there are multiple sources, one can reconstruct state space of each signal source independent from other sources with time delay embedding. Spatial embedding is an alternative method to reconstruct state space. It uses signals from

* Corresponding author at: Abdullah Gül University, Sumer Campus, Kocasinan, Kayseri, Turkey.
E-mail address: fatih.altindis@agu.edu.tr (F. Altındiş).

https://doi.org/10.1016/j.bspc.2020.102196
Received 26 February 2020; Received in revised form 20 July 2020; Accepted 29 August 2020
Available online 17 September 2020

multiple channels to create state space. This allows reduction of data to be processed in the $O(n)$. Both embedding approaches are suitable for reconstruction of state space from EEG signals. Stam shared several studies in which these embedding approaches were tested [9], however, which approach was favorable for which case has remained unknown. Additionally, parameters of both embedding methods need to be well-tuned, otherwise arbitrarily chosen parameters mislead reconstruction of the state space and cause to fade underlying assets and highlight noises. Liebert et al. [10], Mindlin et al. [11], and Rosenstein et al. [12] showed that optimal values for the parameters of time delay embedding can be found using several tests. Even though Pritchard et al. [13] and Celluci et al. [14] presented the applicability of these test methods to EEG signals, it is still not clear how these test results are affected by the artefacts and noise.

Topological structures of the state spaces are expressed by homologies such as connectivity, edge, hole and surface [15]. Vietoris-Rips filtration is a resourceful method to extract homologies of the state spaces [16]. This filtration approach assumes points of the state space as expanding disks with initial radii equal to zero. Disk radii start to grow step by step, and filtration ends when the disk radii reach a predetermined value, τ_{max}. When any two disks overlap during the filtration new connections between them emerge, and when one disk embraces another disk's center, connection between them dies. Fig. 1a shows a simple example of filtration. At the end of filtration, existing homologies are counted and represented by *Betti* numbers (β_n). *Betti*$_0$ (β_0) is the number of connected point groups, *Betti*$_1$ (β_1) corresponds to the number of connections (edges) between points, and *Betti*$_2$ (β_2) indicates surfaces, and so on. For instance, in Fig. 1a all 5 points are connected and there are 2 edges at the end of filtration. The corresponding *Betti* numbers of this state space are $\beta_0 = 1$ and $\beta_1 = 2$.

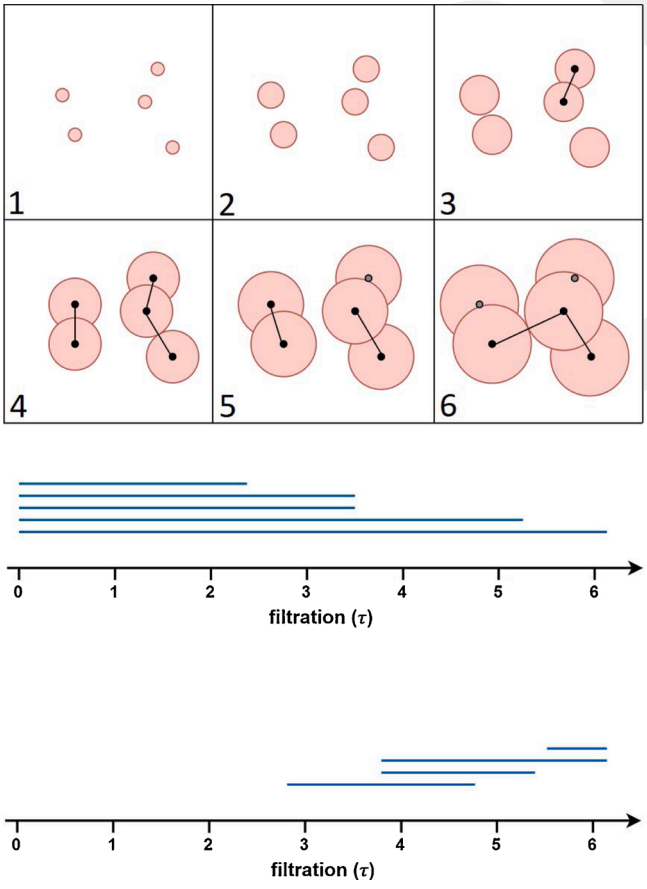

Fig. 1. a) Illustration of a state space that consists of points. Six steps of filtration are visualized. Emerged and engulfed homologies shown through steps. b) Betti barcodes of state space shown for β_0 and β_1 numbers.

It can be seen that some edges emerge and die between step 1 and 6, but *Betti* numbers do not carry this evolution information. The evolution of the state space can be visualized by *Betti* barcodes as shown in Fig. 1b. When a new homology emerges, new bar appears on the graph with the corresponding birth value, and the bar ends when that homology dies. Eventually, *Betti* barcodes show the lifetime of all emerged homologies. Short lived homologies correspond to noise and artefacts. Homologies with long lifetime are referred as persistent homologies, and they possess the underlying features and dominant structures of the state space [17]. Perea et al. proposed an algorithm to extract periodicity of gene expression signals using persistent homology [18], while Wang et al. [19]. and Dabaghian et al. [20] used persistent homology to reveal main features of brain signals. Persistent homology analysis aims to provide robustness against noise, but without the numerical evaluation of a certain application it is not clear how much noise or artefact can be ignored with persistent homology method. Besides, the value of τ_{max} is an important parameter for the extracted persistent homologies. If chosen τ_{max} is too small, then many noise related homologies may appear as persistent homologies. On the contrary, if chosen τ_{max} is too large, feature related homologies disappear and resultant persistent homologies become meaningless. Top it all, computational cost increases exponentially as τ_{max} value gets larger, and it detracts real-time applicability of this method. Thus, the τ_{max} value should be large enough to eliminate noise related homologies; meanwhile it should be small enough to keep feature related homologies of analyzed state space. Moreover, computational cost of the analysis should be feasible to be used in real-time BCI systems.

As mentioned above, determination of suitable/optimal parameters for real-time applications of TDA is required. In the previous studies on TDA of EEG signals, state space reconstruction and filtration steps have no common methodology to follow. There is no information about processing time, and whether it is possible to implement TDA in real-time BCIs. Noise and artefacts are inherent to the EEG signals, and TDA has proven to be robust against noise [2,21]. However, how this approach is affected by noise and artefacts needs to be addressed. In this study, we handled each of these issues systematically. Firstly, in state space reconstruction, determination of optimal parameters for time delay and spatial embedding approaches was investigated using synthetic and real EEG signals with various noise and artefact types. In addition, persistent homologies were extracted as *Betti* numbers with Vietoris-Rips filtration and compared for left- and right-hand motor intensions on two EEG channels located on the left and right sides of the scalp on the motor cortex (C3 and C4 channels). Moreover, processing times of these steps were recorded to demonstrate the feasibility of this approach for real-time BCIs.

2. Materials and methods

2.1. Synthetic EEG

The approach suggested in [22] used real EEG signals for generating synthetic ones to decrease the calibration time of BCI. In our approach, we aimed at imitating mu and beta band power changes during the presence of motor imagery rather than imitating all the characteristics of EEG signals. For that reason our approach, the synthetic EEG signals were composed of summation of pure sine waves generated using MATLAB 2016a built-in 'sin' function. Besides, this approach allowed us to observe and isolate the effects of different types of noise and artefacts on further steps of the topological analysis. Later, noise and artefacts were added gradually to investigate their effects on the state space reconstruction. The resultant signals had frequency components between 0.5 Hz and 100 Hz. Certain band powers of the generated signals were altered in order to simulate motor intention (MI) waves. The band powers of two frequency bands, mu band (7.5–12.5 Hz) and beta band (15–30 Hz), were changed by simply changing the amplitude of sine

waves in association with band frequencies [23]. As depicted in Fig. 2 during imagery period while the subject is imagining the movement of the extremities, for example, without actually moving them, the ratio of beta and mu band powers increases when compared to the resting state, which is called event-related desynchronization (ERD). This concept is defined as the motor intension waves, and is commonly used in BCI applications.

2.2. Real EEG

Here, we used a publicly avaliable EEG dataset that was published by the Technical University of Graz, Austria in 2008 for a BCI competition [24]. This dataset (Graz dataset) included EEG signals that were obtained from 8 subjects. Each subject performed 120 right-hand and 120 left-hand movement imaginations. The experimental paradigm used during the EEG recordings was shown in Fig. 3. The sampling rate in the experiments was 250 samples/s, and each imagery movement trial lasted around 9 s. EEG signals of this dataset did not contain any artefacts such as eye blinks.

2.3. Noise types

Four different noise types were simulated and added to synthetic EEG signals to investigate their effect. These were i) muscle related artefacts in the frequency range of 40−80 Hz, ii) eye blinks, iii) signal discontinuities and iv) linear trends (Fig. 4) [25]. The artefacts lasted between 0.5 and 1 s.

2.4. State space reconstruction

Time delay embedding is a standard procedure in topological analysis to reconstruct state space from single signal source, *i.e.*, time series [11,26,27]. Let time series $x(t),\ t \in Z$ belongs to a single channel EEG signal, where each $x(t)$ corresponds to one sample from the signal. This $x(t)$ signal can be embedded into M *-dimensional* state space with ΔT time delay according to Taken's theorem, where M and $\Delta T \in Z$ [7]. Consequently, coordinates of the points in the state space can be represented with the following equation;

$$y(t) = \{x(t),\ x(t+\Delta T),\ x(t+2\Delta T), \ldots,\ x(t+(M-1)\Delta T)\} \quad (1)$$

These two parameters, ΔT and M, determine the location of each point, $y(t)$, on the state space. Additionally, the number of points that are present in the state space is decided using the length of the time window. Three parameters shape the state space using single channel EEG signal as shown in Fig. 5.

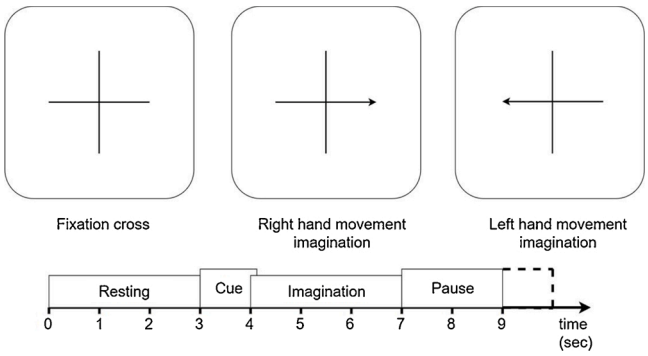

Fig. 3. Experimental paradigm for each trial started with a fixation cross (left), a cue for right or left-hand imagery (middle and right respectively) was shown to the subject, and then hand movement imagery period started.

Each stage of topological analysis, starting from embedding EEG signal into the state space up to extracting persistent homologies, requires delicate approaches to decide parameter values that are crucial on that particular stage. It is highly important to avoid any ambiguities because mistakenly assigned parameter values will have major impacts on the resultant topological features.

The method called "false nearest neighbor" (FNN) test first proposed in [10] has been widely used for determining proper embedding dimension of nonlinear systems [28]. It can be considered as an optimal way to find the lowest embedding dimension to reconstruct the state space. Due to the nonlinear structure of the EEG signals, it is suitable to employ FNN test to find proper embedding dimension. Better yet, FNN test returns the minimum embedding dimension that is possible to reconstruct state space, which directly helps to reduce computational cost of further steps of topological analysis.

Here, we adapted the same test with an extended approach as [31] suggested to find not only the embedding dimension but also delay time and time window parameter values. The methodology of choosing proper values for the optimizing embedding parameters of nonlinear systems should be well-defined; otherwise further analysis may lead to arbitrary results. For instance, if the chosen embedding dimension is

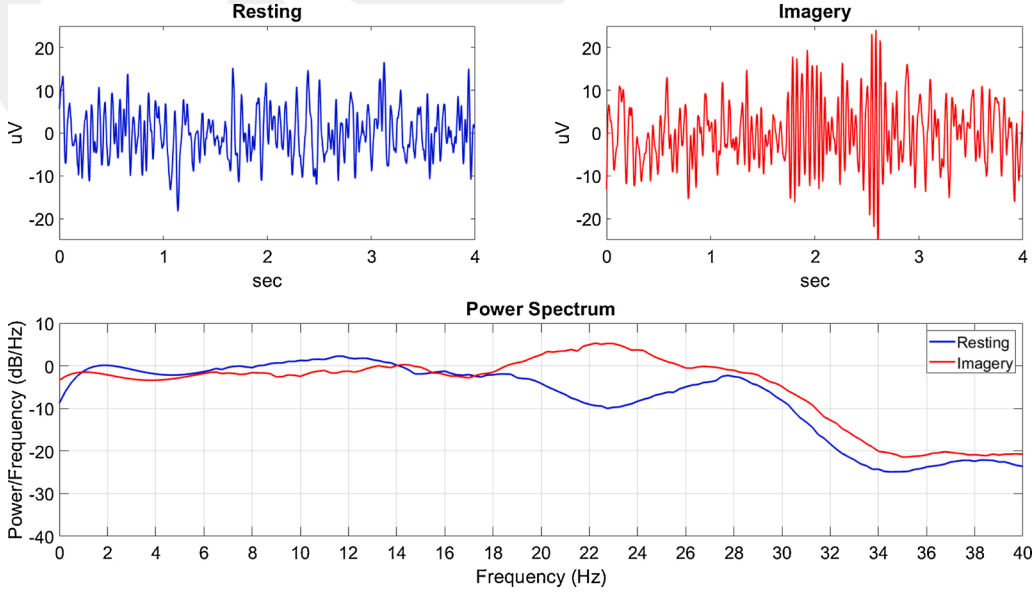

Fig. 2. Synthetic EEG signals for resting (blue) and imagery (red) states, and their power spectra are shown respectively.

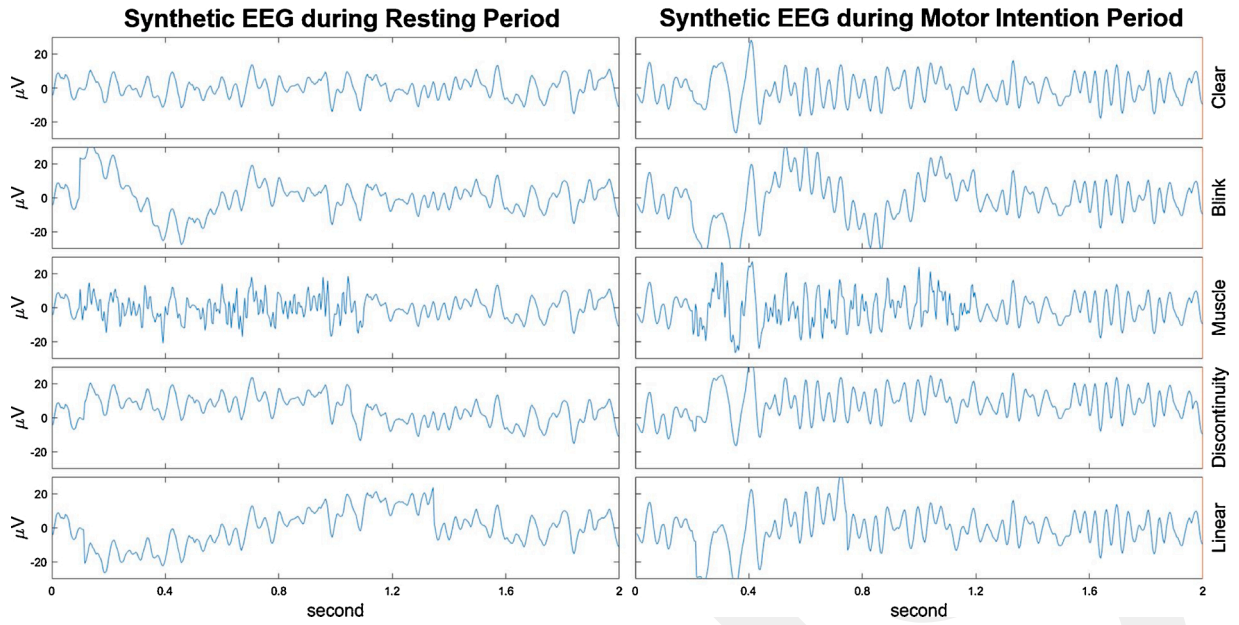

Fig. 4. Synthetic EEG signals with and without MI waves are depicted. Clean signals (first row) and signals with different artefact types (the type is shown on the right side of each panel) are presented.

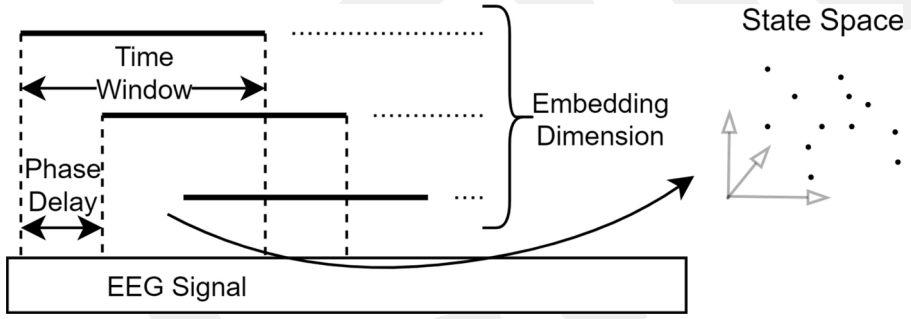

Fig. 5. Time delay embedding uses signal samples as the coordinates of state space points. Time window determines the number of points, phase delay determines the amount of time window shift, and embedding dimension corresponds to the state space dimension.

lower than the optimal value, it yields a densely reconstructed state space. This is not desired because features of the state space may overlap in the dense structure. On the other hand, if the embedding dimension is high, it causes sparsity in the reconstructed state space, because embedded points in the state space land far from each other. Besides, high dimensional state spaces are not cost effective, because they require exponentially increased computational power for further steps of topological analysis.

FNN test examines the changes in the distance between points and their nearest neighbors, while incrementing embedding dimension one by one starting from 1-*dimensional* state space. Let (M-1)-*dimensional* state space be created with time delay embedding method and let the distance between a point and its nearest neighbor be

$$R_{M-1}^2 = \sum_{k=0}^{M-2} [x(t+k\Delta T) - x^r(t+k\Delta T)]^2 \tag{2}$$

where $x(t+k\Delta T)$ are coordinates of a point in the state space and $x^r(t+k\Delta T)$ are the coordinates of the nearest neighbor of that point. Let this state space be folded up into M -dimension by adding Mth coordinate (by adding another data point while creating a point in the state space) to every point. Then, the new distance between a point and its nearest neighbor will be

$$R_M^2 = R_{M-1}^2 + [x(t+(M-1)\Delta T) - x^r(t+(M-1)\Delta T)]^2 \tag{3}$$

Consequently, the amount of distance change between $M-1$ to M dimension can be calculated with the following equation;

$$\frac{|R_M - R_{M-1}|}{R_{M-1}} \tag{4}$$

Firstly, while going from $M-1$ to M dimension if the nearest neighbor changes for a point we accept this neighbor as the false nearest neighbor. Secondly, if the change in distance is below a predefined threshold (R_{tol}), then FNN test accepts corresponding nearest neighbor as a true nearest neighbor, or else it is marked as the false nearest neighbor. We would like to state that, Kennel et al. [28] and Liebert et al. [10] showed that R_{tol} can take a value between 10 and 20 and it is assumed as $R_{tol} = 10$ in this study as suggested in [10]. Total percentage of the FNN is calculated for all dimensions starting from 1 to M. The lowest embedding dimension that has FNN percentage lower than 1% was chosen as the best embedding dimension.

In addition to embedding dimension, time delay (ΔT) needs to be optimized because if ΔT is too low, points in the reconstructed state space may land on same places and cause redundancy in the state space. On the other hand, a high time delay value causes points to land on independently and randomly. Mutual information or autocorrelation methods can be used to find the suitable time delay value, however,

rather than finding suitable values for embedding dimension and time delay parameters independently, Kravoskâ discussed an approach to find suitable ΔT value using FNN test [31]. We utilized the same approach and used FNN test as a basis and repeated the test for sequentially increasing ΔT values. Eventually, a grid of FNN test results for different time delay and embedding dimension pairs was created. Time delay and embedding dimension pair that has the lowest FNN rate was chosen to be the best (most suitable) values for reconstructing the state space. Moreover, we repeated the tests for different time window lengths to find out its best value. Thanks to this approach, not only the most suitable parameter values but also computationally efficient parameter values were found.

Synthetic EEG signals were first used in FNN test to find out the optimal values for embedding dimension, time delay and time window. Each test was executed for 15 consecutive time delay and embedding dimension values starting from 1. The test results were plotted as heat maps that showed the percentage of FNN for every combination of embedding dimension and time delay values. Next, the effect of the number of points (length of the time window) in the state space was examined to find out the best time window length. Three different time window lengths, which have 50, 150, and 250 points in the reconstructed state space, were used. More importantly, different artefact types and motor activity imagination (intension) were simulated on synthetic signals to investigate how FNN test results were affected by them.

2.5. Persistent homology

Topological analysis was performed using a Java based software, called JPLEX, that could also run on MATLAB [29]. This software allows

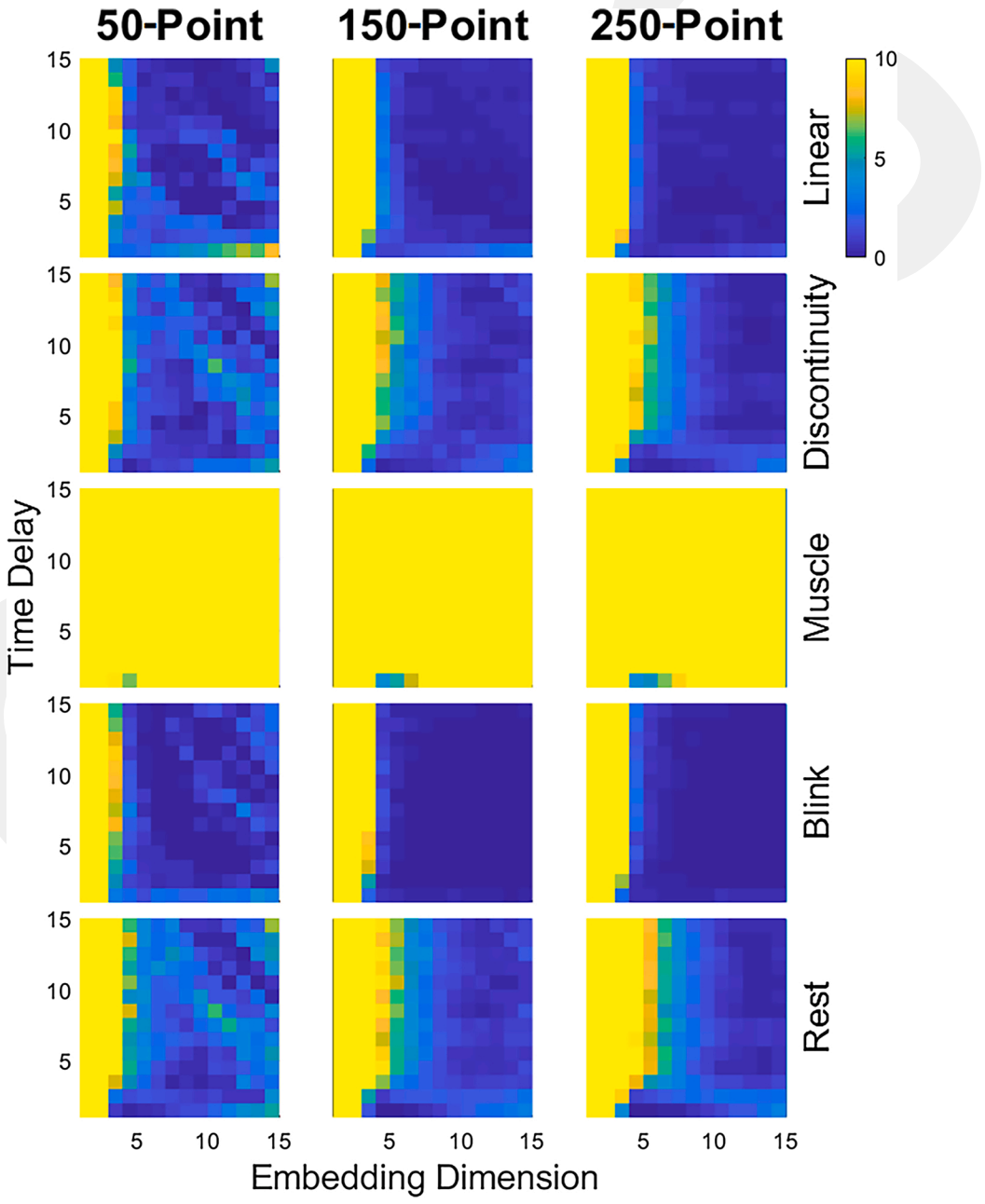

Fig. 6. FNN rates of *synthetic EEG signals* for every combination of time delay and embedding dimension were shown as heat maps. In addition, each column corresponds to FNN rates for different time window lengths (50, 150 and 250 points), and the first row and second to fifth rows correspond respectively to the FNN rates of artefact-free and artefact added (blink, muscle, discontinuity, and linear) *synthetic EEG signals*.

the researcher to reconstruct state space using Euclidean coordinate arrays of points, and can perform Vietoris-Rips stream analysis [17]. Once the state space was created, the radius (R) of spherical volume of M -dimensional state space was calculated, which was also used to determine the maximum filtration ($\tau_{\max}$) value. Since the volume of state space calculated in terms of radius (R), $\tau_{\max}$ could take a value proportional to the radius. Thereby, how much of the state space would be covered at the end of the stream would be known beforehand. For instance, if the $\tau_{\max}$ value was one fourth of the radius ($\tau_{max} = R/4$), then any point of state space would be large enough to cover 25% of the whole state space at the end of the filtration. Even if we considered two distant points, 50% of the state space would be covered just by them at the end. We chose 5 different $\tau_{\max}$ values, R/4, R/10, R/25, R/50, and R/100, for the analysis by taking space coverage into consideration. Hence, the stream analysis became adaptive since $\tau_{\max}$ values were adjusted according to the state space radius.

3. Results

3.1. FNN tests

First, we examined synthetic EEG signals simulating the resting state. Four different artefact types (blink, muscle, discontinuity, and linear) were added to the signal to investigate their effect on FNN test. Each heat map in Fig. 6 corresponds to FNN test results of a pair of an artefact type and a time window length. Apart from the muscle artefact, test results were slightly affected by the artefacts. The lowest embedding dimension and time delay pair that resulted lower than 1% FNN rate (depicted dark blue on the heat maps) was selected as the best values for reconstructing the state space. Results showed that 3 or 5-dimensional state spaces with 1 sample time delay (the lowest row on each heat map) were sufficient to represent EEG signal in the reconstructed state space irrespective of the time window and artefacts. The dark blue regions on the heat maps demonstrated this result visually. Even for the muscle artefact case, time delay and embedding dimension values that provided the lowest FNN rate turned out to be the same values found for other artefact types.

Secondly, motor imagery (MI) waves were simulated as synthetic EEG signals, and FNN test results were shown in Fig. 7 as heat maps. Results indicated that low embedding dimension (3–5) with 1 sample time delay provided less than 1% FNN rate similar to the former test on resting state. Meanwhile, increased time delay and embedding dimension caused higher FNN rates (exceeding 1% threshold).

The FNN test findings of the real EEG signals were similar to the former two FNN test results. In Fig. 8, it can be seen that 1 sample time delay and 3–5 embedding dimension provided the lowest FNN rate, and those were chosen as the most suitable values. Distinctively, FNN rates for optimal values that were shown in Fig. 9 highlighted the importance of time window for reconstructing state space for real EEG signals. While 150 and 250-point time windows provided less than 1% FNN rate with optimal parameters, 50-point time window failed to comply with the findings of synthetic EEG signals. This meant that wider time windows should be preferred to reconstruct state spaces.

3.2. Topological features

In this part of the study, we employed C3 and C4 channels (corresponding to the motor cortex regions of the right- and left-hand movements respectively) of the real EEG data coming from 8 subjects as described in the materials and methods section. Persistent homologies were extracted for the left- and right-hand movement imaginations to be used in brain-computer interfaces. In order to demonstrate the importance of having the most suitable values that can provide less than 1% FNN rate, we used 4 different sets of M and ΔT values to construct state spaces: i) the optimal values of M and ΔT, ii) arbitrary values of M (greater than 4) and ΔT, iii) arbitrary M and the optimal ΔT, and iv) the

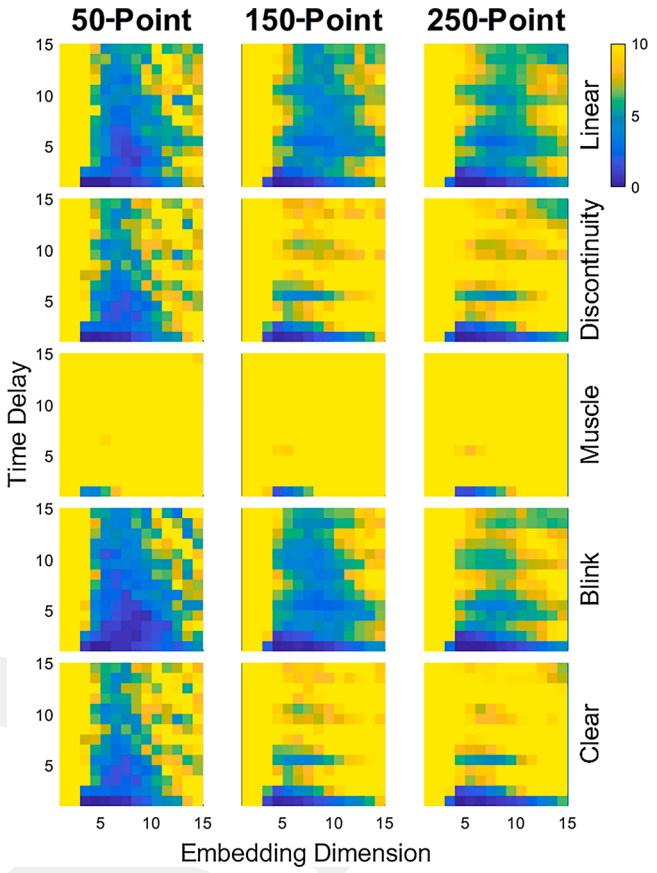

Fig. 7. FNN rates of *synthetic EEG signals* with *MI waves* for every combination of time delay and embedding dimension were shown as heat maps. In addition, each column corresponds to FNN rates for different time window lengths (50, 150 and 250 points), and the first row and second to fifth rows correspond respectively to the FNN rates of artefact-free and artefact added (blink, muscle, discontinuity, and linear) *synthetic EEG signals*.

optimal M and arbitrary ΔT (greater than 2). Extracted topological features, β_1 numbers, of the left- and right-hand movement imaginations were compared using the paired t-test with a significance level of $p < 0.05$. bar graphs in Fig. 10 show the number of persistent homologies extracted from C3 and C4 channels for each case. As shown in Table 1, state spaces created with the optimal parameters on the channel C4 were significantly different ($p < 0.05$) while other combinations were not.

Apart from the time delay and embedding dimension parameters of the reconstruction, we tested the effect of time window length together with the maximum filtration, $\tau_{\max}$,amount. State spaces were analyzed with 5 different $\tau_{\max}$ values and 3 different time window lengths in order to compare their effects.

The effect of $\tau_{\max}$ values on the extracted topological features was presented in twofold. First, the change of extracted Betti numbers through the filtration steps was shown. Initially, filtration started with $\tau = 0$, and it incrementally reached τ = R/4. Persistent homologies of resting and imagery period (left hand) were extracted and plotted at 5 different points ($\tau_{\max}$ = R/100, R/50, R/25, R/10, R/4) in Fig. 11. Stepwise investigation of filtration enabled us to understand how Betti numbers evolved through the filtration steps. As seen in that figure, Betti numbers converged to 1 with increasing filtration level. Another important point we should note from the figure is that Betti numbers of imagery period differed for 3 filtration values (R/25, R/50, R/100) while they were similar for other 2 filtration values. Secondly, paired t-test of the channels for resting and imagery periods that were shown in Table 2 supported this finding.

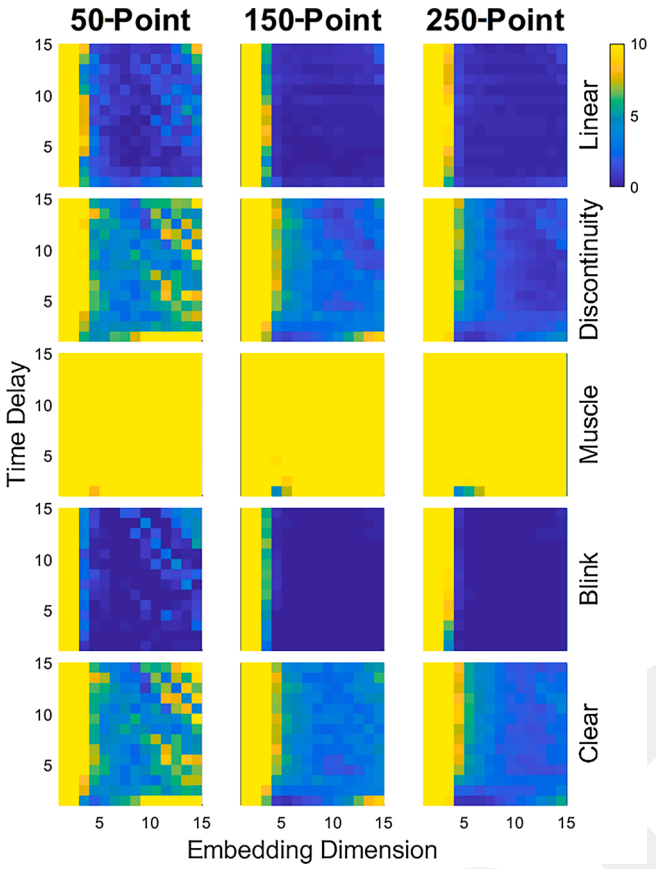

Fig. 8. FNN rates of *real EEG signals* for every combination of time delay and embedding dimension were shown as heat maps. In addition, each column corresponds to FNN rates for different time window lengths (50, 150 and 250 points), and the first row and second to fifth rows correspond respectively to the FNN rates of artefact-free and artefact added (blink, muscle, discontinuity, and linear) *real EEG signals*.

Moreover, we investigated the effect of time window length on the extracted persistent homologies. We used the same time window lengths (50, 150 and 250 points) that were used in the state space reconstruction phase. Extracted persistent homologies that belonged to resting and imagery periods, were compared, and paired *t*-test results are depicted in Table 2. Extracted persistent homologies were significantly different only for 250-point time window. Other two time window lengths failed to provide significantly different persistent homologies for C3 and C4 channels.

Finally, we analyzed EEG signals of all subjects using optimal parameter values that were found above. Extracted persistent homologies of EEG signals coming from C3 and C4 channels were compared for left- and right-hand imagery states. Paired *t*-test results showed that features obtained from of at least one of the channels were significantly ($p < 0.05$) different during imagery state of all subjects (Table 3).

3.3. Computation time

Computation time of EEG trials was recorded for different parameter combinations. For the same time window reconstructed state spaces were evaluated at 5 different filtration levels. Means of the total elapsed time for all trials were plotted. This process was repeated for 3 time window lengths, and the results were shown on the same figure (Fig. 12). Computations were carried out on Windows 10 desktop computer with 3.4 GHz processor and 8 GB RAM (Table 4).

4. Discussion

The first aim of the present study was to clarify ambiguous parameter values of state space reconstruction for EEG signals to be used in topological data analysis. We found that 3–5 embedding dimension and 1 sample time delay provided the lowest FNN rate for synthetic EEG signals. Our FNN test results supported the fact that any increase in embedding dimension or time delay may cause redundancy in the reconstructed state space. Since any misrepresented data will lead to redundant or erroneous topological features on the further steps of the topological analysis, it is crucial to have true representation of the embedded data in multidimensional state space [28].

The FNN test was examined for noise and artefact included synthetic EEG signals to find out its robustness. Apart from the muscle artefacts, robustness against the different types of artefact was high, *i.e.*, the test results remained unchanged. More importantly, wider time window usage enhanced the robustness of the state space. This was expected because as we embedded more signal samples into the state space, ratio of the artefact related points became smaller thereby distortions diminished. FNN test on real EEG signals supported the importance of having a wide time window for overcoming the distortions of artefacts, because while test results of 150- and 250-point time windows were below 1% FNN rate, 50-point time window failed to reach this threshold.

For clarity and convenience, we reported the lengths of time windows used in terms of the number of points rather than seconds. However, we chose these values based on their time equivalents considering the sampling rate, because the number of points time window has solely depends on sampling rate of the signal. Chosen time windows of 50, 150

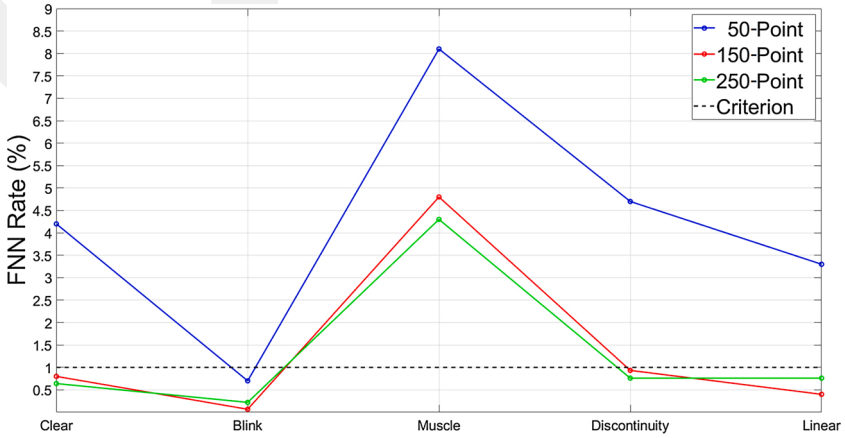

Fig. 9. FNN rates of state spaces that were reconstructed with the best embedding dimension and time delay pairs on *real EEG signals* for different time window lengths.

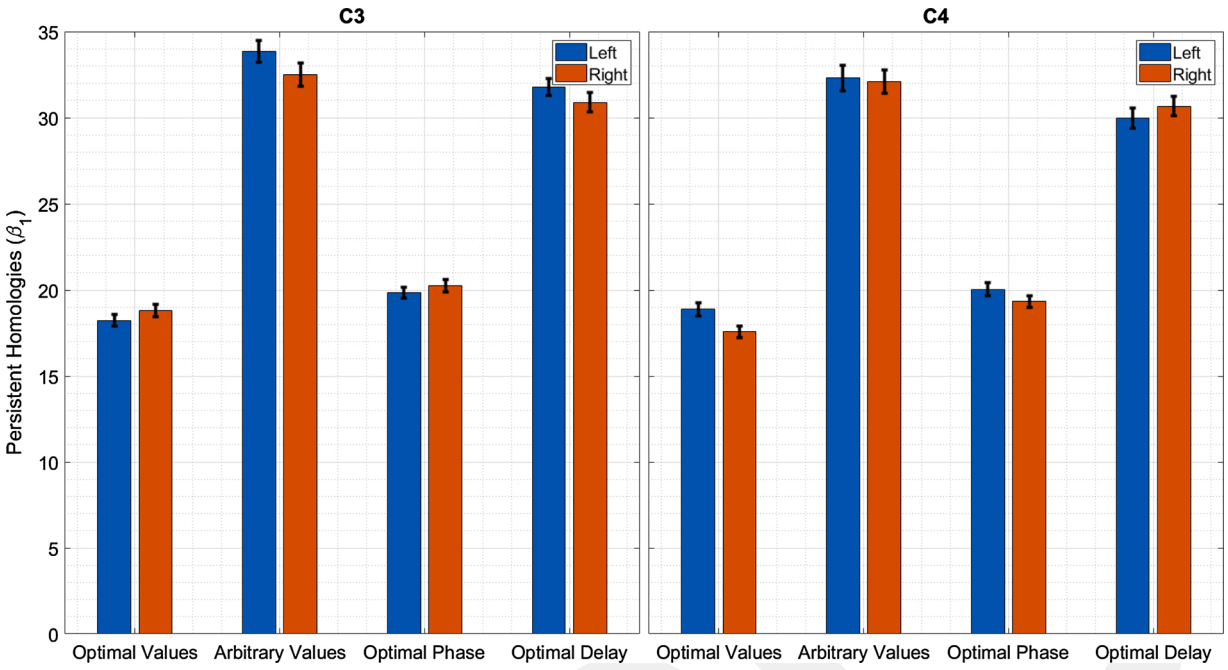

Fig. 10. Extracted β_1 numbers of signals that were taken from C3 and C4 channels showed on bar graph. The cases of left and right hand imaginations were compared. State spaces that were constructed using the optimal parameters showed significantly different β_1 numbers.

Table 1
Paired *t*-test results of extracted persistent homologies that were constructed with different combination of time delay and embedding dimension values. Numbers indicate the significance levels, p.

	Optimal ΔT& Optimal M	Arbitrary ΔT& Arbitrary M	Optimal ΔT& Arbitrary M	Arbitrary ΔT&Optimal M
C3	0.1021	0.2511	0.1500	0.2372
C4	**0.0027**	0.8763	0.0754	0.5005

and 250 points were equal to 0.2, 0.6 and 1 s respectively. Paired *t*-test results in Section 3.2 showed that the features extracted from 250-point time windows were significantly different during motor imagery period.

Finding the optimal embedding values (embedding dimension, time delay and time window) using FNN test was the first step towards extracting distinctive topological features from EEG signals. In fact, exploring the extracted features whether they were significantly different ($p < 0.05$) or not, showed that using the optimal parameters were crucial to have distinctive topological features. Moreover, comparisons depicted that how drastically the topological structure changed when the time delay or embedding dimension was arbitrarily chosen. Our findings demonstrated the parallelism with the studies aimed at

Table 2
Paired *t*-test results of the channels for resting and imagery periods and for different filtration values. Numbers indicate the significance levels, p.

	R/4	R/10	R/25	R/50	R/100
Resting	0.8986	0.2039	0.6481	0.8280	0.5061
Imagery	0.4719	0.0797	**0.0128**	**0.0094**	**0.0073**

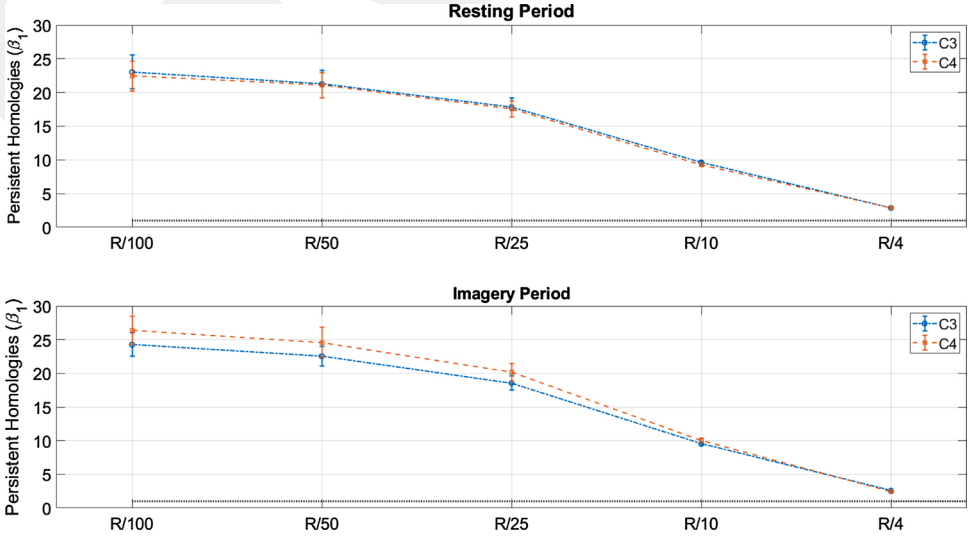

Fig. 11. Extracted β_1 numbers of resting (top) and imagery (bottom) periods were shown for signals coming from C3 and C4 channels, demonstrating the effect of filtration level.

Table 3
Paired *t*-test results of the channels for resting and imagery periods and for different time window lengths. Numbers indicate the significance levels, p.

	50 points	150 points	250 points
Resting	0.9379	0.2674	0.6481
Imagery	0.2688	0.5445	**0.0128**

determining the optimal embedding dimensions [30,31]. Krakovska et al. discussed roles of embedding dimension and time delay for the FNN test of Lorenz and Rössler system [31]. In another study [30], Kennel et al. shared the FNN test results of Lorenz model and synthetic periodic time series and they discussed the reliability of FNN test. Both studies concluded that FNN test method offers to find minimum embedding dimension from observed data.

Reconstruction of the state space with optimal parameters was not enough to extract statistically meaningful features. We should note that when the filtration level was chosen arbitrarily, the computed homologies tended to converge. As we mentioned in Section 1, when the radii of the disks around the points of the state space grow, they overlap and create connections through filtration. This means that the space coverage of each point increases correspondingly. However, when the radii of the disks around the points grow too much, they start to engulf feature related homologies, and the state space converges to a singular structure [2,32]. From this perspective, filtration level should be high as much as possible to eliminate non-persistent homologies (noise related perturbations), but at the same time it should be low enough to avoid convergence. Our findings showed that, filtration level providing 4% coverage for each point was the highest filtration level that was good enough to extract distinctive features related to the persistent homologies.

When it comes to the computational cost, our results demonstrated that the time window and filtration level were needed to be tuned to operate in real-time BCI systems. Our observation was that when the filtration level was high the computational cost of the analysis exploded, and the TDA approach could not respond quickly enough to perform real-time analysis. On the other hand, when our approach was utilized to extract persistent homologies computation time of the analysis went down to 1 s as we have shared in the Section 3.3.

Finally, we compared the extracted features of topological analysis that used optimal values found by above-mentioned tests and comparisons ($_{max}$ = R/25, $\Delta T = 1$, $M = 5$, Time Window Length = 250). It was depicted that the features extracted from at least one channel were significantly different for left- and right-hand imagination. These findings suggest that when the signals are embedded into state space using optimal parameters, and analyzed with optimal filtration level, extracted persistent homologies of the state spaces provide distinctive features for left- and right-hand movement imaginations to be used in motor intension wave based BCIs.

The most important benefit of our approach for future studies is that it provides mathematically sound test results. Thus, it removes ambiguities from the embedding dimension, delay time and time window parameters and enables reproducibility of future studies on topological analysis of EEG signals. Hence, step by step investigations showed that optimal parameters can be obtained by using FNN test in order to extract topological features from EEG signals in a fast and reliable manner.

Additionally, we found that the filtration level of the Rips stream analysis should be chosen carefully in order to avoid singularity of topological structures. For this purpose, filtration level should be determined based on the space coverage of each point in the state space so that reproducibility of the analyses on different EEG signals can be attained. Moreover, avoiding from singularity of topological structure by decreasing filtration level reduces computational cost of the analyses.

The major limitation of the study was the analysis capacity and the speed of the topological analysis toolbox (JPLEX). Our aim was to find the optimal parameters of the TDA, while keeping the analysis time short enough to operate in real-time BCI systems. We were unable to widen time window more than 250 points, because toolbox failed to complete computation of the homologies. In fact, Bauer et al. [33] showed that JPLEX toolbox showed the penultimate performance between 5 commonly used toolbox in terms of running times (in seconds) and memory usage. Nevertheless, JPLEX can be called from MATLAB and Simulink, and it can be embedded into custom MATLAB functions to run on real-time systems. Apart from that, if there were more channels in the dataset, it would be possible to collect more information to model general topological structure of the resting and imagery states.

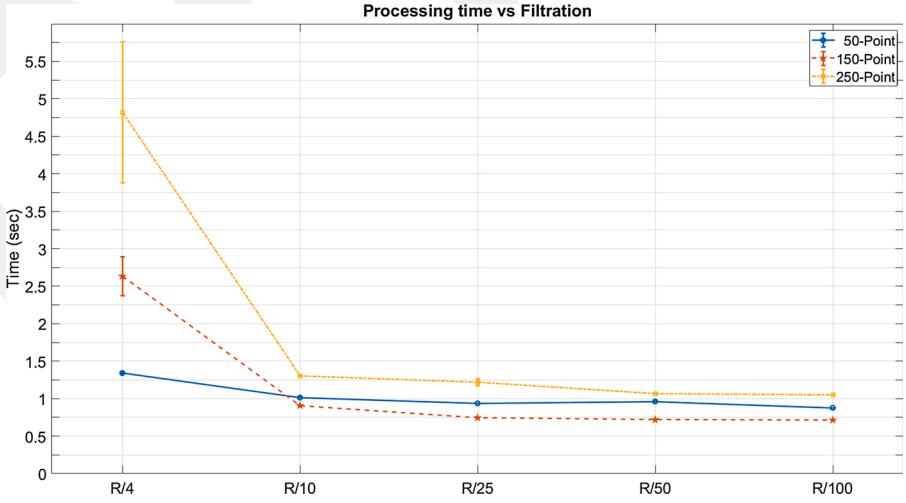

Fig. 12. Elapsed time to process state spaces that were constructed with different time window lengths are plotted for different filtration levels.

Table 4
Paired *t*-test results of persistent homologies of left- and right-hand imagery states using single channel signals. Numbers indicate the significance levels, p.

	B01T	B02T	B03T	B04T	B05T	B06T	B07T	B09T
C3	**0.0000**	**0.0000**	0.4421	0.4684	**0.0156**	**0.0060**	**0.0000**	0.4286
C4	0.3767	**0.0000**	**0.0119**	**0.0028**	**0.0413**	0.1003	**0.0000**	**0.0003**

Moreover, it would be possible to investigate the spatial embedding for reconstruction of the state space if there were more channels. However, due to the number of channels in this dataset, spatial embedding method was excluded from this study.

This study demonstrated the delicate issues of the topological analysis for extracting persistent homologies from EEG signals. For the first time in the literature, we showed how different types of artefacts affected the optimization tests and analysis results. In the future, it will be possible to focus on reconstructing the state space with larger time windows. If more signal samples were embedded into the space, it may better highlight the changes between resting and imagery states. Furthermore, if the computational efficiency of the analysis toolbox were improved, multiple time windows might be used to extract persistent homologies, and the connection between the changes in short and long time windows could be discovered in real-time.

CRediT authorship contribution statement

Fatih Altındiş: Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Bülent Yılmaz:** Conceptualization, Validation, Investigation, Writing - review & editing, Supervision, Project administration. **Sergey Borisenok:** Conceptualization, Methodology, Writing - review & editing. **Kutay İçöz:** Conceptualization, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This study was supported by Abdullah Gul University Scientific Research Projects Coordination Department. Project No: TOA-2015-31.

Special thanks to Prof. Ibrahim Erkutlu at Gaziantep University, Faculty of Medicine, Neurosurgery Department, for his contributions in the conceptualization of the study.

References

[1] G. Carlsson, Topology and data, Bull. Am. Math. Soc. 46 (2) (2009) 255–308.
[2] A. Zomorodian, Topological data analysis, Inverse Probl. 27 (12) (2011) 120201.
[3] L. Wasserman, Topological data analysis, Annu. Rev. Stat. Appl. 5 (March 1) (2018) 501–532.
[4] D.J. McFarland, L.A. Miner, T.M. Vaughan, J.R. Wolpaw, Mu and beta rhythm topographies during motor imagery and actual movements, Brain Topogr. 12 (3) (2000) 177–186.
[5] M. Saggar, O. Sporns, J. Gonzalez-Castillo, P.A. Bandettini, G. Carlsson, G. Glover, A.L. Reiss, Towards a new approach to reveal dynamical organization of the brain using topological data analysis, Nat. Commun. 9 (1) (2018). Article number: 1399.
[6] X. Wang, J. Meng, G. Tan, L. Zou, Research on the relation of EEG signal chaos characteristics with high-level intelligence activity of human brain, Nonlinear Biomed. Phys. 4 (2010).
[7] F. Takens, Detecting strange attractors in turbulence, in: D. Rand, L. Young (Eds.), Dynamical Systems and Turbulence, Springer, Berlin, Heidelberg, 1981, pp. 366–381.
[8] L.M. Seversky, S. Davis, M. Berger, On time-series topological data analysis: New data and opportunities, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (2016) 1014–1022.
[9] C.J. Stam, Nonlinear dynamical analysis of EEG and MEG: review of an emerging field, Clin. Neurophysiol. 116 (October 10) (2005) 2266–2301.
[10] W. Liebert, K. Pawelzik, H.G. Schuster, Optimal embeddings of chaotic attractors, Europhys. Lett. 14 (6) (1991), 521–126.
[11] G.M. Mindlin, R. Gilmore, Topological analysis and synthesis of chaotic time series, Phys. D Nonlinear Phenom. 58 (1–4) (1992) 229–242.
[12] M.T. Rosenstein, J.J. Collins, C.J. De Luca, Reconstruction expansion as a geometry-based framework for choosing proper delay times, Phys. D Nonlinear Phenom. 73 (May 1–2) (1994) 82–98.
[13] W.S. Pritchard, D.W. Duke, Measuring chaos in the brain - a tutorial review of EEG dimension estimation, Brain Cognit. 27 (3) (1995) 353–397.
[14] C.J. Cellucci, A.M. Albano, P.E. Rapp, Comparative study of embedding methods, Phys. Rev. E 67 (June 6) (2003) 13.
[15] H. Edelsbrunner, J. Harer, Persistence. Computational Topology: An Introduction, Duke University, Durham, NC: American Mathematical Society, 2010, pp. 149–174.
[16] D.R. Sheehy, Linear-size approximations to the vietoris-rips filtration, Discret. Comput. Geom. 49 (4) (2013) 778–796.
[17] H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification, Discret. Comput. Geom. 28 (4) (2002) 511–533.
[18] J.A. Perea, A. Deckard, S.B. Haase, J. Harer, SW1PerS: sliding windows and 1-persistence scoring; Discovering periodicity in gene expression time series data, BMC Bioinf. 16 (1) (2015) 257.
[19] Y. Wang, H. Ombao, M.K. Chung, Topological data analysis of single-trial electroencephalographic signals, Ann. Appl. Stat. 12 (September 3) (2018) 1506–1534.
[20] Y. Dabaghian, F. Mémoli, L. Frank, G. Carlsson, A topological paradigm for hippocampal spatial map formation using persistent homology, PLoS Comput. Biol. 8 (8) (2012).
[21] J.M. Kilner, K.J. Friston, Topological inference for EEG and MEG, Ann. Appl. Stat. 4 (3) (2010) 1272–1290.
[22] F. Lotte, Generating artificial EEG signals to reduce BCI calibration time, 5th International Brain-Computer Interface Workshop (2011) 176–179.
[23] G. Pfurtscheller, A. Berghold, Patterns of cortical activation during planning of voluntary movement, Electroencephalogr. Clin. Neurophysiol. 72 (3) (1989) 250–258.
[24] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlogl, BCI Competition 2008 Graz DataSet B, 2018.
[25] A. Delorme, T. Sejnowski, S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis, Neuroimage 34 (February 4) (2007) 1443–1449.
[26] S. Emrani, T. Gentimis, H. Krim, Persistent homology of delay embeddings and its application to wheeze detection, IEEE Signal Process. Lett. 21 (4) (2014) 459–463.
[27] R. Hajiloo, H. Salarieh, A. Alasty, Chaos control in delayed phase space constructed by the takens embedding theory, Commun. Nonlinear Sci. Numer. Simul. 54 (2018) 453–465.
[28] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A 45 (6) (1992) 3403–3411.
[29] H. Adams, A. Tausz, M. Vejdemo-Johansson, javaPlex: a research software package for persistent (co)homology, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8592 (2014) 129–136. LNCS.
[30] M.B. Kennel, H.D.I. Abarbanel, False neighbors and false strands: a reliable minimum embedding dimension algorithm, Phys. Rev. E 66 (August 2) (2002) 26209.
[31] A. Krakovská, K. Mezeiová, H. Budáčová, Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction, J. Complex Syst. 2015 (2015) 932750.
[32] P. Bubenik, Statistical topological data analysis using persistence landscapes, J. Mach. Learn. Res. 16 (2015) 77–102.
[33] U. Bauer, J. Reininghaus, H. Wagner, Phat – persistent homology algorithms toolbox, J. Symb. Comput. 78 (January) (2017) 76–90.