Biomedical Signal Processing and Control 59 (2020) 101899

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density

Mohamad Amin Bakhshali, Morteza Khademi, Abbas Ebrahimi-Moghadam*, Sahar Moghimi

Department of Electrical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

Article history:
Received 3 June 2019
Received in revised form
17 December 2019
Accepted 15 February 2020
Available online 27 February 2020

Keywords:
Imagined speech
Brain–computer interface
Riemannian distance
Correntropy spectral density
EEG signal

ABSTRACT

Several current brain–computer interface (BCI) systems are based on imagined speech. This means that these systems are controlled only by thinking about a speech without verbally expressing it. Imagined speech recognition using electroencephalogram (EEG) signals is much more convenient than other methods such as electrocorticogram (ECoG), due to its easy, non-invasive recording. So, we proposed an approach for EEG classification of imagined speech with high accuracy and efficiency. In this work, correntropy spectral density (CSD) matrices are evaluated for EEG signals obtained from different channels, and the distances between these matrices are considered as measures for imagined speech recognition. Riemannian distance benefits from simplicity and accuracy and it has achieved high scores in BCI competitions. Also, in this work, channel selection and frequency band detection during imagined speech is evaluated with statistical methods. The "Kara One" database is used in this research that includes EEG signals of eight subjects during imagined speech of four English words. We evaluated the proposed approach in comparison with the results of other imagined speech classification methods. Average classification accuracy of the proposed method is 90.25% during imagined speech for all subjects in KARA One database. The simulation results of this paper show the efficiency and accuracy of Riemannian distance of CSD and the superiority of the proposed method over other methods for imagined speech classification.

1. Introduction

An imagined mental activation generates neural activation patterns comparable with the actual perception. For example, thinking of moving a limb activates the motor cortex, also, internal object visualization activates the visual cortex, with similar effects observed for each sensory modality [1,2]. Auditory imagery is defined as the mental representation of sound perception in the absence of external auditory stimulation [3]. However, neural substrate of imagined speech is known less. Imagined speech (inner speech, silent speech, covert speech, or auditory verbal imagery) is an activity in which the brain sends a signal for expressing the words without any auditory stimulation, and tongue, lips, or hands expression [4]. Therefore, it can be said that imagined speech is the preliminary stage of speech production. This concept is most associated with signal processing and detection within electroencephalograph (EEG) data as well as data obtained using alternative non-invasive, brain–computer interface (BCI) devices. Also, imagined speech is applicable for synthetic telepathy and military applications [5]. In this work, BCI applications were considered.

BCI provides a direct communication pathway between the human brain and an external device. BCI receives collected signals from the brain and converts them into one or several activities. This area of advanced research has been widely developed in recent years. The results of these developments have not only provided a very important channel for the disabled and cerebrospinal injured patients to communicate with the surrounding environment, but they have also helped improve the quality of life by enabling them to exert control over their surrounding living facilities. BCI research studies are increasingly growing all around the world considering the reduced cost of electronic equipment and the technological advancements in the field of brain signal recording [6].

Several signal recording methods are used in BCI systems. They include: EEG, Magnetoencephalography (MEG), Electrocorticography (ECoG), functional Magnetic Resonance Imaging (fMRI), and Near-Infrared Spectroscopy (NIRS) [7–11]. EEG is a non-invasive technology and the most cost-effective one among these methods.

* Corresponding author.
E-mail address: a.ebrahimi@um.ac.ir (A. Ebrahimi-Moghadam).

https://doi.org/10.1016/j.bspc.2020.101899
1746-8094/

Moreover, the other advantage of EEG is its better temporal resolution compared with the other methods [12]. EEG-based BCI systems have been extensively studies and applied in recent years [13].

Feasibility studies investigated imagined speech in order to using in EEG-based BCI systems and showed that imagined speech can be inferred using syllables with high discriminative pronunciation [14,15]. Therefore, BCI-based equipment can be controlled by processing brain signals and inferring the imagined speech [16]. Extensive studies have been conducted so far in order to develop BCI systems using imagined speech in addition to motor imagery [17–20]. In order to investigate the feasibility of using EEG signals for imagined speech recognition, research study has been conducted and promising results were reported on imagined speech classification [21,22]. In addition, a similar research study was conducted where feasibility of using EEG signals for imagined speech recognition and increasing the efficiency of such use were investigated [23].

Afterward, EEG classification for imagined speech was conducted by several methods [24–27]. The common method is use of a band-pass frequency filter for preprocessing, a spatial filter for feature extraction, and support vector machine (SVM) for classification [24]. Common spatial pattern (CSP) filters are used for this purpose. Improved CSP filters have been developed in various forms [24,25]. Overall, EEG classification methods can be divided into four main categories: adaptive classifiers, matrix and tensor classifiers, transfer learning methods, and deep learning [27]. Among these approaches, matrix and tensor classifiers were very promising to enhance BCI reliability; specifically, classifiers based on Riemannian geometry were the current state-of-the-art for many BCI designs [27].

In this study, we considered matrix classifiers for imagined speech classification. Considering that the spectral density of a signal is another important feature that can be used in EEG signal classification, since signals that result from dynamic systems such as the brain, contain various different frequency components [28,29]. The power spectral density (PSD) matrix includes information about the correlation of different signals from different channels, and classification of EEG signals based on this feature is very useful for cognitive applications [30].

In addition to PSD, correntropy spectral density (CSD) was used as an appropriate nonlinear operator for signals with non-Gaussian noise [31]. The PSD is based on Fourier transform of the correlation function, but CSD is based on Fourier transform of correntropy function. EEG signals can be contaminated to Gaussian and non-Gaussian noise during the recording process [32]. One of the advantages of using correntropy measures in the classification of EEG signals is robustness against non-Gaussian noise like impulsive noise due to the use of kernel. Since PSD and CSD matrices are positive-definite, they have a constrained representation and belong to a manifold. Under such conditions, calculation of distance between these matrices by using Euclidean distance implies some drawbacks (like swelling effect) as discussed in [33]. So, it must be measured on the manifold's surface [33].

Thus, Riemannian geometry and Riemannian distance are taken into consideration and are used to determine the distance between matrices. Among the advantages of Riemannian methods is the establishment of an equivalence between sensor space (spatial space) and source space (samples' space) [34]. If the dimension of sensor space is smaller than the dimension of source space, then a projection will be found in a source sub-space that will improve the separation of the classes by using Riemannian geometry. Unlike this condition, if the dimension of source space is small, then the number of noisy components will be small and the achieved classification improvement will be small as well, due to the robustness of Riemannian distance against the noisy components. One of the main advantages of Riemannian methods over the other methods is that they are highly robust against outlier data and noise [34]. Therefore, Riemannian methods have obtained very high scores in many of the competitions on BCI data modelling [34]. It should be noted that there is not a unique Riemannian geometry and it depend on the choice of inner product in the tangent space. Until now, EEG signal classification frameworks based on Riemannian geometry were introduced for BCI and non-BCI applications [35]. EEG signal classification approaches for BCI systems based on Riemannian geometry were proposed for motor imagery in [36–38]. Also, reaction time estimation for BCI application was conducted based on Riemannian geometry features [39].

The main objective of the present study is to investigate imagined speech for BCI applications. For this purpose, EEG signals classification has been conducted based on CSD matrices. Based on the properties of these matrices, Riemannian geometry is used as the framework of feature extraction. In this research study, a novel CSD-based Riemannian distance is presented, and it is used as a feature for classification of EEG signals during imagined speech. The second section introduces basic concepts including correntropy and CSD. In the third section, CSD-based Riemannian distance and the proposed method based on it are expressed. In the fourth section, the database is introduced, and the experimental results are presented and discussed. Conclusions are presented in the fifth section.

2. Basic concepts

2.1. Correntropy spectral density

For a stationary and discrete signal $x(n)$ $n = 1, 2, \ldots, L$ the correntropy function $V(m)$ is defined as follows:

$$V(m) \triangleq E[K(x(n), x(n-m))] \tag{1}$$

where L is the number of signal samples, and m is the time delay [35]. Moreover, $E[\cdot]$ is the expected value operator, and $K(\cdot,\cdot)$ is symmetric, positive-definite kernel. Also, by using Eq. (1), the centered correntropy function is defined as follows [40]:

$$\begin{aligned} V_c(m) \triangleq\ & E[K(x(n), x(n-m))] \\ & -E_{x(n)}E_{x(n-m)}[K(x(n), x(n-m))] \end{aligned} \tag{2}$$

where $E_{x(n)}$ and $E_{x(n-m)}$ denote the marginal expected values of the kernel $K(\cdot,\cdot)$ regarding to $x(n)$, and $x(n-m)$, respectively. Gaussian kernel function is used in this study due to its wide range of applications in previous research studies. It is shown in (3) as follows:

$$K(x(n), x(n-m)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x(n)-x(n-m))^2}{2\sigma^2}\right] \tag{3}$$

Furthermore, σ (bandwidth) is determined using the following equation (Silverman's density estimation rule of thumb) in this study [31]:

$$\sigma = 0.9AL^{-\frac{1}{5}} \tag{4}$$

where A is a small value (obtained from the standard deviation and number of samples). Based on Mercer's theory, any symmetric, positive-definite kernel can be written as the inner product of two vectors in feature space, such as the reproductive kernel Hilbert space (RKHS) [41]. As a result, Eqs. (1) and (2) can be rewritten as follows [41]:

$$\begin{aligned} V(m) =\ & E\left[\left\langle \Phi(x(n)), \Phi(x(n-m))\right\rangle\right] \\ V_c(m) =\ & E\left[\left\langle \Phi(x(n)) - E[\Phi(x(n)], \Phi(x(n-m)) \right.\right. \\ & \left.\left. -E[\Phi(x(n-m))]\right\rangle\right] \end{aligned} \tag{5}$$

In Eq. (5), $\langle \cdot, \cdot \rangle$ is the inner product depends on the RKHS operator and $\Phi(x(n))$ is the mapping of signal $x(n)$ into feature space. The above equations, in the form of cross-correntropy equation, can also be generalized for two signals. Correntropy function, similar to an ordinary correlation function, is positive-definite. Thus it can be used in many signal processing techniques. CSD is based on Fourier transform of centered correntropy function and is defined as follows [42]:

$$\hat{P}(\omega) = \sum_{m=-(L-1)}^{L-1} V_c(m) e^{-j\omega m} \geq 0 \tag{6}$$

It was showed that the CSD was positive-definite matrix in [43]. Positive-definiteness of CSD will be used in Section 2.2. CSD is an appropriate tool for quantifying the frequency characteristics of signals. The main advantage of CSD over PSD is that CSD exhibits very favorable performance for signals with Gaussian and non-Gaussian statistics and with low SNR, while PSD is only suitable for Gaussian signals with high SNR [42]. Another advantage of CSD is the ease of spectral estimation method as very few design parameter values need to be set [42]. Due to the robustness of the correntropy function against outliers, the spectral resolution of CSD is much better than PSD [42]. Moreover, the correntropy function extract information of higher-order statistical moments present in EEG signals, thus potentially increases the efficiency of CSD [42].

2.2. Riemannian distance and its importance

Imagined speech recognition requires multi-channel (M channel) signals. Therefore, matrices (such as covariance, PSD, and CSD matrices) with $M \times M$ dimensions are generated in order to analyze and characterize similarity and interdependencies of signals in frequency or spatial domain. It has been proved that all of these matrices are symmetric, positive-definite [42,43].

These matrices form a subset of $M \times M$ complex matrices where locate on a manifold Γ. The CSD matrices of an epoch of multi-channel EEG signal were calculated in a frequency range. So, CSD matrices form a curve on Γ at different frequencies. A distance measure between the two curves was established at the same frequency. Using Euclidean measurement of distance would not be appropriate, because of CSD matrices constrained representation (positive-definite), so they form a manifold in the signal space. Instead of using Euclidean distance, a measurement method should determine the distance along the surface of the manifold. So, we chose to measure distance between CSD matrices by using Riemannian geometry.

The assumed space in which all symmetric matrices are located on a manifold is called the tangent space. Moreover, if there is only one short and unique pathway between two points on a manifold, it is called geodesic distance. The Riemannian distance between two points on the Riemannian manifold is defined as the geodesic length [44]. The closed form of this length will be obtained for CSD matrices in the next section.

3. Materials and methods

Fig. 1 demonstrates the overall block diagram of the proposed method for imagined speech classification based on EEG signal. The raw EEG signal for F channels is represented as $\mathbf{x}_1 \ldots \mathbf{x}_F$. However, all of the F channels were not used for recognition of imagined speech. Out of the F channels only M channels, with the most significant CSD values during imagined speech, were selected by statistical analysis. This is explained in the following section.

3.1. Preprocessing

First, ocular and muscular artifacts were eliminated using Independent Component Analysis (ICA) using EEGLAB toolbox [45]. Signal samples greater than 80V were also removed from the signal. EEG signals were filtered in the 0.5–57 Hz frequency range using a 5th order band-pass Butterworth filter. Studies have shown that cognitive functions modulate neural oscillations at multiple frequencies simultaneously, and this modulation take place at frequency, spatial, and temporal dimensions [46–48]. For the multiple comparisons conducted in the EEG data analysis, a nonparametric statistical test was adopted from the Fieldtrip toolbox [49,50]. Non-parametric cluster-based permutation test is employed to determine the statistical significance of the extracted features and control the false positive rate (Type I error) caused by multiple comparisons [51]. The permutation test uses a test statistic that is based on clustering of adjacent spatial locations that exhibit a similar P feature (sum of CSD values in each frequency band) difference between imagined speech and baseline. The calculation of the test statistic involves the following steps [51]:

1. Compute the t-score for each P (sample) from all spatial locations. That $\overline{X}_b$ and $\overline{X}_I$ are sample means of baseline and imagined speech. s_b^2 and s_I^2 are sample standard deviation of baseline and imagined speech. N_b and N_I are sample size of baseline and imagined speech.

$$t = \frac{\left(\overline{X}_b - \overline{X}_I\right)}{\sqrt{\frac{s_b^2}{N_b} + \frac{s_I^2}{N_I}}} \tag{7}$$

2. Select all P features with absolute value of uncorrected t-score greater than a threshold (2.6). This step identified a set of "candidate positives", of which a high proportion is likely to be true.
3. Cluster the selected P in connected sets based on spatial adjacency. In this study, we define two spatial locations as neighbors if their distance is less than 3.5 cm [51]. Clustering is performed for P features with positive and negative t-scores separately.
4. The cluster-level statistics is defined as the sum of t-scores within a cluster.
5. After getting the cluster-level test statistics for all clusters, the significance of these clusters is obtained by calculating the p-values.

After cluster-based permutation test, the appropriate frequency band and channels during imagined speech were defined. M number of EEG channels were selected and regarded to these channels, feature extraction and classification were conducted.

3.2. Riemannian distance classification

The outputs of the M EEG channels after preprocessing are represented as $\mathbf{y}_1 \ldots \mathbf{y}_M$. In the next step, the signal of each channel was segmented into W epochs, each of which included L samples. Thus, at the kth epoch, and ith channel, the preprocessed signal is represented by $\mathbf{y}_i^{(k)}$ $i = 1, 2, \ldots, 64$ and $k = 1, 2, \ldots, W$, a $L \times 1$ vector. At a certain time sample $n = 1, 2, \ldots, L$, for M channels, an M-element column vector is obtained as $\mathbf{y}^{(k)}(n) = \left[y_1^{(k)}(n), y_2^{(k)}(n), \ldots, y_M^{(k)}(n)\right]^T$. We generalize the correntropy of a two-channel signal into the correntropy of a multi-channel signal, in the form of the correntropy matrix. First, the multi-channel signal was mapped into the fea-

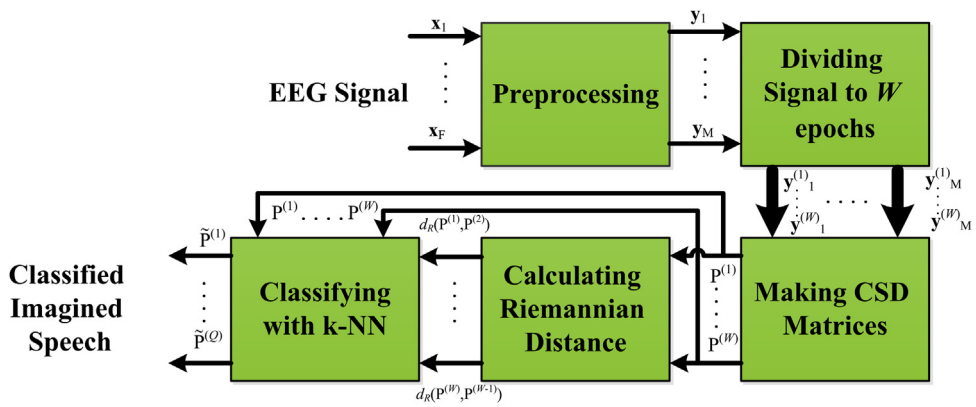

Fig. 1. Block diagram of the proposed method for imagined speech classification with EEG signals. (All vectors and matrices are defined in Section 3).

ture space by nonlinear mapping $\Phi(\cdot)$, then the M-element column vector $\Phi\left(\mathbf{y}^{(k)}(n)\right)$ was obtained as:

$$\Phi\left(\mathbf{y}^{(k)}(n)\right) = \left[\Phi\left(\mathbf{y}_1^{(k)}(n)\right), \Phi\left(\mathbf{y}_2^{(k)}(n)\right), \ldots, \Phi\left(\mathbf{y}_M^{(k)}(n)\right)\right]^T \tag{8}$$

According to Eq. (5), the centered correntropy matrix of the multi-channel signal at the kth epoch can be rewritten as:

$$\begin{aligned} V_c^{(k)}(m) = & E\left[\left\langle \Phi(\mathbf{y}^{(k)}(n)) - E\left[\Phi\mathbf{y}^{(k)}(n)\right], \right.\right. \\ & \left.\left. \Phi(\mathbf{y}^{(k)}(n-m)) - E\left[\Phi(\mathbf{y}^{(k)}(n-m))\right]\right\rangle\right] \\ = & E\left[\left\{\Phi(\mathbf{y}^{(k)}(n)) - E\left[\Phi(\mathbf{y}^{(k)}(n)\right]\right\}\right. \\ & \left.\left\{\Phi(\mathbf{y}^{(k)}(n-m)) - E\left[\Phi(\mathbf{y}^{(k)}(n-m))\right]\right\}^T\right] \end{aligned} \tag{9}$$

Considering Eq. (8) and the definition of $\mu^{(k)} = E\left[\Phi(\mathbf{y}^{(k)}(n)\right]$, the above equation can be expressed in a matrix form as:

$$V_c^{(k)}(m) = E\left[\begin{bmatrix} \Phi(\mathbf{y}_1^{(k)}(n)) - \mu_1^{(k)} \\ \Phi(\mathbf{y}_2^{(k)}(n)) - \mu_2^{(k)} \\ \vdots \\ \Phi(\mathbf{y}_M^{(k)}(n)) - \mu_M^{(k)} \end{bmatrix}_{M\times 1}^{T} \begin{bmatrix} \Phi(\mathbf{y}_1^{(k)}(n-m)) - \mu_1^{(k)} \\ \Phi(\mathbf{y}_2^{(k)}(n-m)) - \mu_2^{(k)} \\ \vdots \\ \Phi(\mathbf{y}_M^{(k)}(n-m)) - \mu_M^{(k)} \end{bmatrix}_{M\times 1}\right]_{M\times M} \tag{10}$$

The correntropy matrix from Eq. (10) can be expressed as follows:

$$V_c^{(k)}(m) = \begin{bmatrix} V_{c11}^{(k)}(m) & \cdots & V_{c1M}^{(k)}(m) \\ \vdots & \ddots & \vdots \\ V_{cM1}^{(k)}(m) & \cdots & V_{cMM}^{(k)}(m) \end{bmatrix}_{M\times M} \tag{11}$$

where $V_{cij}^{(k)}(m)$ is the centered correntropy between the two channels I and j at the kth epoch. In practice, centered correntropy is estimated as follows [41]:

$$\begin{aligned} \hat{V}_{cij}^{(k)}(m) = & \frac{1}{L-m+1}\sum_{n=m}^{L} K(y_i^{(k)}(n), y_j^{(k)}(n-m)) \\ & -\frac{1}{L^2}\sum_{n=1}^{L}\sum_{l=1}^{L} K(y_i^{(k)}(n), y_j^{(k)}(l)) \end{aligned} \tag{12}$$

To calculate $y_j^{(k)}(n-m)$, $y_j^{(k)}(n)$ is shifted by m samples (from $-(L-1)$ to $(L-1)$), and the neighbour matrices are used for substituting the samples so that their shifted form can be used for calculating Eq. (12). Next, for W epochs, CSD matrices $P^{(1)}(\omega)\ldots P^{(k)}(\omega)\ldots P^{(W)}(\omega)$ are obtained using Eq. (6) as:

$$P^{(k)}(\omega) = \sum_{m=-(L-1)}^{L-1} \hat{V}_c^{(k)}(m) e^{-j\omega m} \geq 0 \tag{13}$$

The Nuttal-Strand algorithm was employed to estimate the CSD matrix [52]. In practice, this algorithm is very efficient in spectral density estimation of multivariate signals [53]. These matrices were considered as inputs to the classification block (Fig. 1).PSD matrices of EEG signal epochs were formed like CSD matrices that was described before, but nonlinear mapping into the feature space was excluded in PSD matrices. The CSD matrices are $M \times M$ positive-definite and Hermitian, so structure of these matrices in this work is suitable for the Riemannian framework. The elements of the matri-

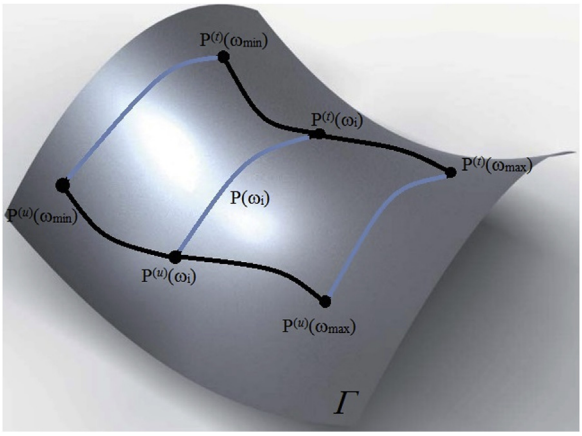

Fig. 2. The distance between CSD matrices of two signals at specified frequency ($P^{(t)}(\omega)$ and $P^{(u)}(\omega)$) on Γ manifold.

ces were calculated in the frequency range of $[\omega_{\min}, \omega_{\max}]$. Thus, CSD matrices of a signal included a set of points at different frequencies that were located on manifold Γ as a curve. The set of all CSD matrices was marked by H_H. Γ was assumed that a manifold in the real linear vector space H_H. Due to considering real analysis of the geometry, this property is important [30]. For example, the tth and uth CSD matrices specify two separate curves on the manifold Γ as $P^{(t)}(\omega)$ and $P^{(u)}(\omega)$ in the interval $\omega \in [\omega_{\min}, \omega_{\max}]$. In order to find the distance between two curves, first, we considered a distance between two points on the curve ($P^{(t)}(\omega_i)$ and $P^{(u)}(\omega_i)$) at the same frequency ω_i. The distance between CSD matrices of two signals can be seen in Fig. 2. For the sake of simplicity of expression, $P^{(t)}(\omega_i)$ and $P^{(u)}(\omega_i)$ are represented as $P^{(t)}$ and $P^{(u)}$, respectively.

We may connect two points $P^{(t)}$ and $P^{(u)}$ with smooth path (P) on the manifold Γ. $l(P)$ is the length of the path between the two points. The curve on the Riemannian manifold that has the shortest distance is geodesic. Thus, distance is written as:

$$d_R(P^{(t)}, P^{(u)}) \triangleq \min_P \left\{ l(P) \right\} \tag{14}$$

Riemannian distance can be calculated with different methods [34]. In this work, a closed-form expression of Riemannian distance (d_R) on manifold Γ was defined using the principles of Riemannian geometry. First, let us assume that each matrix P can be written as $P = \tilde{P}\tilde{P}^H$, where $\tilde{P} = \sqrt{P}U$, and matrix U is a unitary matrix. It has been proven that the distance on manifold Γ between two points $P^{(t)}$ and $P^{(u)}$ is equal to the Euclidean distance between two points $\tilde{P}^{(t)}$ and $\tilde{P}^{(u)}$ [30]. Therefore, distance (d_R) can be written as:

$$\begin{aligned} d_R(P^{(t)}, P^{(u)}) &= \min_{\tilde{P}} l(\tilde{P}) = \min_{\tilde{P}} \left\| \tilde{P}^{(t)} - \tilde{P}^{(u)} \right\| \\ &= \min_{\tilde{P}} \left[tr \left\{ \left(\tilde{P}^{(t)} - \tilde{P}^{(u)} \right)^H \left(\tilde{P}^{(t)} - \tilde{P}^{(u)} \right) \right\} \right]^{\frac{1}{2}} \end{aligned} \tag{15}$$

where tr is the matrix trace. We can expand the trace term in the above equation using identities $\tilde{P}^{(t)} = (P^{(t)})^{1/2}U^{(t)}$, $\tilde{P}^{(u)} = (P^{(u)})^{1/2}U^{(u)}$. The trace of a Hermitian matrix is always real, since the elements on the main diagonal are real. Subsequently, we have

$$\begin{aligned} d_R^2(\tilde{P}^{(t)}, \tilde{P}^{(u)}) = \min_P \Big[& trP^{(t)} + trP^{(u)} \\ & -2\mathrm{Re} \left\{ tr \left(U^{(t)}(U^{(u)})^H (P^{(u)})^{1/2} (P^{(t)})^{1/2} \right) \right\} \Big] \end{aligned} \tag{16}$$

In order to minimize the left-hand side of Eq. (16), the second term on the right-hand side of the equation must be maximized. The singular value decomposition of matrix $(P^{(u)})^{1/2}(P^{(t)})^{1/2}$ is equal to $U^{(u)}\Sigma(U^{(t)})^H$, where Σ is the eigenvalues of the matrix [54]. Also, it has been proved that $\max_{\Sigma_1, \Sigma_2} \mathrm{Re}\left\{ tr(\Sigma_1 \Sigma_2) \right\} = tr(\Sigma_1 \Sigma_2)^{1/2}$, if Σ_1 and Σ_2 are diagonal matrices with eigenvalues on the main diagonal [54].

$$\begin{aligned} & \max_P \mathrm{Re} \left\{ tr \left(U^{(t)}(U^{(u)})^H (P^{(u)})^{1/2} (P^{(t)})^{1/2} \right) \right\} \\ & = \max_p \mathrm{Re} \left\{ tr \left(U^{(t)}(U^{(u)})^H U^{(u)} \Sigma (U^{(t)})^H \right) \right\} \\ & = \max \mathrm{Re} \left\{ tr \left(U^{(t)} \Sigma (U^{(t)})^H \right) \right\} \\ & = tr \left[\left((P^{(t)})^{1/2} (P^{(u)})^{1/2} (P^{(u)})^{1/2} (P^{(t)})^{1/2} \right)^{\frac{1}{2}} \right] \\ & = tr \left[\left((P^{(t)})^{1/2} P^{(u)} (P^{(t)})^{1/2} \right)^{\frac{1}{2}} \right] \end{aligned} \tag{17}$$

As a result, a Riemannian distance for two CSD matrices is obtained as follows:

$$d_R(P^{(t)}, P^{(u)}) = \sqrt{trP^{(t)} + trP^{(u)} - 2tr \left[\left((P^{(t)})^{1/2} P^{(u)} (P^{(t)})^{1/2} \right)^{1/2} \right]} \tag{18}$$

which can be simplified as below:

$$d_R(P^{(t)}, P^{(u)}) = \sqrt{trP^{(t)} + trP^{(u)} - 2tr \left[(P^{(t)}P^{(u)})^{1/2} \right]} \tag{19}$$

Riemannian distance (Eq. (19)), has all the characteristics of a metric, i.e.

$$\begin{cases} d_R(P^{(t)}, P^{(u)}) \geqslant 0; \quad d_R(P^{(t)}, P^{(u)}) = 0 \quad \text{if} \quad P^{(t)} = P^{(u)} \\ d_R(P^{(t)}, P^{(u)}) = d_R(P^{(u)}, P^{(t)}) \\ d_R(P^{(t)}, P^{(u)}) \leqslant d_R(P^{(t)}, P^{(q)}) + d_R(P^{(q)}, P^{(u)}) \end{cases} \tag{20}$$

We employed the Riemannian distance in order to compare CSD matrices. We used k-NN classification algorithm as was employed in [55]. According to the classification algorithm, from W CSD matrices, (W-Q) matrices were labeled and the remaining Q matrices were unlabeled. Labeling the matrices was performed based on the "*leave-one-out*" validation algorithm [56].

As shown in Fig. 1, CSD matrices were considered as inputs to the algorithm and their Riemannian distances were considered as features. In comparison with other classification methods [21,24], the k-NN algorithm is simpler, and it is more appropriate for the features extracted in this work, due to complexity of proposed method. The more samples that there were, the higher was the efficiency of the algorithm. This algorithm was supervised and some of CSD matrices (80% of all matrices) were considered as training data while the other CSD matrices are selected as testing data. Accordingly, each class was classified as matrices $\tilde{P}^{(1)} \ldots \tilde{P}^{(Q)}$, which correspond to signals of epoch 1 to Q (Q from W classified epochs).

4. Experimental results

4.1. Description of data

The database used in this research was the "Kara One" database, which had been made with the aim of investigating imagined speech in the form of syllables and words, and involved combining three signals including EEG, audio, and face tracking during imagined speech [57]. We employed only the EEG signals modality. This database includes the EEG signals of four women and eight men (27.4 ± 5 years). In order to acquire data for this database, each subject sat on a chair, and a monitor was put in front of him/her. Then, an appropriately-sized EEG cap, which included 64 channels, was put on the subject's heads. The electrodes were placed

in accordance with International 10–20 system [58]. To control eye movement artifacts, four electrodes were placed on each side of the eyes. The EEG device sampled brain signals at a rate of 1000 Hertz. The subject made an imagined speech of either of the four words (/*gnaw*/, /*knew*/, /*pot*/, /*pat*/) and seven phonemic/syllabic prompts (/*iy*/, /*uw*/, /*piy*/, /*tiy*/, /*diy*/, /*m*/, /*n*/) displayed on the screen. Then, subject had to vocalize an overt speech. Each trial consisted of four stages: (1) ***rest (baseline) stage***: five seconds of resting was considered in order for the subject to relax and clear his/her mind of any thought; (2) ***stimulation stage***: a text appeared on the monitor for two seconds; (3) ***overt speech stage***: the displayed word or syllable was vocalized loudly; and (4) ***imagined speech stage***: text was imagined for five seconds. For each subject, two sessions of data recording had been performed, and each syllable or word was repeated 12 times, so that there existed a total number of 132 trials. After every 40 trials, a break was given to the subject to rest. Data from four subjects were excluded because of problem in acquisition of data.

4.2. Frequency analysis and channel selection results

From Fig. 3, we can observe several spatial clusters associate with imagined speech and baseline. Normalization was made trial by trial with respect to baseline (pre-stimulus) and then averaged over trials for all subjects. Normalization was calculated by dividing of CSD value to mean CSD of baseline. At β frequency band, two clusters passed permutation test, cluster one including: F7, F5, F3, F1, FT7, FC5, FC3, FC1, T7, C5, C3, C1, TP7, CP5. Average t-score and corrected p-value were −2.975 and 0.027 respectively. Also cluster two including: FT8, FC6 with average t-score and corrected p-value were −3.174 and 0.034. At δ frequency band, only one cluster passed permutation test, including: T7, C5, FC5. Average t-score and corrected p-value were −3.457, 0.012 respectively. At the other frequency bands, no cluster passed permutation test. So, β frequency band was selected as frequency band of imagined speech. δ frequency band was not considered, because of very low frequency values and short time windows of imagined speech. Channel of imagined speech were selected based on cluster test permutation: T7, C5, C3, C1, CP5, FC1, FC3, FC5, FC6, FT7, F1, F3, F5, F7, FT8, TP7. These channels have neural activity in language production and cover the brain region (left inferior frontal lobe (Broca's area) and left superior temporal lobe (Wernicke's area)) that are responsible for language processing [59].

Fig. 4 shows the frequency representation for the selected channels from all trials of all subjects in the database. Fig. 4 demonstrates normalized CSD for imagined and baseline at different frequency bands. Normalization was made trial by trial with respect to baseline (pre-stimulus) and then averaged over trials. Normalization was calculated by dividing of CSD value to mean CSD of baseline. Also 1/f normalization (spectral flattening) was done. Imagined speech had significant CSD values mostly in the frequency range of 13–30 Hz (β frequency band).

4.3. Classification results

For evaluating the efficiency of Riemannian distance, Euclidean distance and the biggest component of CSP were compared. In Fig. 5(a), (b), Riemannian and Euclidean distances of each trials from the other CSD for two word with high pronunciation distinction (/*pat*/-/*knew*/) and two word with high pronunciation similarity (/*pat*/-/*pot*/) were demonstrated. The solid line represent the decision border whereupon distances are equal. The trials are distributed around the solid line regardless of their labels. Distribution of trials in local and separable area for Riemannian distance is considerable. But, wide and inseparable distribution of trials around the solid line for Euclidean distance is disadvantage. This fact implies that the relevant information about labels is not accessible using the Euclidean distance and the associated mean. In all feature extraction methods, some trials are misclassified, but only a small percentage of trials are misclassified in using by Riemannian distance.

In order to evaluate efficiency of the proposed method from classification aspect, the obtained results were compared with the results of other imagined speech classification methods. For this purpose, the proposed method was compared with the LDA method (2011) [60], the k-NN method with Frobenius distance (2014) [61], the SVM method along with CSP (2009) [25], as well as the PSD-based Riemannian distance method (2013) [30].

The results reported in Table 1 indicate that the average classification accuracy in word pairs with high pronunciation similarity, such as (/*pat*/-/*pot*/) and (/*gnaw*/-/*knew*/), was low, while on the contrary it was very high in word pairs with high pronunciation distinction, such as (/*pat*/-/*gnaw*/), (/*pat*/-/*knew*/), (/*pot*/-/*gnaw*/), and (/*pot*/-/*knew*/). This form of comparison and analysis is very common in experiments in cognitive sciences as mentioned in [21,24].

Classification of six pairs of imagined speech for 4 words (/*pat*/, /*pot*/, /*knew*/, /*gnaw*/) was performed with different methods for all subjects from "Kara One" dataset over 48 times repetition. The average accuracy and standard deviation of classification is reported in Table 1. According to this table, LDA classification method had the worst average classification accuracy, while CSP along with SVM classification, k-NN classification with Frobenius distance, Riemannian classification using PSD, and the proposed method had higher average accuracy for imagined speech classification. The experimental results indicated the superiority of the proposed method, which used Riemannian distance and the CSD matrix, over the other methods that either used Euclidean distance or did not use CSD.

SVM classification along with CSP feature extraction is one of the common methods in EEG signal classification for mental tasks [23]. However, average classification accuracies (standard deviation) for imagined speech pairs (/*pat*/-/*pot*/), (/*pat*/-/*gnaw*/), (/*pat*/-/*knew*/), (/*pot*/-/*gnaw*/), (/*pot*/-/*knew*/), and (/*gnaw*/-/*knew*/) were equal to 66.22% (7.9), 69.38% (5.9), 69.7% (5.5), 70.7% (5.8) 69.9% (6.1), and 67.5% (7.4), respectively in this work. The relatively lower classification accuracy of the aforementioned approach in comparison with the proposed method, k-NN, and Riemannian classification may be due to lack of consideration of time and frequency features of the data. The use of CSD in the proposed method improved the classification accuracy. A comparison of the results of Euclidean geometry-based methods with those of Riemannian methods indicated that the use of Riemannian geometry leads to improved accuracy and efficiency. As for the PSD-based Riemannian classification method, the average classification accuracy with standard deviation for imagined speech pairs (/*pat*/-/*pot*/), (/*pat*/-/*gnaw*/), (/*pat*/-/*knew*/), (/*pot*/-/*gnaw*/), (/*pot*/-/*knew*/), and (/*gnaw*/-/*knew*/) was equal to 76.42% (5.4), 78.47% (3.9), 78.50% (4.2), 78.50% (4.2), 78.2% (4.5), and 76.4% (5.4), respectively. This accuracy was lower than the proposed method with average accuracy of 89.07% (2.8), 90% (2.8), 90.01% (2.41), 90.97% (2.8), 91.9% (3), and 89.6% (3.5), respectively. Use of correntropy for finding the spectral density, with the features that were introduced earlier, resulted in higher accuracy compared to the PSD-based method. Also, the standard deviation in Riemannian classification methods was less than that of other methods.

Furthermore, we evaluated the proposed method for EEG signal classification during imagined speech in KARA One database. This comparison approach was applied based on [57]. Different phonemic and phonological classes were introduced with EEG modality of data. Five binary classification tasks were considered [57]: (1) vowel-only vs. consonant (C/V), (2) presence of nasal (Nasal), (3) presence of bilabial (Bilabial), (4) presence of high-front vowel

Table 1
Average classification accuracy (standard deviation) over 48 times repetition for 5 methods in all subjects in "Kara One" database (in percentage) (bold numbers are the biggest values).

Pair	*/pat/ – /pot/*					*/pat/ – /gnaw/*					*/pat/ – /knew/*				
Methods Subjects	LDA [60]	k-NN with Frobenius distance [61]	CSP + SVM [25]	Riemannian with PSD [30]	Proposed Method	LDA [60]	k-NN with Frobenius distance [61]	CSP + SVM [25]	Riemannian with PSD [30]	Proposed method	LDA [60]	k-NN with Frobenius distance [61]	CSP + SVM [25]	Riemannian with PSD [30]	Proposed method
Subject 1	62.2 (4.7)	65.7 (10.3)	68.2 (10.2)	75.7 (4.7)	**85.3 (3.0)**	63.7 (6.4)	68.7 (9.5)	71.2 (8.2)	77.8 (2.6)	**87.9 (2.0)**	60.4 (6.7)	69.8 (8.6)	71.3 (8.2)	76.9 (3.6)	**88.3 (1.2)**
Subject 2	69.3 (8.3)	72.3 (11.5)	71.5 (7.3)	76.3 (4.5)	**87.6 (1.2)**	68.3 (6.3)	74.2 (10.4)	73.7 (6.5)	79.3 (3.6)	**89.9 (1.9)**	69.3 (7.3)	74.5 (8.3)	74.8 (5.3)	78.3 (3.5)	**89.5 (3.0)**
Subject 3	72.1 (5.2)	71.3 (9.7)	72.3 (8.4)	78.2 (3.9)	**92.7 (2.7)**	72.5 (4.2)	76.2 (8.5)	74.4 (4.8)	79.3 (3.9)	**91.6 (1.7)**	72.3 (2.3)	76.9 (7.8)	73.9 (5.2)	78.2 (4.4)	**91.7 (1.5)**
Subject 4	68.7 (8.2)	74.2 (9.5)	71.1 (3.9)	80 (6.6)	**91.2 (2.2)**	67.9 (6.2)	76.4 (8.6)	74.9 (3.6)	82.1 (5.9)	**91.3 (2.8)**	67.3 (8.2)	76.2 (7.7)	75.1 (5.2)	81.2 (7.5)	**92.6 (2.4)**
Subject 5	57.5 (8.2)	63.3 (11.4)	67.2 (9.6)	74.4 (6.2)	**88.3 (3.9)**	57.9 (8)	71.2 (10.1)	68.9 (8.3)	78.4 (5.1)	**88.7 (3.7)**	59.4 (8.2)	71.3 (10.2)	70.2 (8.3)	77.6 (4.2)	**90.1 (3.7)**
Subject 6	68.7 (8.4)	75.1 (7.6)	63.2 (7.9)	76.4 (7.3)	**92.3 (5.8)**	69.1 (7.2)	77.3 (6.6)	69.5 (6.8)	78.1 (4.3)	**92.9 (6.9)**	67.9 (5.2)	76.3 (8.7)	70.4 (6.7)	78.9 (4.6)	**92.8 (4.5)**
Subject 7	65.9 (11.3)	72.1 (12.8)	65.7 (9.6)	78 (6.2)	**89.0 (2.6)**	67.1 (9.3)	74.8 (10.8)	68.9 (5.6)	80.3 (4.2)	**90.1 (2.0)**	67.4 (10.1)	75.8 (10.1)	68.9 (6.5)	80.3 (4.2)	**90.2 (1.2)**
Subject 8	47.3 (11.2)	57.2 (12.8)	50.6 (6.3)	72.4 (4.2)	**86.2 (1.7)**	49.4 (9.9)	61.2 (10.7)	53.6 (4.1)	72.5 (2.3)	**87.6 (1.8)**	48.2 (10.3)	61.4 (9)	53.7 (4.3)	76.6 (2.1)	**84.9 (1.8)**
Average	63.96 (7.1)	68.9 (10.7)	66.22 (7.9)	76.42 (5.4)	**89.07 (2.8)**	64.48 (7.1)	72.50 (9.4)	69.38 (5.9)	78.47 (3.9)	**90 (2.8)**	64.0 (7.2)	72.77 (8.8)	69.7 (5.5)	78.50 (4.2)	**90.01 (2.41)**

Pair	*/pot/ – /gnaw/*					*/pot/ – /knew/*					*/gnaw/ – /knew/*				
Methods Subjects	LDA [60]	k-NN with Frobenius distance [61]	CSP + SVM [25]	Riemannian with PSD [30]	Proposed method	LDA [60]	k-NN with Frobenius distance [61]	CSP + SVM [25]	Riemannian with PSD [30]	Proposed method	LDA [60]	k-NN with Frobenius distance [61]	CSP + SVM [25]	Riemannian with PSD [30]	Proposed method
Subject 1	62.4 (3.2)	68.9 (6.7)	71.2 (8.7)	78.6 (2.9)	**90.2 (2.3)**	63.2 (10.7)	71.7 (7.6)	71.9 (8.3)	78.6 (2.7)	**89.4 (2.9)**	60.2 (3.2)	66.3 (9.9)	69.8 (9.4)	74.8 (4.7)	**85.9 (1.8)**
Subject 2	67.3 (5.4)	74.5 (8.3)	74.9 (5.2)	79.8 (3.2)	**91.7 (1.9)**	69.7 (7.9)	74.5 (8.4)	73.9 (6)	78.2 (4)	**91.0 (1.8)**	65.3 (6.4)	71.5 (11.8)	71.4 (7.5)	75.3 (5.8)	**88.7 (2.9)**
Subject 3	74.5 (5.2)	75.3 (8.5)	76.9 (5.4)	80.2 (2.8)	**92.3 (1.5)**	73.2 (6.2)	75.9 (7.6)	74.9 (3.4)	78.4 (2.8)	**91.8 (2.2)**	70.2 (6.2)	74.2 (8.9)	72.6 (6.8)	78.2 (4.8)	**91.4 (1.7)**
Subject 4	67.2 (4.2)	77.4 (9.5)	77.9 (3.6)	81.2 (5.5)	**93.4 (2.3)**	68.1 (5.4)	76.8 (9)	74.3 (4.7)	81.3 (6.7)	**93.3 (3.0)**	64.3 (6.4)	74.4 (9)	71.9 (5.6)	79.9 (6.4)	**90.8 (3.1)**
Subject 5	59.8 (7.2)	72.4 (11.2)	72.8 (7.3)	77.5 (4.1)	**90.0 (4.8)**	59.4 (8.3)	71.3 (8.2)	70.2 (9.2)	76.6 (4.2)	**91.3 (3.5)**	54.3 (10.2)	70.1 (11.4)	67.3 (9.6)	75.6 (5)	**87.9 (5.8)**
Subject 6	69.9 (9.3)	76.4 (8.4)	70.2 (6.3)	77.9 (6)	**92.2 (5.9)**	70.1 (5.3)	77.3 (8.4)	70.4 (6.7)	78.5 (6.8)	**92.9 (5.8)**	64.3 (9.2)	74.3 (8.9)	68.4 (7.9)	76.7 (6.4)	**92.8 (5.7)**
Subject 7	67.9 (11.4)	75.7 (12.1)	68.8 (6.5)	80 (3.4)	**90.1 (2.0)**	67.8 (9.4)	75.8 (10.7)	70.3 (6.5)	80.1 (4.3)	**91.6 (3.2)**	63.2 (11.4)	72.9 (12.1)	67.8 (8.6)	78 (6.4)	**89.7 (2.7)**
Subject 8	49.4 (10.3)	61.4 (9.6)	53.2 (4)	72.8 (2.1)	**87.9 (1.9)**	49.6 (13)	62.2 (8.7)	53.7 (4.3)	74.3 (5.2)	**94.2 (1.9)**	45.1 (13.1)	60.1 (10)	51.4 (4.3)	73.4 (4.4)	**90.1 (4.7)**
Average	64.8 (7.0)	72.7 (9.2)	70.7 (5.8)	78.5 (3.7)	**90.97 (2.8)**	65.1 (8.2)	73.1 (8.5)	69.9 (6.1)	78.2 (4.5)	**91.9 (3.0)**	62.8 (8.2)	70.4 (10.2)	67.5 (7.4)	76.4 (5.4)	**89.6 (3.5)**

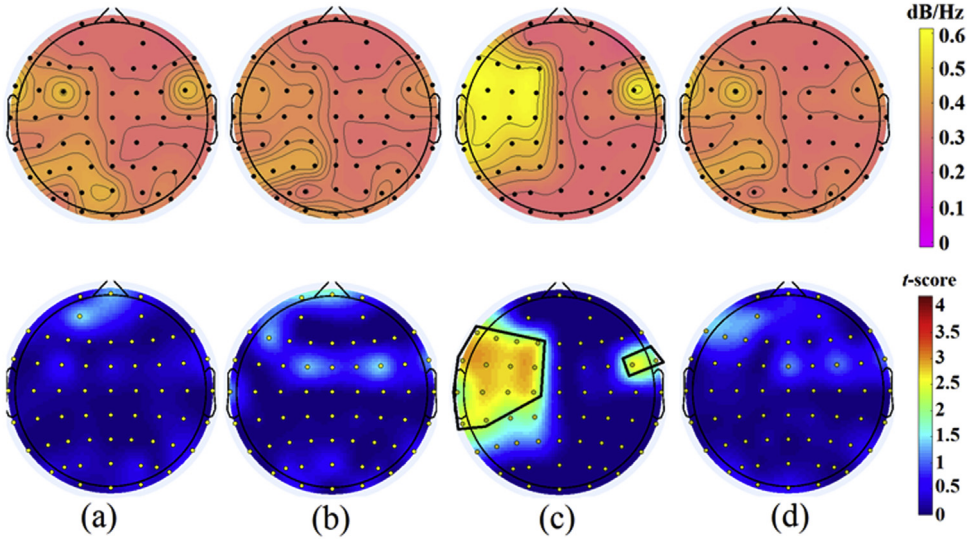

Fig. 3. (First row) Difference of normalized CSD values of imagined and baseline topography map at different frequency bands in average of all subjects in KARA One database. (Second row) t-score topography map at different frequency bands. (a) θ frequency band, (b) α frequency band, (c) β frequency band, (d) γ frequency band.

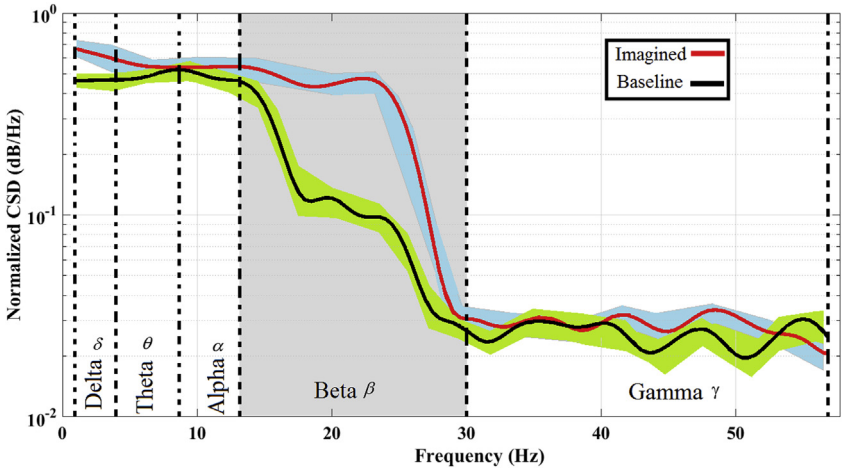

Fig. 4. Normalized CSD (dB/Hz) for imagined and baseline at different frequency bands in average of all subjects in KARA One database over selected channels.

Table 2
Average classification accuracy (in percentage) for 4 methods in all subjects in "Kara One" database (bold numbers are the biggest values.)

Tasks Methods	Bilabial	Nasal	C/V	/uw/	/iy/
Method of [57]	56.64	63.5	18.08	79.16	59.6
Method of [62]	53	47	25	74	53
Method of [63]	**75.55**	**73.45**	85.23	81.99	73.30
Proposed method	69.08	72.10	**86.52**	**83.98**	**75.27**

(/*iy*/), and presence of high-back vowel (/*uw*/) using EEG-only modality.

In Table 2, results were compared to the other methods in [57,62,63]. The classification accuracy of the other methods in this table was reported in [63]. Average accuracy for proposed method was 69.08%, 72.10%, 86.52%, 83.98%, and 75.27% for bilabial, nasal, C/V, /*uw*/, and /*iy*/ respectively. As shown in the table, average accuracy of method [63] was higher than proposed method in bilabial and nasal tasks. However, average accuracy of proposed method was higher in all tasks in comparison with the methods of [57,62]. Therefore, in summary, all simulation results of Table 1 and 2 show the superiority of the proposed method accuracy over other methods for imagined speech classification. Using CSD for frequency analysis and Riemannian geometry for feature extraction are advantages of the proposed method in comparison with the other methods.

5. Conclusion

This work has investigated the EEG multichannel signal classification problem using CSD matrix and Riemannian geometry. This matrix has been evaluated in this study and used as a feature for signal classification. The CSD matrix not only contains the spectral information of each channel, but it also includes some information about the similarity between different channels. This work also showed that using CSD may yield frequency analysis with proper resolution, and classification accuracy outperforms other competing methods. As the advantages of using CSD, results of channel selection are in corresponding to brain areas during imagined speech that mentioned in [59]. In this work, it is shown that Riemannian distance can be defined on the basis of CSD matrices. Furthermore, this study showed that the use of Riemannian geometry instead of Euclidean geometry for EEG signal classification

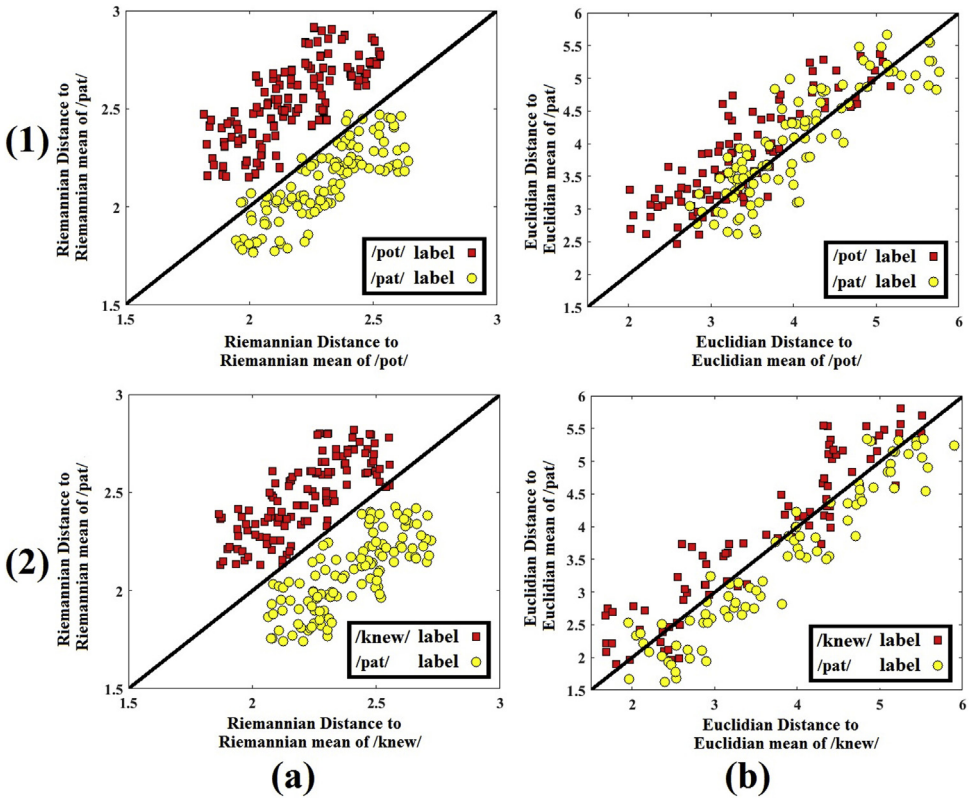

Fig. 5. (a) Riemannian distance of each trial from the two class-related mean CSD matrices (/*pat*/ vs. /*pot*/) and (/*pat*/ vs. /*knew*/) to corresponding Riemannian mean. (b) Euclidean distance of each trial from the two class-related mean CSD matrices (/*pat*/ vs. /*pot*/) and (/*pat*/ vs. /*knew*/) to corresponding Euclidean mean.

would yield higher accuracy compared to other common methods due to the structure of CSD matrices. The reason for this can be attributed to the nature of Riemannian geometry and positive-definite matrices. In summary, the novelties of this paper are as below:

- A new correntropy spectral density matrix in multichannel EEG signals has been introduced.
- A new Riemannian distance for CSD matrices in EEG signals is proposed.

Besides, this work has the following contributions as below:

- Frequency band and EEG channels during imagined speech is investigated.
- Proposed method for imagined speech classification is evaluated.
- Proposed method is compared with the other well-known approaches and the superiority of proposed method is showed.

Moreover, in this study a feasibility study was conducted for imagined speech classification on the “Kara One” database. The results obtained from imagined speech classification and recognition with the proposed method showed promising accuracy. Accordingly, this work can provide a pathway for designing more accurate experiments with more conformity to classification algorithms that could be used for implementation of BCI systems.

However, in this study, preprocessing procedure of the proposed method is a typical approach for noise removing and preparing signals. Using ICA method for eliminating ocular and muscular artifacts and frequency filtering are common method in the preprocessing stage. But, using non-parametric cluster-based permutation test in this paper has various advantages. Nevertheless, using more efficient preprocessing methods like data-driven decomposition methods has been suggested for future works. Due to non-stationarity of EEG signals, data-driven decomposition methods based on the Empirical Mode Decomposition (EMD) technique can be useful. Several researches were conducted based on this technique for BCI applications and showed high performance in EEG classification [17–20,64].

Nevertheless, the proposed method and Riemannian geometry-based methods require more computational time than the other methods, but this higher computational time requirement can be ignored regarding their high efficiency. However, future works may try to reduce the computational cost using other Riemannian metrics or an optimized implementation of matrix algebra.

CRediT authorship contribution statement

Mohamad Amin Bakhshali: Conceptualization, Methodology, Software, Formal analysis, Validation, Data curation, Investigation, Writing - original draft. **Morteza Khademi:** Methodology, Validation, Data curation, Writing - review & editing, Supervision, Project administration. **Abbas Ebrahimi-Moghadam:** Methodology, Validation, Data curation, Writing - review & editing, Supervision. **Sahar Moghimi:** Conceptualization, Software, Formal analysis, Resources.

Declaration of Conmpeting Interest

The authors declare no conflicts of interest.

References

[1] R.J. Stevenson, T.I. Case, Olfactory imagery: a review, Psychon. Bull. Rev. 12 (2005) 244–264.
[2] S.M. Kosslyn, Mental images and the brain, Cogn. Neuropsychol. 22 (2005) 333–347.

[3] T.L. Hubbard, Auditory aspects of auditory imagery, in: S. Lacey, R. Lawson (Eds.), Multisensory Imagery, Springer New York, New York, NY, 2013, pp. 51–76.
[4] S. Martin, P. Brunner, C. Holdgraf, H.-J. Heinze, N.E. Crone, J. Rieger, G. Schalk, R.T. Knight, B.N. Pasley, Decoding spectrotemporal features of overt and covert speech from the human cortex, Front. Neuroeng. 7 (2014).
[5] R. Bogue, Brain–computer interfaces: control by thought, Ind. Robot: Int. J. 37 (2010) 126–132.
[6] K. Casimo, K.E. Weaver, J. Wander, J.G. Ojemann, BCI use and its relation to adaptation in cortical networks, IEEE Trans. Neural Syst. Rehabil. Eng. PP (2017) 1.
[7] J. Xu, S. Mitra, C.V. Hoof, R. Yazicioglu, K.A.A. Makinwa, Active electrodes for wearable EEG acquisition: review and electronics design methodology, IEEE Rev. Biomed. Eng. PP (2017) 1.
[8] J. Zhang, G. Sudre, X. Li, W. Wang, D.J. Weber, A. Bagic, Clustering linear discriminant analysis for MEG-based brain computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 19 (2011) 221–231.
[9] G. Schalk, E.C. Leuthardt, Brain–computer interfaces using electrocorticographic signals, IEEE Rev. Biomed. Eng. 4 (2011) 140–154.
[10] R. Sitaram, N. Weiskopf, A. Caria, R. Veit, M. Erb, N. Birbaumer, fMRI brain–computer interfaces, IEEE Signal Process. Mag. 25 (2008) 95–106.
[11] K. Watanabe, H. Tanaka, K. Takahashi, Y. Niimura, Y. Kurihara, NIRS-based language learning BCI system, IEEE Sens. J. 16 (2016) 2726–2734.
[12] E. Donchin, K.M. Spencer, R. Wijesinghe, The mental prosthesis: assessing the speed of a P300-based brain–computer interface, IEEE Trans. Rehabil. Eng. 8 (2000) 174–179.
[13] J. Minguillon, M.A. Lopez-Gordo, F. Pelayo, Trends in EEG-BCI for daily-life: requirements for artifact removal, Biomed. Signal Process. Control 31 (2017) 407–418.
[14] K. Brigham, B.V.K.V. Kumar, Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy, 2010th International Conference on Bioinformatics and Biomedical Engineering (2010) 1–4.
[15] K. Brigham, B.V.K.V. Kumar, Subject identification from electroencephalogram (EEG) signals during imagined speech, 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) (2010) 1–8.
[16] K. Mohanchandra, S. Saha, G.M. Lingaraju, EEG based brain computer interface for speech communication: principles and applications, in: A.E. Hassanien, A.T. Azar (Eds.), Brain–Computer Interfaces: Current Trends and Applications, Springer International Publishing, Cham, 2015, pp. 273–293.
[17] P. Gaur, R.B. Pachori, H. Wang, G. Prasad, Enhanced motor imagery classification in EEG-BCI using multivariate EMD based filtering and CSP features, in: International Brain–Computer Interface (BCI) Meeting, June, 2016.
[18] P. Gaur, G. Kaushik, R.B. Pachori, H. Wang, G. Prasad, Comparison analysis: single and multichannel EMD-based filtering with application to BCI, in: Machine Intelligence and Signal Analysis, Springer, Singapore, 2019, pp. 107–118.
[19] P. Gaur, R.B. Pachori, H. Wang, G. Prasad, An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain–computer interface, in: 2015 International Joint Conference on Neural Networks (IJCNN), July, 2015, pp. 1–7.
[20] P. Gaur, R.B. Pachori, H. Wang, G. Prasad, An automatic subject specific intrinsic mode function selection for enhancing two-class EEG based motor imagery-brain computer interface, IEEE Sens. J. 19 (2019) 6938–6947.
[21] D. Siyi, S. Ramesh, L. Tom, D.Z. Michael, EEG classification of imagined syllable rhythm using Hilbert spectrum methods, J. Neural Eng. 7 (2010) 046006.
[22] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, R. Srinivasan, Toward EEG sensing of imagined speech, in: J.A. Jacko (Ed.), Human–Computer Interaction. New Trends: 13th International Conference, HCI International 2009, San Diego, CA, USA, July 19–24, 2009, Proceedings, Part I, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 40–48.
[23] M. Matsumoto, J. Hori, Classification of silent speech using support vector machine and relevance vector machine, Appl. Soft Comput. 20 (2014) 95–102.
[24] L. Wang, X. Zhang, X. Zhong, Y. Zhang, Analysis and classification of speech imagery EEG for BCI, Biomed. Signal Process. Control 8 (2013) 901–908.
[25] C.S. DaSalla, H. Kambara, M. Sato, Y. Koike, Single-trial classification of vowel speech imagery using common spatial patterns, Neural Netw. 22 (2009) 1334–1339.
[26] C. Nguyen, G.K. Karavas, P. Artemiadis, Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features, J. Neural Eng. 15 (2018) 016002.
[27] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain–computer interfaces: a 10-year update, J. Neural Eng. 15 (2018), 031005 (28pp.).
[28] P. Herman, G. Prasad, T.M. McGinnity, D. Coyle, Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (2008) 317–326.
[29] Q. Lei, H. Bin, A wavelet-based time-frequency analysis approach for classification of motor imagery for brain–computer interface applications, J. Neural Eng. 2 (2005) 65.
[30] Y. Li, K.M. Wong, Riemannian distances for signal classification by power spectral density, IEEE J. Sel. Top. Signal Process. 7 (2013) 655–669.
[31] I. Santamaria, P.P. Pokharel, J.C. Principe, Generalized correlation function: definition, properties, and application to blind equalization, IEEE Trans. Signal Process. 54 (2006) 2187–2197.
[32] X. Zou, L. Feng, H. Sun, Robust compressive sensing of multichannel EEG signals in the presence of impulsive noise, Inf. Sci. 429 (2018) 120–129.
[33] F. Yger, M. Berar, F. Lotte, Riemannian approaches in brain–computer interfaces: a review, IEEE Trans. Neural Syst. Rehabil. Eng. PP (2017) 1.
[34] M. Congedo, A. Barachant, R. Bhatia, Riemannian geometry for EEG-based brain–computer interfaces; a primer and a review, Brain–Comput. Interfaces 4 (2017) 155–174.
[35] Y. Li, K.M. Wong, H. deBruin, EEG signal classification based on a Riemannian distance measure, in: 2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH), 26–27 September, 2009.
[36] S. Guan, K. Zhao, S. Yang, Motor imagery EEG classification based on decision tree framework and Riemannian geometry, Comput. Intell. Neurosci. 2019 (2019) 13.
[37] P. Gaur, R.B. Pachori, H. Wang, G. Prasad, A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry, Expert Syst. Appl. 95 (2018) 201–211.
[38] P. Gaur, et al., Tangent space features-based transfer learning classification model for two-class motor imagery brain–computer interface, Int. J. Neural Syst. (2019) 1950025.
[39] D. Wu, B.J. Lance, V. Lawhern, S. Gordon, T.P. Jung, C.T. Lin, EEG-based user reaction time estimation using Riemannian geometry features, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (2017) 2157–2168.
[40] J.W. Xu, H. Bakardjian, A. Cichocki, J.C. Principe, A new nonlinear similarity measure for multichannel signals, Neural Netw. 21 (2008) 222–231.
[41] J.C. Principe, Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives, Springer Publishing Company, Incorporated, 2010.
[42] A. Garde, So, x.L. rnmo, R. Jane, B.F. Giraldo, Correntropy-based spectral characterization of respiratory patterns in patients with chronic heart failure, IEEE Trans. Biomed. Eng. 57 (2010) 1964–1972.
[43] J. Xu, Nonlinear Signal Processing Based on Reproducing Kernel Hilbert Space, University of Florida, 2007, PhD Thesis.
[44] P. Fletcher, S. Joshi, Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors, in: Proceeding Computer Vision and Mathematical Methods in Medical Biomedical Image Analysis, 2004, pp. 87–98.
[45] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (2004) 9–21.
[46] T. Xu, M. Stephane, K.K. Parhi, Multidimensional analysis of the abnormal neural oscillations associated with lexical processing in schizophrenia, Clin. EEG Neurosci. 44 (2013) 135–143.
[47] T. Harmony, T. Fernández, J. Silva, J. Bosch, P. Valdés, A. Fernández-Bouzas, L. Galán, E. Aubert, D. Rodríguez, Do specific EEG frequencies indicate different processes during mental calculation? Neurosci. Lett. 266 (1999) 25–28.
[48] M. Stephane, a. Leuthold, M. Kuskowski, K. McClannahan, T. Xu, The temporal, spatial, and frequency dimensions of neural oscillations associated with verbal working memory, Clin. EEG Neurosci. 43 (2012) 145–153.
[49] E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG and MEG-data, J. Neurosci. Methods 164 (2007) 177–190.
[50] R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data, Comput. Intell. Neurosci. 2011 (2010) e156869.
[51] T. Xu, M. Stephane, K. Parhi, Abnormal neural oscillations in schizophrenia assessed by spectral power ratio of MEG during word processing, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (2016) 1148–1158.
[52] T. Ning, C. Nikias, Multichannel AR spectrum estimation: the optimum approach in the reflection coefficient domain, IEEE Trans. Acoust. Speech Signal process. 34 (1986) 1139–1152.
[53] A. Schl“ogl, A comparison of multivariate autoregressive estimators, Signal Process. 86 (2006) 2426–2429.
[54] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 2012.
[55] A. Sharmila, P. Geethanjali, DWT based detection of epileptic seizure from EEG signals using naive bayes and k-NN classifiers, IEEE Access 4 (2016) 7716–7727.
[56] Z. Shao, M.J. Er, N. Wang, An efficient leave-one-out cross-validation-based extreme learning machine (ELOO-ELM) with minimal user intervention, IEEE Trans. Cybern. 46 (2016) 1939–1951.
[57] S. Zhao, F. Rudzicz, Classifying phonological categories in imagined and articulated speech, 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015) 992–996, Available from: http://www.cs.toronto.edu/complingweb/data/karaOne/karaOne.html.
[58] R. Oostenveld, P. Praamstra, The five percent electrode system for high-resolution EEG and ERP measurements, Clin. Neurophysiol. 112 (2001) 713–719.
[59] G. Hickok, Computational neuroanatomy of speech production, Nat. Rev. Neurosci. 13 (2) (2012) 135–145.
[60] B. Blankertz, S. Lemm, M. Treder, S. Haufe, K.-R. M”uller, Single-trial analysis and classification of ERP components: a tutorial, NeuroImage 56 (2011) 814–825.
[61] O. Aydemir, T. Kayikcioglu, Decision tree structure based classification of EEG signals recorded during two dimensional cursor movement imagery, J. Neurosci. Methods 229 (2014) 68–75.

[62] P. Sun, J. Qin, Neural Networks Based Eeg-Speech Models, 2016 (arXiv preprint) arXiv:1612.05369.
[63] P. Saha, S. Fels, M. Abdul-Mageed, Deep learning the EEG manifold for phonological categorization from active thoughts, in: ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April, 2019, pp. 2762–2766.
[64] P. Gaur, R.B. Pachori, H. Wang, G. Prasad, A multivariate empirical mode decomposition based filtering for subject independent BCI, in: 2016 27th Irish Signals and Systems Conference (ISSC), June, 2016, pp. 1–7.