Applied Medical Informatics
Original Research Vol. 43, No. 1/2021, pp: 24-42

Using Mismatch Negativity to Detect Selective Auditory Attention by Convolutional Neural Network

Masoud GERAVANCHIZADEH, Sahar ZAKERI*

Faculty of Electrical and Computer Eng., University of Tabriz, Tabriz 51666-15813, Iran

*E-mail: s_zakeri@tabrizu.ac.ir

* Corresponding Author: Tel.: +984133340081

Received: January 6, 2021 /Accepted: March 15, 2021 / Published online: March 30, 2021

Abstract

Background: In every moment of life, the brain processes a lot of combinations of several sounds. This processing includes stream separation and attended selection, among others. Recent studies show that listener's attention could be decoded by analyzing electroencephalography (EEG). *Method:* In this research, a new method for classifying EEG signals is introduced when 40 subjects were asked to listen to two concurrent speech signals and attend to one of them in a dichotic scenario. The mismatch negativity (MMN) component, as one of the important evoked related potentials (ERPs), plays a crucial role during the attentional process which can be extracted by the non-negative Tucker decomposition method. Then, linear and nonlinear features are extracted. Most of these features are significant according to the analysis of variance (ANOVA) test. Finally, a combination of selected significant features is employed to train and test the convolutional neural network (CNN) classifier. The proposed auditory attention detection method is compared with the three recent and common methods as the baseline systems. *Results:* The proposed method detects the attended speaker by MMN with a classification accuracy of 98.21%. To introduce a practical application of the proposed method, six near-ear electrodes are selected to detect attended speech. The classification performance equal to 71.75% shows that using only these electrodes in hearing-aids could be a promising strategy in detecting attentional behavior. *Comparison with Existing Methods:* Comparing to three conventional auditory attention detection methods, we find that the proposed approach shows higher accuracy with MMN as input data. *Conclusion:* The results open a new perspective to design neural-based brain-computer interfaces (BCI) using selective auditory attention.

Keywords: Selective Attention; Mismatch Negativity; Electroencephalography (EEG); Tensor Decomposition; Feature Extraction; Convolution Neural Network

Introduction

One of the human abilities is attending a specific speaker in a multi-talker scenario, such as a cocktail party. It is yet unclear what happens in the brain to facilitate the attentional process and separate a specific voice or sound. Attention is a cognitive process and plays a salient role in the vision and auditory perception, helping the human concentrate on specific objects of the environment while discarding others.

Auditory scene analysis (ASA) is a basic auditory cortex's ability, which allows us to detect and understand sound events in an acoustic environment [1]. When listening to a person in a cocktail party or searching for one musical instrument such as a violin in a symphonic orchestra, we depend on our ear's exceptional ability to analyze complex acoustic scenes into auditory streams of target and interferences [2]. It is known that the auditory scene analysis (ASA) mechanism is achieved based on

two types of auditory processing: 1) bottom-up attention, often driven by salient differences between target and background, and 2) a top-down cognitive attention, driven by prior knowledge [3]; although it is a challenge to model these types of information processing mathematically [4]. The investigation of auditory selective attention was introduced first by Cherry [5]. Since then, other researchers have presented several dichotic [6] and binaural [7] methods to inspect the mechanism of the auditory attention detection (AAD) in real adverse conditions [8]. There are numerous applications regarding AAD modeling in ASA. Some studies have shown the use of auditory attention methods in the brain-computer interface (BCI) systems [9, 10] and robotics [11]. Other applications concern the use of AAD to control other devices, such as sound recording devices [12]. As a golden aim, the notion of AAD can also be employed in a neuro-steered hearing prosthesis, where the device can amplify the attended speech of a hearing-impaired (HI) subject placed in a competitive talker scenario.

The auditory attention process takes place on the level of the sensory cortex in the brain [13] and influences several biological activities, including facial expression, eye movement, and especially many interconnected neural networks [14]. This process can be tracked by recording brain activities [15], using electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and magnetoencephalography (MEG). Among all these techniques, fMRI and MEG have some limitations, such as low temporal resolution, portability, and price, making them very impractical [16]. However, due to its accessibility and applicability in real-time measurements, EEG has been considered an interesting tool in cognitive neuroscience studies [17].

External or internal stimuli create a particular type of EEG signal is known as event-related potentials (ERPs) [18]. It is known that ERPs reveal the neural activation in the primary sensory cortex and its associative cortical regions, which are related to higher levels of cognitive processes. Their analysis includes the computation of the amplitude, latency, scalp topographic distribution, polarity (positive/negative peaks), and time course, which helps to understand human psychological behavior, specifically auditory attention [19]. Usually, the early ERP components (e.g., P100, N100, and P200) are related to the selective attentional mechanism, whereas the late components such as P300 are often associated with formation and interpretation of the stimulus [20]. Among these components, it has been proved that the mismatch negativity (MMN), known also as small ERP, occurs in response to auditory attention stimuli [21] (see Figure 1). This type of ERP appears when repetitive standard auditory stimuli are occasionally interspersed with deviant stimuli, which are audibly different from the standards [21].

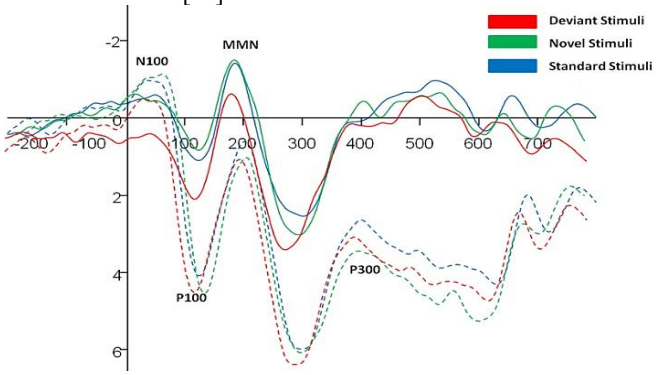

Figure 1. An example of event-related potentials and its components (adapted from [28])

Various methods have been developed in the recent decade to detect selective auditory attention. In one of these methods, a forward mapping from the amplitude envelopes of speech signals to EEG signals was created. Here, the aim was to estimate a function that characterizes the way stimuli are mapped onto neural responses. The attentional state of a listener could then be deducted from the estimated EEG signals [22]. Miran et al. [23] proposed a real-time method based on Bayesian filtering. However, the system's classification accuracy to the attended and unattended speaker was near to previous works [22] with little performance improvement. Other researchers reconstructed attend and unattended speech in a dichotic listening scenario in the low-frequency range (1-8 Hz) by the

EEG signals [24, 25]. In this frequency range, EEG corresponds to the spectrum of speech envelope. Here, the subject's attention is detected based on the correlation between the reconstructed speech envelope and the actual attended and unattended speech envelopes of the two ears, a process called backward mapping. The reconstructed envelope is obtained by using a decoder which is a function that maps the electrode's responses to the speech stimuli [26, 27].

In similar research, Aroudi and Doclo [29] investigated the potential of using mixture signals instead of clean speech signals, as reference signals, to decode auditory attention in near realistic (i.e., anechoic, reverberant, noisy, and reverberant noisy) conditions.

Another method in detecting attention concentrated on the extracted informative features from EEG. Machine learning techniques classify the extracted features corresponding to the attended and unattended speakers. Haghighi et al. [28] calculated the cross-correlation coefficients between EEGs and target/distractor acoustic envelopes at different time lags to classify attended and unattended speech in offline mode. Furthermore, they employed principle coefficient analysis (PCA) to reduce feature dimensions and regularized discriminant analysis (RDA) as the classifier. In another study, online EEGs recorded by mobile were utilized to compute the P300 component, which was then used to detect auditory attention in a dual speaker scenario [30]. In this approach, the fast classification with a supervised and cross-validation linear discriminant analysis (LDA) achieves an accuracy near 80% for all experimental conditions.

Few studies focused on employing powers of EEG bands and some auditory ERPs measures (i.e., N1, P1, and P2) in the training of classifiers [31] to detect selective attention. The template matching classification technique achieved near 70% accuracy [31] with all three ERP components or the single N1 component when listeners were involved in a visual task. Wang et al. [32] examined EEG signals from healthy subjects during a dual attention task. Their distraction-detection model was based on independent component analysis (ICA). This research's classification performance using support vector machines (SVM) gained an accuracy of ~ 85% in real-time. Based on a new protocol to record the EEG signals, EEG was decomposed to its subbands (i.e., delta, theta, alpha, beta, and gamma) to classify attention and non-attention states in educational environments [33]. The results showed that the best effective features were related to the beta band and the energy of the signals in a mathematical operation task. It was also verified that the c-SVM (i.e., SVM classifier type 1) and LDA classifiers had higher performance in attention detection than other methods with 92% accuracy. Using the empirical decomposition technique (EMD), Looney et al. [34] estimated selective attention features by modeling the degree of the gamma-band synchronization between attended stimuli and neural activity in EEG with 71% accuracy. Lu et al. [35] extracted four types of entropy (i.e., approximate entropy, sample entropy, composite multi-scale entropy, and fuzzy entropy) as informative features of EEGs to classify three kinds of tasks, namely, rest and two different auditory objects attention. They obtained an identification accuracy of ~ 80% among the tasks with the LDA and SVM classifiers.

It is clear that the first two AAD methods (i.e., forward and backward mapping) require specifically "clean" speech to compute the correlation between stimuli envelopes and EEG data; a situation which never happens in realistic acoustic environments, especially in a cocktail party condition. It has also been indicated that the accuracy of decoders depends on temporal resolution or the trial length of stimuli (e.g., shorter trial lengths such as 10 s were preferred over the most reported 60 s trial length [20]). As a preliminary solution, multiple analysis and preprocessing steps have been considered to optimize the decoder estimation by deep neural networks [36]. Furthermore, using auditory attention decoders to detect attended speech has an essential limitation in real-life scenarios. The AAD system does not know *a priori* which speaker attends to select the appropriate decoder [36]; an important issue ignored in these researches.

In contrast to these approaches, informative features do not require access to clean auditory stimuli, which makes them applicable in real-life situations such as a cocktail party. Although this can be regarded as a benefit, the approaches employing informative features in AAD still have some shortcomings as to their performance. This can be due to inefficient classification procedures or features extracted from the raw EEG data to train classifiers.

In this study, a novel approach is introduced to detect attended speech based on the MMN data to overcome the aforementioned limitations of AAD methods. To this aim, first, the technique of

Tucker decomposition is employed to extract the desirable MMN component. To alleviate the deficiencies of informative feature-based approaches in AAD, some linear and nonlinear candidate features are obtained in the next stage. Then, the known statistical analysis of variance (ANOVA) test is used to select significant features for training a classifier. Finally, a feature set based on the combination of significant features is used by the convolutional neural network (CNN) to classify the attended and unattended speech.

Material and Method

Data Acquisition

The EEG signals were recorded from forty subjects (age 27.3 ± 3.2; 32 males; 8 females; seven left-handed) in 30 trials (approx. 1800 s of data per subject), each having a length of ~1 min [37]. Two different stories were presented simultaneously via headphone to each subject; one to the left ear and one to the right ear. Twenty subjects were asked to attend to the left ear stimuli and the rest to the right ear stimuli. The recording of EEGs was performed with 128 electrodes with a sampling frequency of 512 Hz. Then, the EEG signals were down-sampled to 128 Hz to reduce the processing time. After each trial, 4-6 multiple choice questions on both stories were asked from the subjects. None of the subjects had hearing impairments or neurological disorders. To eliminate the effects of 50 Hz power line noise, eye blinking, and muscle movement artifacts, the data were passed through a band-pass filter over the frequency range of 0-134 Hz using a BioSemi Active Two system.

Two audio stories about nature, each with a length of ~1 min, were used as the auditory stimuli. While the participants focused their attention on the specific auditory stimulus, the corresponding auditory responses in the form of EEG signals were generated. A different male speaker read each story, and the native language was Dutch. The audio stories were dichotically presented with Sennheiser HD650 headphones, and the graphical user interface (GUI) from the Neurobehavioral System was designed to guide the subjects. The participants were asked to keep their full attention on a pre-determined auditory stimulus while their EEGs were being recorded. For each subject, the data acquisition protocols were randomly performed to prevent the EEG signals from being contaminated by a fixed order of auditory tasks or the dominance of ears.

Proposed Auditory Attention Detection using Informative Features and Convolutional Neural Network

This study introduces a novel algorithm that detects the attended speech at the left or right ear by classifying the distinguishing features extracted from the EEG signals. To this purpose, first, a segment of the EEG signal (~1000 ms) is used to derive the MMN component by the non-negative Tucker decomposition method. Then, the MMN data are utilized to extract linear (amplitude, peak time, maximum, and minimum) and nonlinear (approximate and sample entropies, Lyapunov exponent, fractal dimension, and Hurst exponent) features. At the next step, the ANOVA test is used to select significant features for training a classifier. At the end of the procedure, a feature set based on the combination of significant features is applied to the CNN (Convolutional Neural Network) classifier to detect the attended and unattended stimuli. The block diagram representation of the method used in this study is given in Figure 2.

Non-negative Tucker decomposition method. The ERP data, which is often produced by averaging the EEG signals, include different modes (e.g., time, space, stimulus, and participant). The collection of these modes is commonly named as a tensor [38]. In the case that data are matrices (i.e., having two dimensions), the desirable ERP components (e.g., MMN) could be obtained by the decomposition of ERP data using the principal component analysis (PCA) or independent component analysis (ICA) techniques. However, when the ERP data are represented by a tensor (i.e., a data structure with more than two dimensions), tensor decomposition methods could be used to extract desirable components [39]. Two popular models for tensor decomposition are, respectively, Tucker decomposition [40] and canonical polyadic decomposition (CPD) [41]. Due to some advantages of the Tucker decomposition [42], in this paper, we adopt this model for obtaining MMN.

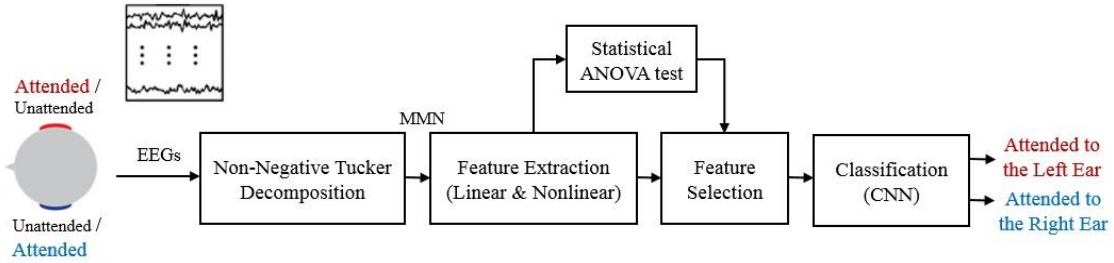

Figure 2. The block diagram of the proposed AAD algorithm to detect attended/unattended speech. After the decomposition of the data by the Tucker decomposition method, the MMN component is computed and processed to extract appropriate linear and nonlinear features. Then, an optimal feature set specified by the ANOVA test is selected and given to the CNN classifier to determine auditory attended/unattended stimuli.

For the N^{th}-order tensor, $\underline{\boldsymbol{Y}} \in \underline{\Re}^{I_1 \times I_1 \times \cdots \times I_N}$, the Tucker decomposition is represented as follows [43]:

$$\underline{\boldsymbol{Y}} = \underline{\boldsymbol{G}} \times {}_1\boldsymbol{A}^{(1)} \times {}_2\boldsymbol{A}^{(2)} \times \ldots {}_N\boldsymbol{A}^{(N)} + \underline{\boldsymbol{E}} = \underline{\widehat{\boldsymbol{Y}}} + \underline{\boldsymbol{E}}, \quad (1)$$

where N is the number of modes (e.g., frequency, time, space, etc.), $\underline{\boldsymbol{G}} \in \Re^{R_1 \times R_2 \times \ldots \times R_N}$ is the core tensor, ${}_n\underline{\mathbf{A}}^{(n)} = \left[\mathbf{a}_{r_1}^{(1)}, \mathbf{a}_{r_2}^{(2)}, \ldots, \mathbf{a}_{r_N}^{(N)}\right] \in \Re^{I_n \times R_n}$ expresses the mode-n components matrix, r is the rank of the matrix, $\underline{\widehat{\boldsymbol{Y}}}$ approximates tensor $\underline{\boldsymbol{Y}}$, and $\underline{\boldsymbol{E}} \in \Re^{R_1 \times R_2 \times \ldots \times R_N}$ represents error or noise. Boldface capital letters denote matrices, boldface lowercase letters denote vectors, and lowercase letters denote scalars. Figure 3 illustrates a typical decomposition of data into three modes of frequency, time, and space. The matrices ${}_n\underline{\mathbf{A}}^{(n)}$ and $\underline{\boldsymbol{G}}$ are solutions of Eq. (1) which are obtained by the following least-square problem:

$$min\left\|\boldsymbol{Y} - {}_n\boldsymbol{A}^{(n)}\ \boldsymbol{G}^{(-n)}\right\|^2,\ n = 1,\ 2,\ \ldots,\ N. \quad (2)$$

In theory, the Tucker decomposition does not lead to particular solutions. However, when additional constraints are imposed on different modes, it results in a unique solution [44]. In order to overcome the sparsity of topography, independency among sources, and nonnegativity of time-frequency representation, the nonnegative Tucker factorization (NTF) was used to generate appropriate components. The specific core tensor **G**, symbolized as, $\underline{\boldsymbol{G}}(3,8,4,:)$ represents the desired MMN [45], where the first, second, and third entries of **G** correspond, respectively, to spectral (3rd component), temporal (8th component), spatial (4th component) modes, and ':' means considering all samples of the fourth entry without change.

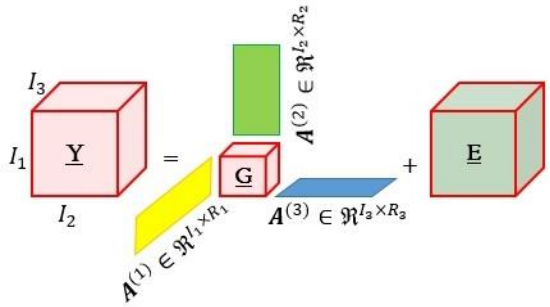

Figure 3. 3D illustration of the Tucker tensor decomposition model. I_1, I_2, and I_3 denote, respectively, the frequency, time, and space

Feature extraction. Feature extraction plays a crucial role in extracting beneficial information from the EEG time series $(y(t))$ which are measured as the output of the brain system. In this study,

effective linear and nonlinear features are extracted from the MMN data. The linear features include amplitude (Amp), maximum (Max), minimum (Min), and peak time (PT) (peak time is the time when the signal level reaches 90% of its maximum value) [46]. In Figure 4, a sample of MMN signal obtained from the Tucker decomposition is shown with its linear features. In the following, the nonlinear features used in the implementations are described in detail.

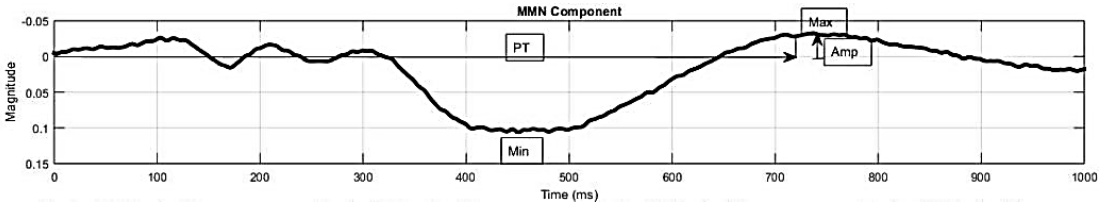

Figure 4. A sample of MMN signal obtained from the Tucker decomposition with the specified linear features of 'Min', 'Max', 'PT', and 'Amp'

a) *Power spectral density*. Power spectral density (PSD) analysis of EEGs is a reliable method of describing brain functionality [31]. This method helps to determine which frequencies of the EEG signal have significant differences between two groups of left- and right-ear attended listeners. In other words, PSD is an indicator to find which of the subbands of EEG signals are related to attention. To investigate this, the EEG signal is decomposed to its subbands, including delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-20 Hz), and gamma (20-70 Hz) bands [47]. To examine which frequencies of the EEG signal are affected by the attentional mechanism, PSD is computed for the two groups of subjects. In one experiment, the subjects attend to the left story and ignore the right one and in the other, the subjects attend to the right story and ignore the left one.

b) *Approximate and sample entropies*. Entropy indicates the randomness or regularity, and predictability of a system [48]. Approximate entropy (ApEn) is a popular tool in quantifying the amount of regularity and the unpredictability of fluctuations over time series $y(n)=\left[y(i), y(i+1),\ldots, y(i+m-1)\right], (i = 1,\ 2,\ \ldots,\ N-m+1)$,which has been used for analyzing many nonlinear dynamic signals such as hormone release rate, heart rate variability, and EEGs [49]. ApEn is driven from the correlation integral $C_r^m(i)$ as:

$$ApEn(N,m,r) = (N-(m-1))^{-1}\sum_{i=1}^{N-(m-1)} ln\, C_r^m\,(i) - (N-m)^{-1}\sum_{i=1}^{N-m} ln\ \ C_r^m(i) \quad (3)$$

where, $N$is the number of data points, m is the embedding dimension, and r is the tolerance window. Larger values of ApEn means the signal has more complexity. In this study, the Lorenz model with $N = 122$ (corresponding to ~1000 ms),$m = 2$, and $r = 0.2$ [50] is used to compute the ApEn values.

c) *Sample entropy*. This feature is a modification of ApEn. Sample entropy (SampEn) has two advantages over ApEn: independence from data length (i.e., *N*) and relatively easy implementation. For the given *N* data points from a time series $y(n) = [y(i),\ y(i+1),\ \ldots,\ y(i+m-1)]$, SampEn can be defined as:

$$SampEn(N,m,r) = \lim_{n\to\infty}\left\{-ln\left(\frac{A^m(r)}{B^m(r)}\right)\right\}, \quad (4)$$

where $B^m(r)B^m(r)$ and $A^m(r)$ are the mean values of the similarity between patterns or templates of length *m* and *m*+1 points, respectively. It is clear that $A^m(r)$ is always smaller or equal to $B^m(r)$. Therefore, SampEn is always either be zero or positive value. A smaller value of SampEn also indicates more self-similarity in the data set or less noise. Generally, $m = 2$ and $r = 0.2 \times\ std$ are chosen to compute SampEn, where std denotes the standard deviation.

d) *Fractal dimension*. A fractal is a set of data points that resembles the whole set when looking at smaller scales. An essential characteristic of a fractal is self-similarity. This means that the details of a fractal are similar to each other at a certain scale, but not necessarily identical to those of the

structure seen at larger or smaller scales. Fractal dimension (FD) of a waveform represents a powerful tool for transient detection. This feature has been used in the analysis of EEG to identify and distinguish specific states of physiological function [51]. According to the Katz method [52], the FD of a signal is defined as follows:

$$FD^{katz} = \frac{log(L)}{log(d)}, \tag{5}$$

where L is the total length of the signal and d is the estimated distance between the first and the most distant points of the sequence. Mathematically, d can be expressed as $d = max\|y(1) - y(i)\|, \forall i,$ where $y(i)$ represents the time series.

e) *Hurst exponent.* The Hurst exponent, H, is used to measure long-term dependency and its extent in a time series [53]. H evaluates the smoothness of a fractal time series based on the asymptotic behavior of the rescaled range of the process. In EEG analysis, H is often used to characterize the non-stationary behavior of the signal episodes. The Hurst exponent is defined as:

$$E\left[\frac{R(n)}{Std(n)}\right] = Cn^H \quad \text{as} \quad n \to \infty \tag{6}$$

where $R(n)$ is the range of the first n cumulative deviations from the mean, $Std(n)$ is their standard deviation, $E[.]$ the expected value, n is the time span of observation (number of data points in a time series), and C is a constant assumed to be 1. The range $0.5 < H < 1$ indicates the signals with long-term positive correlations, $0 < H < 0.5$ indicates the signals with long-term switching between high and low values in adjacent pairs of the data and $H = 0.5$ shows completely uncorrelated signals.

f) *Lyapunov exponent.* There are some features of a system to determinate deterministic chaos from random or periodic behavior. A chaotic system can be recognized by its sensitive dependence on initial conditions. In such systems, two adjacent points in the trajectory at time 0 diverge widely from each other at time t. The distance between the times 0 and t in the i^{th} direction is shown, respectively, by $\|\delta y_i(0)\|$ and $\|\delta y_i(t)\|$ for the time series $y(t)$. By quantifying the separation rate, λ_i, Lyapunov exponent (LE) is defined as [54]:

$$\frac{\|\delta y_i(t)\|}{\|\delta y_i(0)\|} = 2^{\lambda_i t} \quad (t \to \infty) \tag{7}$$

The separation rate can be variant for different orientations of the initial separation vector. Thus, there is a spectrum of Lyapunov exponents equal in number to the dimensionality of the phase space. It is common to refer to the largest Lyapunov exponent (LLE) because it determines a notion of predictability for a dynamical system [55]. A negative exponent implies that the orbits approach a common fixed point. A zero exponent means the orbits maintain their relative positions, i.e., they are on a stable attractor. Finally, a positive exponent implies the orbits are on a chaotic attractor.

Convolutional neural network (CNN) classifier. Convolutional neural network as a developed and improved neural network, consists of a number of different layers stacked together in a deep architecture: an input layer, a group of convolutional and pooling layers (which can be combined in various ways), a limited number of fully connected hidden layers, and an output (loss) layer [56]. The architecture of CNN used in this research is presented in Fig. 5. It is designed with two convolutional and two max-pooling layers. The strength of CNNs depends on extracting information or features with kernel filters from a given data. The feature map of convolutional layers is computed as:

$$a(t) = (y * h)(t) = \int_{\tau=0}^{N-1} y(\tau)h(t-\tau), \tag{8}$$

where a is the feature map, $y, h,$ and N are the input signal, filter, and the number of elements in $y,$ respectively. The pooling layer or down-sampling layer reduces the dimension of the former output of convolutional layer to decrease computational load and prevent overfitting. In this study, the max-pooling operation is used to select the maximum value in each feature map. The fully connected layer has full connection to all the activations in the previous layer. Rectified linear activation unit (ReLU) and Softmax are the two types of activation functions employed in the CNN structure. They are chosen to impart nonlinearity to the neural network structure. Softmax is used to predict which class

(left attended or right attended) the extracted feature belongs to. Drop out layer is considered to reduce data length. Figure 5 shows the structure of the CNN classifier used in this work. In Table 1, the layers and the corresponding parameters of the proposed CNN network are given in detail. The hyper-parameter settings of the network are presented in Table 2.

Simulations and Evaluations

Experimental Setup

For the purpose of evaluating the performance of the proposed selective AAD method, three recently developed attention detection systems are simulated and used as baselines.

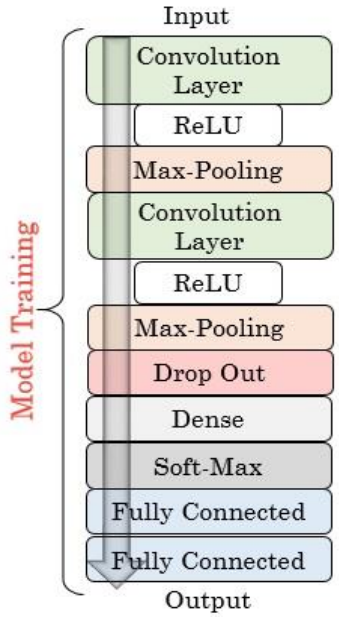

Figure 5. The deep pyramidal CNN architecture used for the classification of left- and right-attended speech.

Table 1. Detailed information about the proposed CNN network.

Number of layers	Number of Neurons	Kernel size	Parameters of layers
Conv1D	2042	5×1	Strides=1, Activation=ReLU
Max-Pooling	1021	2	Strides=2
Conv1D	1018	3×1	Strides=1, Activation=ReLU
Max-Pooling	509	2	Strides=2
Drop Out	255	-	Rate= 0.5
Dense	100	-	Unit size=2, Activation= Soft-Max
Fully connected	50	-	-
Fully connected	2	-	-

Table 2. Different parameter settings used to create the CNN network.

Parameters	Values
Loss Function	Categorical cross entropy
Optimizer	Adam
Learning rate	0.001
Batch size	100
Epochs	50

The baseline systems have an inherently different structure in the detection of attended speech and are denoted as "O'Sullivan et al." [25], "Akram et al." [26], and "Lu et al." [35]. While the methods of "O'Sullivan et al." and "Akram et al." use a backward mapping technique to reconstruct the envelope of the attended speech, the approach of "Lu et al." employs the technique of informative feature to extract entropies (i.e., approximate, sample, composite multiscale, and Fuzzy entropies) for learning the classifier.

In the training of the first and second baseline systems ("O'Sullivan et al.", "Akram et al."), two of the three available trials for each subject are taken randomly in the training of decoders. In this way, in each run of the training procedure, 40 trials (2 (trials) × 20 (subjects)) are used in the design of the decoder. In the testing phase, the remaining data (20 trials) of each run are given to the designed decoder to reconstruct the attended speech envelope. The third baseline's training and testing phases ("Lu et al.") and the proposed methods are as follows. First, the trials of all subjects are put in succession to form an array of 60 trials. Then, in each run of the algorithms, 50 trials (83% of the whole dataset) are selected randomly for the training and validation of the classifiers. The remaining trials are taken for the test procedure. The random splitting and a 10-fold cross-validation approach are used to evaluate CNN.

Evaluation Criteria

Three parameters, namely, Accuracy, Sensitivity, and Specificity, are considered to determine the performance of the CNN classifier. The value of the Accuracy shows the overall detection accuracy. Sensitivity is defined as the rate of correctly classified trials while Specificity indicates the rate of correctly rejected trials. These parameters are defined as [57].:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (9)$$

$$Sesitivity = \frac{TP}{TP+FN} \times 100 \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (11)$$

where, TP (true positive) and TN (true negative) are, respectively, the number of correctly identified and correctly rejected subjects. On the other hand, FP (false positive) and FN (false negative) denote the number of incorrectly identified and incorrectly rejected subjects, respectively.

Statistical Analysis

The statistical analysis method is a multiple-sample parametric or non-parametric analysis of variance which is used to evaluate linear and nonlinear features. The analysis of variance (ANOVA) test [58] is utilized to examine significant differences between two or more data groups. Clearly, ANOVA checks the impact of one/more factor(s) by comparing the means of different samples. The *p*_value is the parameter to determine a significant difference. In this work, the threshold 0.005 (i.e., p-value < 0.005 is used to select significant features extracted from MMN.

Results and Discussion

EEG Subbands Analysis

Figure 6 shows the variations of these subbands for a subject during the time he/she attends to the right or left ear stimuli. All group-level statistical comparisons are conducted using the ANOVA test and a significance threshold of 0.5% is chosen. In Table 3, the average of the mean and standard deviations of PSD between the two groups of subjects attending to the left-and right-ear stimuli are reported in different subbands of the delta, theta, alpha, beta, and gamma. The numerical *p*-values for different EEG subbands are also illustrated in this table, where significant differences between groups of listeners are depicted with the symbol "*****".

Table 3. The average mean ± std values of PSD obtained in different EEG subbands between the two groups of subjects attending to the left-and right-ear stimuli. The asterisk indicates a significant difference $(p\_value < 0.005)$ between both groups of subjects in the attention task.

EEG subbands	Delta	Theta	Alpha	Beta	Gamma
(mean±std)×10^{-3}	**12.5±3.5***	**6.9±3.1***	**2.4±0.8***	**2.9±1.1***	5.2±2

std = standard deviation

As shown in Table 3, there are significant differences between the groups of participants (attending to the left-and right-ear stimuli) in the frequency ranges up to 30 Hz. For frequencies above 30 Hz (i.e., gamma-band), significant differences are not observed. Knowing that the MMN data lies in the frequency ranges under 30 Hz [60], our experiments' outcome confirms the choice of MMN as a reliable and reasonable EEG component to detect and classify auditory attention state.

Results for Scalp EEG

This study focuses on the MMN changes to detect attended speech. To achieve this goal, MMN is obtained from the non-negative Tucker decomposition. Figure 7 shows Tucker decomposition

results in different modes (i.e., spectral, temporal, and spatial) and their corresponding components, respectively. Referring to Figure 7, MMN is represented completely by the specific core tensor (i.e., $\underline{\boldsymbol{G}}(3,8,4,:)$; see Sec. 2.2.1). The EEGLAB toolbox is used to analyze data in spatial mode [59]. After the decomposition procedure, several linear and nonlinear features are extracted from the MMN data. Table 4 represents the mean and standard deviation values of all features. Here, by observing the *p*-values of the ANOVA test, all features except HE are selected as significant which are given to the classifier.

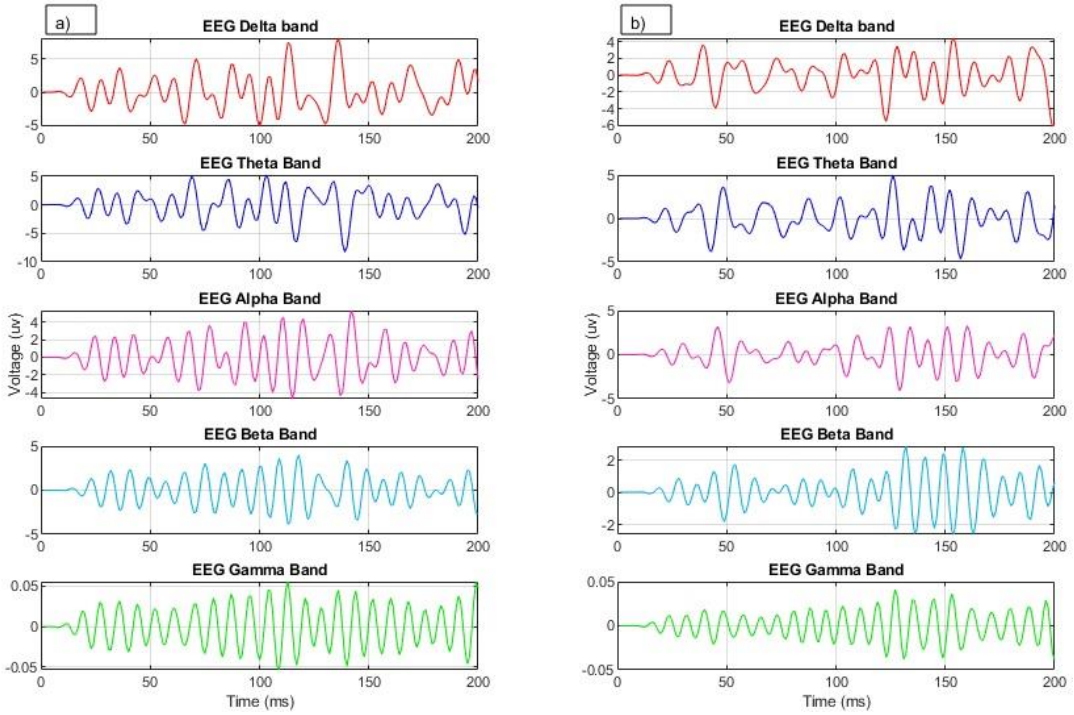

Figure 6. The variations of EEG subbands, delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-20 Hz), and gamma (20-70 Hz), for a subject attending to a) the left-ear or b) the right-ear stimuli

Table 4. The average mean ± std values of linear and nonlinear features obtained from the MMN data between the two groups of subjects attending to the left-and right-ear stimuli. The asterisk indicates a significant difference (p-value < 0.005) between both groups of subjects in the attention task.

Features	**FD**	**HE**	**LE**	**ApEn**	**SampEn**
mean ± std	1.06 ± 0.02*	0.49 ± 0.28	29.47 ± 9.3*	0.19 ± 0.09*	0.83 ± 0.24*
Features	**Amp (μV)**	**PT (ms)**	**Max (μV)**	**Min (μV)**	
mean ± std	5.25 ± 11.9*	361.3 ± 38.3*	4.19 ± 4.9*	-4.35 ± 5.5*	

std = standard deviation

In the next stage, attention patterns are classified by CNN using different combinations of significant features specified by the ANOVA test. The CNN classifier is learned to separate data corresponding to the subjects attending to the left or to right stimuli. To this end, a 10-fold cross-validation approach is adopted. For each fold of the data, the train of the classifier is repeated 50 times. The assessment of the classifier performance for various combination of features is shown in Table 5. According to this table, the feature set LE+SampEn+Amp yields the best classification accuracy with acceptable sensitivity and specificity values. Figure 8 depicts the performance measures

of Accuracy and Loss for the CNN classifier during training for different epochs. By observing the graphs of Accuracy and Loss, the classifier reaches its best performance after 20 epochs of each fold.

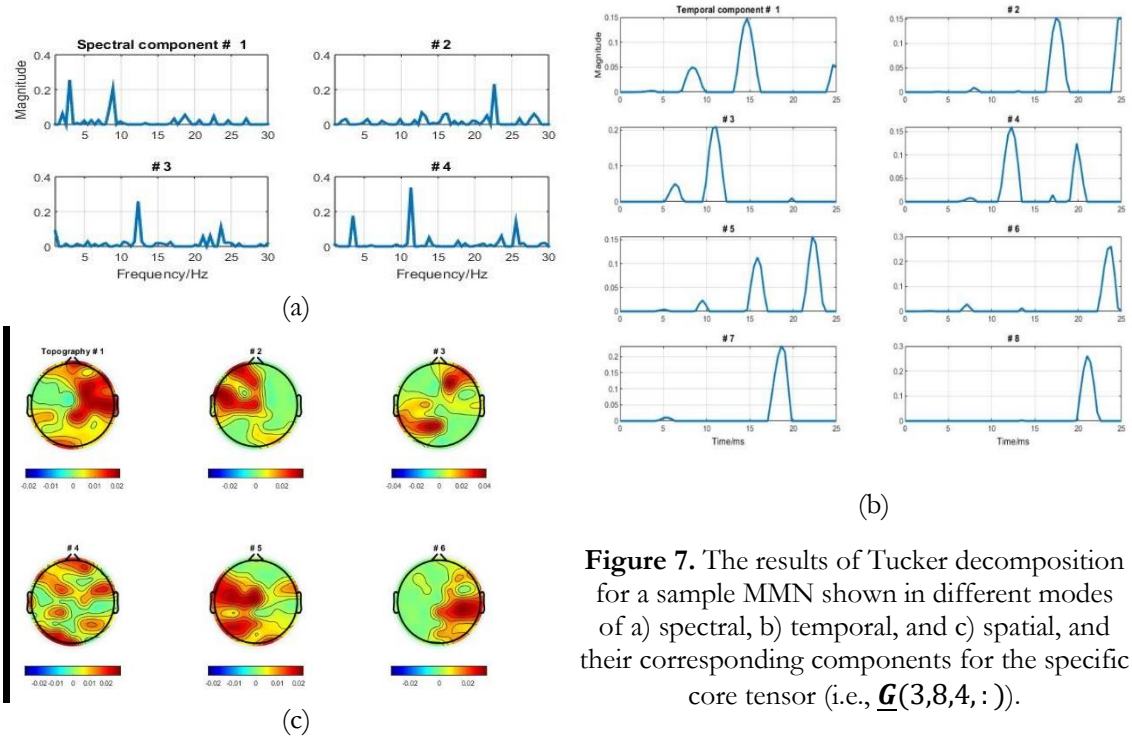

Figure 7. The results of Tucker decomposition for a sample MMN shown in different modes of a) spectral, b) temporal, and c) spatial, and their corresponding components for the specific core tensor (i.e., $\underline{\boldsymbol{G}}(3,8,4,:)$).

Table 5. The performance evaluation of the CNN classifier in terms of Accuracy, Sensitivity, and Specificity for different combinations of features obtained from 128 electrodes. The highest values are depicted as boldface.

Features	**Accuracy**	**Sensitivity**	**Specificity**
Amp	67.73	66.75	67.72
Max	54.19	53.01	54.45
Min	51.75	60.73	67.21
PT	59.63	75.13	62.96
LE	65.00	65.12	64.42
FD	61.74	62.84	64.19
ApEn	64.64	66.21	65.41
SampEn	66.85	64.68	57.22
LE + FD	79.94	79.58	79.40
LE + ApEn	79.37	77.23	76.59
LE + SampEn	80.17	87.41	89.30
SampEn + FD	71.59	76.77	72.29
SampEn + Amp	80.58	82.35	82.15
Amp + PT	79.92	80.02	78.50
SampEn + ApEn	71.99	70.94	70.44
LE + FD + SampEn	89.46	88.98	89.19
Amp + PT + Max	71.15	76.75	72.95
LE + SampEn + Amp	**98.21**	**97.45**	**98.40**
LE + FD + SampEn + Amp	97.83	**97.77**	**98.61**
LE + FD + SampEn + ApEn	91.32	91.29	90.83
Amp + PT+ Max + Min	76.71	77.55	69.59
LE + FD + SampEn + ApEn + Amp	96.91	96.09	97.10

LE + FD + SampEn + ApEn + PT	94.10	94.02	95.34
LE + FD + SampEn + ApEn + Amp + PT + Max + Min	89.48	90.69	89.99

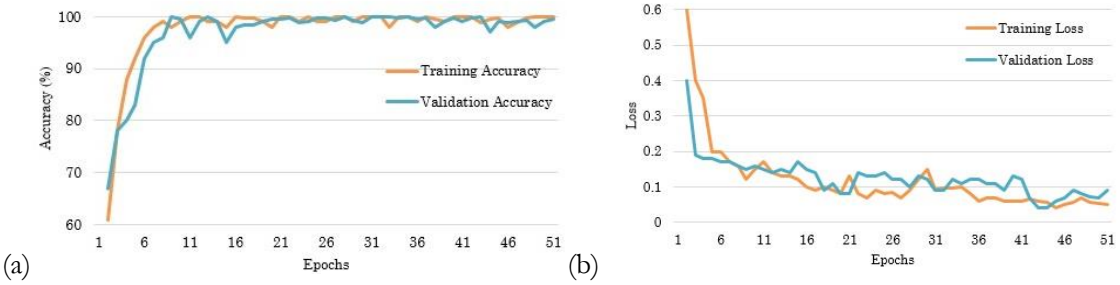

Figure 8. The training procedure of the proposed CNN classifier in terms of Accuracy (a) and Loss (b) for the data obtained from 128 electrodes

The comparison of the proposed AAD method with three baseline classification approaches in terms of mean classification accuracy, for three types of input data, raw EEG, raw MMN, and the feature set LE+SampEn+Amp, is shown in the Figure 9 (panel (a)). As inferred from the figure, the proposed method gives the highest classification rates as compared with "O'Sullivan *et al.*"**,** "Akram *et al.*", and "Lu *et al.*" for all types of inputs. Furthermore, CNN attains the highest accuracy (98.21%) in classifying two groups of left- and right-ear attended subjects when trained with LE+SampEn+Amp.

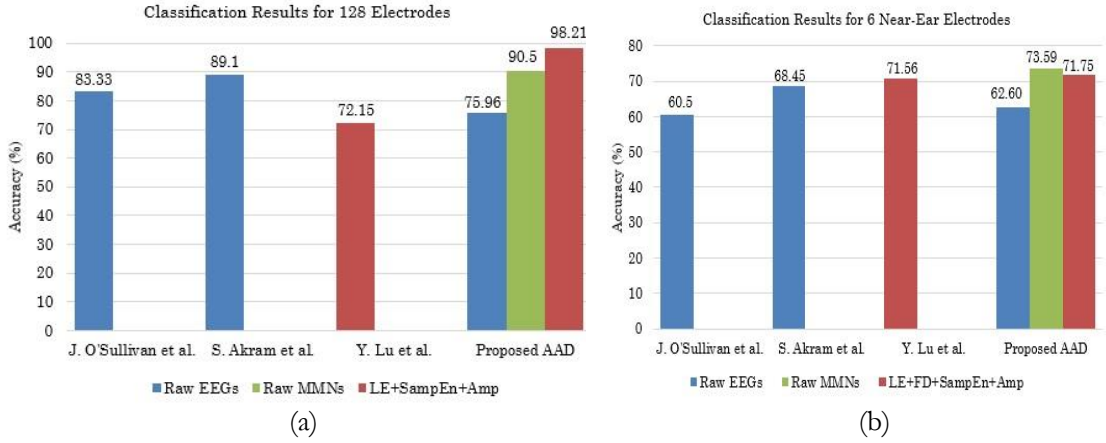

Figure 9. The comparison of the proposed and baseline AAD methods in terms of mean classification accuracy for three types of input data obtained from 128 electrodes, (b) obtained from 6 near-ear electrodes

As inferred from Figure 9, the proposed method gives the highest classification rates, 98.21%, as compared with baseline systems: "O'Sullivan *et al.*", "Akram *et al.*", and "Lu *et al.*" for all types of inputs (i.e., raw EEG, raw MMN, and significant feature set) with the feature set of LE+SampEn+Amp.

Results for Near-Ear Electrodes

An important application of the proposed AAD concerns developing a method to detect auditory attention for hearing-aids listeners. Since hearing-aids are designed as around-the-ear systems, the proposed AAD method should be implemented in such a form that EEGs are recorded only from the electrodes close to the ears. This removes the requirement for scalp EEG measurements and decreases the computational load as well. To achieve this aim, EEGs of six nearest electrodes (TP8,

FT8, and T8 for the right ear and TP7, FT7, and T7 for the left ear, with Cz as the reference electrode) are chosen and analyzed.

Figure 10 shows the results of statistical analysis for features of Amp, PT, Max, Min, LE, FD, SampEn, ApEn, and HE extracted from the MMN component during the auditory attention tasks. The values of these measurements are shown as the mean±standard deviation. The small *p*-values of the ANOVA test, marked with symbol "*", indicate the significant differences between the two states of auditory attention tasks. For LE, HE, FD, Amp, and SampEn, there are significant differences in all six channels in which the *p*_values are less than 0.005. For ApEn, TP8, TP7, T8, FT8, FT7, for Max, TP8, TP7, T8, FT8, for Min, TP8, T8, FT8, FT7, and for PT, TP8, T8, FT8 electrodes have the small *p*-values. As inferred from Figure 10, TP8, T8 and FT8 have significant differences with all nine features. Therefore, it could be possible to use just these three electrodes in order to design neuro-steered hearing aids with attention detection capability at the right ear. The results of the ANOVA test suggest that LE, HE, FD, Amp, and SampEn are important features to be considered in the attention detection. The impact of electrode reduction on AAD using different combinations of selected features is explored by the CNN classifier.

The classification performance of various feature combinations is shown in Table 6.

Table 6. The performance evaluation of the CNN classifier in terms of Accuracy, Sensitivity, and Specificity for different combinations of features obtained from 6 near-ear electrodes. The highest values are depicted as boldface

Features	**Accuracy**	**Sensitivity**	**Specificity**
Amp	58.22	57.91	57.55
LE	61.96	60.47	62.32
FD	59.45	59.27	59.14
HE	58.32	59.44	56.98
SampEn	62.17	63.57	63.16
LE + FD	62.86	62.33	61.25
HE + FD	61.05	60.50	63.11
LE + HE	59.14	57.96	58.78
LE + SampEn	64.28	68.27	63.45
SampEn + FD	63.30	64.09	64.37
SampEn + Amp	63.76	62.26	61.85
HE + SampEn	63.61	61.15	61.31
LE + FD + HE	67.50	68.98	69.85
LE + FD + SampEn	68.43	66.32	67.72
LE + SampEn + Amp	71.12	**72.65**	70.25
LE + FD + HE + SampEn	70.58	71.85	69.46
LE + FD + SampEn + Amp	**71.75**	**72.41**	**71.90**
LE + FD + SampEn + Amp + HE	70.63	71.54	**71.96**

According to this table, the feature set LE+FD+SampEn+Amp yields high classification accuracy and sensitivity. The classification results are demonstrated in Figure 9 (panel (b)) for three different input data, raw EEG, raw MMN, and LE+FD+SampEn+Amp. As it is evident, the proposed AAD method achieves high accuracy in the classification of two groups of auditory attention tasks as compared with the baseline systems. It is observed that the classification result for the selected feature set is near to that of "Lu *et al.*" method, however, raw MMN data results in the highest accuracy. This could be due to the fact that in contrast to the selected feature set, raw MMN carries a large amount of data which is required by the CNN classifier for its best performance.

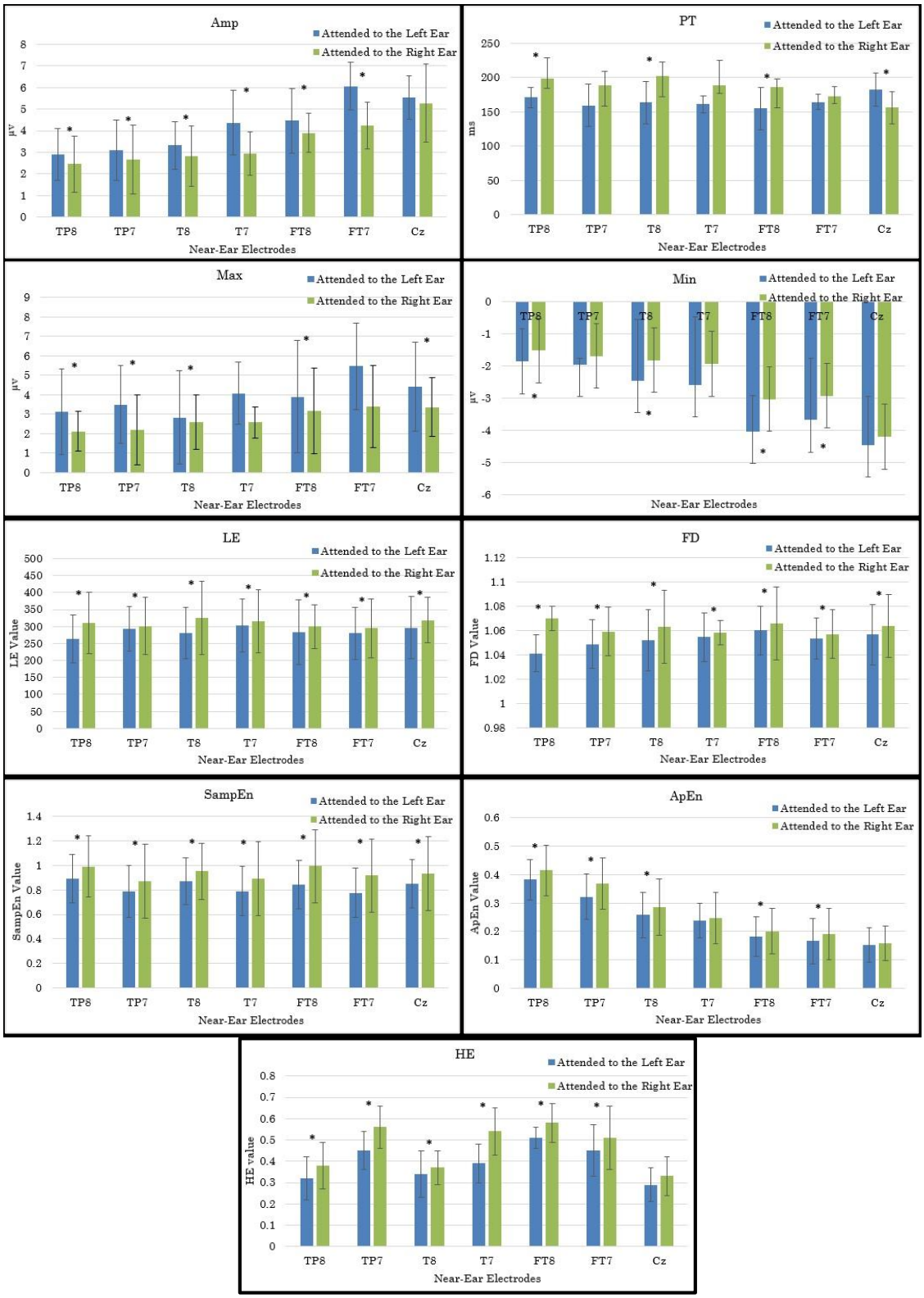

Figure 10. The statistical results (i.e., mean±standard deviation) obtained from 6 near-ear electrodes for nine features extracted from the MMN component during the attention to the left- and right-ear stimuli. The symbol "*" denotes significant differences between the states of two auditory attention tasks.

As compared with the classification performance of 128 electrodes data, the performance of the classification results with 6 near-ear elactrodes is in an acceptable range. These results could be considered for using auditory attention detection in order to design neuro-steered hearing aids with attention detection capability at the right ear where our need to scalp EEG recording is removed.

This research could be considered as a case study with only two speakers which limits the generality of the results to be applied to a more realistic condition. Creating a real acoustic environment with more than two speakers is suggested as further future work. Also, an important aspect that remains unclear is whether the spatial location of speakers influences the classification accuracy. Furthermore, the role of working memory of subjects is unknown in the attention detection in a cocktail party scenario with multiple speakers. As a technical issue, handling the influence of working memory on AAD with CNN is impossible, since this kind of neural network considers only the current input for the classification. Nowadays, powerful and efficient deep hybrid neural networks, using CNN and the long-short term memory (LSTM), have been developed which could handle sequential data and memorize previous inputs due to their internal memory. As a future plan, the authors plan to implement the proposed AAD method with the CNN-LSTM model.

Currently, noninvasive EEG-based AAD methods are considered as important tools for improving hearing-aids. This means that AADs could be employed in the design of neuro-steered hearing prostheses to amplify the attended speech in a competitive talker scenario for HI subjects.

The previous studies have confirmed that MMN, as a small part of EEG, is affected by auditory attention. This has motivated us to develop a new AAD approach which is both accurate and optimal in the classification of two groups of subjects attending to the left- and right-ear stimuli. In this paper, the MMN data are extracted from the EEGs by the Tucker decomposition algorithm. For the first time, significant linear and nonlinear features are extracted from MMN to examine distinguishable differences between the two auditory attention task using the ANOVA test. Then, an optimal feature set is selected and applied to the CNN classifier to detect the attended and unattended stimuli.

Few studies have been conducted to process and analyze the simultaneous EEG and fMRI data using tensor decomposition method. This lies in the fact that this method is based on complex mathematical relations and, therefore, its implementation imposes a high computational load. The proposed AAD method solves this problem by employing fewer samples of EEG (i.e., ~1 s corresponding to the length of MMN) and smaller temporal resolutions in extracting MMN. In other words, compared to the conventional AAD methods which use approx. 60 s of EEG signals, the proposed algorithm requires only a small portion of EEG for the classification task, which makes it suitable for real-time tracking of attention.

The evaluation results of the current study showed that the CNN-based auditory attention classifier is capable of achieving high accuracy (98.21%) as compared with the baseline systems. Implemented as around-the-ear systems, the current technology of hearing-aids incorporates fewer electrodes to make the processing time as small as possible. To achieve this goal, EEGs of six nearest electrodes (i.e., TP8, TP7, T8, FT8, T7, and FT7) are considered in the extraction of optimal features to detect auditory attention. The classification performance (71.75%) shows that using only these electrodes in hearing-aids could be a promising strategy in detecting attentional behavior.

List of abbreviations

ASA = auditory scene analysis
AAD = auditory attention detection
BCI = brain-computer interface
EEG = electroencephalography
fMRI = functional magnetic resonance imaging
MEG = magnetoencephalography
ERPs = event-related potentials
MMN = mismatch negativity
ICA = independent component analysis
SVM = support vector machines

EMD = empirical decomposition
CNN = convolutional neural network
GUI = graphical user interface
PCA = principal component analysis
CPD = canonical polyadic decomposition
ApEn = Approximate entropy
SampEn = Sample entropy
FD = Fractal dimension

Conflict of Interest

The authors declare that there are no financial and/or personal relationships with other people or organizations that could inappropriately influence their work.

Authors' Contributions

Masoud Geravanchizadeh and Sahar Zakeri carried defined the aim of research and the design of experiment. SZ carried out the experiments. SZ participated in the design of the study and performed the statistical analysis. MG coordinate and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Bregman AS. Auditory scene analysis: The perceptual organization of sound. The Journal of the Acoustical Society of America 1994;95:1177. doi: 10.1121/1.408434.
2. Hausfeld L, Riecke L, Valente G, Formisano E. Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. NeuroImage 2018;181: 617-626.
3. Elhilali M, Xiang J, Shamma SA, Simon JZ. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biology 2009;7(6): e1000129. doi: 10.1371/journal.pbio.1000129
4. Kaya EM, Elhilali M. Investigating bottom-up auditory attention. Frontiers in Human Neuroscience 2014;8:327. doi: 10.3389/fnhum.2014.00327
5. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. The Journal of the Acoustical Society of America 1953;25(5):975-979. doi: 10.1121/1.1907229
6. Broadbent DE. Perception and communication. Elsevier, 2013.
7. Das N, Bertrand A, Francart T. EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. Journal of Neural Engineering 2018;15(6):066017. doi: 10.1088/1741-2552/aae0a6
8. Oberem J, Lawo V, Koch I, Fels J. Intentional switching in auditory selective attention: Exploring different binaural reproduction methods in an anechoic chamber. Acta Acustica United With Acustica 2014;100(6): 1139-1148. doi:10.3813/AAA.918793
9. Kallenberg M, Desain P, Gielen S. Auditory selective attention as a method for a brain computer interface. Masters Thesis, Radboud University Nijmegen, 2006.

10. Papanastasiou G, Drigas A, Skianis C, Lytras MJH. Brain computer interface based applications for training and rehabilitation of students with neurodevelopmental disorders. A literature review. Helyon 2020;6(9):e04250. doi:10.1016/j.heliyon.2020.e04250
11. Baek SC, Chung JH, Lim YJS. Implementation of an Online Auditory Attention Detection Model with Electroencephalography in a Dichotomous Listening Experiment. Sensors 2021;21(2):531. doi: 10.3390/s21020531
12. Enriquez-Geppert S, Huster RJ, Herrmann CS. EEG-neurofeedback as a tool to modulate cognition and behavior: a review tutorial. Frontiers in Human Neuroscience 2017;11:51. doi: 10.3389/fnhum.2017.00051
13. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 2012;485(7397):233. doi:10.1038/nature11020
14. Hindarto H, Sumarno S. Feature extraction of electroencephalography signals using fast fourier transform. CommIT Journal 2016;10(2):49-52. doi: 10.1109/ICSPS.2010.5555506
15. Peelen MV, Kastner S. Attention in the real world: toward understanding its neural basis. Trends in Cognitive Sciences 2014;18(5):242-250. doi: 10.1016/j.tics.2014.02.004
16. Liu Z, Ding L, He B. Integration of EEG/MEG with MRI and fMRI. IEEE Engineering in Medicine and Biology Magazine 2006;25(4):46-53.
17. Chan HL, Kuo PC, Cheng CY, Chen YS. Challenges and future perspectives on electroencephalogram-based biometrics in person recognition. Frontiers in neuroinformatics 2018;12:66. doi: 10.3389/fninf.2018.00066
18. Luck SJ. An introduction to the event-related potential technique. (Second Ed.) MIT press, 2014, 416 pp.
19. Vazquez-Marrufo M. Event-Related Potentials for the Study of Cognition. Event-Related Potentials and Evoked Potentials 2017;1. doi: http://dx.doi.org/10.5772/intechopen.69308
20. Zink R, Proesmans S, Bertrand A, Van Huffel S, De Vos M. Online detection of auditory attention with mobile EEG: closing the loop with neurofeedback. BioRxiv 2017; 218727. doi: 10.1101/218727
21. Sussman ES. A new view on the MMN and attention debate: the role of context in processing auditory events. Journal of Psychophysiology 2007;21(3-4):164-175.
22. Wu MCW, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. Annual review of neuroscience 2006;29:477-505. doi: 10.1146/annurev.neuro.29.051605.113024
23. Miran S, Akram S, Sheikhattar A, Simon JZ, Zhang T, Babadi B. Real-time tracking of selective auditory attention from M/EEG: A bayesian filtering approach. Frontiers in neuroscience 2018;12:262. doi: 10.3389/fnins.2018.00262
24. O'Sullivan J, Chen Z, Herrero J, McKhann GM, Sheth SA, Mehta AD, et al. Neural decoding of attentional selection in multi-speaker environments without access to clean sources. Journal of neural engineering 2017;14(5):056001. doi: 10.1088/1741-2552/aa7ab4
25. O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, et al. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cerebral cortex 2015;25(7):1697-1706. doi: https://doi.org/10.1093/cercor/bht355
26. Akram S, Presacco A, Simon JZ, Shamma SA, Babadi B. Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. NeuroImage 2016;124:906-917. doi: 10.1016/j.neuroimage.2015.09.048
27. Ding N, Simon JZ. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. Journal of neurophysiology 2012;107(1):78-89. doi: 10.1152/jn.00297.2011
28. Haghighi M, Moghadamfalahi M, Akcakaya M, Erdogmus D. EEG-assisted modulation of sound sources in the auditory scene. Biomedical signal processing and control 2018;39:263-270. doi: 10.1016/j.bspc.2017.08.008
29. Aroudi A, Doclo S. EEG-based auditory attention decoding using unprocessed binaural signals in reverberant and noisy conditions?. Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2017;484-488. doi: 10.1109/EMBC.2017.8036867

30. Zink R, Hunyadi B, Van Huffel S, De Vos M. Tensor-based classification of an auditory mobile BCI without a subject-specific calibration phase. Journal of neural engineering 2016;13(2):026005. doi: 10.1088/1741-2560/13/2/026005
31. Choi I, Rajaram S, Varghese LA, Shinn-Cunningham BG. Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. Frontiers in human neuroscience 2013;7:115. doi: 10.3389/fnhum.2013.00115
32. Wang YK, Jung TP, Chen SA, Huang CS, Lin CT. Tracking attention based on EEG spectrum. International Conference on Human-Computer Interaction 2013;450-454. doi: 10.1007/978-3-642-39473-7_90
33. Alirezaei M, Sardouie SH. Detection of Human Attention Using EEG Signals. 24th National and 2nd International Iranian Conference on Biomedical Engineering 2017;1-5. doi: 10.1109/ICBME.2017.8430244
34. Looney D, Park C, Xia Y, Kidmose P, Ungstrup M, Mandic DP. Towards estimating selective auditory attention from EEG using a novel time-frequency-synchronisation framework. The International Joint Conference on Neural Networks 2010;1-5. doi: 10.1109/IJCNN.2010.5596618
35. Lu Y, Wang M, Zhang Q, Han Y. Identification of auditory object-specific attention from single-trial electroencephalogram signals via entropy measures and machine learning. Entropy 2018;20(5):386. doi: 10.3390/e20050386
36. Akram S, de Cheveigné A, Diehl PU, Graber E, Graversen C, Hjortkjaer J, et al. Telluride Decoding Toolbox. 2015. Available from: http://www.ine-web.org/software/decoding
37. Power AJ, Foxe JJ, Forde EJ, Reilly RB, Lalor EC. At what time is the cocktail party? A late locus of selective attention to natural speech. European Journal of Neuroscience 2012;35(9):1497-1503. doi: 10.1111/j.1460-9568.2012.08060.x
38. Cichocki A, Zdunek R, Phan AH, Amari SI. Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. John Wiley & Sons, 2009.
39. Kroonenberg PM. Applied multiway data analysis. John Wiley & Sons, 2008.
40. Tucker LR. Some mathematical notes on three-mode factor analysis. Psychometrika 1966;31(3):279-311. doi: 10.1007/BF02289464
41. Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics 1927;6(1-4):164-189. doi: https://doi.org/10.1002/sapm192761164
42. Kolda TG, Bader BW. Tensor decompositions and applications. Society for Industrial and Applied Mathematics review 2009;51(3):455-500. doi: 10.1137/07070111X
43. Cichocki A, Mandic D, De Lathauwer L, Zhou G, Zhao Q, Caiafa C, et al. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE signal processing magazine 2015;32(2):145-163. doi: 10.1109/MSP.2013.2297439
44. Zhou G, Cichocki A. Fast and unique Tucker decompositions via multiway blind source separation. Bulletin of the Polish Academy of Sciences. Technical Sciences 2012;60(3):389-405. doi: 10.2478/v10175-012-0051-4
45. Cong F, Lin QH, Kuang LD, Gong XF, Astikainen P, Ristaniemi T. Tensor decomposition of EEG signals: a brief review. Journal of neuroscience methods 2015;248:59-69.
46. da Silva FL, Pijn JP, Boeijinga P. Interdependence of EEG signals: linear vs. nonlinear associations and the significance of time delays and phase shifts. Brain topography 1989;2(1):9-18. doi: 10.1007/BF01128839
47. Miltner WH, Braun C, Arnold M, Witte H, Taub E. Coherence of gamma-band EEG activity as a basis for associative learning. Nature 1999;397(6718):434-436. doi: 10.1038/17126
48. Ke Y, Chen L, Fu L, Jia Y, Li P, Zhao X, et al. Visual attention recognition based on nonlinear dynamical parameters of EEG. Bio-medical materials and engineering 2014;24(1):349-355. doi: 10.3233/BME-130817
49. Wu SD, Wu CW, Lin SG, Wang CC, Lee KY. Time series analysis using composite multiscale entropy. Entropy 2013;15(3):1069-1084. doi: 10.3390/e15031069

50. Burioka N, Miyata M, Cornélissen G, Halberg F, Takeshima T, Kaplan DT, et al. Approximate entropy in the electroencephalogram during wake and sleep. Clinical EEG and neuroscience 2005;36(1):21-24. doi: 10.1177%2F155005940503600106
51. Finotello F, Scarpa F, Zanon M. EEG signal features extraction based on fractal dimension. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2015;4154-4157. doi: 10.1109/EMBC.2015.7319309
52. Katz MJ. Fractals and the analysis of waveforms. Computers in biology and medicine 1988;18(3):145-156. doi: 10.1016/0010-4825(88)90041-8
53. Natarajan K, Acharya R, Alias F, Tiboleng T, Puthusserypady SK. Nonlinear analysis of EEG signals at different mental states. Biomedical engineering online 2004;3(1):1-11. doi: 10.1186/1475-925X-3-7
54. Güler NF, Übeyli ED, Güler I. Recurrent neural networks employing Lyapunov exponents for EEG signals classification. Expert systems with applications 2005;29(3):506-514. doi: 10.1016/j.eswa.2005.04.011
55. Murugavel AM, Ramakrishnan S, BalasamyK, Gopalakrishnan T. Lyapunov features based EEG signal classification by multi-class SVM. World Congress on Information and Communication Technologies 2011;197-201. doi: 10.1109/WICT.2011.6141243
56. Bologna G. A Simple Convolutional Neural Network with Rule Extraction. Applied Sciences 2019;9(12):2411. doi: https://doi.org/10.3390/app9122411
57. Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG proceedings: health care and life sciences 2010;19:67. Available from: https://pdfs.semanticscholar.org/d1e5/c3097daf99db2c8dce3ac0edc3c5ade41460.pdf
58. Wang R, Wang J, Yu H, Wei X, Yang C, Deng B. Power spectral density and coherence analysis of Alzheimer's EEG. Cognitive neurodynamics 2015;9(3):291-304. doi: https://doi.org/10.1007/s11571-014-9325-x
59. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods 2004;134(1):9-21. doi: 10.1016/j.jneumeth.2003.10.009
60. Javitt DC, Lee M, Kantrowitz JT, Martinez A. Mismatch negativity as a biomarker of theta band oscillatory dysfunction in schizophrenia. Journal of schizophrenia research 2018;191:51-60. doi: https://doi.org/10.1016/j.schres.2017.06.023