EEG-based Speech Activity Detection

Marianna Koctúrová and Jozef Juhár

Department of Electronics and Multimedia Communications,
Faculty of Electrical Engineering and Informatics, Technical University of Košice,
Letná 9, 042 00 Košice, Slovak Republic
marianna.kocturova@tuke.sk, jozef.juhar@tuke.sk

Abstract: The brain-computer interface is one of the most up-to-date communication options. The advances made in this area open up opportunities to help mentally or physically disadvantaged people. The brain-computer interface offers the possibility of re-acquiring communication skills by deaf individuals. Electroencephalography (EEG) based speech recognition is, therefore, a novel research topic, which is an important component in communication technologies. In this article, we propose a speech activity detector algorithm, which, as expected, should improve the performance of the EEG based speech recognition system. EEG data uploaded while pronouncing 50 different phrases were classified using a feed-forward neural network. As a result of detection, a 0.82 F1 score was achieved.

Keywords: electroencephalography; speech activity detection; EEG-based speech recognition; brain-computer interface

1 Introduction

Electroencephalography (EEG) is a method for measuring brain electrical signals on the head surface. The advantage of EEG is that it is a non-invasive technique that facilitates the use of the brain-computer interface (BCI) technology without implantation of electrodes by neurosurgery. We are convinced that speech investigation from the brain's electrical impulses leads to the BCI communication enablement [1].

Speech activity detection from EEG signals is based on the speech information look up in the brain's electrical signals. EEG devices can capture part of these signals on the head surface. However, they are weakened, degraded, or mixed with several sources or artifacts after crossing the skull [2]. Another problem of searching for such a speech signal from EEG is that we are still unable to record a clear EEG signal that consists solely of speech information. The presence of EEG signals from different sources, which are formed and transmitted simultaneously, makes speech signal detection very difficult. Although the EEG subject focuses

only on producing speech during the scanning of his brain, the recorded EEG is still a mixture of impulses from many sources.

Speaking is a complex process that requires the involvement of several brain areas and articulatory organs to create a specific sound. Verbal language is created in the brain for several hundred milliseconds before the speech. A study [3] suggests that the brain needs an average of 600 ms to produce a word. Words and sentences include several kinds of abstract information that are lexical, grammatical, phonological, and graphic information. These components are stored in brain speech centers. Before the word is formed, the individual components are linked together and sent information about the articulation to the motor center, which controls the correct movement of the articulatory organs. Because speech is represented in the human brain as a cluster of information that is transmitted by nerve cells by electrical impulses, we can investigate the speech from the nerve perspective using the brain-computer interface [4].

Our speech recognition system from EEG signals was discussed in [5]. The study described the effort to find a suitable speech recognition algorithm. The experiment consisted of EEG signals processing from 10 subjects that read 50 different phrases. When recognizing these 50 phrases using the Hidden Markov model, the best result was achieved when training a single subject model at 53.4% accuracy. The cross-subject experiment showed a significantly dropped accuracy of recognition to 6%. Based on the results of the experiment we started to create an algorithm for speech activity detection, an analogy to Voice Activity Detector used in ASR, which could help to achieve higher results of speech recognition from EEG signals. An appropriately designed speech activity detector from EEG signals could be an important part of the EEG based speech recognizer.

Speech activity detection from EEG, and generally from BCI, has so far been very poorly reviewed. In addition to many BCI-based speech recognition kinds of research, speech activity detection from brain waves is still a very little solved and published problem, although it is very close to this field.

One of the few BCI speech activity detection studies has been presented in [6]. Functional Near-Infrared Spectroscopy (fNIRS) signals to detect speech from the brain were used. The study was executed on normal audible speech, silent speech, imagery speech. The result of the study was an F1 score of 0.7. This study confirms that speech can also be detected using BCI devices, which can greatly help develop speech recognition in the BCI field.

Another study discusses speech activity detection from EEG, suggests the use of such a system to improve classical speech recognition in a noisy environment. The study [7] demonstrates adding EEG-based speech activity detection can improve the performance of a voice activity detector that could be used to detect an acoustic noisy signal. The use of an EEG speech activity detector can also help predict whether a person wants to speak or not.

EEG based speech research contributes significantly to CogInfoCom discipline. Speech detection from EEG signals can greatly aid in the development of new forms of communication, combining computer science and cognitive science. [8].

In this paper, we present an EEG-based speech activity detection algorithm and a comparison of different EEG signal processing methods and their impact on improving the detection of speech in the EEG signal. The proposed algorithm was evaluated and tested on a pattern recognition neural network with a speech EEG corpus with a larger dictionary. Speech activity detection searches for a brain signal pattern that is generated before the word is formed or when it is spoken. The speech activity detector can help improve the performance of the aforementioned EEG-based speech recognition by eliminating false alarms in word recognition.

2 Methodology

2.1 EEG Recording

The spread of electrical pulse from the brain through the skull and scalp causes the electrical signal captured on the surface of the head to be attenuated, distorted, and dispersed. EEG signals on the scalp are diffused because the secondary electrical currents are spreading between different mediums of the head such as cerebrospinal fluid, skull, and scalp tissues, which have noticeably different conductivities. The lower SNR in the EEG is justified by the fact that the distribution of electrical potential on the head surface is more dispersed [9]. Due to the diffusing of the signal on the scalp, we used the electrode placement over the entire head and record signals from more scalp areas.

2.2 Brain Waves

In the published experiment [5] it was stated that the results of speech recognition were better in dividing the signal into wave frequencies. In this experiment, it has been shown that potentially the greatest amount of speech information is carried by the beta alpha and gamma waves. From the study [10] we can assume that different frequencies carry different types of information. Slow waves were more represented in simpler or calmer brain processes. Faster waves have been observed in processes requiring concentration and greater cognitive activity. When designing the EEG signal processing methods, we were inspired by these findings. In the experiment, we used the decomposition of EEG signals on 5 frequency bands. If the different frequency components are responsible for different brain activities, we can assume that the speech activity we are looking for will be easier or more clearly detectable in the distributed EEG waveform [11].

3 Experiment

In the experiment, we designed an algorithm for speech activity detection from EEG signals. In the signal processing part, we divided the experiment into two parts: in the first one, we focused on signal processing from the perspective of frequency waves. In the second part, we decided to describe the EEG signal from the cepstral region. Both of these signal processing approaches were finally trained and tested using a neural network and the results of such processing compared. Figure 1 shows a block diagram of a speech activity detection experiment.

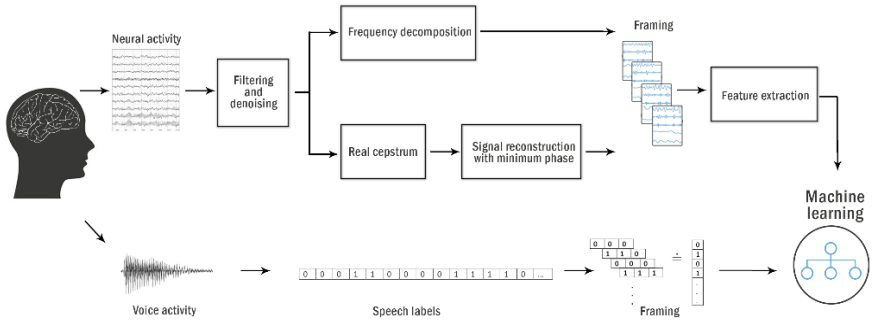

Figure 1
Block diagram of speech activity detector. Signal processing as performed in two separate parts: frequency decomposition and reconstruction from real cepstrum. These parts were considered separately.

3.1 Data Acquisition

EEG signals were recorded with a 16-channel OpenBci Ultracortex Mark III EEG cap [12]. The EEG device is based on sensing brain signals using dry electrodes configured according to the 10/20 system [13] [14]. The EEG device is mobile and the signal is transmitted from the head surface to the computer via Bluetooth. It captures the brain waves with 250 Hz sampling frequency. For recording brain signals, this device incorporates 16 channels namely Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2 according to the international 10/20 system as shown in Figure 2, where two reference electrodes are positioned above the ears. For signal recording, open-source graphic user interface OpenBCI GUI compatible with the used EEG cap was used [12]. All signals were recorded to format .txt, which simplifies the work of further signal processing.

The participant of the experiment to which EEG was measured was right-handed speaking Slovak, without any reported neurological disease. The participant followed the pre-agreed instructions during the recording. He was sitting motionless and tried to focus solely on speech processes. After starting the EEG recording, the participant read loudly phrases showed him on single charts one by one.

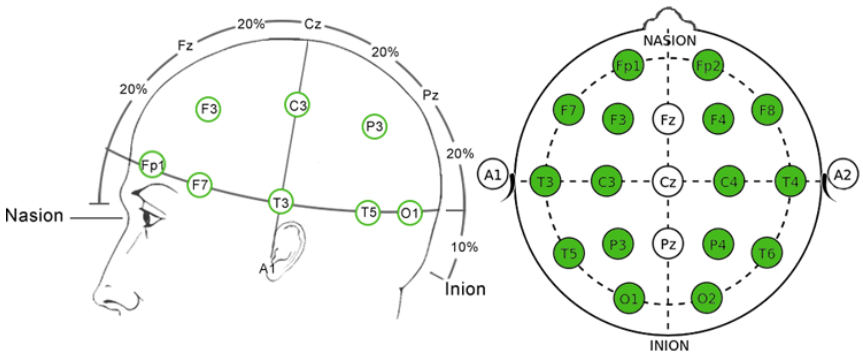

Figure 2
Electrode placement according to 10/20 international system

A dictionary created for the experiment contains 50 various phrases, there were 14 one-word phrases, 20 two-word phrases, and 16 three-word phrases. The EEG signals were recorded in 9 sessions. To create speech labels, we also recorded an audio signal while recording EEGs.

3.2 Signal Preprocessing

The EEG signal was browsed to remove excessive noise at first. Evident noise was found mainly in the sections at the beginning and the end of the sessions. These parts were manually removed from the EEG as well as from audio recordings. In Figure 3 we see noise-infected parts of the EEG recording of one session.

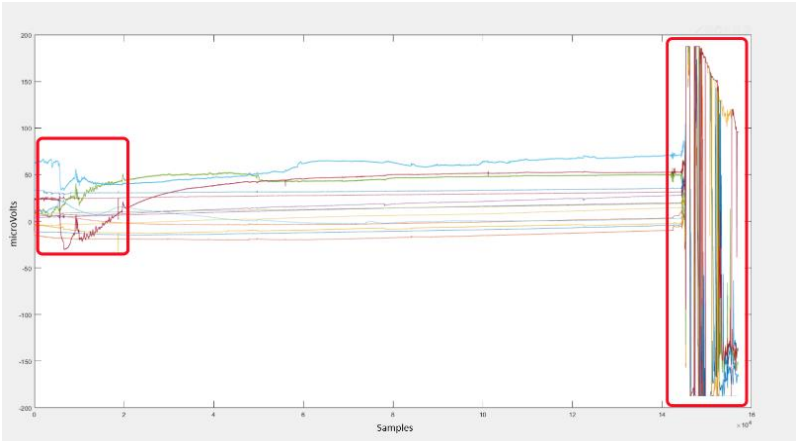

Figure 3
Example of noise in the EEG signal of one session

During EEG recording, the signal is mixed with artifacts. Some types of artifacts can be prevented by using appropriate methods to record brain activity, but obtaining a perfect signal without artifacts is not yet possible.

An artifact of power line interference is commonly found in the EEG signal. It has a frequency of 50 Hz, and it is easy to remove using frequency filtering [1]. For this purpose Butterworth filter was used because it has it an extremely flat frequency characteristic in passband range, and compared to other filters it contains lower passband ripples [15] [16]. Unlike the experiments [5] [17], other artifacts filtering was omitted, which could cause potential loss of speech information in the EEG signal.

3.3 Frequency Decomposition of the EEG Signal

The EEG signal from each channel into 5 frequency components in the first part of the experiment was divided. These separate components were combined into a matrix with a width of 16x5. Figure 4 shows a portion of the signal waveform from channel F7 in separate frequency bands.

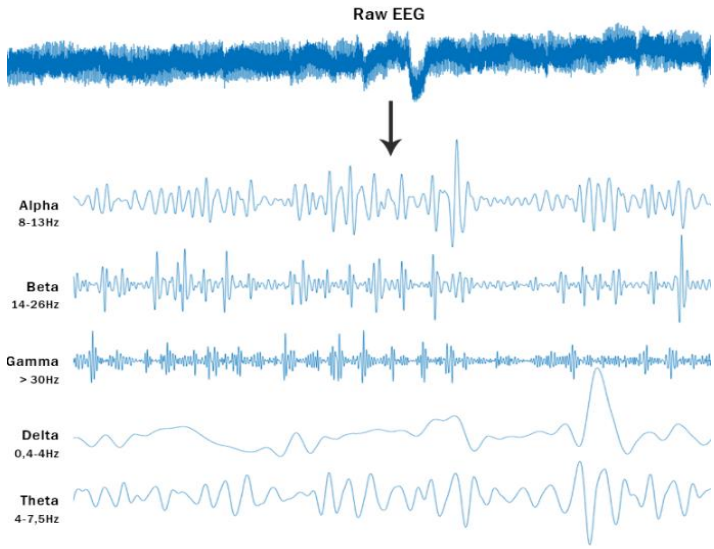

Figure 4
EEG signal decomposition into five frequency bands according to normal brain rhythms

3.4 Signal Reconstruction from Real Cepstrum

The second part of the experiment consists of a cepstral reconstruction signal. Real cepstrum was calculated from the EEG signal for each channel. The reconstructed signal had a minimum phase property [18]. Figure 5 is the

reconstructed EEG signal in comparison with the original raw EEG signal from channel F7.

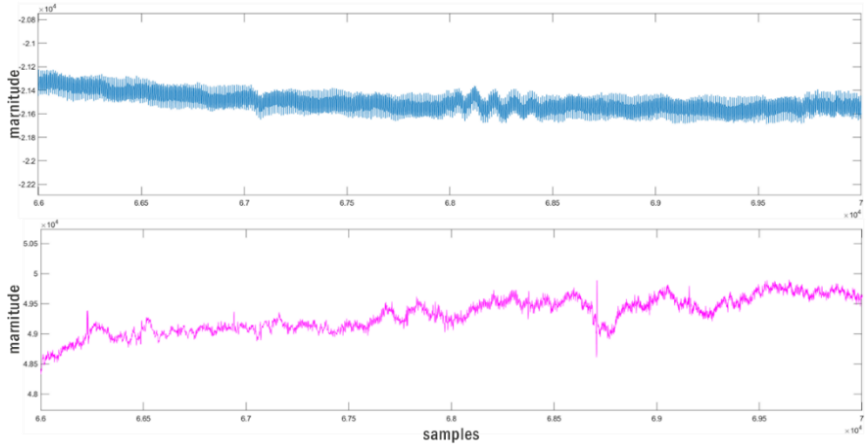

Figure 5
EEG signal after reconstruction. Blueshape- raw EEG signal. Magenta shape – reconstructed EEG signal with minimum phase

3.4 Feature Extraction

The reconstructed EEG signal was split into 240 ms frames with a 50% overlap. For each frame following features were calculated. The set of features was extracted from the frequency and time domains. Table 1 shows the features extracted from the EEG signal. The feature selection suitable for speech activity detection was influenced by the studies [19] [20]. The publication [19] describes methods of selecting features for describing EEG seizures. The study used the feature used in ASR from an audio signal.

The *mean* value of the frame is an average value of signal partis. A *standard deviation* is a measure of the amount of variation or dispersion of a set of values. The *skewness* factor indicates the measured lack of symmetry of the distribution. We can say that the data set is symmetrical when it looks equal on the right and left sides of a given central point. Using the *kurtosis* coefficient we determine whether the data are peaked or flat in comparate to the normal distribution. The average *band power* summarizes the contribution of the frequency band to the overall power of the signal [21]. When dealing with information content, the *Shannon entropy* is often considered as the foundational and most natural one. Entropy, regarded as a measure of uncertainty, is the most paradigmatic example of these information quantifiers. [22]

Table 1

Features extracted from the EEG signal and their descriptions.

μ is the mean of the signal x(n), δ is the standard deviation of the signal x(n), E represents statistical expectation, x_i is the coefficient of the signal in an orthonormal basis.

Feature	Description	Domain	Equation
Mean	Mean value of a local	Time	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
STD	Standard deviation	Time	$\delta = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$
Skewness	Local skewness in a frame	Time	$skew = E\left[\left(\frac{(x(n)-\mu)}{\sigma}\right)^3\right]$
Kurtosis	Local kurtosis in a frame	Time	$kurt = E\left[\left(\frac{(x(n)-\mu)}{\sigma}\right)^4\right]$
RMS	Root mean square	Time	$rms = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert^2}$
Band power	The average band power	Frequency	$P = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\lvert x(t)\rvert^2\,dt$
Shannon entropy	Amount of information in a variable	Information theory	$Shannon = \sum_{i=1}^{n} x_i \log_2 x_i$

Based on the set of features, EEG signals were described. Features were calculated for each signal channel from 60 samples, which at a sample rate of 250 hz represents a frame of 240 ms. Hence, overlap was set to 50%, the features were recalculated for the 30 previous samples. These 7 feature values calculated for each recording channels were merged into a row matrix of shape (112.1). In the case of the decomposed signal, it was the matrix of shape (560.1), since we calculated the symptoms for 16 channels divided into 80 waves.

3.5 Speech Labels

Speech and non-speech segments in EEG data were labeled manually. Finding speech activity in EEG data is not possible due to the complexity of the signal. Therefore, audio was recorded and synchronized with EEG data. Speech labels were labeled from the audio signal, which is a simple task.

For this purpose, we utilized the tool Transcriber for segmenting, labeling, and transcribing speech [23]. The audio record was divided into time segments. Each

segment by the start and end times of the pronounced word or the silence was bound.

The proposed EEG speech activity detection model consisted of EEG data recorded during the speech activity generation. The speech of a subject consisting of 50 different phrases has been tagged with values 1 and 0 indicating segments of speech activity or non-speech activity of brain waves.

3.6 Neural Network Training

The EEG signal described by selected features was fed to the input layer to the neural network. We have compared two neural network models with two different signal processing approaches. The first was the frequency decomposition of the EEG signal into 5 waveforms of normal brain rhythms. The second one consisted of signal reconstruction from real cepstrum.

In the experiment, the 2-layer feed-forward neural network was used. Various numbers of hidden neurons were tested. The best results were achieved with 60 neurons on a single hidden layer. The neural network consist of a single tanh activated hidden and binary output sigmoid activated output layer. We trained the network with Scaled conjugate gradient backpropagation, with binary cross-entropy as the loss function. Output pseudo probabilities were thresholded with a 0.5 decision boundary.

4 Results

The proposed EEG speech activity detection model consisted of EEG data recorded during the speech activity generation. The speech of a subject consisting of 50 different phrases has been tagged with values 1 and 0 indicating segments of speech activity or non-speech activity of brain waves.

In the experiment, we tried to get the best possible result of speech activity detection. Therefore, we compared different approaches to EEG signal processing to find information about speech using the Neuron network.

In the baseline experiment published in [16], the same method of Feed-forward neural network training was used. The EEG signals were processed in different methods. The best result reported for the F1 (F-speech) score was 0.77. Despite relatively good F1 results, the trained model had a problem with the prediction of zero segments (non-speech).

Table 2 shows the results of speech activity detection by the neural network in comparison with various signal processing approaches. These results were achieved on the same EEG database. We thought that decomposing the EEG signal into different frequency bands could result in an improvement. However,

the detection results achieved by this approach were comparable to those of the preceding experiment.

The speech activity detection best result with signal reconstruction with minimum phase was achieved. The F1 score value was 0.82. An improvement of 0.05 absolute in the F1 score compared to the preceding experiment was achieved. The inclusion of feature extracting functions in signal processing has yielded a better analysis of the speech EEG signal.

Table 2

Features extracted from the EEG signal and their descriptions

	Accuracy	**Precision**	**Recall**	**F1 score**
Baseline experiment	0.45	0.69	0.87	**0.77**
Wave decomposition	0.67	0.72	0.82	**0.77**
Reconstruction with minimum phase	0.77	0.79	0.84	**0.82**

Figure 6 shows an example of a graphical representation of speech activity detection output compared to input targets. The blue shape shows input targets indicating speech activity and the red shape shows output segments of speech activity predicted by our proposed algorithm.

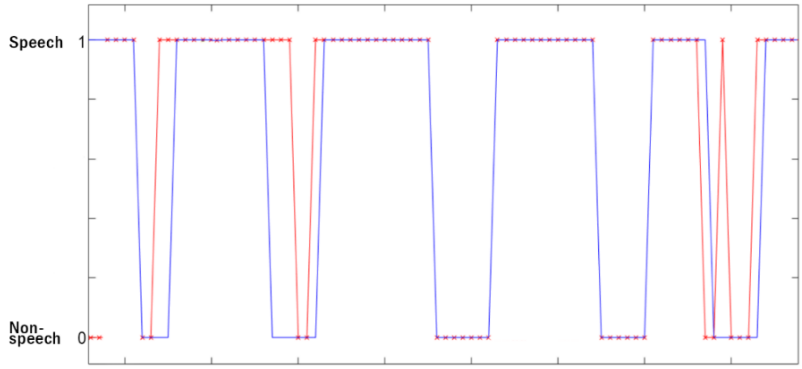

Figure 6

A portion of output targets in comparison with input labels. The blue shape represents input segmentation. Red crossed shaper represents ground truth

Conclusions

This study examined speech activity in EEG signals. The results show that speech can be detected using EEG technology. A significant difference in EEG signal processing has been demonstrated in creating a speech detector model using a reconstructed minimum phase signal. From the results, we can assume that the cepstral area of EEG signals contains a greater amount of speech information. In this study, we used, among other methods, the method of reconstructing the EEG

signal from real cepstrum with minimum phase, which has not yet been investigated in a relation to finding a speech pattern.

The results obtained in this paper can, in collaboration with other research dealing with speech and brain signals, contribute to solving the problem of EEG speech recognition. The method of signal processing based on cepstral reconstruction brought interesting results in our research. Although research has focused on speech detection, signal processing methods can be used and investigated in other research related to EEG and speech, such as [24] .

In our future work, we would like to explore this area of EEG signal processing. In the course of further research, we would like to execute also more experiments with speech detection using RNN and CNN models, which could deliver better results in this field.

Acknowledgment

The research in this paper was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the project KEGA 009TUKE-4/2019 and by the Slovak Research and Development Agency under the projects APVV-15-0731 and APVV-15-0517.

References

[1] J. Wolpaw and E. W. Wolpaw, Brain-computer interfaces: principlesand practice. OUP USA, 2012

[2] Ray, Amit, and Naoum P. Issa.:EEG 11, Understanding Epilepsy: A Study Guide for the Boards 1, 2019, p. 203

[3] Sahin, Ned T., et al.: Sequential processing of lexical, grammatical, and phonological information within Broca's area, Science 326.5951, 2009, 445-449

[4] Wang, M., Chen, Y., and Schiller, N. O.: Lexico-syntactic features areactivated but not selected in bare noun production: Electrophysiologicalevidence from overt picture naming, Cortex, 2018

[5] Rosinová, Marianna, et al.: Voice command recognition using EEG signals, 2017 International Symposium ELMAR. IEEE, 2017

[6] Herff, C., et al. "Self-paced BCI with NIRS based on speech activity." International BCI Meeting, 2013

[7] Krishna, Gautam, et al.:Voice Activity Detection in presence of background noise using EEG. arXiv preprint arXiv:1911.04261, 2019

[8] Baranyi, Péter, Adam Csapo, and Gyula Sallai. Cognitive Infocommunications (CogInfoCom) Springer, 2015

[9] Destoky, Florian, et al: Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope, Neuroimage 184, 2019, 201-213

[10] Jones, Stephanie R.: When brain rhythms aren't 'rhythmic': implication for their mechanisms and meaning, Current Opinion in Neurobiology 40, 2016, pp. 72-80

[11] Sanei, Saeid, and Jonathon A. Chambers: EEG signal processing. John Wiley & Sons, 2013

[12] Tools, OpenBCI-Open Source Biosensing. "Openbci. com. Retrieved 24 February 2018" 2018

[13] Jurcak, V., Tsuzuki, D., and Dan, I.,: 10/20, 10/10, and 10/5 systemsrevisited: their validity as relative head-surface-based positioning sys-tems, Neuroimage, Vol. 34, No. 4, 2007, pp. 1600-1611

[14] Oostenveld, Robert, and Peter Praamstra.: The five percent electrode system for high-resolution EEG and ERP measurements. Clinical neurophysiology 112.4 2001, pp. 713-719

[15] Widmann, Andreas, Erich Schröger, and Burkhard Maess. "Digital filter design for electrophysiological data–a practical approach." Journal of neuroscience methods 250, 2015, pp. 34-46

[16] R. M. Rangayyan: Biomedical signal analysis, Vol. 33, John Wiley &Sons, 2015

[17] Koctúrová, Marianna and Juhár, Jozef: Speech Activity Detection from EEG using a feed-forward neural network, 10th IEEE International Conference on Cognitive Infocommunications Proceedings, 2019, pp. 147-152

[18] Pei, S-C., and H-S. Lin.: Minimum-phase FIR filter design using real cepstrum. IEEE Transactions on Circuits and Systems II: Express Briefs, 2006, pp. 1113-1117

[19] Temko, A., et al.: Speech recognition features for EEG signal description in detection of neonatal seizures, Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010

[20] Subasi, Abdulhamit. Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques: A MATLAB Based Approach. Academic Press, 2019

[21] Sanei, Saeid.: Adaptive processing of brain signals. John Wiley & Sons, 2013

[22] Rosso, Osvaldo A., Raydonal Ospina, and Alejandro C. Frery. "Classification and verification of handwritten signatures with time causal information theory quantifiers." PloS one 11.12, 2016

[23] Geoffrois, Edouard, et al.: Transcribing with Annotation Graphs, LREC, 2000

[24] Kovács, Annamária, István Winkler, and Klára Vicsi. "EEG correlates of speech: Examination of event related potentials elicited by phoneme classes." 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2017