

Auditory evoked potential EEG-Biometric dataset

Nibras Abo Alzahab , Angelo Di Iorio , Luca Apollonio , Muaaz Alshalak , Alessandro Gravina , Luca Antognoli , Marco Baldi , Lorenzo Scalise , Bilal Alchalabi 

Published: Dec. 1, 2021. Version: 1.0.0

When using this resource, please cite: ([show more options](#))

Abo Alzahab, N., Di Iorio, A., Apollonio, L., Alshalak, M., Gravina, A., Antognoli, L., Baldi, M., Scalise, L., & Alchalabi, B. (2021). Auditory evoked potential EEG-Biometric dataset (version 1.0.0). *PhysioNet*. <https://doi.org/10.13026/ps31-fc50>.

Please include the standard citation for PhysioNet: ([show more options](#))

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.

Abstract

This data set consists of over 240 two-minute EEG recordings obtained from 20 volunteers. Resting-state and auditory stimuli experiments are included in the data. The goal is to develop an EEG-based Biometric system.

The data includes resting-state EEG signals in both cases: eyes open and eyes closed. The auditory stimuli part consists of six experiments; Three with in-ear auditory stimuli and another three with bone-conducting auditory stimuli. The three stimuli for each case are a native song, a non-native song, and neutral music.

Background

The recoded dataset is multipurpose. It is helpful for simple analysis of resting-state, comparison between resting state with eyes closed and with eyes open, studying the effect of auditory stimuli in terms of both methods of conduction by determining the neuronal mechanisms that follow the auditory stimulation and analysis of how different languages may affect the behaviour of human evoked potentials.

Additionally, it could be used for a simple EEG-based user authentication system.

The data was recorded at Marche Polytechnic University (UNIVPM) - Faculty of Engineering.

Recording Tools

- OpenBCI Ganglion Board, 200 Hz sampling rate, four channels: T7, F8, Cz, and P4.
- Gold Cup Electrodes.
- Ten20 Conductive Paste.
- Software: OpenBCI GUI · v5.0.3

Methods

Determination of eligibility

The subjects filled out a questionnaire that contains the subjects' information. Additionally, the subjects read and agreed to a detailed informed consent.

Preparation and installation of equipment

The experiments start with the installation of the equipment needed. This includes four gold-cup electrodes with Ten20 Conductive Paste on the scalp. The electrodes were placed on the T7, F8, Cz, and P4 positions according to the 10/10 international EEG system [1], which were chosen according to these publications [2-4]. Additionally, two electrodes were placed in the left and right ears as a reference electrode and ground electrode, respectively. Once all of the electrodes were installed, a simple calibration was performed to ensure that everything worked correctly. The calibration included checking the connectivity of the electrodes by measuring the skin

impedance. The non-noisy two-minutes were segmented from the raw data.

Testing and data recording

The subject was asked to sit down and relax on a comfortable chair. The recording was performed in a single day per subject with the same order of tasks. A non-noisy two minutes recording was segmented from the raw data. The recording involves the acquisition of the electroencephalographic signal as follows:

1. Three minutes of resting-state, eyes open for three sessions.
2. Three minutes of resting-state, eyes closed for three sessions.
3. *Non-Related experiment (Not provided in the dataset).*
Three minutes of resting-state, eyes open for three sessions using noise isolation headset.
4. *Non-Related experiment (Not provided in the dataset).*
Three minutes of resting-state, eyes open for three sessions using noise isolation headset.
5. Three minutes of listening to a song in their native language using in-ear headphones.
6. Three minutes of listening to a song in a non-native language using in-ear headphones.
7. Three minutes of listening to neutral music using in-ear headphones.
8. Three minutes of listening to a song in their native language using bone-conducting headphones.
9. Three minutes of listening to a song in a non-native language using bone-conducting headphones.
10. Three minutes of listening to neutral music using bone-conducting headphones.

Note 1: If the person is Italian: the Arabic song was used for the non-Native song.

Note 2: If the person is not Italian: the Italian song was used for the non-Native song.

Note 3: Neutral music is a musical genre that emphasizes tranquillity, relaxation, and peaceful soundscapes. It is typically composed of instrumental music (music with no or wordless vocals), and it may involve acoustic instruments, electric synthesizers, and even recorded nature sounds.

Data Description

The Data are provided in .csv and WFDB formats. The recordings are named as follows:

sXX_exXX_sXX: s(Subject Number)_ex(Experiment Number)_s(Session Number. Just for ex01 and ex02).

Example: s03_ex02_s01: Subject 03 Experiment 02 Session.

The Data is provided in three folders:

1. **Raw Data:** Each record contains the dataset as recorded without segmenting and without filtering.
 2. **Segmented Data:** Each record contains two minutes of the EEG signals having the lowest possible noise, i.e. the data when the subject was moving and blinking their eyes less. Data were segmented manually, and the initial and final positions of the trim points are presented in the attached excel file: data_trim.csv.
 3. **Filtered Data:** Each record contains two minutes of the EEG signals which includes the segmented data after applying a 1st order 1-40 Hz Butterworth filter and a 50 Hz notch filter with a quality factor of 30.
- *Data_trim.csv:* Contains the initial and final point where the Segmented Data was trimmed from the Raw Data. The points are provided in seconds and samples.
 - *Subjects.csv:* Contains the metadata of the subjects and the questionnaire's answers.
 - *Songs.csv:* Contains links for the songs used in the recordings

Note 1: Each file in filtered and segmented data contains five columns: Sample index and four EEG channels.

Note 2: Each file in the raw data contains additional columns that can be discarded, but it was kept for the purpose of providing raw recording files. The important columns are the first five: Sample index and four EEG channels.

Note 3: For the comments in *Subjects.csv*: file. Each subject was asked whether he is fluent in the non-native language.

Usage Notes

The data was used in a Master's thesis at UNIVPM titled: Design and Implementation of Techniques for the Secure Authentication of Users Based on Electroencephalogram (EEG) Signals.

The reuse of the dataset involves, but is not limited to, studying the effect of the different auditory stimuli on EEG signals of both

methods of conduction by determining the neuronal mechanisms that follow the auditory stimulus and analysis of how different languages may affect the behaviour of human evoked potentials. Additionally, studying the resting state differences between eyes open and eyes closed states, and comparing resting state with auditory stimuli cases.

The limitation of the data is the low number of electrodes and subjects.

Acknowledgements

This data set was created and contributed to PhysioNet by Nibras Abo Alzahab and his colleagues, Angelo Di Iorio, Luca Apollonio, Muaaz Alshalak, Alessandro Gravina, Luca Antognoli, Lorenzo Scalise, Marco Baldi and Bilal Alchalabi.

The recordings were conducted at Marche Polytechnic University (UNIVPM) - Faculty of Engineering: Via Brecce Bianche, 12, 60131 Ancona AN, Italy. The subjects were asked to read and sign a detailed informed consent. The experiments were conducted in compliance with the World Medical Association (WMA) Declaration of Helsinki [5]. All the individually identifiable health information of the subjects and their identity are protected and can not be accessed with non-authorised users in order to preserve privacy for research participants.

Conflicts of Interest






The author(s) have no conflicts of interest to declare

References

1. Jurcak V, Tsuzuki D, Dan I. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*. 2007 Feb 15;34(4):1600-11.
2. Altaht SH. Robust EEG Channel Set for Biometric Application (Doctoral dissertation, University of Canberra).
3. Ravi KV, Palaniappan R. A minimal channel set for individual identification with EEG biometric using genetic algorithm. In *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)* 2007 Dec 13 (Vol. 2, pp. 328-332). IEEE.
4. Marcel S, Millán JD. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE transactions on pattern analysis and machine intelligence*. 2007 Feb 20;29(4):743-52.
5. World Medical Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*. 2001;79(4):373.

Contents

Share

Access

Access Policy:
Anyone can access the files, as long as they conform to the terms of the specified license.

License (for files):
[PhysioNet Contributor Review Health Data License 1.5.0](#)

Data Use Agreement:
[PhysioNet Contributor Review Health Data Use Agreement 1.5.0](#)

Discovery

DOI (version 1.0.0):<https://doi.org/10.13026/ps31-fc50>**DOI (latest version):**<https://doi.org/10.13026/s5s6-a546>**Topics:**[eeg](#)[biometric](#)[electroencephalogram](#)[auditory stimuli](#)[resting-state](#)**Corresponding Author**

You must be logged in to view the contact information.

Files



















Total uncompressed size: 2.9 GB.

Access the files

- [Download the ZIP file](#) (786.1 MB)
- Download the files using your terminal: `wget -r -N -c -np https://physionet.org/files/auditory-eeg/1.0.0/`
- Download the files using AWS command line tools: `aws s3 sync s3://physionet-open/auditory-eeg/1.0.0/ DESTINATION`

[Visualize waveforms](#)

Folder Navigation: <base>

Name		Size	Modified
 Filtered_Data			
 Raw_Data			
 Segmented_Data			
 WFDB_Files			
 Data_trim.csv		8.4 KB	2021-06-28
 LICENSE.txt		14.4 KB	2021-11-30
 README.txt		7.8 KB	2021-11-08
 RECORDS		24.4 KB	2021-11-16
 SHA256SUMS.txt		211.2 KB	2021-12-01
 Songs.csv		876 B	2021-11-30
 Subjects.csv		1.1 KB	2021-11-29

PhysioNet is a repository of freely-available medical research data, managed by the MIT Laboratory for Computational Physiology.

Supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH grant number R01EB030362.

For more accessibility options, see the [MIT Accessibility Page](#).

[Back to top](#)