Psychoacoustic Annoyance and the EEG

A Deep Learning Approach

Jonas Egeler

TH Ingolstadt — Labor für Akustik, 85049 Ingolstadt, Deutschland, E-Mail: joe4206@thi.de

Introduction

Jury testing and the application of psychoacoustic metrics to sound pressure data are two established methods for the assessment of sound quality. So far, little is known to what extend the results of these methods correlate with biosignals, especially the electrical activity of the brain.

A growing number of manufacturers offers wearable biosensors at a relatively low cost. In this paper, it will be discussed whether the annoyance of a particular sound (Zwicker's psychoacoustic annoyance index PA) can be predicted based on measurements with semi-professional EEG, ECG and EDA sensors. An auditory experiment with six participants was conducted to collect data. Eight airplane ambience sounds of $180\,s$ length were presented to the test persons. EEG, ECG and EDA data was recorded synchronous with the acoustic stimulation. A deep learning model based on wavelet power and wavelet entropy signal features combined with a CNN was implemented and trained with the recorded data.

The validation results show that the model performs well in the classification and regression for data of individual experimental runs, but is not able to generalise between multiple experimental runs or multiple subjects.

Related Research

Various studies investigated the effects of sound exposure on the EEG. Li et al. examined the relationship between EEG variation and subjective annoyance under noise exposure [3]. They presented white noise and pure tones of $160\,Hz$, $500\,Hz$ and $4000\,Hz$ at $70\,dB(A)$ SPL. The study found the sum of the relative alpha and theta power in the frontal brain region to be most discriminative for long time noise exposure ($5\,min$). For short-time exposure ($5\,s$), the study found no regularities.

In a recent study, Chen et al. examined correlations between Zwicker's metric for psychoacoustic annoyance and EEG signal parameters for the acoustic environment of a high-speed passenger train [4]. The EEG of 30 participants was measured under ten operating conditions ranging from $250\,\frac{km}{h}$ to $350\,\frac{km}{h}$ with one electrode placed at the frontal lobe. Time of noise exposure was $5\,min$ for each condition. Subjective annoyance of the noise was recorded with a questionnaire and was found to correlate with Zwicker's psychoacoustic annoyance metric which is given by [5]:

$$PA = N_5 \left(1 + \sqrt{(\omega_S)^2 + (\omega_{FR})^2}\right). \qquad (1)$$

N_5: percentile loudness

S, F, R: sharpness, fluctuation strength, roughness

ω_S: $\omega_S = \left(\frac{S}{acum} - 1.75\right) \cdot 0.25 \lg\left(\frac{N_5}{sone} + 10\right)$

ω_{FR}: $\omega_{FR} = \frac{2.18}{(N_5/sone)^{0.4}} \left(0.4\frac{F}{vacil} + 0.6\frac{R}{asper}\right)$.

The results of the experiment show that EEG alpha and theta band powers increase with rising psychoacoustic annoyance, while beta power decreases. Based on these findings an annoyance index calculated based on band power features is proposed as an electrophysiological metric for perceived annoyance. This index is reported to predict psychoacoustic annoyance with 93.30% accuracy in a k-NN machine learning model.

The results of Chen et al. indicate that the EEG is indeed suited for the task of acoustic annoyance recognition based on physiological features. Recent advances in EEG emotion recognition show, however, that EEG band power might not be the most discriminative feature for EEG classification tasks and that approaches with more refined features like time-frequency information or spectral entropy as inputs for deep neural networks are more sensitive to the subtle changes in brain potentials [9].

Heart rate variability (HRV), the variance of the inter-beat intervals, and electrodermal activity (EDA), as measured by the skin conductivity, are further physiological markers which are considered to be potential indicators for noise-induced stress. Walker et al. [1] found a negative correlation between HRV and noise exposure, in a study of Darvishi et al. [7], a significant linear relationship between noise SPL and EDA was found.

In a prequel study to this project, an EEG experiment with 11 participants was conducted. One hundred machine sounds and nature sounds of $4\,s$ length were presented to the test persons. An eight channel EEG headset was used to record the EEG data synchronous with the acoustic stimulation. PA was calculated from sound presssure measurements with a binaural headset. Multiple methods for EEG signal classification were tested. The analysis of the ERP (event related potential) showed a stronger N1 and P1 reaction for sounds with high PA values than for sounds with low PA values ($p = 0.11$). A deep learning approach for EEG classification proposed by Yang et al. [6] was implemented to learn the class affiliations $PA < \overline{PA}$ and $PA > \overline{PA}$ based on spectral entropy features. The model achieved a maximum validation accuracy of 58.8 % for one subject and a two-class distinction.

The overall poor classification results of the prequel study indicate that the stimulation time of $4\,s$ is suited for an ERP study design, but not for an analysis based on time-frequency EEG features, which is in congruence with the findings of Li et al. Therefore, an approach oriented by the experimental design of [4] based on a longer stimulation period is taken in this project. In addition to the EEG, EDA and ECG will be recorded. The investigation of inter-experiment and inter-subject prediction qualities of the proposed models will be a focal point of the study. As a benchmark model for EEG classification, a deep learning architecture inspired by Mokatren et al. [8] will be implemented. The architecture combines both spectral energy and spectral entropy signal features in 3D tensors and feeds it into a CNN. It yields a cross-subject two-class validation accuracy of 91.06 % for the DEAP dataset. Data normalisation and data augmentation methods will be employed, in order to improve the generalisation capability of the model.

Experiment Setup

The test group for the experiment consisted of six participants (five males and one female). The individual recording sessions were all conducted in the acoustic semi-free field room of THI in order to reduce interfering noise to a minimum. An Electro-Voice SX 300 speaker and a Synq Audio Digit 3K6 amplifier were used for the playback of the sounds. The participants were seated in a distance $d = 2.85\,m$ away from the speaker. A HEAD acoustics BHS I headset was used to measure the sound pressure at a position slightly behind the participant's ear in order to record the effective physical stimulation. The signal of the BHS I headset and a trigger signal marking stimulus onset were recorded simultaneously with a four channel HEAD acoustics SQuadriga recording system.

An Empatica E4 wristband was used to measure HR, HRV and EDA. The EEG was recorded with a g.tec g.GAMMAcap with g.SCARABEO wet electrodes and a g.Nautilus Multi-Purpose eight channel biosignal amplifier. The placement of the ground electrode, the reference electrode and the eight active electrodes is shown in figure 1. The EEG data was recorded at a sample rate of 500 Hz. A Butterworth bandpass filter ($[0.1, 100]\,Hz$) was applied to the data directly while recording. The complete experimental setup is shown in figure 2.

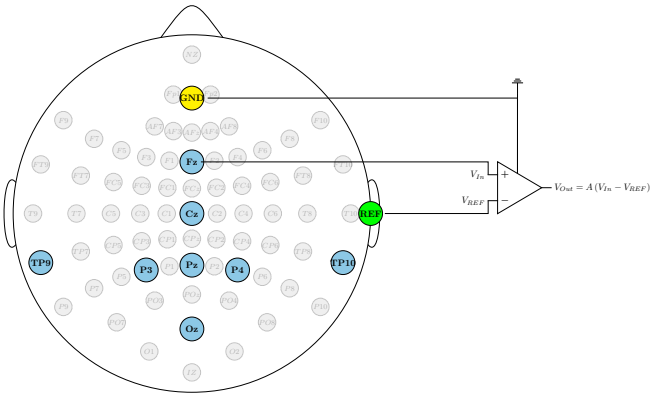

Figure 1: Electrode placement

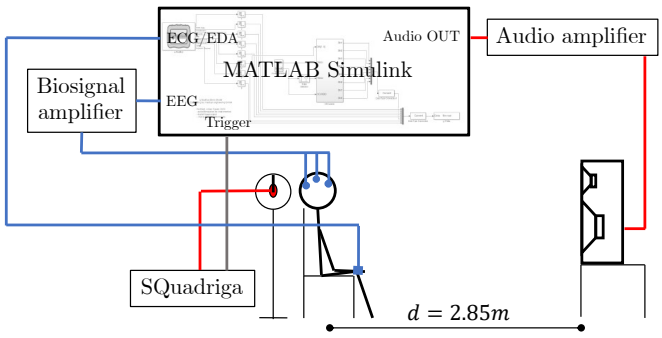

Figure 2: Experimental setup

Sounds and Paradigm

Due to its relevance for passenger comfort, the soundscape of a commercial airplane was chosen for the study. A set of eight stationary turbine sounds with 3 min length was presented to the participants. The sounds were synthesised in Ableton Live from real-life cabin interior recordings and simulations of airplane turbine sounds at different fan rotation speeds [1]. In figure 3, the statistical distribution of the psychoacoustic parameters of the sound set is visualised. It can be seen that the sounds were selected to have a uniformly distributed PA.

In order to ensure the comparability of the listening situations, the level of the sound source was calibrated at the beginning of each recording session with a test tone (sine, frequency 1 kHz, amplitude 0.5) for the sound pressure level at the position of the participant's ear to be 94 dB. The PA value for each trial was calculated from the sound pressure measurements of the binaural headset in MATLAB.

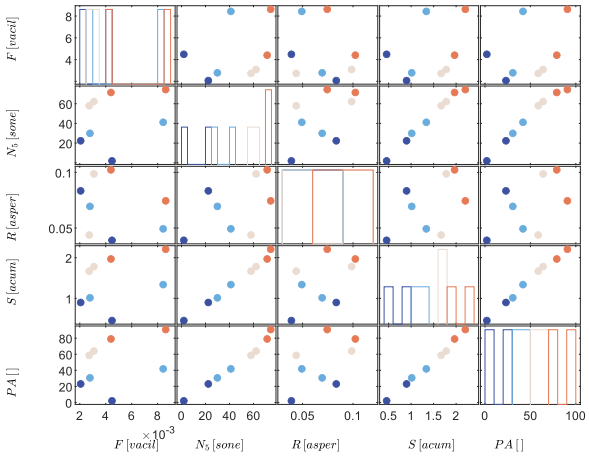

Figure 3: Psychoacoustic properties of the sound set

The experimental paradigm is shown in figure 4. At the beginning of each trial, a short beep sound (sine, 1 kHz, 0.1 s) is played back as an indicator for the participants to close their eyes and freeze all movements. The beep is followed by a 30 s period to record the baseline signal. After that, the sound stimulus with 180 s length is played back. After each trial, a pause of 60 s is planned, to avoid fatigue. This sequence is repeated eight times in total to present the whole set of eight sounds.

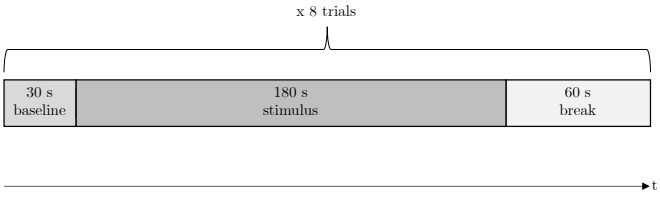

Figure 4: Experimental paradigm

Data Preprocessing

Data preprocessing was performed with MATLAB and the EEGLAB toolbox[2]. The EEG signals were bandpass-filtered ($[0.1, 100]\,Hz$) and 50 Hz line noise was removed via narrow-band filtering. The EEG signals were re-referenced to the common average. The physiological signals were upsampled to match the sample rate of the EEG recordings and the tonic component of the EDA was removed via the cvxEDA algorithm[3]. With help of the trigger signal, the preprocessed data was cut into the the eight trials.

[1] https://www.boomlibrary.com/sound-effects/turbine/
[2] https://sccn.ucsd.edu/eeglab/index.php
[3] doi:10.1109/TBME.2015.2474131

Data Analysis

Subjective Response and PA

In the prequel study, the averaged subjective responses on percieved annoyance showed a strong correlation ($r = 0.81$) with Zwicker's PA metric. Therefore, PA was used as a label for the classification task and as the target value for the regression task.

3DWPWH + CNN Model

The model is partially adapted from [8] and combines wavelet power (WP) and wavelet entropy (WH) features of the EEG signals and relative mean (RM) and entropy (H) features of the PHYSIO signals in 3D tensors. Spatial information about the EEG electrode positions is encoded in the mapping of the values into the tensors. A convolutional neural network (CNN) is used to extract spatio-temporal features from the input tensors, in order to learn a mapping between the physiological features and the PA. Both a classification and a regression variant of the model were implemented. Various normalisation and data augmentation methods were applied to improve the generalisation capability of the model. The preprocessed EEG data is further processed as follows:

1. EEG time window cutting: The signals ${}^{tr}_{ch}x$ of each of the eight trials $tr \in [1, ..., 8]$ and eight EEG channels $ch \in [1, ..., 8]$ ($[FP1, ..., O2]$) are cut into a baseline signal ${}^{tr}_{ch}x^{baseline}$ ($8\,s$ after beep to $1\,s$ before stimulation) and 30 stimulation epochs ${}^{tr}_{ch}x^{ep}, ep \in [1, ..., 30]$ of $3\,s$ length.

2. PHYSIO time window cutting: The signals ${}^{tr}_{pch}x$ of each of the eight trials $tr \in [1, ..., 8]$ and three physiological signal channels $pch \in [1, ..., 3]$ is cut into a baseline signal ${}^{tr}_{pch}x^{baseline}$ ($8\,s$ after beep to $1\,s$ before stimulation) and 30 stimulation epochs ${}^{tr}_{pch}x^{ep}$, $ep \in [1, ..., 30]$ of $3\,s$ length.

3. The data is augmented by noise addition (3 copies + 1/f filtered Gaussian noise, $SNR = 10\,dB$).

4. All signals are downsampled to $f_s = 125\,Hz$.

5. Discrete wavelet decomposition: A wavelet decomposition of the signals ${}^{tr}_{ch}x^{ep}$ is performed in order to obtain the wavelet detail coefficients ${}^{tr}_{ch}x^{ep}_{lv}$, $lv \in [1, ..., 5]$ ($[\delta, ..., \gamma]$). The decomposition level is set to 4 for the decomposition frequency ranges to roughly match the EEG bands of interest.

6. EEG feature calculation: For each trial, each channel, each decomposition level (alias frequency band) and each epoch, the relative wavelet power (WP) is calculated by taking the squared RMS value of the wavelet detail coefficients ${}^{tr}_{ch}x^{ep}_{lv}$ and normalising it with the total energy of the respective epoch:

$$ {}^{tr}_{ch}WP^{ep}_{lv} = \frac{\left[\mathrm{rms}\left({}^{tr}_{ch}x^{ep}_{lv}\right)\right]^2}{\left[\mathrm{rms}\left(\sum_{lv=1}^{5} {}^{tr}_{ch}x^{ep}_{lv}\right)\right]^2}. \qquad (2) $$

   Wavelet entropy (WH) is calculated by taking the entropy of the epoch signal and normalising it with the entropy of the respective baseline signal:

$$ {}^{tr}_{ch}WH^{ep}_{lv} = \frac{\log_2\left(\left[\mathrm{var}\left({}^{tr}_{ch}x^{ep}_{lv}\right)\right]^2\right)}{\log_2\left(\left[\mathrm{var}\left({}^{tr}_{ch}x^{baseline}_{lv}\right)\right]^2\right)}, \qquad (3) $$

7. PHYSIO feature calculation: For each trial and each physio channel, the relative mean (RM) is calculated by dividing the epoch mean by the respective baseline mean:

$$ {}^{tr}_{pch}RM^{ep} = \frac{\mathrm{mean}\left({}^{tr}_{pch}x^{ep}\right)}{\mathrm{mean}\left({}^{tr}_{pch}x^{baseline}\right)}. \qquad (4) $$

   Entropy (H) is calculated by taking the entropy of the epoch signal and normalising it with the entropy of the respective baseline signal:

$$ {}^{tr}_{pch}H^{ep} = \frac{\log_2\left(\left[\mathrm{var}\left({}^{tr}_{pch}x^{ep}\right)\right]^2\right)}{\log_2\left(\left[\mathrm{var}\left({}^{tr}_{pch}x^{baseline}\right)\right]^2\right)}. \qquad (5) $$

8. 2D topographic mapping: In order to include information about the relative electrode position in the model, the calculated EEG WP and WH features are mapped to 9×9 matrices according to the following scheme:

$$ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Fp1 & 0 & Fp2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & T7 & 0 & C3 & 0 & C4 & 0 & T8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & O1 & 0 & O2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (6) $$

   The missing values (zero entries) are interpolated by inverse distance weighted (IDW) interpolation.

9. The PHYSIO RM and H features are mapped to the centre of the matrices:

$$ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & PHYSIO & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (7) $$

10. Matrix stacking: For every trial-epoch combination, a feature tensor is created by stacking the five WP matrices (one for each decomposition level), the five WH matrices (one for each decomposition level), the three RM matrices (one for each PHYSIO channel) and the three H matrices (one for each PHYSIO channel). The final feature tensors have a size of $9 \times 9 \times 16$.

11. Z-score normalisation of all feature tensors along the epoch dimension.

The whole processing pipeline is summarized in figure 5.

The 3D input tensors can be interpreted as colour images with 16 colour channels. Convolutional neural networks (CNN) have proven to be good at extracting meaningful patterns from such data structures. Mokatren et al. propose a CNN with the following architecture to extract information from the input tensors [8]: (1) Input layer ($9 \times 9 \times 16$) | (2) 2D convolutional layer 1 (32 filters, kernel size 3×3) | (3) Batch normalization layer 1 | (4) ReLU layer 1 | (5) Max-pooling layer 1 (pool size 2×2) | (6) 2D convolutional layer 2 (64 filters, kernel size 3×3) | (7) Batch normalization layer 2 | (8) ReLU layer 2 | (9) Max-pooling layer 2 (pool size 2×2) | (10) Dropout layer 1 (dropout rate 0.45) | (11) Fully connected layer 1 (64 neurons) | (12) ReLU layer 3 | (13) Dropout layer 2 (dropout rate 0.2) || (14 classification) Fully connected layer 2 (8 neurons) | (15 classification) Softmax classification layer (8 output classes) || (14 regression) Fully connected layer 2 (1 neuron) | (15 regression) Regression layer (1 output value).

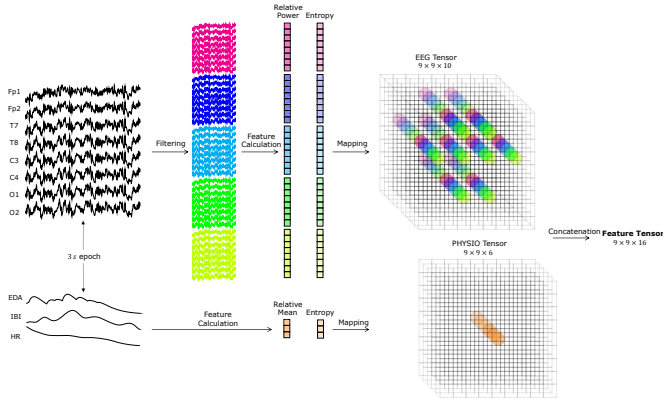

Figure 5: 3DWPWH: feature tensor calculation

The training of both the classification and the regression model was performed in MATLAB. The training parameters for were: 'sgdm', 'MiniBatchSize', 128, 'MaxEpochs', 100, 'InitialLearnRate', 1e-4, 'LearnRateSchedule', 'piecewise', 'LearnRateDropFactor', 0.1, 'LearnRateDropPeriod', 20, 'Shuffle', 'every-epoch'.

Levels of Generalisation

The recorded data allows for analyses at different levels of generalisation:

1. Individual level: The feature tensors generated from the individual recording sessions were split into 70 % training data and 30 % test data. 25 % of the training data were held out for model validation during training. The average accuracy for classification and average RMSE for regression is given in the results.
2. Intra-individual level: The feature tensors generated from three consecutive recordings of subject s01 were used for training. Holdout validation was performed with the feature tensors of one complete recording (all eight trials), to examine the intra-individual generalisation capability of the model. The average accuracy and average RMSE across all possible training-holdout permutations is given in the results.
3. Inter-individual analysis: The feature tensors generated from six recordings with different subjects were used for training. Holdout validation was performed with the feature tensors of the recording of one individual subject which were not used for the training, in order to examine the inter-individual generalisation capability of the model.

Results and Discussion

The average classification accuracy and the average regression RMSE across all permutations of training and holdout data at the different levels of generalisation are summarized in table 1.

Table 1: Results

Generalisation Level	∅ Validation Accuracy	∅ Validation RMSE
Individual	98.6 %	6.0
Intra-individual	12.7 %	37.7
Inter-individual	13.8 %	53.2

It can be seen that the classification and regression performance of the proposed model is high when part of the data of the same subject or experimental run which was used for the training is also used for the validation of the model. This result strenghtens the finding in [4], that it is possible to predict the PA based on EEG features and non-parametric machine learning methods with classification accuracies up to 93 % using a simple train-test split validation.

A question that was neglected in [4] and was addressed here, is the generalisation capability of the models. Looking at the results, it becomes obvious that as soon as the model is validated with completely new data (data from experimental runs or subjects which were held out from the training), both the classification and regression performance drop to the range of the guessing probability (accuracy 12.5 % or RMSE 35.5). The model could not find experiment-overarching coherences between the calculated signal features and the the PA.

Some issues that came up during the evaluation phase of this project and could be taken as starting points for future research are: focus on data quality (movement artifacts and low SNR aggravated the extraction of meaningful signal features), higher data quantity (to mitigate model overfitting), reduced complexity of the auditory stimuli (focus on individual psychoacoustic parameters), randomization of the order of the sound presentation (to prevent the model from learning temporal trends) and the implementation of results from auditory attention decoding research (e.g. correlation-based auditory attention techniques).

Acknowledgment

I would like to thank all participants in the study. Also, I would like to thank the HEAD acoustics company for providing a license of their Artemis SUITE software. Most notably, I would like to thank Prof. Dr.-Ing. Jörg Bienert and Dr.-Ing. Xiaojing Zhao for giving me support in the lab as well as guidance and feedback throughout this project.

References

[1] Erica D. Walker, Anthony Brammer, Martin G. Cherniack, Francine Laden, and Jennifer M. Cavallari. Cardiovascular and stress responses to short-term noise exposures-a panel study in healthy males. Environmental research, 150: 391–397, 2016. doi: 10.1016/j.envres.2016.06.016.

[2] M. Trimmel, M. Kundi, G. Binder, E. Groll-Knapp, and M. Haider. Combined effects of mental load and background noise on cns activity indicated by brain dc potentials. Environment International, 22(1):83–92, 1996. doi: 10.1016/0160-4120(95)00106-9.

[3] Zheng-Guang Li, Guo-Qing Di, and Li Jia. Relationship between electroencephalogram variation and subjective annoyance under noise exposure. Applied Acoustics, 75:37–42, 2014. doi: 10.1016/j.apacoust.2013.06.011.

[4] Xieqi Chen, Jianhui Lin, Hang Jin, Yan Huang, and Zechao Liu. The psychoacoustics annoyance research based on eeg rhythms for passengers in highspeed railway. Applied Acoustics, 171:107575, 2021. doi: 10.1016/j.apacoust.2020.107575.

[5] Hugo Fastl and Eberhard Zwicker. Psychoacoustics: Facts and models, volume 22 of Springer series in information sciences. Springer, Berlin, 2007. doi: 10.1007/978-3-540-68888-4.

[6] Yilong Yang, Qingfeng Wu, Yazhen Fu, and Xiaowei Chen. Continuous convolutional neural network with 3d input for eeg-based emotion recognition. In Neural Information Processing, pages 433–443, Cham, 2018. Springer International Publishing.

[7] Ebrahim Darvishi, Rostam Golmohammadi, Javad Faradmal, Jalal Poorolajal, and Mohsen Aliabadi. Psychophysiological responses to medium levels of occupational noise: An exposure–response relationships. Acoustics Australia, 47(3): 217–228, 2019. doi: 10.1007/s40857-019-00159-0.

[8] Lubna Shibly Mokatren, Rashid Ansari, Ahmet Enis Cetin, Alex D. Leow, Olusola A. Ajilore, Heide Klumpp, and Fatos T. Yarman Vural. Eeg classification by factoring in sensor spatial configuration. IEEE Access, 9:19053–19065, 2021. doi: 10.1109/ACCESS.2021.3054670.

[9] Alexander Craik, Yongtian He, and Jose L. Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. Journal of neural engineering, 16(3):031001, 2019. doi: 10.1088/1741-2552/ab0ab5.