Did You Hear That? Detecting Auditory Events with EEGNet

George Ramzi[1], Ian McLoughlin[2], and Ramaswamy Palaniappan[3]

***Abstract*— The behavioural nature of pure-tone audiometry (PTA) limits those who can participate in the test, and therefore those who can access accurate hearing threshold measurements. Event Related Potentials (ERPs) from brain signals has shown limited utility on adult subjects, and a neural response that can consistently be identified as a result of pure-tone auditory stimulus has yet to be identified. The in doing so challenge is worsened by the nature of PTA, where stimulus amplitude decrease to a patient's lower threshold of hearing. We investigate whether EEGNet, a compact Convolutional Neural Network, could help in this domain. We trained EEGNet on a dataset collected whilst patients underwent a test designed to mimic a pure-tone audiogram, then assessed EEGNet performance in the detection task. For comparison, we also trained Support Vector Machines (SVMs) and Common Spatial Patterns + Linear Discriminant Analysis (CSPLDA) on the same task, with the same training paradigms. The results show that EEGNet is capable of detecting hearing events with 81.5% accuracy on unseen participants, outperforming SVMs by just over 5%. Whilst EEGNet outperformed SVMs and CSPLDA, it did not, however, always show a statistically significant improvement. Further analysis of EEGNet predictions revealed that, with sufficient test repetition, EEGNet has the potential to accurately ascertain hearing thresholds. The implication of these results is for a brain-signal based hearing test for those with physical or mental disabilities that limit their participation in a PTA. While this research is promising, future research will be needed to address the complexity of test setup, the duration of testing, and to further improve accuracy.**

I. INTRODUCTION

Pure-Tone Audiometry (PTA) is a tool for the diagnosis and measurement of hearing acuity. Unlike some alternative methods of audiometry, PTA allows for the accurate diagnosis of frequency-specific hearing loss. This is important for the calibration of hearing aids, since modern devices allow for frequency-specific amplification to accommodate a patient's specific pattern of hearing loss.

A. *Pure-Tone audiometry*

The PTA procedure is straightforward and can be performed relatively quickly [1]. Patients are presented with a short tonal stimulus to one of their ears, and are asked to press a response button when they perceive a tone. Should they respond, the volume is decreased and the test repeated until they can no longer detect the tone. Stimuli may then be repeated to confirm that it was inaudible, and the amplitude recorded as the patient's hearing threshold for that frequency in that ear. The process repeats with a selection of tone frequencies, before being repeated with the alternate ear (for more detail on this, see Section II). The test has the benefits of being both frequency-specific and easy to explain to patients.

However, there are a number of reasons why a patient may not be able to participate in such a test; Notably, physical disabilities that limit the patient's ability to press a response button, psychological or cognitive disabilities that make following the instructions of the test difficult. Extremes of age may introduce both limitations. The Auditory Brainstem Response (ABR) is an auditory evoked response commonly used to diagnose hearing loss in infants [2], and is capable of estimating frequency specific hearing thresholds using brief tone "clicks" [3]; however the response diminishes rapidly with age and is most effective on infants [4]. Auditory Steady State Response (ASSR) has also been used to derive frequency specific thresholds, but using frequency modulated tones as opposed to pure tones [5]. Both ABR and ASSR are capable of use in adults, but hearing thresholds estimated through both methods differ from those measured through a standard behavioural PTA [6]. Notably, neither ABR nor ASSR use pure tones; both methods use a modified tone that changes the frequency specificity of the PTA. With this in mind, we aim to create a test that utilised ERPs, but based on unmodified pure-tones.

B. *EEGNet*

We investigated whether EEGNet, a Convolutional Neural Network designed for classifying EEG data with minimal hyperparameters, is capable of classifying evoked potentials from PTA-like auditory stimuli. For an overview of the base network and it's architecture, see [7]. EEGNet has seen several non-ERP applications including the detection of auditory attention [8]. It has also found use in a number of modified network designs, including classifying a motor imagery task using Temporally Constrained Sparse Group Lasso [9], detecting

[1]George Ramzi is with Department of Computing, University of Kent, Canterbury, UK G.B.Ramzi@outlook.com, formerly George Langroudi
[2]Ian McLoughlin is with the ICT Cluster, Singapore Institute of Technology, 10 Dover Drive, Singapore, ian.mcloughlin@singaporetech.edu.sg
[3]Ramaswamy Palaniappan is with Department of Computing, University of Kent, Canterbury, UK R.Palani@kent.ac.uk

SSVEP events using Filter Banks [10], the assessment of depression using Long-Short term memory [11], and detecting objects that were being observed in a Virtual Reality environment with Gated Recurrent Units [12]. Whilst these authors all compared their systems against an unmodified EEGNet it should be noted that the authors did not all note the increased number of parameters their modifications introduced (increasing the number of trainable parameters conflicts with one of EEGNet's original design intentions). In our system, the base EEGNet network is adopted, as shown in Fig 1. Inputs are 1 second of preprocessed 125Hz stacked 32-channel EEG data. An output fully connected layer and softmax function are added to provide a binary classification output. Training uses the Adam optimiser with an initial learning rate of 3e-03. All activations are ELU (exponential linear units), dropout is applied for both blocks, and batch normalisation between each layer are applied as per [7].

We note that comparison between publications to assess the usefulness of EEGNet in real-world applications is exceedingly limited across most sensory domains.

II. Data Collection

This work builds upon our previous research outlined in [13] with a similar process of data collection, as outlined below.

A. Tone Generation

As with our previous work, this paper uses our *pseudo*-audiometer system to generate tones. The design of *pseudo*-audiometer was designed to closely mimic the procedure of clinical audiometry, as outlined in [1]. This system generates tones from 40dB HL to -5dB HL, decreasing in 5dB increments. The tones were presented at 1000Hz, 2000Hz, 4000Hz, 8000Hz, 250Hz, and 500Hz in that order for a duration of between 1 and 3 seconds, with a sampling frequency of 48kHz. Generated tones are pure sine-waves, presented with a minimum of 4 seconds of silence between presentations.

B. Participants

After our experimental design was approved by the University of Kent Research Ethics Advisory Group (approval reference 0481819), we recruited 9 participants for the experiment (6=M), with a mean age of 38 ± 10.2. Instructions were provided both vocally and in writing, and all participants declared they were not aware of any pre-existing health or hearing conditions that may interfere with the experiment.

All participants participated in 8 trials spread over 2 days, 4 trials per day, with at least 5 minutes between trials for participants to relax. Participants were sat in a comfortable chair with their dominant arm resting comfortably, and their dominant hand used to press a response button during *active* trials, but not during *passive* trials. All tones were presented through Sennheiser HD 380 pro circumaural headphones. Subjects were isolated from visual or audible distraction.

C. Dataset

Data for all subject was collected using a StarStim R32 32-channel EEG kit sampling at 500Hz. Each trial contained exactly 60 tone presentations per ear. With 9 participants, each participating in 8 trials, our final dataset accumulated 8640 events, with half being *active* and half being *passive*. Each event consists of 1 second of post-stimulus activity. The montage used for recording contained the following electrode locations: P8, T8, CP6, FC6, F8, F4, C4, P4, AF4, Fp2, Fp1, AF3, Fz, FC2, Cz, CP2, PO3, O1, Oz, O2, PO4, Pz, CP1, FC1, P3, C3, F3, F7, FC5, CP5, T7, and P7.

Tone presentations that a participant responded to are designated as "positive" events, whilst tone presentations without response are "negative" events. Due to the PTA methodology, there are naturally far more positive events than negative ones. This would lead to an unbalanced data set, and hence we included a number of pre-stimulus events (1 second of data pre-stimulus, where no tone was presented). These are guaranteed to be true-negative events, since no tone was presented, and ensured a balanced data set for training. All data was band-pass filtered from 0.5Hz to 40Hz to remove noise, before being downsampled to 125Hz.

D. Cross Validation

For completeness, when assessing each classifier, we performed cross validation using 3 different methods.

- **K-folds** Used as a benchmark, K-Folds is very commonly used in machine learning and is helpful for making comparisons; However, it does not reflect any real-world scenario where such a system could be used clinically.
- **Leave-One-Participant-Out (LOPO)** This closely emulates a likely real-world scenario, in which a participant has never participated in the test before and therefore has no existing data to train with. We use this to assess how well the classifier generalises to unseen participants.
- **Leave-One-Trial-Out (LOTO)** This emulates a scenario in which a participant has participated in the test before and has previous test data saved in the training data. It is worth noting that this scenario, whilst useful for measuring and comparing classifiers, results in the classifier training on almost the entire dataset.

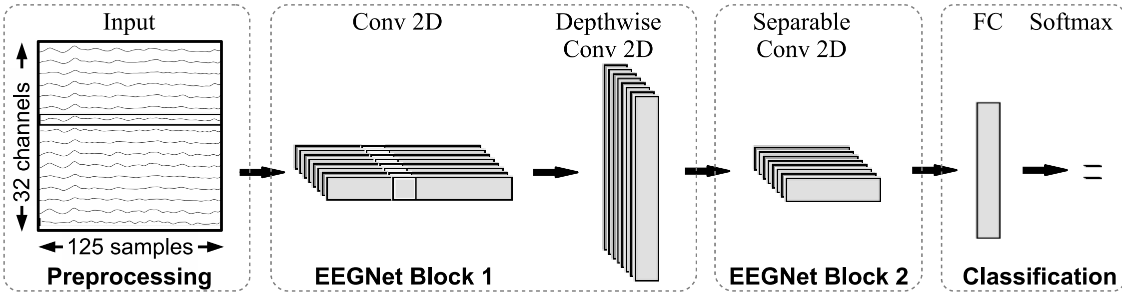

Fig. 1. An overview of the architecture incorporating EEGNet. Each convolution in the architecture operates over a single dimension, with convolutions operating either spatially or temporally.

E. *Classifiers*

Our goal was to assess how viable EEGNet is for the detection of auditory evoked potentials. To do this, we chose to compare against two other common EEG classifiers: Support Vector Machines (SVM), and Common Spatial Patterns paired with Linear Discriminant Analysis (CSPLDA). Both are well used for EEG classification tasks and BCI applications, and are also suited to live classification tasks [14].

Every effort was made to ensure the comparison against these classifiers was as fair as possible; each classifier was trained with the same volume of data, with identical splits in the dataset during cross validation.

It is worth noting that, unlike SVMs and CSPLDA, EEGNet benefits from having some data separated for "validation". This separate data is used to evaluate the network after each training epoch. During K-Folds, 1 fold was separated for use as validation data, in Leave-One-Participant-Out 1 participant was used for validation, and for Leave-One-Trial-Out 1 trial was used for validation. To ensure the selection was fair, for every cross-validation the participant/trial/fold used for validation changed, just as the one used for testing changed. For SVMs and CSP+LDA this validation data was simply included in the training data.

III. Results

After performing parameter optimisation on each classifier the optimal parameters were used to train and test each classifier, following the cross validations mentioned in Section II-D. The results are outlined in Table I and visualised in Fig 2.

Across all cross validation methods EEGNet outperformed the SVMs with a consistently higher accuracy. This may be due to the fact that EEGNet trained across all channels, whilst SVMs only trained across a single channel. We thus also trained an SVM for each paradigm on the entire dataset (named 'Flat'); since SVMs expect a single vector, all channels were simply appended one after the other. We see that the Flat SVMs performed better than SVMs trained on individual channels, but were still outperformed by EEGNet across all paradigms.

TABLE I

Average accuracy and standard deviation of cross validation results (best performance in bold). SVM scores are the average score of all channels. Flat, in this context, refers to SVMs which were trained with the signal from all channels. σ refers to the standard deviation of classifier performance across folds.

Paradigm	Classifier	Accuracy	σ
KFolds	LDA	0.5672	0.0128
	SVM	0.7239	0.1678
	SVM Flat	0.8093	0.0092
	EEGNet	**0.8267**	0.0143
LOPO	LDA	0.5411	0.0573
	SVM	0.7650	0.0332
	SVM Flat	0.7888	0.0287
	EEGNet	**0.8153**	0.0696
LOTO	LDA	0.5578	0.0597
	SVM	0.7707	0.0404
	SVM Flat	0.8038	0.0363
	EEGNet	**0.8107**	0.0651

IV. Discussion and conclusions

We found that EEGNet consistently outperformed both SVM's and CSPLDA. However, a Wilcoxon signed-rank test reveals that the improvement was not always statistically significant. To test the performance of EEGNet, we compared the results of EEGNet against the results of both CSPLDA and SVMs.

It's important to note that, due to the nature of pure-tone audiometry, the distribution of errors is just as important as accuracy; how errors are distributed directly affects how these classifiers could be used in real-world applications. As an example, if errors were distributed mainly around the subject's hearing threshold, the classifier would only be useful if the errors near the threshold were consistent across all subjects.

When examining the predictions made by EEGNet, the errors are well distributed across all volumes. Since the error distributed is even, test repetition can potentially improve threshold accuracy. In fact, if *pseudo-*

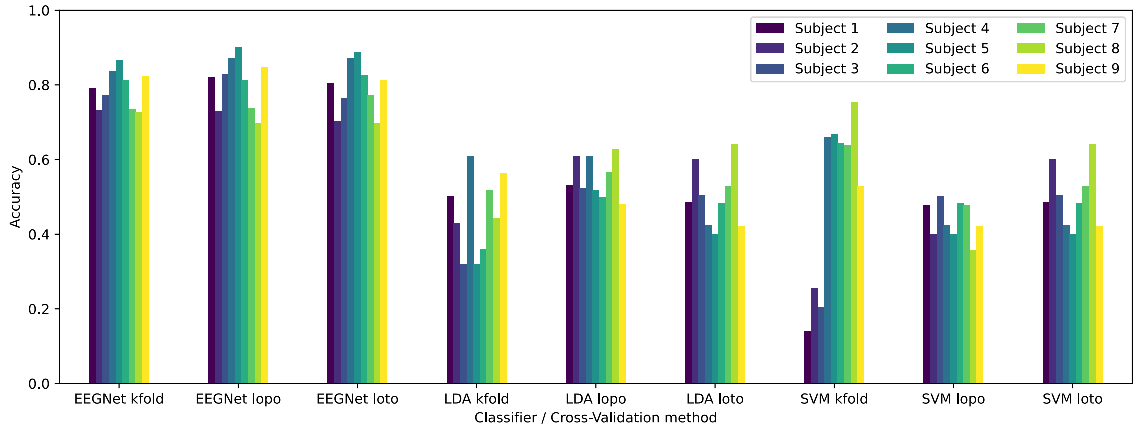

Fig. 2. The optimal accuracy for each classifier on individual subjects. From this, we note that LDA+CSP demonstrates less inter-subject variation than SVMs, while EEGNet has less overall and inter-subject variance than either SVM or CSPLDA. It should also be noted that LDA and SVMs classified some subjects with *below* chance accuracy, with overall classification accuracy being elevated by other subjects.

audiometer was modified to more closely mimic an audiogram, when it detects a subject did not hear a tone it can increase volume and retest, mimicing the standard procedure of an audiogram, and obtaining additional data around the threshold.

The application of EEGNet for audiometry could allow for audiometry to be performed on people who cannot participate in behavioural audiometry, such as those with physical disabilities limiting their ability to press a response button. However, it has not been tested with such participants, and also has some disadvantages. Firstly, standard pure-tone audiometry requires very little setup, and can be performed rapidly. Audiologists are capable of making the pure-tone audiogram a relatively fast test; by comparison our test is significantly slower. Setup time is currently a 30-45 minute process, cleanup is another 15 minutes, and each test takes approximately 12 minutes. With repetitions, test duration could be 1.5 hours or more. It may be possible to reduce the setup time drastically using more EEG recorders with a simpler setup such as the Emotiv Epoc [15], although further work would be needed to allow classification with the reduced number of channels provided by an Epoc. Future research needs to investigate the ability to predict thresholds with as few electrodes as possible to speed up setup/cleanup times and reduce participant discomfort.

REFERENCES

[1] British Society of Audiology, "Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking," *Read. Br. Soc. Audiol.*, 2011.

[2] K. Hecox and R. Galambos, "Brain stem auditory evoked responses in human infants and adults," *Archives of Otolaryngology*, vol. 99, no. 1, pp. 30–33, 1974.

[3] D. L. Jewett, M. N. Romano, and J. S. Williston, "Human auditory evoked potentials: possible brain stem components detected on the scalp," *Science*, vol. 167, no. 3924, pp. 1517–1518, 1970.

[4] D. R. Stapells, "Frequency-specific ABR and ASSR threshold assessment in young infants," *Phonak. Com*, pp. 409–448, 2011.

[5] C. Petitot, L. Collet, and J. D. Durrant, "Auditory steady-state responses (ASSR): effects of modulation and carrier frequencies: respuestas auditivas de estado estable (ASSR): efectos de la modulación y de la frecuencia portadora," *Int. J. Audiology*, vol. 44, no. 10, pp. 567–573, 2005.

[6] Y.-H. Lin, H.-C. Ho, and H.-P. Wu, "Comparison of auditory steady-state responses and auditory brainstem responses in audiometric assessment of adults with sensorineural hearing loss," *Auris Nasus Larynx*, vol. 36, no. 2, pp. 140–145, 2009.

[7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018.

[8] C. Ahuja and D. Setia, "Measuring human auditory attention with EEG," in *2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2022, pp. 774–778.

[9] X. Deng, B. Zhang, N. Yu, K. Liu, and K. Sun, "Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces," *IEEE Access*, vol. 9, pp. 25 118–25 130, 2021.

[10] H. Yao, K. Liu, X. Deng, X. Tang, and H. Yu, "FB-EEGNet: a fusion neural network across multi-stimulus for SSVEP target detection," *J. Neuroscience Methods*, p. 109674, 2022.

[11] X. Song, D. Yan, L. Zhao, and L. Yang, "LSDD-EEGNet: An efficient end-to-end framework for EEG-based depression detection," *Biomedical Signal Processing and Control*, vol. 75, p. 103612, 2022.

[12] H. Zhao, Y. Yang, P. Karlsson, and A. McEwan, "Can recurrent neural network enhanced EEGNet improve the accuracy of ERP classification task? an exploration and a discussion," *Health and Technology*, vol. 10, no. 4, pp. 979–995, 2020.

[13] G. Langroudi, R. Palaniappan, and I. McLoughlin, "Auditory evoked potential detection during pure-tone audiometry," in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021, pp. 919–922.

[14] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, p. 031005, 2018.

[15] K. Smitha, K. P. Thomas, A. Vinod *et al.*, "Online electroencephalogram (eeg) based biometric authentication using visual and audio stimuli," in *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2016, pp. 454–459.