2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)
Virtual Conference, May 4-6, 2021

Auditory Evoked Potential Detection During Pure-Tone Audiometry

George Langroudi[1], Ramaswamy Palaniappan[1], and Ian McLoughlin[2]

***Abstract*— Modern audiometry is largely a behavioural task, with the pure-tone audiogram (PTA) being the gold standard for evaluating frequency-specific hearing thresholds in adults. The nature of behavioural audiometry makes estimating accurate hearing thresholds difficult in infants and people with disabilities, where following instructions or interacting with the test may be difficult or impossible. We propose a method in which Auditory Evoked Potentials (AEPs) are used as an alternative to behavioural audiometry for detecting frequency-specific thresholds. Specifically, P300 responses elicited by the tones of a PTA are automatically detected from electroencephalogram (EEG) data, to evaluate hearing acuity. To assess the effectiveness of this method, we created a dataset of EEG recordings from participants presented with a series of pure tones at 6 different frequencies with steadily decreasing volumes, during a PTA test. This dataset was used to train a support vector machine (SVM) to identify when a participant was played a tone, whether they perceived it or not using their EEG. Results demonstrate that detecting hearing events can be very accurate for participants for whom the classifier has been trained a-priori. However, accuracy drops significantly for unseen participants – when the classifier has not been trained on any prior data from a given participant before classifying their EEG. However, by establishing that AEP response-based audiometry is viable for detecting tones, future work will explore the ability of more powerful deep neural networks to accurately estimate for unseen participants.**

I. INTRODUCTION

Over nearly 100 years of use, the Pure Tone Audiogram has remained the most reliable and trusted method of obtaining hearing thresholds [1]. The PTA is a behavioural test where patients are presented a series of tones at steadily decreasing volumes, and are asked to respond when they perceive a tone by pressing a response button. The tones presented are typically pure sine waves, and as such can be presented at any frequency; in practice, the tones presented typically range from 250Hz to 8kHz, and the intermediate octave in between [2]. Frequencies between these octaves can also be used, as well as frequencies as low as 125Hz and as high as 16kHz. The high frequency specificity and speed of the PTA has made it indispensable for the assessment of hearing loss.

However, the PTA suffers from some limitations that can make it either inconvenient, unreliable, or unusable. Notably, PTA requires patient interaction, which excludes certain minority groups such as the young or disabled. Additionally the PTA relies on patient honesty, with no way for audiologists to detect inaccuracies caused by any dishonesty. While some alternative methods for behavioural audiometry in infants exist [3, Chapter 12], they are mostly used to diagnose hearing loss as opposed to assessing the severity of such loss.

While the clinical or audiometric application of AEP is mostly in the form or Auditory Brainstem Response [4], AEP has been the target of research in other domains. For example, in Brain Computer Interface (BCI) design [5], [6]. The visual P300 response has generated a great deal of interest in this realm, often as a method of spelling out characters [7].

While the use of P300 in response to auditory events is less documented, there has been research into auditory P300 use for BCI scenarios. An adapted auditory version of the P300 speller has shown promise [8], [9], although it has a lower throughput and higher peak latency than the visual paradigm.

II. EXPERIMENTAL METHODOLOGY

To ensure applicability to clinical audiometry, our experiment was designed to closely mimic the operation of a PTA with no modifications made to the tones being presented or the interaction of the participant. We recruited participants to undertake an audiometric test. We then simultaneously recorded their EEG data while they undertook a PTA test. EEG and auditory data was recorded for offline analysis.

A. Pseudo-audiometer design

As noted above, we designed our experiment to closely mimic the methodology of a PTA. While we acknowledge that limiting the paradigm to that of a PTA will result likely in weaker potentials than would be possible with unconstrained test design (see Section IV), our intention was to: (a) ensure the method is compatible with use in real-world audiometry settings, (b) change as little as feasibly possible, to ensure that any results

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

2021 10th International IEEE/EMBS Conference on Neural Engineering (NER) | 978-1-7281-4337-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/NER49283.2021.9441417

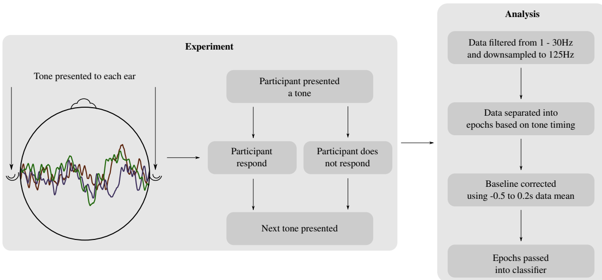

Fig. 1. An overview of our experimental methodology

would not compromise the frequency specificity and integrity of the PTA. With this in mind, we designed and implemented a *pseudo*-Pure Tone Audiometer (${}_P$PTA) that could be used to perform a rudimentary PTA-style hearing test while recording the necessary timing markers alongside EEG data for post-experimental analysis.

Participants were presented with tones at 6 frequencies in the following order: 1000Hz, 2000Hz, 4000Hz, 8000Hz, 250Hz, and 500Hz. Each tone was played at 10 different volumes, starting at 40dB HL and descending in 5dB increments to -5dB HL. Starting with the left ear, participants were played 10 tones at 1000Hz, and then 10 tones at 2000Hz, and so on before repeating the sequence in the right ear.

Tones were spaced at random intervals of between 4 and 8 seconds, and had a random duration between 1 and 3 seconds. Additionally, the duration of the tones was random, between 1 and 3 seconds. The tones were generated using MATLAB at a sample rate of 48kHz and presented to participants using Sennheiser HD 380 pro circumaural headphones.

B. Participants

Participants were recruited once we had received ethics approval from the Faculty of Science Ethics Committee at the University of Kent. Each participant gave their informed, written consent to participate in the experiment after having the details explained both verbally and in writing. Nine participants, reporting no pre-existing hearing conditions, were recruited; three female and 6 male, aged between 22 and 51, with a median age of 32. Each participant completed 2 sessions with a 5 minute break between sessions.

Each participant was tested individually, while sitting comfortably in a low-noise test room without external distraction. Prior to the experiment, participants were fitted with an EEG device (described in section II-C) and asked to keep their eyes closed during the experiment to reduce visual artefacts in the resultant data.

C. Data Collection

Simultaneous to the participant listening to tones during the experiment, we recorded their EEG data using a 32 channel Starstim tES 32 with a 500Hz sampling rate and configured in a 10-10 montage. Each EEG epoch in the dataset represents 1 second of data, starting when the participant is presented the tone; this was chosen since most documented evoked potentials occur within the first second of stimulus onset. Recorded EEG data was bandpass filtered between 1-30Hz and downsampled to 125Hz prior to analysis. Baseline correction was done using the data 500ms to 200ms preceding the onset of the each tone, and EEG data recorded from every participant was individually normalised. With 9 participants recording 2 sessions, and 60 tones presented in each ear per trial, we had 2160 epochs to use.

D. Auditory Evoked Responses

As discussed previously in section I our intention was to elicit AEPs that could be detected by a classifier, and specifically we expected to elicit an auditory P300 response. Whilst the P300 response is strong, it is expected that it would elicit other auditory responses (N1, P1, etc.) that can be detected. We were able to elicit a response on the centre-line electrodes as expected, especially when the participants were presented the higher volume tones, with a weaker response when epochs were averaged across positive epochs at all volumes. It is also worth noting that the highest volume tone is also the first tone presented at each frequency, so the novelty of the stimulus is higher and thus the response would be expected to be better defined.

E. Data analysis

When training our classifier we experimented with three different forms of cross-validation to emulate three possible usage scenarios:

- K-Folds - A simple baseline, best-case scenario for comparison. We used stratified 10-fold cross validation (CV).
- Leave-One-Trial-Out (LOTO) - The classifier was trained on epochs from all other participants plus one of the trials for a given participant, and is tested against the remaining trial. This emulates a scenario in which a participant has had a recording performed previously, and a classifier was trained to include their recorded data.
- Leave-One-Participant-Out (LOPO) - The classifier is trained on all available epochs except for those belonging to a single participant and is tested on that remaining participant, emulating a scenario in which a participant has not contributed any pre-existing data, so the classifier has never seen data from this participant before.

We chose these three arrangements as they are most applicable to real-world contexts. After processing, we interpret each instance of data into one of three classes:

- ***Positive participant response*** - The participant perceived the tone, responded by pressing the response button while the tone was presented, released it after the tone finished.
- ***Negative participant response*** - The participant did not respond during the tone.
- ***No sound presented*** - The participant was not presented with any sound.

When training we used the positive response epochs, and those where the participant was not presented with a sound. As one of the future directions of this work is to detect ambiguous epochs, it was necessary that the training data only contained epochs that we have confidence in. Epochs where the participant was presented a sound but did not respond can be considered ambiguous, as patients are prone to making mistakes at these volumes close to their hearing threshold[1].

F. Support Vector Machines

When classifying our data we used exclusively Support Vector Machines with polynomial kernels. Each classification involved training multiple SVM's with a combination of parameters to find the optimal settings for that data. Each SVM was trained on a single channel: 125 time samples were given to the classifier for each epoch, and will output either a Positive or a Negative; whether the classifier believes the participant has perceived a tone or not. Rather than having a single set of parameters for all channels, we chose to allow the classifier to select unique parameters for each individual channel. This allows the classifier to specifically adjust to the unique responses that occur across the scalp.

All SVM's were created using libsvm and Scikit-learn, and all used a Radial Bias Function kernel. These SVMs take 2 parameters: γ and C. To find the optimal parameters for each channel, hundreds of SVMs were trained using every valid combination of γ and C on the training sets and then tested against the testing set, with the testing and training sets being defined by the cross validation method used. Results were compared and measured using the resultant receiver operating characteristic area under curve (AUC) score, and then the parameters yielding the optimal AUC were selected.

III. RESULTS

When optimising the parameters for training an SVM on each channel we achieved accuracies as high as 95.34% with 10-fold CV. This high accuracy is not unexpected, since there are multiple trials from each participant represented in the training set. While testing

[1]We recognise that this leads to a situation where the classifier does not see ambiguous epochs during training, but encounters them during inference. However, identifying which ambiguous epochs were positive and which were negative is itself a goal of this work.

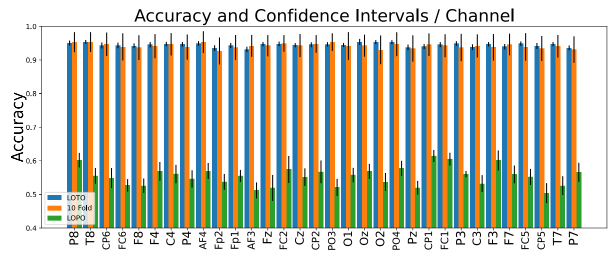

Fig. 2. The average accuracy across folds for each channel with optimised parameters, and the confidence intervals for each channel.

data never appears in the training set, *participants* in the testing dataset have also contributed data to the training dataset. Thus k-folds does not accurately reflect one likely real-world situation in which test participants did not contribute to the enrolled training dataset.

When using Leave-One-Trial-Out cross validation we achieved even higher accuracy across a majority of channels than with 10-fold CV. While it is likely that the higher accuracy is a result of the classifier being trained on more data due to the higher number of folds, it also demonstrates that the classifier can perform remarkably well when presented with only a small amount of a participant's data beforehand (i.e. a single EEG recording). This lends itself well to a paradigm in which patient data is stored between audiometric evaluations; if a patient can participate in even a single behavioural test as baseline, their future tests could be expected to achieve high accuracy.

The classifiers trained with Leave-One-Participant-Out CV achieved notably lower accuracy than either k-fold or LOTO CV, with some channels having accuracy no higher than chance. This suggests that, at least with the number of participants currently involved, the tested classifier cannot generalise to unseen participants. Further work is needed to identify a method of accurately detecting events on unseen participants, as this is especially useful in audiometry.

IV. DISCUSSION

As discussed in Section II-A our $_P$PTA was designed to mimic a standard PTA, but this constrains us in terms of methods of eliciting stronger potentials. One such obvious method would be to use a "distractor" stimulus [10], where the participant is frequently presented a stimulus they are informed of before and asked not to respond to. If a suitably differentiable tone could be itentified for use as a distractor stimulus, it may be possible to elicit stronger potentials during a PTA. However, an improvement could be made without a distractor is using a suitably large inter-stimulus interval [11]. The novelty could also be increased by randomising the frequency order of tones.

Our results show that while the generalisation accuracy using an SVM was poor with our current dataset, we have demonstrated that the use of even a single training recording from a participant enables the resulting model to provide high classification accuracy on subsequent trials. These results, while highly encouraging, however reveal that the current $_P$PTA + classifier design is unsuitable for classification of unseen participants, which would clearly limit its application as noted above. Further work will be needed to identify a more generalisable approach for widespread clinical application.

Some novel approaches to auditory P300 detection have shown promise by combining the Auditory Steady-State response (ASSR) [12], another auditory AEP, with the P300 speller paradigm to improve the throughput [13]. Whilst the difference between ASSR and PTA thresholds has been demonstrated to be stable, ASSR has the strongest effect on frequencies between 500Hz and 4000Hz. Due to the nature of ASSR tone modulation it is yet to be seen whether this technique can be applied for use in audiometry, however it represents a possible avenue to achieve more generalisable hearing event detection in audiometry.

Future practical implementations of this work would ideally work with a reduced number of electrodes, however improved generalisation is likely to be more achievable if a classifier is able to view more electrodes, or has some understanding of the electrode layout. This avenue is something we intend to explore in the future, since a classifier with some awareness of electrode layout may be able to make better use of the higher number of electrodes.

V. CONCLUSION

Whilst clinical usage of EEG in audiology is already established for diagnosis of total hearing loss in infants, this paper has investigated the use of EEG for the diagnosis of frequency specific hearing loss. We have demonstrated that even a relatively simple classifier such as SVM can be used to detect the presence or absence of pure tone hearing events from EEG data. The results show that, when the classifier has ‘seen’ previous trials from a participant during training, it is capable of classifying future hearing events from that participant with very high accuracy, but does not generalise well to classifying unseen participants. The need to baseline a participant with prior data enables certain potential application scenarios, but unfortunately not application as a drop-in replacement for PTA testing. A drop-in replacement for PTA will need to overcome this limitation either though more complex classifiers, or by training with a larger dataset.

Now that this paper has demonstrated the potential viability of auditory tone detection from EEG, it is expected that future research will tackle the issue of generalisation, use of more channels, different classifiers and alternative approaches to P300.

REFERENCES

[1] C. A. Champlin and T. Letowski, “Audiometric calibration: Air conduction,” *Seminars in Hearing*, vol. 35, no. 4, pp. 312–328, 2014.

[2] British Society of Audiology, “Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking,” Tech. Rep., 2011. [Online]. Available: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf

[3] S. A. Gelfand, *Essentials of audiology*, 4th ed. Thieme New York, 2016, no. 562.

[4] D. R. Stapells, “Frequency-Specific Evoked Potential Audiometry in Infants,” *A Sound Foundation Through Early Amplification*, no. November, pp. 13–31, 2000.

[5] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, “BCI2000: A General-Purpose Brain-Computer Interface (BCI) System,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, jun 2004. [Online]. Available: http://ieeexplore.ieee.org/document/1300799/

[6] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak, “Braincomputer interface spellers: A review,” *Brain Sciences*, vol. 8, no. 4, 2018.

[7] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, “A Brain-Computer Interface Controlled Auditory Event-Related Potential (P300) Spelling System for Locked-In Patients,” *Annals of the New York Academy of Sciences*, vol. 1157, no. 1, pp. 90–100, mar 2009. [Online]. Available: http://doi.wiley.com/10.1111/j.1749-6632.2008.04122.x

[8] A. Furdea, S. Halder, D. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, “An auditory oddball (P300) spelling system for brain-computer interfaces,” *Psychophysiology*, vol. 46, no. 3, pp. 617–625, may 2009. [Online]. Available: http://doi.wiley.com/10.1111/j.1469-8986.2008.00783.x

[9] D. Klobassa, T. Vaughan, P. Brunner, N. Schwartz, J. Wolpaw, C. Neuper, and E. Sellers, “Toward a high-throughput auditory P300-based braincomputer interface,” *Clinical Neurophysiology*, vol. 120, no. 7, pp. 1252–1261, jul 2009. [Online]. Available: http://dx.doi.org/10.1016/j.clinph.2009.04.019https://linkinghub.elsevier.com/retrieve/pii/S1388245709003344

[10] J. Polich, “Updating P300: An integrative theory of P3a and P3b,” *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, oct 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1388245707001897

[11] D. Strüber and J. Polich, “P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus interval effects,” *International Journal of Psychophysiology*, vol. 45, no. 3, pp. 187–196, sep 2002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167876002000715

[12] R. Galambos, S. Makeig, and P. J. Talmachoff, “A 40-Hz auditory potential recorded from the human scalp.” *Proceedings of the National Academy of Sciences*, vol. 78, no. 4, pp. 2643–2647, apr 1981. [Online]. Available: http://www.pnas.org/cgi/doi/10.1073/pnas.78.4.2643

[13] N. Kaongoen and S. Jo, “A novel hybrid auditory BCI paradigm combining ASSR and P300,” *Journal of Neuroscience Methods*, vol. 279, pp. 44–51, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.jneumeth.2017.01.011