Auditory Tagging: Improving Performance of Auditory Brain-Computer Interfaces by Modulating Stimuli

Michal Robert Žák[1,2] and Moritz Grosse-Wentrup[1,3,4]

Abstract— **We propose auditory tagging, a novel method to enhance decoding performance in auditory brain-computer interface (BCI) paradigms. Drawing inspiration from steady-state visually evoked potentials (SSVEPs), auditory taggers involve embedding a steady frequency onto an auditory stimulus with the goal of eliciting a detectable neuronal response. In this work, we introduce three such approaches and evaluate them on the auditory intention decoding (AID) paradigm. In AID, subjects are primed with a question and potential target and non-target answer options are provided for this question. The BCI then decodes whether a given sample is a target or non-target. Despite the conceptual promise of the auditory taggers, our experiment results did not reveal statistically significant improvements in decoding accuracy using the proposed tagging approaches. We discuss potential explanations for this observation and highlight possible avenues of improvement for future research.**

I. Introduction

Auditory brain-computer interfaces (BCIs) represent a promising, albeit less explored, direction within BCI research compared to more traditional paradigms. In auditory paradigms, subjects are typically presented with multiple auditory stimuli and select one of them by directing their attention to the stimulus of choice. For instance, Hill et al. [1] developed a BCI that plays two distinct sounds to the subject's left and right ears. The subject controls the BCI by shifting their attention to the left or right auditory stimulus. To decode the subject's focus, the authors primarily used event-related potentials (ERPs) generated by the paradigm, though other approaches were also explored. Other studies extend this idea further and place the subject in a multi-source environment, where auditory sources are spatially separated rather than played to each ear individually. These paradigms employ Auditory Attention Decoding (AAD) to determine which source the subject is focusing on [2], a challenge known as the cocktail-party problem [3]. AAD is also referred to as auditory attention identification [2] or auditory attention detection [4] in the literature. The auditory sources in these experiments range from simple tones to more natural audio stimuli such as speech. One notable application of this research is neuro-steered hearing-aid devices, as proposed by Belo et al. [4] and demonstrated by Geirnaert et al. [5]. These devices detect and enhance the user-attended source, thereby improving hearing-aid performance.

[1]Research Group Neuroinformatics, Faculty of Computer Science, University of Vienna, Vienna, Austria

[2]Doctoral School Computer Science, Faculty of Computer Science, University of Vienna, Vienna, Austria

[3]Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

[4]Research Network Data Science, University of Vienna, Vienna, Austria

However, there are several inherent limitations to auditory paradigms. Tones and other simplistic stimuli generate strong brain responses but can feel unnatural and uncomfortable to the subject. In addition, auditory paradigms that present multiple stimuli concurrently can feel overwhelming to subjects if there are too many sources, limiting the scalability of the paradigm to multiple targets. Using more naturalistic sources like speech can mitigate this issue to some extent but may be more difficult to decode due to more complex and varied brain responses.

Presenting stimuli sequentially could offer a more comfortable and natural conversational scenario for subjects. Further, presenting one stimulus at a time would allow scaling up paradigms to an arbitrary number of stimuli as they could chain any number of them together. This approach facilitates a shift from active BCI paradigms, where the subject must perform an action to select among multiple sources, to reactive or passive BCI paradigms, which observe the subject's brain activity and act accordingly. Removing the need for the subject to actively engage with the system enables more seamless interactions. This approach is less common in the literature though. For example, Roebben et al. [6] tested whether they could distinguish if a subject is listening to the currently played auditory stimulus. They instructed the subject to focus on an auditory story while occasionally performing mental tasks such as mathematics, which caused the subject to ignore the auditory input. The authors demonstrated statistically significant decoding accuracies well above 70% in distinguishing whether a data window contained attended or unattended audio.

In an effort to expand the field of reactive auditory BCIs, we developed the Auditory Intention Decoding (AID) paradigm [7]. The AID paradigm extends our work on conversational brain-AI interfaces [8] from the visual- to the auditory domain: A question posed to a subject using the AID-BCI, such as ”Hello, this is Pizzeria Romano. How can I help you?” is extended by an AI, that is listening to the conversation, to ””Hello, this is Pizzeria Romano. How can I help you? Can I take your order, show you our menu, or make a reservation?”. Decoding from brain recordings which of the presented options the BCI user attends to would then allow another AI to generate an answer sentence, such as ”I would like to make a reservation, please!”, and thus continue the conversation. However, our previous analysis in [7] indicated that the effects of attention to individual words on EEG responses were insufficient for decoding the subject's intention reliably in a single-trial setting.

In this work, we attempt to address this limitation by introducing the concept of auditory taggers. These taggers aim to encode a specific target frequency in the brain response by modulating the played audio in various ways. We hypothesize that attending to stimuli tagged with a specific frequency will highlight brain regions involved in processing the stimulus, potentially amplifying Event-Related Potential (ERP) (and other brain) responses, thereby enhancing the decodability of attention via neural network classifiers. These taggers are inspired by steady-state visual evoked potentials (SSVEP), which encode a steady frequency with each visual stimulus (see [9] for a review). SSVEP has been shown to be a superior marker for visual attention compared to other traditional metrics [10]; we aim to elicit a similar effect with auditory stimuli. Another significant inspiration for the tagging paradigm was the auditory steady-state response (ASSR), a neuroscience technique that studies the brain's response to rapidly presented auditory stimuli. First introduced by [11], ASSR involves modulating stimuli onto a carrier frequency to map auditory processing pathways. The most robust ASSR responses occur at 40 Hz in adults and 20 Hz in children [11]. In neuroscience, ASSR has been applied to study hearing thresholds [12] and as a biomarker for schizophrenia [13]. Some studies have attempted to apply ASSR techniques to BCIs, typically presenting a cocktail party problem-like scenario where the BCI decodes which source the user focuses on (e.g., [14]). The distinction between BCIs built on ASSR and AAD is subtle, and it could be argued that ASSR-BCIs are a special case of AAD-BCIs.

In the following, we introduce three different auditory taggers and compare their performance to a baseline with no modification of auditory stimuli in a feasibility study with ten subjects. In a post-hoc analysis, we compare the performance of these paradigms to assess their relative contributions to the decoding accuracy within the AID framework.

II. METHODS

The following subsections detail the technical aspects of the auditory taggers, which encode frequency information into audio segments to enhance decodability. Further, we describe the designed experiment structure to test the taggers and outline the methods used to evaluate our results.

A. Taggers

We employ auditory taggers to modulate audio stimuli, aiming to elicit a detectable response in the brain. We hypothesized that this modulation would facilitate the decoding of brain areas involved in processing the stimuli, similar to the ASSR or SSVEP paradigms. Our goal is to identify differences between the processing of desired (target) and non-desired (non-target) options or to enhance the decodability of ERPs evoked by the paradigm.

A key distinction between BCIs based on the SSVEP/ASSR paradigm and the AID paradigm is the presentation of stimuli. In SSVEP/ASSR, all stimuli are presented simultaneously (e.g., flashing multiple parts of the screen with different target patterns at the same time). In contrast, the AID paradigm requires playing stimuli sequentially. This changes the decoding target of the underlying classifier from distinguishing which stimulus a subject is focusing on to determining whether the subject is focusing on the currently played stimulus or not. Thus, the classification problem becomes binary: target vs. non-target for each stimulus. Consequently, methods to decode the selection in SSVEP/ASSR may not be directly applicable to AID taggers.

We designed three tagging paradigms in addition to a baseline and explain them in detail below: Raw, Amplitude Modulation (AM), Frequency Modulation (FM), and Binaural.

1) Raw: This paradigm does not modify the stimuli, serving as a baseline for comparison.

2) AM: Amplitude Modulation encodes the desired frequency as a signal envelope of the speech signal. We computed the modulated signal $x_m[n]$ by multiplying the original signal $x[n]$ with a sine wave of the desired frequency (in our case 40 Hz) $m[n]$: $x_m[n] = x[n] \cdot m[n]$

3) FM: Frequency Modulation encodes the desired frequency by computing the instantaneous frequency of a carrier signal (the word stimulus audio) and adding the amplitude of a modulating signal to the instantaneous frequency (a 40 Hz sine wave of equal length). The amplitude of the modulating sine wave ranges from $[-1; 1]$, allowing the instantaneous frequency to be modified by ± 1 at most.

4) Binaural: The binaural beat effect occurs when two sine waves of different frequencies are played in separate ears, eliciting a perceived oscillation of the mean difference frequency between the two tones. This effect was first described by H.W. Dove in [15]. Utilizing this effect, the binaural tagger plays the original signal in the left ear and a 40 Hz shifted signal in the right ear.

B. Experiment Design

We designed a streamlined experiment to test the core ideas of the AID paradigm and incorporated the tagging paradigms into it. The experiment presents stimuli composed of three-digit numbers, striking a balance between ease of generation, ease of remembering, and sufficient length for the tagger to have a strong enough effect.

Each trial presents the subject with a generated number (the target) and instructs them to focus on it. After allowing enough time for the subject to read the number, we played an audio containing the target and two other numbers in random order. Specifically, the sentence *Concentrate on your number. <num1>, <num2>, <num3>* is played, where *<num1-3>* are the generated numbers. We created the audio by dynamically stitching together voice lines generated using the text-to-speech service *Crikk* on March 15, 2024, with the artificial voice *Natasha*. To extend the duration of the stimuli, we played the audios back at 75% speed. Additionally, we tagged all numbers in a given sentence with one of the aforementioned taggers (all numbers within one sentence always used the same tagger).

The generated trials are further grouped into blocks of nine, two for each defined tagger, plus one additional attention check trial, collectively called experiment rounds. Each round shares a common primer number to enable an analysis of the effects of repeated presentation of targets in the future. The attention check is generated via the same procedure as regular stimuli but does not contain the primer number and is tagged with a random tagger. When the subject recognizes that they are listening to an attention check, they are instructed to press the spacebar, indicating that they are paying attention to the experiment. This gives the subject a task during the experiment while giving us a way to decide whether the participant was actively participating. After each block, the subject can take as long a break as they need to help them remain focused on the task, as trial runs of the experiment have shown it to be very tiring. The experiment has a total of 20 rounds, which results in an approximate one-hour paradigm recording, depending on the length of the breaks the subject takes. The experiment structure is further illustrated in Fig. 1.

C. Data Acquisition

We recorded the subject's EEG activity throughout the described experiment. Healthy, English-speaking participants took part in the research group's EEG laboratory and were compensated for their time. The experiment was approved by the ethics committee of the University of Vienna.

In the recorded signal, we marked key events, such as the presentation of non-target and target stimuli, using triggers. We analyzed the data offline, with no real-time (online) decoding performed. We acquired EEG recordings using a Bittium NeurOne Tesla system with a sampling frequency of 1,000 Hz and 128-channel EASYCAP caps.

D. Classification Model Training

To assess whether auditory taggers influence classification performance, we utilized EEGNet. EEGNet is a deep, convolutional neural network designed to work well on EEG classification tasks [16]. We have utilized deep learning to enable us to capture more complex temporal and spatial features as opposed to more traditional approaches, with the trade-off of the amount of needed training data and explainability. Given the exploratory nature of this study, model performance was prioritized over model explainability.

We configured the network with the Adam optimizer [17], a learning rate of 0.001, a batch size of 32, and 200 training epochs. We used 10-fold stratified cross-validation (CV), as there is a class imbalance of training data (two non-target samples to one target sample), with an 80/20 train/test split.

We generated samples for classification by segmenting EEG data from -1 to 1.5 seconds relative to the end of each word stimulus. It remains unclear whether it is better to segment data around the offset or onset of the stimulus. We have chosen to use the offset, as this guarantees the subject had time to process the word. To avoid overlapping data segments, we discarded numbers with a total duration shorter than 2 seconds (there is an additional 0.5 s break between two numbers). We labeled each sample as non-target (0) or target (1) and annotated it with the corresponding tagger identifier. Due to EEGNet implementation constraints, we downsampled the data to 128 Hz and normalized each sample using the median power computed from the training set.

A separate classifier was trained for each tagger for all subjects pooled together. It would be ideal to train a classifier for each subject and each tagger individually; however, this results in too little training data (approximately 80 training samples and 20 test samples). We see training the model on all subjects together (group-trained model) as a valid compromise, which enables us to analyze the effects of the tagger labels.

E. Performance Analysis and Statistical Testing

Three types of performance evaluations were conducted to assess classification significance and compare tagger effects. The procedures of these analyses are detailed in the following sections.

1) Significance of Auditory Tagger Performances: We conducted permutation testing with $N = 10,000$ iterations to evaluate whether the observed classification accuracies

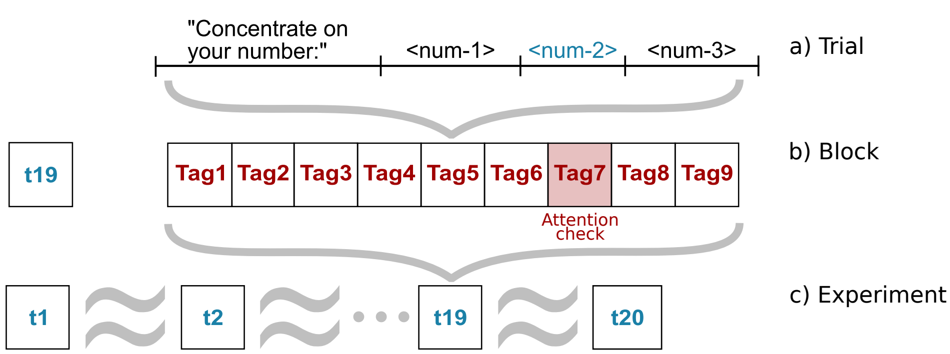

Fig. 1. Structure of the conducted experiment. The experiment consists of three hierarchical stages: **a) Trial**: A single trial involves playing one sentence for the subject. Each trial contains three numbers, one randomly designated as the target number (highlighted in blue) that the subject must focus on. **b) Block**: Trials are grouped into blocks of nine, with one trial per block serving as an attention check (containing no primed number). In the example shown, the block's target number is t_{19}. Each trial in a block is assigned a tagger, with three distinct taggers plus one baseline tagger defined in this study. We used each tagger twice per block, except for the attention check trial, in which we assigned a tagger at random. **c) Experiment**: The full experiment consists of 20 blocks, each with a randomly assigned target number. Participants can take breaks between blocks to minimize fatigue, resulting in an approximate total runtime of one hour.

exceeded chance. For each tagger, predicted labels were randomly shuffled within the test set to generate a distribution of classification accuracies. The p-value was computed as the proportion of permuted accuracies exceeding the actual accuracy.

As classifiers were trained on pooled data across participants, subject-level performance was assessed by extracting subject-specific test trials from each fold. While this approach does not yield fully independent subject-level models, it enables estimation of individual contributions within a group-trained model. For group-level significance, permutations were restricted within subjects to control for inter-subject variability.

2) Comparison of Tagger Performances: To determine whether classification performance varied across tagger conditions, we employed a permutational analysis of variance (PERMANOVA) [18]. The test assessed whether the distribution of accuracies differed significantly across the four tagger types. The F-statistic was computed as:

$$F = (SS_A/SS_R) \cdot [(N-g)/(g-1)] \tag{1}$$

where SS_A is the among-group normalized sum of square differences, SS_R is the within-group (residual) normalized sum of square differences, N is the total number of samples, and g is the number of classes (four taggers in our case).

A null distribution of F-values was constructed by permuting tagger labels $10,000$ times within each subject, and the p-value was calculated as the proportion of permuted F-values exceeding the observed F-statistic.

3) Pairwise Tagger Comparisons: To assess differences between individual tagger pairs, we conducted a pairwise permutation test. The baseline difference in mean classification accuracy between each tagger pair was computed. Next, labels between the two taggers were permuted $10{,}000$ times within subjects to construct a null distribution of mean differences. P-values were derived from the proportion of permuted differences greater than or equal to the observed value.

Both subject-level and across-subject analyses were conducted, with all permutations performed within-subject to control for inter-individual variability.

III. Results

A total of eleven subject recordings were conducted for this study. However, due to poor data quality arising from subject S3's dense hair, this subject was excluded from further analysis, leaving a final cohort of ten subjects.

A. EEGNet Performance

Figure 2a summarizes the performance of the individual tagger conditions. Each box represents a separately trained classifier, with the distribution reflecting the accuracies across individual cross-validation (CV) folds.

Figure 2b presents classification performance at the subject level. As noted in the Methods section, no subject-specific classifiers were trained. Instead, test-set samples from the classifier trained on the combined data from all subjects were grouped by subject to estimate individual-level performance, despite the limited sample sizes.

Table I reports the mean classification accuracies for each tagger condition and for each subject, along with p-values from the permutation analysis. Statistically significant results ($p < 0.05$) are highlighted in bold.

As shown in Table I, few per-subject results reached statistical significance. However, all conditions trained on pooled data across subjects (the "All" columns) achieved statistically significant performance.

B. Comparison of Tagger Conditions

The PERMANOVA analysis did not provide evidence to reject the null hypothesis that all tagger conditions originate from the same sample-generating distribution. When performance was assessed across subjects, the resulting p-value was 0.652. Similarly, no statistically significant effects ($p < 0.05$) were observed at the individual subject level.

Pairwise comparisons between tagger conditions likewise failed to identify any cases in which a tagging paradigm significantly outperformed the RAW condition. This finding held true both when data were pooled across all subjects and when analyzed on a per-subject basis.

IV. Conclusion

In this work, we introduced auditory taggers as a novel approach to enhance the performance of auditory BCI systems. However, we did not find any statistically significant evidence that auditory taggers systematically influence the performance of auditory attention decoding in the AID paradigm.

We emphasize that our analysis was foundational in nature and is limited by the trade-off between the number of tagging methods and the sample size per tagger. In particular, we did observe a large variation between the performance of the taggers within individual subjects. Whether these variations are estimation errors due to the small sample sizes or reflect actual preferences of individual subjects to specific tagging methods would need to be explored in a study with more trials per subject, e.g., by focusing on a comparison between one tagging method and a baseline. At the same time, the development of more robust tagging paradigms capable of eliciting stronger neural responses may also improve signal detectability and enhance overall system performance.

In addition to these high-level conclsuions, several directions for minor improvements of the paradigm design also emerged over the course of this study. One limitation of the current setup is the variability in stimulus duration (e.g., ”one hundred” vs. ”one hundred seventy-five”), which complicates the windowing of training samples. Future studies could normalize stimulus length to facilitate more consistent and effective decoding. Additionally, incorporating longer pauses between words may allow participants more time to process each stimulus, potentially resulting in clearer and more decodable effects in the data.

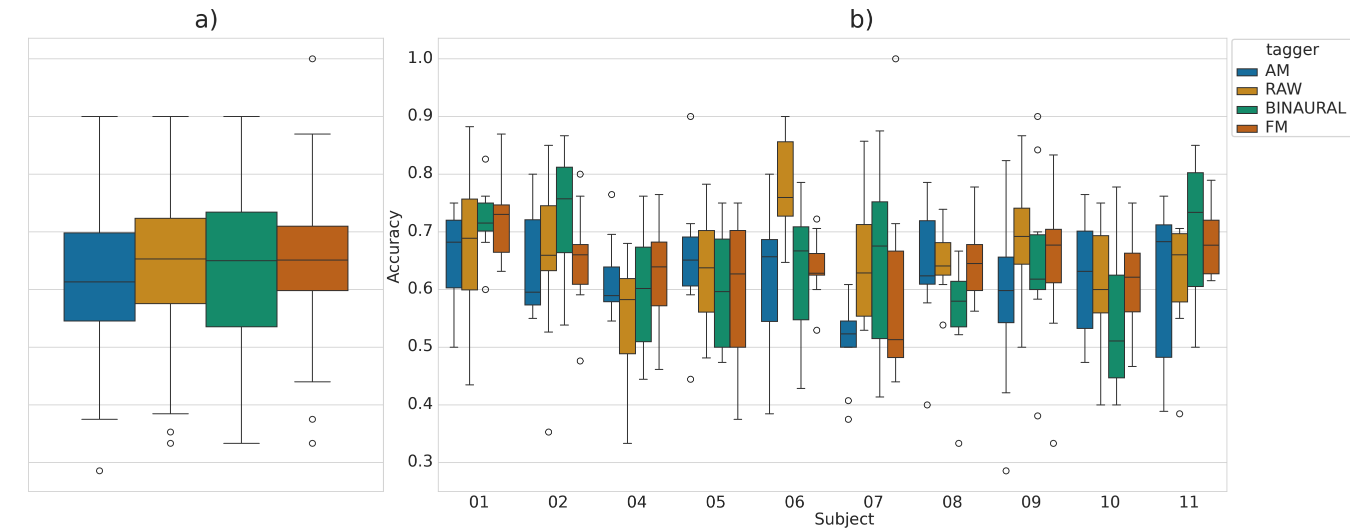

Fig. 2. Classification accuracies obtained using EEGNet across 10 cross-validation folds. **a)** Accuracies for classifiers trained separately on each tagging condition using data pooled across all subjects. Each box plot represents a distinct classifier trained under one condition. **b)** Subject-wise classification accuracies derived from the pooled-data classifier. As noted in the Methods section, individual classifiers were not trained per subject due to insufficient data; instead, performance for each subject reflects how well the pooled classifier generalizes across individuals.

TABLE I
EEGNet accuracy results

		S1	S2	S4	S5	S6	S7	S8	S9	S10	S11	All
AM	μ	0.659	0.643	0.617	**0.656**	**0.619**	0.508	0.644	0.584	0.625	0.613	**0.617**
	σ	0.080	0.087	0.065	0.109	0.134	0.066	0.108	0.145	0.101	0.128	0.114
	p-value	0.294	0.249	0.263	0.014	0.032	0.998	0.145	0.382	0.161	0.140	0.005
FM	μ	**0.727**	0.653	0.628	0.598	0.638	0.595	0.645	0.645	0.611	**0.684**	**0.642**
	σ	0.070	0.085	0.088	0.117	0.051	0.163	0.063	0.128	0.090	0.058	0.105
	p-value	0.034	0.801	0.064	0.782	0.315	0.545	0.062	0.106	0.572	0.001	0.001
Binaural	μ	**0.721**	0.732	0.595	0.598	0.630	0.644	0.564	0.652	0.545	**0.701**	**0.638**
	σ	0.056	0.100	0.103	0.105	0.112	0.138	0.089	0.137	0.118	0.126	0.127
	p-value	0.001	0.058	0.191	0.375	0.592	0.482	1.000	0.111	0.905	0.010	0.028
Raw	μ	0.673	0.661	0.552	0.628	**0.774**	**0.645**	0.646	**0.683**	0.600	**0.622**	**0.648**
	σ	0.124	0.138	0.101	0.094	0.086	0.103	0.052	0.101	0.111	0.095	0.117
	p-value	0.594	0.192	0.943	0.653	0.000	0.042	0.094	0.030	0.052	0.018	0.000

Although the tagging paradigms tested here did not yield significant effects, we believe there remain unexplored potential in auditory BCI research in general and auditory tagging methods in particular, because auditory BCIs represent a relatively unobtrusive alternative to their visually-based counterparts. Future research building on this groundwork may uncover more nuanced or condition-specific benefits of taggers and further advance the field of auditory BCI and the AID paradigm in particular.

APPENDIX

Data:
`https://zenodo.org/records/15332758`

Data analysis repository:
`https://gitlab.cs.univie.ac.at/neuroinformatics/publications/aid-data-analysis-smc-2025`

Experiment repository:
`https://github.com/michalrzak/auditory-intention-decoding`

REFERENCES

[1] N. J. Hill and B. Schölkopf, "An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026011, Feb. 2012, publisher: IOP Publishing.

[2] E. Alickovic, T. Lunner, F. Gustafsson, and L. Ljung, "A tutorial on auditory attention identification methods," *Frontiers in Neuroscience*, vol. 13, Mar. 2019, publisher: Frontiers.

[3] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, Sep. 1953.

[4] J. Belo, M. Clerc, and D. Schön, "EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI," *Frontiers in Computer Science*, vol. 3, Apr. 2021, publisher: Frontiers.

[5] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. de Cheveigne, E. Lalor, B. T. Meyer, S. Miran, T. Francart, and A. Bertrand, "Electroencephalography-based auditory attention decoding: Toward neurosteered hearing devices," *IEEE Signal Processing Magazine*, vol. 38, no. 4, pp. 89–102, Jul. 2021, conference Name: IEEE Signal Processing Magazine.

[6] A. Roebben, N. Heintz, S. Geirnaert, T. Francart, and A. Bertrand, "'are you even listening?' - eeg-based decoding of absolute auditory attention to natural speech," *Journal of Neural Engineering*, vol. 21, no. 3, p. 036046, Jun. 2024, publisher: IOP Publishing.

[7] M. R. Žák and M. Grosse-Wentrup, "Auditory conversational bai: A feasibility study," Apr. 2025, arXiv:2504.13937 [cs].

[8] A. Meunier, M. R. Žák, L. Munz, S. Garkot, M. Eder, J. Xu, and M. Grosse-Wentrup, "A conversational brain–artificial intelligence interface," Mar. 2024, arXiv:2402.15011 [cs, eess].

[9] D. Zhu, J. Bieger, G. Garcia Molina, and R. M. Aarts, "A survey of stimulation methods used in ssvep-based bcis," *Computational Intelligence and Neuroscience*, vol. 2010, no. 1, p. 702357, 2010.

[10] S. Ladouce, J. J. T. Tresols, K. L. Goff, and F. Dehais, "Eeg-based assessment of long-term vigilance and lapses of attention using a user-centered frequency-tagging approach," Dec. 2024, pages: 2024.12.12.628208 Section: New Results.

[11] 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, publisher: Proceedings of the National Academy of Sciences.

[12] P. Korczak, J. Smart, R. Delgado, T. M. Strobel, and C. Bradford, "Auditory steady-state responses," *Journal of the American Academy of Audiology*, vol. 23, no. 3, pp. 146–170, Mar. 2012, publisher: Thieme Medical Publishers.

[13] B. F. O'Donnell, J. L. Vohs, G. P. Krishnan, O. Rass, W. P. Hetrick, and S. L. Morzorati, "Chapter 6 - the auditory steady-state response (assr): A translational biomarker for schizophrenia," in *Supplements to Clinical Neurophysiology*, ser. Application of Brain Oscillations in Neuropsychiatric Diseases, E. Başar, C. Başar-Eroĝlu, A. Özerdem, P. M. Rossini, and G. G. Yener, Eds. Elsevier, Jan. 2013, vol. 62, pp. 101–112.

[14] H. Bharadwaj, A. K. C. Lee, and B. Shinn-Cunningham, "Measuring auditory selective attention using frequency tagging," *Frontiers in Integrative Neuroscience*, vol. 8, 2014.

[15] H. W. Dove, "Über die combination der eindrke beider ohren und beider augen zu einem eindruck," in *Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königl. Preuss. Akademie der Wissenschaften zu Berlin*. Druckerei der Königlichen Akademie der Wissenchaften., 1841.

[16] 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," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, Jul. 2018, publisher: IOP Publishing.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Jan. 2017, arXiv:1412.6980 [cs].

[18] M. J. Anderson, "Permutational multivariate analysis of variance (permanova)," in *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2017, pp. 1–15.