Target of selective auditory attention can be robustly followed with MEG

Dovilė Kurmanavičiūtė^{a,*}, Antti Rantala^a, Mainak Jas^{a,c}, Anne Välilä^a,
Lauri Parkkonen^{a,b}

^a*Department of Neuroscience and Biomedical Engineering, Aalto University, P.O. Box 12200, FI-00076 Aalto, Finland*

^b*Aalto NeuroImaging, Aalto University, FI-00076 Aalto, Finland*

^c*CNRS LTCI, Télécom Paris Tech, Université Paris-Saclay, Paris, France*

Abstract

Selective auditory attention enables filtering relevant from irrelevant acoustic information. Specific auditory responses, measurable by electro- and magnetoencephalography (EEG/MEG), are known to be modulated by attention to the evoking stimuli. However, these attention effects are typically demonstrated in averaged responses and their robustness in single trials is not studied extensively.

We applied decoding algorithms to MEG to investigate how well the target of auditory attention could be determined from single responses and which spatial and temporal aspects of the responses carry most of the information regarding the target of attention. To this end, we recorded brain responses of 15 healthy subjects with MEG when they selectively attended to one of the simultaneously presented auditory streams of words “Yes” and “No”. A support vector machine was trained on the MEG data both at the sensor and source level to predict at every trial which stream was attended.

Sensor-level decoding of the attended stream using the entire 2-s epoch resulted in a mean accuracy of $93\% \pm 1\%$ (range 83–99% across subjects). Time-resolved decoding revealed that the highest accuracies were obtained 200–350 ms after the stimulus onset. Spatially-resolved source-level decoding indicated that the cortical sources most informative of the attended stream were located primarily in the auditory cortex, especially in the right hemisphere.

*Corresponding author dovile.kurmanaviciute@aalto.fi

Our result corroborates attentional modulation of auditory evoked responses also to naturalistic stimuli. The achieved high decoding accuracy could enable the use of our experimental paradigm and classification method in a brain–computer interface.

Keywords: selective attention, single-trial analysis, magnetoencephalography, MEG, auditory system, dichotic listening

1. Introduction

Selective auditory attention is a cognitive function which enables filtering of relevant information from irrelevant. The need of such a selection mechanism has been illustrated by the cocktail party problem in which the listener has to concentrate his/her auditory attention to one speaker while suppressing the voices of the irrelevant speakers to follow that one speaker (Cherry, 1953). Electroencephalographic measurements during dichotic listening have shown that selective auditory attention modulates brain responses generated in auditory cortex (Hillyard et al., 1973; Woldorff et al., 1993).

In the last decade, machine-learning methods have been applied to test whether the target of selective attention can be detected from electro- and magnetoencephalographic (EEG/MEG) data (Nijboer et al., 2008; Furdea et al., 2009; Halder et al., 2010, 2016; Schreuder et al., 2010; Hhne et al., 2011; Hill et al., 2012; Nambu et al., 2013; Hbner et al., 2018). EEG and MEG are well suited for monitoring attention effects as they provide a high temporal resolution in the order of milliseconds, enabling the detection and classification of evoked responses (e.g. auditory or visual P300; McCane et al., 2015; Yeom et al., 2014; Curtin et al., 2012), steady-state responses (e.g. SSRs or mixed SSR/P300; Kaongoen and Jo, 2017; Kim et al., 2011) and oscillatory brain activity (e.g. sensory-motor rhythm (SMR); Geronimo et al.).

The ability to detect the target of auditory attention from brain signals has been exploited to improve the performance of hearing aids (Kidd, 2017) as well as in brain-computer interfaces (BCI) e.g. to re-enable communication in paralyzed patients (Sellers and Donchin, 2006; Astrand et al., 2014; McCane et al., 2015). However, attentional effects are not equally easy to detect from all response types. Hill and colleagues (2012) argue that attention-based classification on ERPs is more reliable than that on steady-state evoked potentials (SSEPs) in a dichotic listening task due to the limited attentional modulation of auditory SSEPs.

Many BCI approaches employ a secondary mental task artificially connected to the primary task because of the poor signal-to-noise ratio of the responses related to it; for example, a primary task of communicating a “yes” or “no” answer could be linked to a secondary task of imagining moving the right or left hand, respectively. Here, we will test the use of spoken-word stimuli in BCI that only comprises the primary task and thus requires minimal training of the subjects.

2. Materials and methods

2.1. Participants

Fifteen healthy adult volunteers (4 females, 11 males; mean age 28.8 ± 3.8 years, range 23–38 years) participated in our study. Two subjects were left-handed and the rest right-handed. Participants did not report hearing problems or history of psychiatric disorders. The study was approved by the Aalto University Ethics Committee. All participants gave their informed consent prior to the recordings.

Figure 1: Experimental design. **Top:** Block structure and the instructions to the subject. **Bottom:** Stimulus sequence within each block.

46 2.2. Stimuli and experimental protocol

47 The auditory stimulus comprised of two simultaneous word streams; the
 48 word “Yes” was repeatedly presented on the left side and the word “No” on
 49 the right; see Figure 1. In each word stream, high- and low-pitch versions
 50 of the same word stimuli alternated. To control for subjects’ attention, the
 51 sequence contained occasional deviants (violations of the regular alternation),
 52 which comprised three consecutive high-pitch versions of the same word.
 53 Deviant probability was 10% in both streams for the first seven subjects and
 54 5% for the rest subjects in order to reduce the mental load of memorizing
 55 the deviant count.

56 To create a realistic acoustic scene, the stimuli were recorded with a
 57 dummy head at the center of a room with dimensions comparable to those of
 58 the magnetically shielded room where the MEG recordings were performed.
 59 The speakers were standing at about 40 degrees to the left/right of the
 60 dummy head at a distance of 1.13 m.

61 The experiment comprised 8 blocks, each lasting about 5 min. Two sec-
 62 onds before a block started, the subject was instructed to direct his/her
 63 attention to one of the streams by the cues “LEFT-YES” or “RIGHT-NO”
 64 on the screen. The task of the subject was to focus on the indicated word
 65 stream, covertly count the deviants and maintain gaze at the fixation cross
 66 displayed on the screen. The experiment always started with the condition
 67 “Attended Left” and was followed by the condition “Attended Right”. The
 68 order of the remaining six blocks was randomized across subjects. The total
 69 length of the experiment was 50–60 minutes including the breaks between
 70 the blocks.

71 PsychoPy version 1.79.01 (Peirce, 2007, 2008) Python package was used
 72 for controlling and presenting the auditory stimuli and visual instructions.
 73 The stimulation was controlled by a computer running Windows 2003 for
 74 the first nine subjects and Linux Ubuntu 14.04 for the rest. Auditory stim-
 75 ulti were delivered by a professional audio card (E-MU 1616m PCIe, E-MU
 76 Systems, Scotts Valley, CA, USA), an audio power amplifier (LTO MACRO
 77 830, Sekaku Electron Industry Co., Ltd, Taichung, Taiwan), and custom-
 78 built loudspeaker units outside of the shielded room and plastic tubes con-
 79 veying the stimuli separately to the ears. Sound pressure was adjusted to a
 80 comfortable level for each subject individually.

81 2.3. MEG data acquisition

82 MEG measurements were performed with a whole-scalp 306-channel Elekta
83 Neuromag VectorView MEG system (Elekta Oy/MEGIN, Helsinki, Finland)
84 at the MEG Core of Aalto Neuroimaging, Aalto University. During acqui-
85 sition, the data were filtered to 0.1–330 Hz and sampled at 1 kHz. Prior to
86 the MEG recording, anatomical landmarks (nasion, left and right preauricu-
87 lar points), head-position indicator coils, and additional scalp-surface points
88 (around 100) were digitized using an Isotrak 3D digitizer (Polhemus Navi-
89 gational Sciences, Colchester, VT, USA). Bipolar electrooculogram (EOG)
90 with electrodes positioned around the right eye (laterally and below) was
91 recorded. Fourteen of the 15 subjects were recorded with continuous head
92 movement tracking. All subjects were measured in the seated position. The
93 back-projection screen was 1 m from the eyes of the subject. If needed, vision
94 was corrected by nonmagnetic goggles.

95 2.4. Data pre-processing

96 The MaxFilter software (version 2.2.10; Elekta Oy/MEGIN, Helsinki,
97 Finland) was applied to suppress external interference using temporal signal
98 space separation to compensate for head movements (Taulu and Hari, 2009).
99 Further analysis was performed using MNE version 2.7.4 and MNE-Python
100 (version 0.14; Gramfort et al., 2014) and ScikitLearn (version 0.18; Pedregosa
101 et al., 2011) software packages.

102 Finite-impulse-response (FIR) filters were employed to filter the unaver-
103 aged MEG data to 0.1–30 Hz for visualization of the evoked responses and
104 for sensor- and source-level decoding. Ocular artifacts were suppressed by re-
105 moving those independent components (1–4 per subject, on average 3) that
106 correlated most with the EOG signal. 2-s long epochs with a 0.50-s pre-
107 stimulus period were extracted from the data at every word stimulus. The
108 delay in the sound reproduction system was considered in the epoch timing.
109 Epochs were rejected if any of the gradiometer signals exceeded 4000 fT/cm.
110 Responses to deviants were excluded from data analysis.

111 2.5. Evoked responses

112 2.5.1. Sensor-level analysis

113 The trial counts were equalized across the conditions (“Attended Left”,
114 “Attended Right”, “Unattended Left” and “Unattended Right”) and the tri-
115 als were averaged. Only attended attention conditions were used in sensor-
116 and source-level classification.

117 2.5.2. Source-level analysis

118 Head models were constructed based on individual magnetic resonance
119 images (MRIs) when available ($N = 12$) applying the watershed algorithm
120 implemented in the FreeSurfer software (Version 5.3; Dale et al., 1999; Fischl
121 et al., 1999). Using the MNE software, single-compartment boundary ele-
122 ment models (BEM) comprising 5120 triangles were then created based on
123 the inner skull surface. The MRIs of three subjects were not available and
124 these subjects were excluded from the source-level analysis.

125 For the source space, the cortical mantle was segmented from MRIs using
126 FreeSurfer and the resulting triangle mesh was subdivided to 4098 sources
127 per hemisphere. The dynamic statistical parametric mapping (dSPM; Dale
128 et al., 2000) variant of minimum-norm estimation was applied to model the
129 activity at these sources. The noise covariance used in the model was esti-
130 mated for each subject from all epochs' 0.50-s pre-stimulus intervals. dSPM
131 sources "Attended Left" and "Attended Right" attention conditions were es-
132 timated for all subjects individually. The obtained source amplitudes were
133 then normalized for each subject and a group-level dSPM source estimate was
134 calculated by morphing the normalized individual estimates to the FreeSurfer
135 average brain and averaging them. For the group averages individual dSPMs
136 were normalized by putting source peak value to 1.

137 2.6. Classification

138 2.6.1. Sensor-level classification

139 A linear support vector machine (SVM; Cortes and Vapnik, 1995) imple-
140 mented in the ScikitLearn package (Pedregosa et al., 2011) was applied for
141 single-epoch classification of the conditions "Attended Left" vs. "Attended
142 Right". To this end, the pre-processed MEG data (filtered to 0.1–30 Hz) were
143 down-sampled by factor 8 to a sampling rate of 125 Hz. Amplitudes of the
144 planar gradiometer channels were concatenated to form the feature vector.
145 Five-fold cross-validation (CV) was applied with an 80/20 split; 80% of data
146 were used for training and the rest for testing. The empirical chance level
147 was around 55% for our sample size of 500 trials in this two-class decoding
148 task (Combrisson and Jerbi, 2015).

149 Decoding was separately performed on data of 1) the entire epoch (250
150 samples x 204 channels; *entire-epoch decoding*), 2) one time point (1 sample
151 x 204 channels; *spatially-resolved decoding*), and 3) one channel (250 samples
152 x 1 channel; *time-resolved decoding*).

153 *2.6.2. Source-level classification*

154 A linear SVM decoder with five-fold CV (80/20 split) was applied to the
 155 individual source estimates calculated for the conditions “Attended Left”
 156 and “Attended Right”. A spatial searchlight across the source space was
 157 used on the 2-s epochs and the resulting accuracy maps were morphed to the
 158 FreeSurfer average brain (comprising 20484 source points) and averaged. In
 159 addition, the accuracies obtained for the left and right auditory cortex were
 160 compared across the individuals using a paired t-test.

161 3. Results

162 3.1. Behavioral data

163 The average relative absolute error of the reported deviant count was 49%
 164 for the 10-% deviant probability ($N = 5$; subjects S03–S07) and 12% for the
 165 5-% probability ($N = 7$; subjects S09–S15).

Subject	Score (<i>mean</i> $\pm$ <i>SD</i> ; %)
S01	90 $\pm$ 1
S02	95 $\pm$ 2
S03	97 $\pm$ 1
S04	92 $\pm$ 1
S05	99 $\pm$ 1
S06	84 $\pm$ 2
S07	87 $\pm$ 2
S08	96 $\pm$ 0
S09	91 $\pm$ 3
S10	94 $\pm$ 1
S11	96 $\pm$ 1
S12	97 $\pm$ 1
S13	90 $\pm$ 1
S14	99 $\pm$ 1
S15	84 $\pm$ 2
MEAN	93 $\pm$ 1

Table 1: Entire-epoch classification accuracy for all subjects and the group mean accuracy.

166 3.2. Sensor-level analysis

167 Average evoked responses to each attention condition (“Attended Left”,
 168 “Unattended Left”, “Attended Right”, “Unattended Right”) are shown in
 169 Figures 2 and 3.

170 Time-resolved classification revealed that the most informative responses
 171 occurred in 100–400 ms after the stimulus onset (Figure 2 and 3). The
 172 average evoked responses for subject S03 (group) peaked at 185 ms (185
 173 ms), 236 ms (245 ms) and 309 ms (312 ms) for “Attended Left” attention
 174 condition. For the condition “Attended Right”, responses peaked at 193 ms
 175 (195 ms), 273 ms (304 ms) and 361 ms (390 ms). Both in Subject S03 and
 176 in the group, time-resolved classification peaked at 320 ms.

Figure 2: Evoked responses and classification accuracies in a representative subject (S03) to four consecutive stimuli. **A:** Acoustic waveforms of the stimuli and average responses at a planar gradiometer channel low-pass-filtered at 30 Hz in the attended (solid lines) and unattended (dashed lines) conditions. **B:** Spatial patterns (gradient strength maps) of the evoked responses. **C:** Time-resolved classification accuracies with standard deviation (blue shading) computed across the folds. **D:** Spatially-resolved classification accuracies.

Figure 3: Evoked responses and classification accuracies at the group level ($N = 15$). **A:** Acoustic waveforms of the stimuli and average responses at a planar gradiometer channel low-pass-filtered at 30 Hz in the attended (solid lines) and unattended (dashed lines) conditions. **B:** Spatial patterns (gradient strength maps) of the evoked responses. **C:** Time-resolved classification accuracies with standard deviation (blue shading) computed across the folds. **D:** Spatially-resolved classification accuracies.

Figure 4: Source estimates of the evoked responses. Normalized group average ($N = 12$). **A**: “Attended Left” and “Attended Right” at 200 ms after stimulus onset; **B**: “Attended Left” and “Attended Right” at 350 ms after stimulus onset .

Figure 5: Source-level spatial-searchlight classification in a representative subject (S03) morphed to the average brain. Color indicates the accuracy of decoding “Attended Left” vs. “Attended Right” based on the signal from that cortical location; the top 5% of the scores in each hemisphere are shown.

Figure 6: Source-level decoding at the group level ($N = 12$). Color indicates the accuracy of classification of “Attended Left” vs. “Attended Right” based on the signal from that cortical location; the top 5% of the scores in each hemisphere are shown.

177 Spatially-resolved classification indicated that the most informative sig-
 178 nals arose from temporal regions. Both the temporal and spatial decoding
 179 patterns were qualitatively similar across the subjects; see Figure 2 for a
 180 representative subject and Figure 3 for the group result. Using the entire
 181 epochs for decoding “Attend Left” vs. “Attend Right” conditions yielded
 182 scores 84–99% (mean 93%; Table 1) across the 15 subjects.

183 3.3. Source-level analysis

184 Source modeling of the peaks of the evoked responses indicated sources
 185 in both auditory cortices. Paired t-tests showed significant source-amplitude
 186 differences ($p < 0.05$) between the left and right hemisphere sources in the
 187 “Attended Left” condition but not in the “Attended Right” condition ($p >$
 188 0.05 ; $N = 12$). In addition, a paired t-test showed that in the “Attended
 189 Left” condition, the source amplitudes were significantly different from those
 190 in the “Attended Right” condition at 350 ms ($p < 0.05$) (see Figure 6) while
 191 at 200 ms the difference was not significant.

192 Spatial-searchlight decoding applied in source space indicated that audi-
 193 tory cortices were the most informative about attention target; see Fig. 6.
 194 All subjects ($N = 12$ with source estimates) showed the highest accuracy for
 195 source signals arising from the auditory cortices. The across-subjects average
 196 peak value was 74.0% in the left and 77.6% in the right temporal areas. This
 197 difference between the hemispheres was not significant ($p = 0.389$, $N = 12$).

198 4. Discussion

199 In this paper, we showed that the target of selective auditory attention
200 to concurrent streams of naturalistic speech stimuli can be robustly detected
201 from unaveraged MEG responses and that this detection is most accurate for
202 signals arising from the auditory cortices 300–400 ms after stimulus onset.

203 In our data, the earliest clearly-discernible response peaks at about 200 ms
204 after the onset of the spoken-word stimulus. This response – often referred to
205 as N2 or N200 in EEG literature – shows only weak dependence on attention
206 in our results. In contrast, the responses occurring within 300–400 ms are
207 significantly modulated by attention. Several studies have shown that the
208 late component of the P300 response is affected by attention (see e.g. [Chennu](#)
209 [et al., 2013](#); [Picton, 1992](#)) and this component is likely the largest contributor
210 to our classification results.

211 In general, an increased P300 amplitude can be due to unexpected changes
212 in the stimulus sequence (e.g. in an auditory oddball task). As opposed to
213 the mismatch negativity (MMN) response occurring earlier and indexing local
214 deviants ([Näätänen et al., 2007](#)), the P300 appears to reflect mostly global,
215 consciously-perceived changes in the stimulus stream, e.g. an unexpected
216 stimulus sequence ([Bekinschtein et al., 2009](#)). These observations provide
217 further evidence that the P300 response echoes cognitive processes, such as
218 attention, that are closely linked to conscious perception.

219 To elicit brain responses with maximal attentional modulation but with
220 minimal subject training, we employed meaningful stimuli that are easy to
221 attend to even during dichotic listening. As pointed out by Hill and colleagues
222 ([2014](#)), applying naturalistic stimuli as opposed to meaningless tone pips
223 could make dichotic listening more pleasant and thus contribute to stronger
224 attentional modulation of the responses and eventually to higher accuracy in
225 classifying the target of attention.

226 Due to the above factors and the obtained high classification accuracy,
227 our paradigm could be well-suited for a brain–computer interface (BCI).
228 Several studies have exploited selective auditory attention and/or P300 re-
229 sponses to drive a BCI but usually not with real spoken words. For ex-
230 ample, Halder and colleagues ([2018](#)) used five Japanese Hiragana syllables
231 (/ka/, /ki/, /ku/, /ke/, and /ko/) presented at different spatial locations
232 in the auditory scene while measuring EEG, applied shrinkage Linear Dis-
233 criminant Analysis (LDA) to classify the target of attention from the P300
234 responses, and obtained classification accuracy of about 70%. Sugi and col-

leagues (2018) similarly employed spatially distinct sound sources (six in their case) and optimized the stimulus onset asynchrony for maximal information transfer rate; the optimal SOA was found to be 400–500 ms, which yielded over 85% accuracy when classifying the target sound source vs. all others. Heo and colleagues (2017) utilized piano and violin music, sounds of nature as well as pure tones which were all amplitude modulated at 38 and 42 Hz to elicit auditory steady-state responses. LDA classification of the EEG responses to sounds of nature yielded the highest accuracy (83%), and the authors argue that this due to the acceptance, or pleasantness, of these stimuli compared to the other stimuli in that study.

The high classification accuracy we have now obtained offline does not readily indicate high online accuracy. In an online setting, the classifier can only be trained with samples from the beginning of the recording, which may lower the classification accuracy if the responses evolve in the course of the measurement session due to adaptation or change in the mental strategy to maintain attention in one stream. In addition, all the pre-processing that we now perform offline to improve data quality may not be available online due to computational reasons.

Individual differences in response latencies and spatial patterns on the MEG sensor array may limit across-subject generalization of trained classifiers. Future studies could assess these differences and their impact on classification accuracy.

Our current results are based on MEG measurements. As a non-portable and expensive technology, MEG-based BCIs have limited applications beyond neuroscientific experimentation. However, a MEG BCI could assist the development of an eventual EEG-based BCI that could be adopted widely.

Despite the current limitations above, our paradigm and classification approach holds promise for a future BCI. The use of stimuli that directly carry the semantics of the communication or control elements and an intuitive selection task make such a BCI easy to use and likely reduce the training time of both the subject and the classifier.

5. Conclusions

We have shown that the target of auditory attention to one of two concurrent streams spoken words can be robustly decoded from single MEG responses. Our result corroborates attentional modulation of auditory evoked responses also to naturalistic stimuli. The achieved high decoding accuracy could enable the use of our experimental paradigm and classification method in an efficient and intuitive brain-computer interface.

6. Acknowledgements

The measurements were conducted at the MEG Core of Aalto Neuroimaging, Aalto University, Finland. Measurements were financially supported by Aalto Brain Centre. Authors declare no conflicts of interest.

References

- Astrand, E., Wardak, C., Ben Hamed, S., 2014. Selective visual attention to drive cognitive brain-machine interfaces: from concepts to neurofeedback and rehabilitation applications. *Front. Syst. Neurosci.* 8, 144. doi:[10.3389/fnsys.2014.00144](https://doi.org/10.3389/fnsys.2014.00144).
- Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., Naccache, L., 2009. Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U S A* 106, 1672–1677. doi:[10.1073/pnas.0809667106](https://doi.org/10.1073/pnas.0809667106).
- Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., Owen, A.M., Bekinschtein, T.A., 2013. Expectation and attention in hierarchical auditory prediction. *J. Neurosci.* 33, 11194–11205. doi:[10.1523/JNEUROSCI.0114-13.2013](https://doi.org/10.1523/JNEUROSCI.0114-13.2013).
- Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi:[10.1121/1.1907229](https://doi.org/10.1121/1.1907229).
- Combrisson, E., Jerbi, K., 2015. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* 250, 126–136. doi:[10.1016/j.jneumeth.2015.01.010](https://doi.org/10.1016/j.jneumeth.2015.01.010).

- 297 Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20,
298 273–297. doi:[10.1023/A:1022627411411](https://doi.org/10.1023/A:1022627411411).
- 299 Curtin, A., Ayaz, H., Liu, Y., Shewokis, P.A., Onaral, B., 2012. A p300-based
300 EEG-BCI for spatial navigation control. *Conf. Proc. IEEE Eng. Med. Biol.*
301 *Soc.* 2012, 3841–3844. doi:[10.1109/EMBC.2012.6346805](https://doi.org/10.1109/EMBC.2012.6346805).
- 302 Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. i.
303 segmentation and surface reconstruction. *NeuroImage* 9, 179–194. doi:[10.1006/nimg.1998.0395](https://doi.org/10.1006/nimg.1998.0395).
- 304
305 Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine,
306 J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combin-
307 ing fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*
308 26, 55–67. doi:[10.1016/S0896-6273\(00\)81138-1](https://doi.org/10.1016/S0896-6273(00)81138-1).
- 309 Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II:
310 Inflation, flattening, and a surface-based coordinate system. *NeuroImage*
311 9, 195–207. doi:[10.1006/nimg.1998.0396](https://doi.org/10.1006/nimg.1998.0396).
- 312 Furdea, A., Halder, S., Krusienski, D.J., Bross, D., Nijboer, F., Birbaumer,
313 N., Kübler, A., 2009. An auditory oddball (p300) spelling system for
314 brain-computer interfaces. *Psychophysiology* 46, 617–625. doi:[10.1111/
315 j.1469-8986.2008.00783.x](https://doi.org/10.1111/j.1469-8986.2008.00783.x).
- 316 Geronimo, A., Kamrunnagar, M., Schiff, S.J., . Single trial predictors for
317 gating motor-imagery brain-computer interfaces based on sensorimotor
318 rhythm and visual evoked potentials. *Front. Neurosci.* 10. doi:[10.3389/
319 fnins.2016.00164](https://doi.org/10.3389/fnins.2016.00164).
- 320 Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brod-
321 beck, C., Parkkonen, L., Hämäläinen, M.S., 2014. Mne software for pro-
322 cessing meg and eeg data. *NeuroImage* 86, 446 – 460.
- 323 Halder, S., Rea, M., Andreoni, R., Nijboer, F., Hammer, E.M., Kleih, S.C.,
324 Birbaumer, N., Kübler, A., 2010. An auditory oddball brain-computer
325 interface for binary choices. *Clin. Neurophysiol.* 121, 516–523. doi:[10.
326 1016/j.clinph.2009.11.087](https://doi.org/10.1016/j.clinph.2009.11.087).

- 327 Halder, S., Takano, K., Kansaku, K., 2018. Comparison of four control
328 methods for a five-choice assistive technology. *Front. Hum. Neurosci.* 12,
329 228. doi:[10.3389/fnhum.2018.00228](https://doi.org/10.3389/fnhum.2018.00228).
- 330 Halder, S., Takano, K., Ora, H., Onishi, A., Utsumi, K., Kansaku, K., 2016.
331 An evaluation of training with an auditory p300 brain-computer interface
332 for the japanese hiragana syllabary. *Front. Neurosci.* 10, 446. doi:[10.3389/
333 fnins.2016.00446](https://doi.org/10.3389/fnins.2016.00446).
- 334 Heo, J., Baek, H.J., Hong, S., Chang, M.H., Lee, J.S., Park, K.S., 2017.
335 Music and natural sounds in an auditory steady-state response based brain-
336 computer interface to increase user acceptance. *Comput. Biol. Med.* 84,
337 45–52. doi:[10.1016/j.compbiomed.2017.03.011](https://doi.org/10.1016/j.compbiomed.2017.03.011).
- 338 Hill, N.J., Moinuddin, A., Häuser, A.K., Kienzle, S., Schalk, G., 2012. Com-
339 munication and control by listening: toward optimal design of a two-class
340 auditory streaming brain-computer interface. *Front. Neurosci.* 6, 181.
341 doi:[10.3389/fnins.2012.00181](https://doi.org/10.3389/fnins.2012.00181).
- 342 Hill, N.J., Ricci, E., Haider, S., McCane, L.M., Heckman, S., Wolpaw, J.R.,
343 Vaughan, T.M., 2014. A practical, intuitive brain-computer interface for
344 communicating 'yes' or 'no' by listening. *J. Neural. Eng.* 11, 035003.
345 doi:[10.1088/1741-2560/11/3/035003](https://doi.org/10.1088/1741-2560/11/3/035003).
- 346 Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical
347 signs of selective attention in the human brain. *Science* 182, 177–180.
348 doi:[10.1126/science.182.4108.177](https://doi.org/10.1126/science.182.4108.177).
- 349 Hühne, J., Schreuder, M., Blankertz, B., Tangermann, M., 2011. A novel
350 9-class auditory ERP paradigm driving a predictive text entry system.
351 *Front. Neurosci.* 5, 99. doi:[10.3389/fnins.2011.00099](https://doi.org/10.3389/fnins.2011.00099).
- 352 Hübner, D., Schall, A., Prange, N., Tangermann, M., 2018. Eyes-closed in-
353 creases the usability of brain-computer interfaces based on auditory event-
354 related potentials. *Front. Hum. Neurosci.* 12, 391. doi:[10.3389/fnhum.
355 2018.00391](https://doi.org/10.3389/fnhum.2018.00391).
- 356 Kaongoen, N., Jo, S., 2017. A novel hybrid auditory BCI paradigm combin-
357 ing ASSR and p300. *J. Neurosci. Methods* 279, 44–51. doi:[10.1016/j.
358 jneumeth.2017.01.011](https://doi.org/10.1016/j.jneumeth.2017.01.011).

- 359 Kidd, G.J., 2017. Enhancing auditory selective attention using a visually
360 guided hearing aid. *J. Speech Lang. Hear. Res.* 60, 3027–3038. doi:[10.1044/2017_JSLHR-H-17-0071](https://doi.org/10.1044/2017_JSLHR-H-17-0071).
361
- 362 Kim, D.W., Hwang, H.J., Lim, J.H., Lee, Y.H., Jung, K.Y., Im, C.H., 2011.
363 Classification of selective attention to auditory stimuli: toward vision-free
364 brain-computer interfacing. *J. Neurosci. Methods* 197, 180–185. doi:[10.1016/j.jneumeth.2011.02.007](https://doi.org/10.1016/j.jneumeth.2011.02.007).
365
- 366 McCane, L.M., Heckman, S.M., McFarland, D.J., Townsend, G., Mak, J.N.,
367 Sellers, E.W., Zeitlin, D., Tenteromano, L.M., Wolpaw, J.R., Vaughan,
368 T.M., 2015. P300-based brain-computer interface (BCI) event-related po-
369 tentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-
370 matched controls. *Clin. Neurophysiol.* 126, 2124–2131. doi:[10.1016/j.clinph.2015.01.013](https://doi.org/10.1016/j.clinph.2015.01.013).
371
- 372 Nambu, I., Ebisawa, M., Kogure, M., Yano, S., Hokari, H., Wada, Y., 2013.
373 Estimating the intended sound direction of the user: Toward an auditory
374 brain-computer interface using out-of-head sound localization. *PLoS ONE*
375 8, e57174. doi:[10.1371/journal.pone.0057174](https://doi.org/10.1371/journal.pone.0057174).
- 376 Nijboer, F., Furdea, A., Gunst, I., Mellinger, J., McFarland, D.J., Birbaumer,
377 N., Kübler, A., 2008. An auditory brain-computer interface (BCI). *J.*
378 *Neurosci. Methods* 167, 43–50. doi:[10.1016/j.jneumeth.2007.02.009](https://doi.org/10.1016/j.jneumeth.2007.02.009).
- 379 Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch
380 negativity (MMN) in basic research of central auditory processing: a re-
381 view. *Clin. Neurophysiol.* 118, 2544–2590. doi:[10.1016/j.clinph.2007.04.026](https://doi.org/10.1016/j.clinph.2007.04.026).
382
- 383 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel,
384 O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas,
385 J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.,
386 2011. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12,
387 2825–2830.
- 388 Peirce, J.W., 2007. PsychoPy—psychophysics software in python. *J. Neurosci.*
389 *Methods* 162, 8–13. doi:[10.1016/j.jneumeth.2006.11.017](https://doi.org/10.1016/j.jneumeth.2006.11.017).
- 390 Peirce, J.W., 2008. Generating stimuli for neuroscience using PsychoPy.
391 *Front. Neuroinform.* 2, 10. doi:[10.3389/neuro.11.010.2008](https://doi.org/10.3389/neuro.11.010.2008).

- 392 Picton, T.W., 1992. The p300 wave of the human event-
393 related potential. Clin. Neurophysiol. 9, 456–479. doi:[10.1097/
394 00004691-199210000-00002](https://doi.org/10.1097/00004691-199210000-00002).
- 395 Schreuder, M., Blankertz, B., Tangermann, M., 2010. A new auditory multi-
396 class brain-computer interface paradigm: Spatial hearing as an informative
397 cue. PLoS ONE 5, e9813. doi:[10.1371/journal.pone.0009813](https://doi.org/10.1371/journal.pone.0009813).
- 398 Sellers, E.W., Donchin, E., 2006. A p300-based brain–computer interface:
399 Initial tests by ALS patients. Clin. Neurophysiol. 117, 538–548. doi:[10.
400 1016/j.clinph.2005.06.027](https://doi.org/10.1016/j.clinph.2005.06.027).
- 401 Sugi, M., Hagimoto, Y., Nambu, I., Gonzalez, A., Takei, Y., Yano, S., Hokari,
402 H., Wada, Y., 2018. Improving the performance of an auditory brain-
403 computer interface using virtual sound sources by shortening stimulus on-
404 set asynchrony. Front. Neurosci. 12, 108. doi:[10.3389/fnins.2018.00108](https://doi.org/10.3389/fnins.2018.00108).
- 405 Taulu, S., Hari, R., 2009. Removal of magnetoencephalographic artifacts
406 with temporal signal-space separation: demonstration with single-trial
407 auditory-evoked responses. Hum. Brain Mapp. 30, 1524–1534. doi:[10.
408 1002/hbm.20627](https://doi.org/10.1002/hbm.20627).
- 409 Woldorff, M., Gallen, C., Hampson, S., Hillyard, S., Pantev, C., Sobel, D.,
410 Bloom, F., 1993. Modulation of early sensory processing in human auditory
411 cortex during auditory selective attention. Proc. Natl. Acad. Sci. U S A
412 90, 8722–8726. doi:[10.1073/pnas.90.18.8722](https://doi.org/10.1073/pnas.90.18.8722).
- 413 Yeom, S.K., Fazli, S., Müller, K.R., Lee, S.W., 2014. An efficient ERP-
414 based brain-computer interface using random set presentation and face
415 familiarity. PLoS ONE 9, e111157. doi:[10.1371/journal.pone.0111157](https://doi.org/10.1371/journal.pone.0111157).