

ACCEPTED MANUSCRIPT • OPEN ACCESS

Ear-EEG compares well to cap-EEG in recording auditory ERPs: a quantification of signal loss

To cite this article before publication: Arnd Meiser *et al* 2022 *J. Neural Eng.* in press <https://doi.org/10.1088/1741-2552/ac5fcb>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2022 The Author(s). Published by IOP Publishing Ltd..

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Ear-EEG compares well to cap-EEG in recording auditory ERPs: a quantification of signal loss

Arnd Meiser*¹ and Martin G. Bleichner¹

¹Neurophysiology of Everyday Life Group, Department of Psychology, University of Oldenburg, Oldenburg, Germany

March 18, 2022

Abstract

Objective: Ear-EEG (Electroencephalography) allows to record brain activity using only a few electrodes located close to the ear. Ear-EEG is comfortable and easy to apply, facilitating beyond-the-lab EEG recordings in everyday life. With the unobtrusive setup, a person wearing it can blend in, allowing unhindered EEG recordings in social situations. However, compared to classical cap-EEG, only a small part of the head is covered with electrodes. Most scalp positions that are known from established EEG research are not covered by ear-EEG electrodes, making the comparison between the two approaches difficult and might hinder the transition from cap-based lab studies to ear-based beyond-the-lab studies.

Approach: We here provide a reference data-set comparing ear-EEG and cap-EEG directly for four different auditory event-related potentials (ERPs): N100, MMN, P300 and N400. We show how the ERPs are reflected when using only electrodes around the ears.

Main results: We find that significant condition differences for all ERP-components could be recorded using only ear-electrodes. The effect sizes were moderate to high on the single subject level. Morphology and temporal evolution of signals recorded from around-the-ear resemble highly those from standard scalp-EEG positions. We found a reduction in effect size (signal loss) for the ear-EEG electrodes compared to cap-EEG of 21-44%. The amount of signal loss depended on the ERP-component; we observed the lowest percentage signal loss for the N400 and the highest percentage signal loss for the N100. Our analysis further shows that no single channel position around the ear is optimal for recording all ERP-components or all participants, speaking in favor of multi-channel ear-EEG solutions.

Significance: Our study provides reference results for future studies employing ear-EEG.

Keywords: ear-EEG, ear-centered sensing, reference data-set, signal loss, auditory ERPs

Highlights

- Ear-EEG can detect significant differences between conditions in auditory ERPs
- The morphology of the ERPs recorded from around the ear resemble those from standard recording positions in high-density cap-EEG
- No single channel around the ear is sufficient to record all ERPs and from all participants, speaking to a multi-channel ear-EEG

1 Introduction

EEG (Electroencephalography) has been invaluable to shed light on the mechanics of cognition. In the classical use, EEG-electrodes are connected via cable to a signal amplifier. Both the rather bulky setup and the laboratory environment pose a challenge for EEG-researchers interested in more natural behavior. With

*Corresponding author: arnd.meiser@uni-oldenburg.de

the development of mobile EEG, it is now possible to leave the lab and to record EEG during motion in everyday life. However, the conspicuous cap-EEG still hampers natural behaviour and social interaction. For mobile, naturalistic and unobtrusive EEG recordings, ear-EEG poses an alternative (Bleichner et al. (2015); Debener et al. (2015); Mikkelsen et al. (2015)). In ear-EEG, electrodes are worn exclusively in- or around the ear. In combination with a small amplifier that can be clipped to the collar (or even to additional hardware, see Bleichner and Emkes (2020)) and a smartphone that stores the recorded brain signals, it is fully portable and can be used to monitor brain activity for extended periods of time (Bleichner et al. (2015); Debener et al. (2015); Bleichner and Debener (2017); Hölle et al. (2021)). As such, ear-EEG is suited to record during natural behavior where it really happens: in daily life.

Compared to high-density scalp-EEG, ear-EEG has a reduced spatial coverage of the head (i.e. electrodes only around the ears) and a lower number of electrodes. Both factors potentially reduce the sensitivity of ear-EEG for the brain signal of interest. In a recent simulation study on the sensitivity of ear-EEG to cortical sources, we showed that especially temporal areas of the cortex are a promising target for ear-EEG (Meiser et al. (2020), see also Kappel et al. (2019)). Empirical findings have shown that oscillatory as well as different kinds of event-related activity (ERPs) can be recorded with ear-EEG (Zibrandtsen et al. (2016); Bleichner et al. (2016)). However, the reduced coverage of ear-EEG compared to scalp-EEG makes the comparison of the results with either approach difficult. It might therefore hinder the transition from cap-based lab studies to ear-based beyond-the-lab studies.

An important step towards reliable ear-EEG is therefore an understanding of the signals we record from around the ear, even before moving "into the wild" where the signals become more complex and less controllable. One important factor for recordings with ear-EEG is surely the low number of electrodes: compared to classical high-density EEG, we have to expect a certain amount of signal loss (Meiser et al. (2020)). Here we want to empirically investigate what cortical signals ear-EEG is sensitive to with a quality comparable to what we are used to from scalp-EEG. In our work we define ear-EEG as an EEG configuration where all recording electrodes and the reference electrodes are exclusively close to the ear (Bleichner and Debener (2017)). We exclusively take cap-electrodes that are close to the ear and compute bipolar channels to simulate ear-EEG.

We focus our investigation on event-related potentials evoked in well-documented auditory experimental paradigms: the N100, the Mismatch Negativity (MMN), the P300 and the N400.

First, we want to know how well these ERPs can be recorded with ear-EEG compared to scalp-EEG. What differences in morphology and effect size can be expected on group- and on the individual level? That is, for different ERPs, we provide here a reference of what can be expected for the transition from cap- to ear-EEG. With ear-EEG, we want to study people's brain activity in everyday life situations. Hence, unlike in the lab, each person is exposed to a different environment. The focus in signal analysis therefore needs to be on the individual. To this end, the second aim of this study is to investigate how effect sizes recorded with ear-EEG vary between subjects.

Our third goal is to test the reliability of ear-EEG recordings within the individual. From a usability and product development point of view, ear-EEG should have few electrodes at fixed positions to be easily integrated into a hearing aid or headphones (Fiedler et al. (2016)). This would require that the electrode positions should be equally well suited for different people, as well as for different ERPs within a person. In theory, neither is warrantable. Whether or not a signal can be recorded with a set of electrodes depends on the orientation of the dipole relative to the electrodes (Meiser et al. (2020)). ERPs that originate in different parts of the cortex will have differently oriented dipoles that are more or less easily picked up with a set of

electrodes. Furthermore, due to differences in the cortical folding, a signal that originates from the same anatomical structure in different participants may still project differently to the scalp surface (Ahlfors et al. (2010), Kanai and Rees (2011)). We investigate here whether those factors are relevant in practice.

In summary, first we characterize the ERPs recorded with ear-EEG and compare them to scalp-EEG by quantifying the signal loss from scalp to ear. Second, we are looking into the inter-subject variability of ERPs as recorded with ear-EEG. Third, we investigate if a specific subset of ear-electrodes is optimal for recording across individuals and ERPs. Our results show the potential and limitations of ear-EEG.

2 Methods

2.1 Participants

The conduction of this study was approved by the ethics committee of the University of Oldenburg. Participants were recruited via the black board of the online platform of the University of Oldenburg and received financial compensation for their participation. We recorded from 20 right-handed, healthy participants (10 female) with no self-reported history of hearing impairment. Participant's age ranged from 20 to 31 ($M = 25.5$ years, $SD = 3.4$), all gave written informed consent. Data from two participants had to be excluded because of insufficient data quality as defined in 2.5.4, leaving 18 participants (10 female) for analysis.

2.2 Tasks and Stimuli

2.2.1 Task selection

We want to study different aspects of auditory cognition that are also relevant in everyday life. We used standard tasks of attentive and passive listening, selective listening and more complex language processing. The complexity of the stimuli we used ranged from simple tone bursts to spoken sentences.

The first ERP we investigated is the N100. The amplitude of this ERP-component is known to be influenced by attention: when attention is directed towards an auditory stimulus, the N100-response is larger than when the stimulus is ignored (Choi et al. (2013), Bleichner et al. (2016)).

The second ERP-component investigated is the Mismatch Negativity (MMN). It is elicited in a classical auditory oddball task and appears whenever a repeating stream of tones is interrupted by a deviant tone. The MMN appears during passive listening and does not require active attention to an auditory stimulus (Naatanen (1987), Näätänen et al. (1993), Näätänen et al. (2001)). The MMN is even present when the tones are actively ignored and attention is directed elsewhere.

The third ERP-component investigated is the P300. It is elicited in a similar way as the MMN, but unlike it, the P300 reflects active listening to the auditory stimulus played (Halgren et al. (1980); Başar-Eroglu et al. (2001); Bledowski et al. (2004); Watkins et al. (2007); Polich (2007) and therefore requires focused attention.

The last component is the N400. The N400-response is studied in the context of semantic processing (Kutas and Hillyard (1980)) and hence reflects a more in depth analysis of an acoustic stimulus.

2.2.2 Task description

Presentation order of the tasks was quasi-randomized and included a short break after each task. If not indicated otherwise, we instructed participants to direct their gaze to a fixation cross on eye-level during the task. The total duration of all tasks was approximately 90 minutes. Details of the task-procedures can be found in Figure 1. Stimuli were presented using Presentation® (Version 20.1, Neurobehavioral Systems,

Inc., Berkeley, CA, www.neurobs.com). Each task started with a 60s count-down so that participants had time to settle down and take a comfortable sitting position.

For the N100-task, we adapted our paradigm from Bleichner et al. (2016), who adapted it from the original study of Choi et al. (2013). Three concurrent sound streams were played to participants from three different directions (left, center and right). Each stream consisted of a different number of tones: four (left), three (center) and five tones (right), lasting 0.75s, 1.0s and 0.6s, respectively, so that each stream lasted for exactly three seconds (see Figure 1). Each stream was played by a different instrument: cello (left), clarinet (center) and oboe (right). Each instrument played their tones at two different pitches, leading to sequences that were either ascending (e.g. left: low/low/high/high), alternating (e.g. center: low/high/low) or descending (e.g. right: high/high/low/low/low). The differences in number, onset latency and pitch ensured that the streams were perceptually easily distinguishable. The task of the participant was to attend either to the left or to the right stream (never the center stream) and to indicate after each trial whether the sequence was ascending, alternating or descending via button press. Unlike in the original study, where the direction of attention was cued, here the participants indicated their intended direction of attention by pressing one of two buttons prior to each trial. This button press initiated the start of the sound playback. We instructed participants to avoid any obvious pattern in their choice of the attended direction and to instead keep their choices as unforeseeable as possible to an external observer while still maintaining a ratio of approximately 50% of attending the left and right side. The task had a duration of approximately 30 minutes, including 150 trials. To familiarize the participants with the individual tones, we presented the different tone sequences with three examples (ascending, alternating and descending for left, center and right), followed by 10 practice trials. The N100 we see for each tone depends on whether the participant directed their attention to it or not (Choi et al. (2013); Bleichner et al. (2016)). We refer to this ERP-component as N100(att.) from here on out.

The design of this task allows us to extract an additional N100-component that is independent of the direction of attention, i.e. the N100 in response to the onset of the three sound streams. This first onset has the same latency for the three directions. The tones are indistinguishable and therefore not subject to attention processes. This component is referenced to as N100 to separate it from the N100(att.)-component.

To elicit the MMN and the P300, an auditory oddball paradigm was used (adapted from Kappenman et al. (2021)). Participants listened to a stream of two tones: a rare target tone was embedded into a stream of repeated standard tones in a ratio of 20:80. Duration of the tones were 124ms (20ms ramp on/off), played at a frequency of 1000Hz and 800Hz for the target and standard tone, respectively. In total, 1000 tones were played at jittered intervals between 800ms and 1000ms. Participants heard the target and standard tones once prior to the task. Duration of each task was approximately 15 minutes. To elicit the MMN, participants watched a silent short video (Sand Art-”Return to yourself” by Kseniya Simonova, find the video in Kappenman et al. (2021)), with the instruction to ignore all tones.

To elicit the P300 response, participants were instructed to actively attend to the sound stream and silently count the target tones. Afterwards, participants reported the number of counted targets to the experimenter. In the N400-task, subjects heard sentences spoken by a female voice in German that either ended in a semantically meaningful or meaningless way, depending on the last word of the sentence (e.g., “The woman poured the wine into the glass” / “The Woman poured the wine into the wardrobe”), with stimuli taken from Goregliad Fjaellingsdal et al. (2016). Participants indicated for each sentence via left or right button press if the sentence was congruent (i.e., meaningful) or incongruent (i.e., meaningless). Participants had to answer within two seconds to prevent “overthinking” the meaningfulness of a sentence. We presented

participants with 200 sentences (50% congruent) in a randomized order. Before the experiment started, two training sentences (one congruent, one incongruent) were presented to the participants. The task duration was approximately 30 minutes, with a break after two thirds of the task.

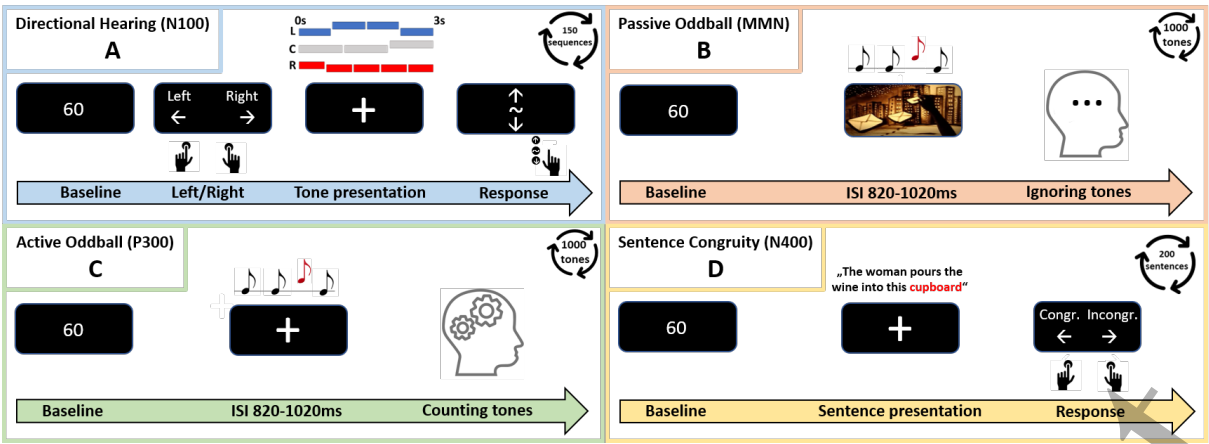


Figure 1: Four auditory paradigms. Each Box depicts a different task from which an ERP is extracted. The arrows indicate the time course of one trial, the circle-shaped arrows in the upper right corner of each box indicate the number of trials presented. In each box, the upper part represents the auditory stimulus, the middle part shows what participants saw on the monitor and the lower part shows participant decisions. Each task started with a 60s count-down. In **A**, participants indicate via left or right button press whether they would attend the left or right sound stream. The tones are then presented, each direction with either a ascending (c), alternating (l) or descending (r) sequence. After tone presentation, participants indicated their response via button press (high, middle or low button). In **B**, participants watched a silent movie while ignoring a stream of rare and frequent tones. In **C**, participants heard the same tones, this time with the instruction to silently count the target tones. In **D**, participants listened to sentences and had to indicate via button press whether the sentence ended either in a meaningful or a meaningless way.

2.3 Procedure

Before the beginning of the EEG recording, participants gave informed, written consent to participate in the study and filled out a questionnaire to confirm their right-handedness (according to the Edinburgh handedness inventory, (Oldfield (1971))). Participants washed their hair in the lab. We prepared the EEG-cap using alcohol swabs and abrasive gel (Abralyt HiCl, Easycap GmbH, Germany). We seated our participants in a soundproof booth with a monitor in front of them in one-meter distance at eye-level. Sounds were presented via two Sirocco S30 loudspeakers (Cambridge Audio, London, United Kingdom), positioned at 45° to the left and right of the participant. Before each task, loudness of the auditory stimuli was adjusted to a comfortable level.

2.4 EEG acquisition

We recorded EEG from a 96-channel Ag/AgCl EEG-cap (Easycap, Hersching, Germany). Electrodes were equidistantly placed with a central fronto-polar site as ground and recorded against a nose-tip reference. Data was collected with a BrainAmp EEG amplifier-system (BrainProducts, Gilching, Germany) on the BrainVision Recorder (BrainVision Recorder, Version 1.20.0506, Brain Products GmbH, Gilching, Germany). As shown in Figure 2, electrodes are labeled from E01 to E96. Ear-EEG was defined as channels around both ears (E18, E24, E71, E72, E85, E92 on the right ear and E22, E28, E80, E81, E90, E95 on the left ear). These electrodes are termed R1 to R6 for the right ear and L1 to L6 for the left ear when we refer to them as ear-electrodes. Recordings were made at 1000Hz sampling frequency. Impedances were kept below 10 kΩ prior to the start of the experiment and improved again when necessary in every break between the tasks.

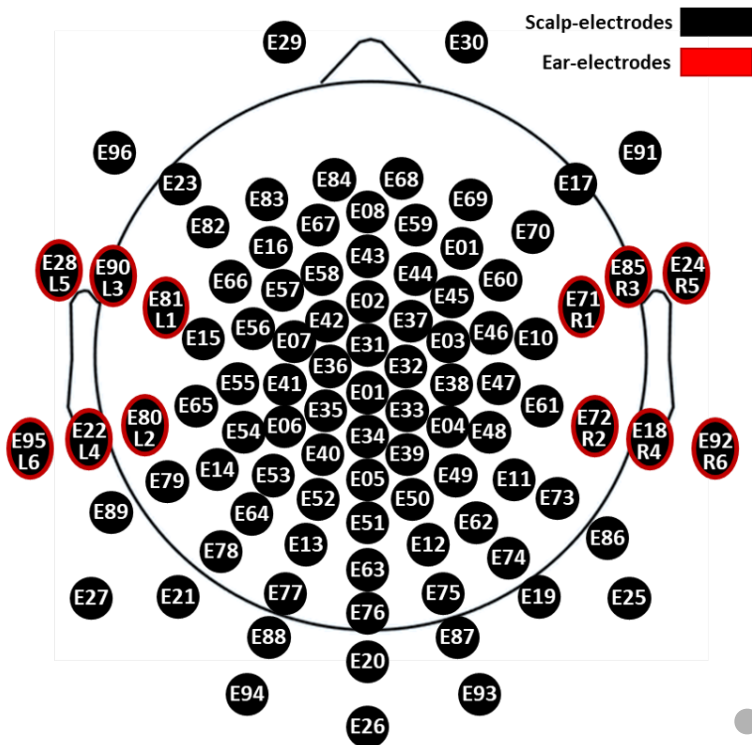


Figure 2: Channel layout of a 96 channel EEG cap. E29 and E30 are placed below the eyes. Red circles indicate the ear electrodes as a sub-selection of scalp-electrodes.

2.5 Analysis

2.5.1 Filtering and ICA

The processing steps from EEG rawdata to ERP- and effect size-measures for ear-EEG and two configurations of scalp-EEG are illustrated in Figure 3. As a first step, data was split into ear- and scalp-data. Ear-data was constructed by removing all non-ear-channels (see Figure 2) from the data-sets. We processed both data variants using the *EEGLAB*-plugin (v2020.0) for MATLAB (Delorme and Makeig (2004)). Continuous data was band-pass filtered offline between 1Hz and 30Hz for all tasks with the *pop.eegfiltnew* function implemented in *EEGLAB* (FIR filter) and down-sampled to 100Hz. Exclusively for the scalp-, but not the ear-configuration, we performed an Independent Component Analysis (ICA) for each task using the *CORRMAP* plug-in for *EEGLAB* (Viola et al. (2009)) to remove eye-blinks, lateral eye movement and heart-beat artefacts. *CORRMAP* requires a manually chosen template for each artefact type and continues to identify components based on topographic similarity. The resulting ICA weights of these components were mapped onto the unprocessed raw data and filtered again with the same band-pass filter described above. The details of this approach can be found in Stropahl et al. (2018).

2.5.2 Event-Related Potentials

For all tasks, data was cut into 1-second epochs, -200ms to 800ms around tone onset. For the N100(att.), epochs were created for each individual tone of the attended and unattended sides (four left, five right tones). We then averaged over all tones, excluding only the first one since tones from every direction appeared simultaneously. From this first, overlapping tone, we extracted a 1-second epoch for the N100-component. For N400, data was epoched around the onset of the critical last word. For the MMN and the P300, data was epoched around the onset of target and standard tones. The resulting ERPs for all tasks were baseline corrected (-200ms to 0ms before tone onset). We removed epochs with unrealistically high amplitudes (± 3 standard deviations) with the *pop.jointprob* function implemented in *EEGLAB*.

For the scalp-EEG, we used standard positions for the investigation of ERPs that allowed us to relate our

results to the existing literature (N100/N100(att.): Choi et al. (2013), E43, corresponding to Fz; MMN: Näätänen and Alho (1995), E31, corresponding to FCz; P300: Katayama and Polich (1999), E05, corresponding to Pz; N400: Perrin and Garcia-Larrea (2003), E34, corresponding to CPz). See Figure 2 for the channel-layout. In the remainder of the manuscript, this configuration of standard channels is termed scalp (sta.). Importantly, despite the fact that these channels are widely used to report ERP results for the respective tasks, these channels are not necessarily optimal for each participant, and there might be other channels that capture the effect of interest better on the group- and the individual level. To identify the channel with the maximum effect size, both for scalp- and ear-EEG, we computed all possible bipolar channel-combinations per configuration. For the scalp-EEG, this results in 4560 unique bipolar channel pairs (i.e. number of unique combinations of two channels). This configuration of individualized bipolar channel pairs is termed scalp(ind.). The same was done for the 12 ear-electrodes. This results in 36 unique bipolar channel pairs, when electrodes around both ears were considered, and in 15 unique combination when only the electrodes around one ear were used (see Figure 2). These configurations are referred to as ear bilateral, ear right, and ear left, respectively.

2.5.3 Performance measures

We calculated Hedges' g, effect size over the time for all ERPs using the *mes* function in MATLAB (R2012a, The Mathworks Inc. Natick, MA, USA). By convention, effect sizes are interpreted as small, medium or high with values up to 0.2, 0.5 and 0.8, respectively. Effect sizes were calculated from the ERPs of two conditions in each task: one target condition and one control. For the MMN and the P300, those were frequent tones (control) and infrequent tones (target). For the N100 (att.) they were attended and unattended tones and for the N400 they were congruent and incongruent sentences. The N100 from the first tones in the directional hearing-task only had the control condition (only tone-bursts without a contrasting condition). Effect sizes were therefore calculated against the baseline before tone-onset. We pre-defined task-specific time windows (see 1) and identified the largest absolute deflection in effect size for both EEG-configurations (ear- and scalp-EEG). We then took the channel with the highest peak in the corresponding time-window (see Figure 4). For scalp(sta.), we used the pre-defined channel for ERPs and effect size measures.

To investigate the experimental effects recorded from the best-performing channel (difference between target- and control condition), we performed mass univariate analysis for all EEG-configurations and all times points using the *tmax* function (*tmax*, Mass Univariate ERP Toolbox, http://openwetware.org/wiki/Mass_Univariate_ERP_Toolbox) implemented for *EEGLAB*. It is a permutation test with 2500 permutations, including a strong control of the family wise error rate (FWER) as described in Groppe et al. (2011).

We further compared the inter-individual spread of effect sizes in ear-EEG (bilateral, right and left) to scalp-EEG. We did so by calculating the signal loss (the decrease of effect size) from scalp(ind.)- to ear-EEG and tested whether this signal loss differs between tasks. To this end, we calculated an *ANOVA* as implemented in the *anovan* function in MATLAB, followed by a post-hoc Tukey-Kramer-test as implemented in the *multcompare* function. The *ANOVA* therefore comprised the two factors of the ERP investigated (N100, N100 (att.), MMN, P300, N400) and the configuration of ear-EEG used (left, right and bilateral).

2.5.4 Exclusion criteria

To assure data quality, we used the default settings of the *pop_clean_rawdata* function in the *EEGLAB*-toolbox. The function identifies bad channels based on correlation with neighbouring channels, a flat-line

criterion and overall variance of the channel-data. Artefactual channels were identified separately for each sub-task of each subject. If one of the ear-electrodes was identified as artefactual, we excluded the complete sub-task from further analysis. A total of 19 sub-tasks from 8 different participants were excluded, leaving 61 sub-tasks for analysis

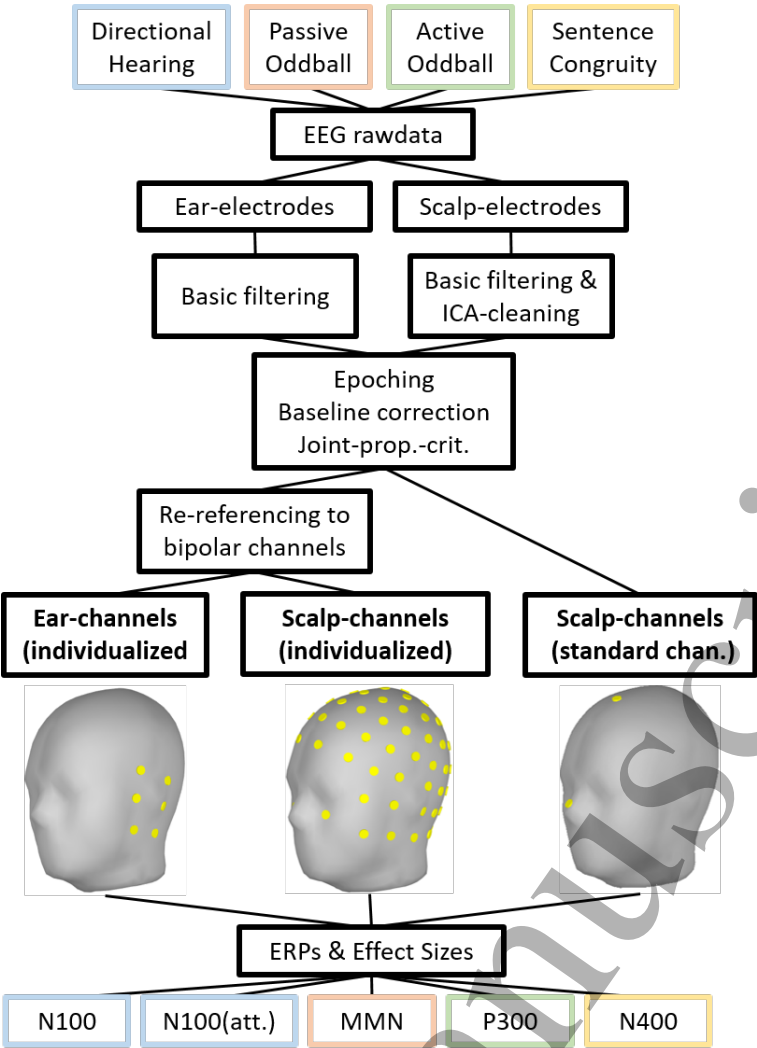


Figure 3: Pre-processing pipeline for three EEG-configurations. Rawdata recorded from each task is split into ear- and scalp-electrodes. Both configurations receive basic filtering. Scalp data is additionally undergoing ICA-cleaning to remove artefacts from eye-blinks, eye-movement and heartbeat. Data is then epoched. For the individualized configurations of ear and scalp, data is re-referenced to bipolar channels. For all three configurations, we extract five Event-Related Potentials and corresponding effect sizes.

3 Results

Our analysis comprised the calculation of five ERPs from auditory tasks: The N100, the N100(att.), the MMN, the P300 and the N400. We calculated effect sizes for the ERPs in target and control conditions and compared them, both on a group- and individual level. We further investigated the effect size of each bipolar channel within and between individuals.

3.1 Behavioral data and task compliance

We evaluated the performance measures (where available) to check that subjects followed the instructions for the tasks. In the N100-task, subjects correctly identified the tone sequences as ascending, alternating or descending in $M = 77\%$ ($SD = 19\%$) of the trials (chance level at 33%). The decision prior to each trial to attend to either the right or to the left tone sequences resulted in 49% attend-right-decisions. This percentage indicates that participants were able to maintain the desired ratio of 50:50. In the P300-task, participants counted on average $M = 197$ ($SD = 8$) of the 200 target tones, indicating that they were able to

maintain focus throughout the task. In the N400 task, subjects correctly identified sentences as congruent or incongruent in $M = 95\%$ ($SD = 12\%$) of the trials. Reaction times in this task were 466ms on average with a standard deviation of 155ms. Judging from both the number of correct answers and the reaction times within the allowed 2 seconds, participants were able to follow the task instructions .

3.2 Event-related potentials

In the pre-processing of the data for the final ERPs, the ICA led to the removal of $M = 4.65$ components ($SD = 0.20$) across tasks, including eye-blinks, lateral eye-movement and heart-beat artefacts. The joint probability criterion led to the removal of an average $M = 4.22\%$ of the epochs ($SD = 0.16\%$) across tasks. In Figure 4, we show the event-related potentials per task on the group level. In the left column, we show topographies for the scalp(sta.) configuration. Topographies and their latencies for each task were taken from the time-points of maximum effect size as described in 2.5.3. In the center- and right columns, we show epochs from -200ms to 800ms after stimulus onset. The center columns show the ERPs of the scalp-configurations (sta. and ind.) and the ear-configuration (bilateral), averaged over subjects. The right columns show the effect sizes corresponding to the ERPs on the group- and on the individual level, again for the three configurations. Key data about the ERPs and effect sizes illustrated in Figure 4 can be taken from 1A and B.

The topographies from the scalp(sta.)-configuration in Figure 4A are similar to what is to be expected from literature. An exception is the right-lateralized negative activation in the temporal region of the scalp in the MMN-task.

In the N100-task, there is a clear N100 visible for all three EEG-configurations, with the largest deflection for the scalp (ind.)-configuration. The ear-configuration shows a smaller, yet pronounced deflection.

In the N100(att.)-task, all three configurations show a pronounced negative deflection for the attended tones at around 200ms, followed by a positive deflection at around 440ms, both in line with Bleichner et al. (2016). While the peak of the target condition in the scalp(sta.)-configuration is larger than in the ear-configuration in the center column (right), the standard deviation is highest for the scalp(sta.)-configuration. In the effect size-measures in the right column (left), both the ear- and the scalp(ind.)-configuration show higher effect size peaks than in the scalp(sta.)-configuration.

In the MMN-task, the ERPs show a visible negative deflection at around 200ms. This deflection is more pronounced for the target tones than for the standard tones, as was to be expected (Näätänen et al. (1993, 2007)). Again, the ear-data shows a larger effect size in the right column (left) than for the scalp (sta.)-configuration, while the scalp(ind.)-configuration shows the largest effect size.

In the P300-task, there is a positive deflection at around 300ms for the target tones, but not for the standard tones for ear-, scalp(sta.)- and scalp(ind.)-ERPs, as is described in common literature (Katayama and Polich (1999); Bennington and Polich (1999); Howard and Polich (1985)). Hedges' g is largest on average for the scalp(ind.)-configuration, second largest for the scalp(sta.)-configuration and lowest for the ear-configuration. It is therefore the only ERP-component for which effect sizes of the scalp (sta.)-configuration are higher than for the ear-configuration.

In the N400-task, we see a negative deflection around 420ms in all configurations that is larger for incongruent than for congruent sentences. Like the other ERPs we investigated, this is in line with previous research (Chwilla et al. (1995); Koivisto and Revonsuo (2001); Perrin and Garcia-Larrea (2003)). Hedges' g is largest for the scalp(ind.)-configurations, followed by the ear-configuration. The scalp(sta.)-configuration shows the lowest effect size.

Importantly, all five ERPs are visible in the ear-channels to a degree that is comparable to the scalp(ind.)-configuration and even better for all but the P300-component compared to the scalp(sta.)-configuration. The differences between target and control configurations all exceed the critical t-scores in the mass univariate analysis in the pre-defined time-windows. For the graphical illustration see the Supplementary Information, Figure 7. Morphology and latencies are highly similar in all three configurations. Average effect sizes range from low (0.2) to moderate (0.5) for the ear-configuration. Effect sizes on the group level are high (greater than 0.8) for all five ERP-components.

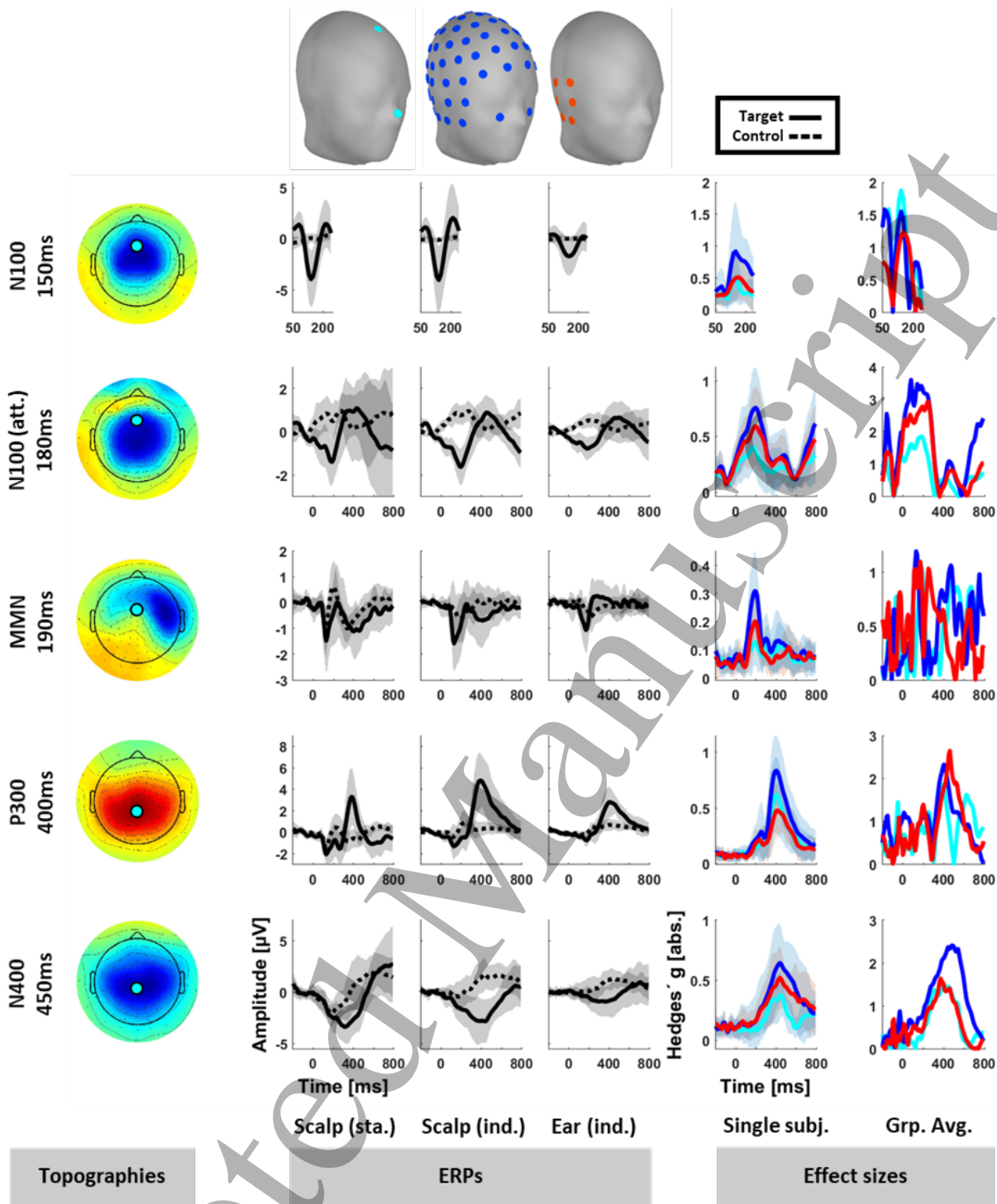


Figure 4: ERP-based results for auditory perception tasks. **Left column:** Topographies of scalp-EEG at the latency of the peak effect size for the N100, N100(att.), MMN, P300 and N400. **Center columns:** ERPs for the standard scalp channel (left, cyan), the individualized scalp channel (center, blue) and ear (right, red). The task condition is shown as a solid black line, the control condition as a dashed black line. The standard deviation is in the shaded area. **Right columns:** Effect sizes based on the single subject (left) and the group average (right).

3.3 Signal loss

Table 1C shows the decrease in effect sizes from the scalp(ind.)-configuration (with the highest effect sizes across ERP-components) to all other configurations. Those include the scalp(sta.)-configuration and the ear-

configurations (bilateral, left and right). Overall, the reduction of effect size from the best scalp-channels to the best ear-channels ranges from 44% for the N100 to 21% for the N400. When using only channels from the left ear, reduction of effect sizes range from 50% for the N100 to 29% for the N400. When using only channels from the right ear, signal losses range from 49% for the N100 to 32% for the N400. The ANOVA shows significant main effects for ERP-component investigated ($F(4,61), p < 0.001$) and ear-EEG used ($F(2,61), p < 0.001$) and no interaction effect ($F(8,61), p = 0.8673$). The analysis shows here that the signal loss from the scalp(ind.)-configuration to the ear-configurations is not the same for all ERP-components. For the different ear-configurations, the post-hoc comparisons indicate a significantly different signal loss between bi- and unilateral ear-EEG, but not between left- and right unilateral ear-EEG. A detailed plot of the comparisons can be found in the Supplementary Information, Figure 8. Signal losses from the scalp(ind.)-configuration to the scalp(sta.)-configuration range from 57% for the N100 to 30% for the P300.

A		Scalp (standard)			Scalp (individualized)			Ear (bilateral)			Ear (unilateral right)			Ear (unilateral left)		
Amplitude [μV]																
Task	Time-window	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
N1	50ms-250ms	-3.84	-4.22	2.74	-2.93	-4.33	3.47	-1.88	-2.18	1.85	-1.95	-2.21	0.87	-1.40	-1.81	1.05
N100 (att.)	50ms-250ms	-2.30	-2.36	1.17	-1.24	-1.06	0.71	-0.50	-0.53	0.47	0.09	0.15	0.60	0.21	0.12	0.48
MMN	130ms-230ms	-1.52	-1.40	1.07	-1.61	-1.77	0.97	-0.84	-1.10	1.08	-0.16	-0.08	0.60	-0.31	-0.34	0.43
P300	250ms-500ms	3.89	4.48	2.42	4.70	5.31	2.97	2.31	2.67	1.11	1.12	1.39	1.17	0.64	0.98	1.19
N400	400ms-600ms	-4.03	-4.08	2.11	-1.43	-1.92	2.14	0.21	-0.16	1.70	0.12	-0.09	1.49	-0.70	-1.02	1.76

B		Scalp (standard)			Scalp (individualized)			Ear (bilateral)			Ear (unilateral right)			Ear (unilateral left)		
Hedges'g [a.u.]																
Task	Time-window	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
N100	50ms-250ms	0.44	0.59	0.33	1.19	1.38	0.46	0.72	0.77	0.28	0.65	0.70	0.30	0.62	0.71	0.30
N100 (att.)	50ms-250ms	0.49	0.46	0.22	0.91	0.85	0.29	0.62	0.66	0.31	0.51	0.60	0.31	0.48	0.54	0.26
MMN	130ms-230ms	0.16	0.17	0.10	0.33	0.38	0.13	0.24	0.25	0.08	0.20	0.22	0.08	0.20	0.21	0.07
P300	250ms-500ms	0.63	0.70	0.32	1.00	1.00	0.28	0.55	0.58	0.21	0.51	0.52	0.19	0.53	0.55	0.22
N400	400ms-600ms	0.37	0.41	0.19	0.75	0.83	0.28	0.57	0.65	0.20	0.48	0.58	0.19	0.54	0.56	0.20

C		Signal loss (Hedges'g) to Scalp (ind.)			
Signal loss [%]		Scalp (sta.)	Ear (bilateral)	Ear (unilateral right)	Ear (unilateral left)
N100		56.97	44.24	49.53	48.47
N100(att)		45.33	22.28	28.95	37.05
MMN		54.63	33.05	41.87	45.78
P300		29.60	41.81	48.22	44.84
N400		50.27	21.20	29.78	32.11

Table 1: Response peak statistics and signal loss (percentage reduction of effect size relative to best scalp configuration). In **A**, we show the amplitude peaks of the time courses for ERPs (difference waves, by subtracting the target from the control conditions) and Hedges'g per task. The task is indicated in the left-most column, followed by the time-window in which the largest peak was searched for. In the following 3 columns, we see the EEG-configuration shown in Figure 4: scalp-EEG (standard channel and individualized channels) and ear-EEG (individualized channels). In addition, the two right-most columns show the peaks for the best channel around only the left and right ear, respectively. For each configuration, we show the median, mean and standard deviation of the largest peak. In **B**, we show Hedges'g for the same configurations. **C** shows per task the signal loss from the scalp (ind.)-configuration to the other four configurations in percent.

3.4 Effect size spread

Figure 4 shows ERPs for all tasks and low to moderate effect sizes for the ear-EEG on the subject average. Yet, the shaded areas in Figure 4 (effect sizes) hint to a considerable spread of effect sizes between subjects. We show in Figure 5 the individual effect sizes recorded from the best channel in a boxplot. From the scalp(ind.)-configuration to the ear-configurations, scores decrease since the ear-channels are a subgroup of the scalp-channels. The few cases where Hedges'g is higher for the ear-configurations can be explained only by the different pre-processing. From the ear-configuration (bilateral) to ear(right), the scores decrease by only $M = 10\%$ ($SD = 1\%$) on average across tasks and by only $M = 12\%$ ($SD = 6\%$) for ear(left). On average, in $M = 43\%$ of the cases over all tasks ($SD = 10\%$), the best channel of the right ear is also the best channel of the bilateral configuration, i.e. the best channel is calculated from electrodes on the right ear. The best channel from the left ear is also the best channel from the bilateral configuration in $M = 38\%$ of the subjects over tasks ($SD = 12\%$). This means that in approximately 40% of the subjects, there was no benefit from using electrodes from around both ears rather than from one.

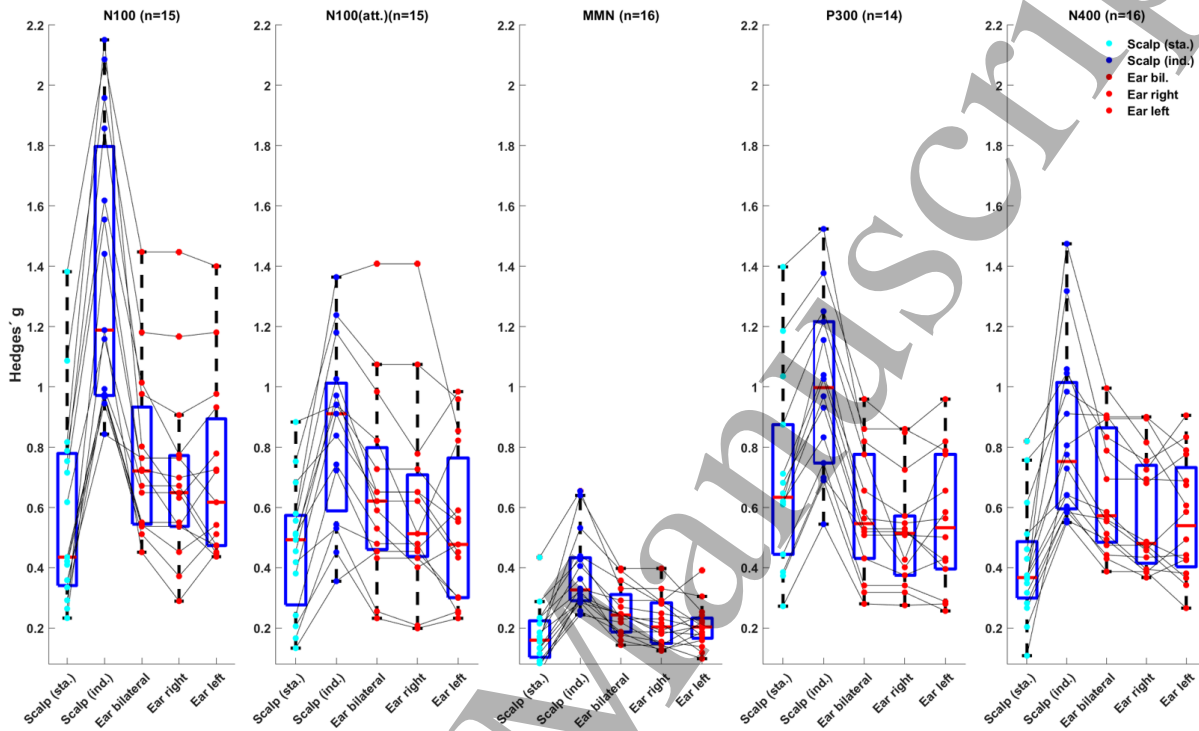


Figure 5: Boxplots of individual effect sizes over tasks and over EEG-configurations. We show maximum effect sizes from the best channel per subject. We plot highest effect sizes from scalp channels (sta. and ind.) against those from from bilateral and from unilateral ear channels (left and right). The number of subjects per task is shown as n above each column. Scalp(ind.) is shown in blue, scalp(sta.) in cyan, the ear-configurations in red. Solid lines connect each subject across configurations. The red lines in each box indicate the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data-points.

3.5 Individual channel orientation

From the unilateral channels (ear(right) and ear(left)), we investigated if channel orientation is suited to explain the performance of the best channels per task. Figure 4A and B show on the right the classification of all 15 bipolar channel combinations per ear as either horizontal (red), vertical (blue), right-diagonal (from lower left to upper right, purple) or left-diagonal (upper left to lower right, green). Within an orientation, channels are sorted from highest to lowest inter-electrode distance. In the illustrations on the left, we show in the upper panels maximum amplitudes in the pre-defined time-windows for each channel (x-axis, labeled according to their orientation) and each subject (y-axis). Each column is one ERP-component. In the lower panels, we show Hedges'g-scores, normalized to values from 0 to 1.

Amplitudes decline as a function of inter-electrode distance; the smaller the distance, the lower the amplitude. Within a group, this pattern is clearly visible in each task and for both left- and right unilateral ear-EEG. For the effect sizes, this pattern is not apparent: in none of the tasks we see a pronounced relation between channel orientation and effect size. Overall, neither distance nor orientation of the ear-channels show a visible pattern to explain effect size.

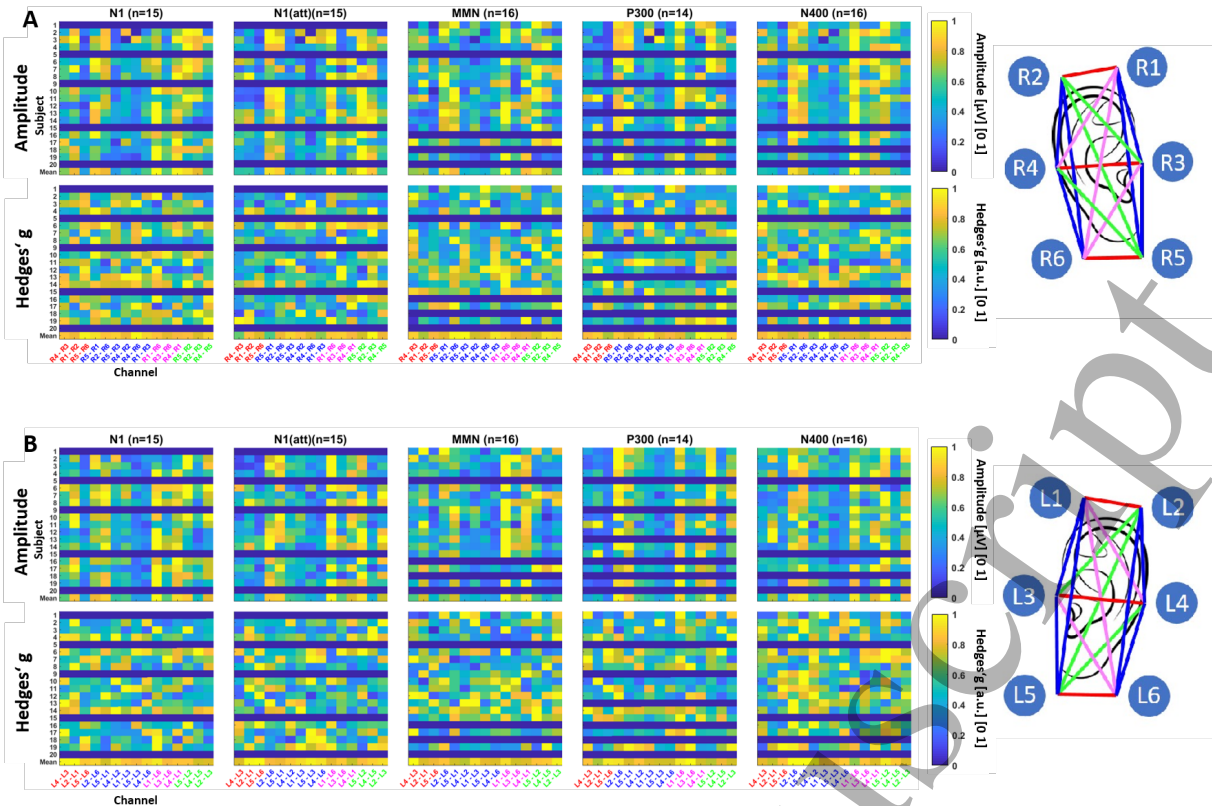


Figure 6: Response peak amplitudes and effect sizes of the best channels, sorted by orientation and distance of the channel. We show maximum amplitude values from all bipolar channels per participant for five ERPs. **A** shows combinations of channels from the right ear, **B** the left ear. Amplitudes are normalized to values between 0 and 1. The graphic displays subjects on the y-axis with the last row as the mean over all subjects. Channel combinations are shown on the x-axis. Values are color-coded from blue (low) to yellow (high). Dark blue lines over all channel combinations indicate that we rejected the subject from analysis for the particular task. Channel labels on the x-axis are color-coded according to their orientation. Orientation is displayed on the upper right of A and B. Labeled electrodes forming a bipolar channel are connected with lines. Red, blue, purple and green color indicate horizontal, vertical, right-diagonal (from lower left to upper right) and left-diagonal (upper left to lower right) channels, respectively.

4 Discussion

Ear-EEG records significant ERPs

We provide here a reference data-set for ear-EEG to quantify how recordings relate to the well-understood scalp-EEG. As we know that ear-EEG is feasible for auditory studies (Debener et al. (2015); Bleichner et al. (2016); Mirkovic et al. (2016)) and that primarily signals from temporal areas are recorded (Meiser et al. (2020); Kappel et al. (2019)), we study here auditory attention tasks. With ear-EEG, our results show significant differences between target and control ERPs in all five investigated ERP-components. Effect sizes found in ear-EEG range from medium to high on average over individuals, which is smaller than what is found for scalp-EEG with individualized channels, but is larger than for standard recording positions. The results for ear-EEG in our study are in agreement with the literature comparing ear- and scalp-setups (e.g. Bleichner et al. (2016); Gu et al. (2018); Kappel et al. (2017); Mikkelsen et al. (2015)). In these studies, specialized equipment for the recording of ear-EEG was used, rather than a sub-selection of electrodes from scalp-EEG. Our results are independent of specialized hardware and therefore can be generalized more easily.

The results converge with those reported in existing literature.

Ear-EEG recordings of similar quality compared to scalp-EEG

In ear-EEG, the morphology of ERP-components resembles those of the scalp-EEG, suggesting that ear-EEG is sensitive to similar sub-components. This is in accordance with previous work where features found in scalp-EEG were also present in ear-EEG, such as ERPs (Blechner et al. (2016)), oscillations (Zibbrandtsen et al. (2016)) and physiological artefacts (Kappel et al. (2017)).

In our study, amplitudes recorded from around the ear are smaller than those recorded from a standard position or from the individualized bipolar channels from the entire scalp. That is trivially explained by the larger inter-electrode distances for the scalp configurations. More importantly, our results show that the effect sizes of ERPs can be recorded equally well and sometimes even better using individualized ear-EEG compared to scalp-EEG. In fact, ear-EEG always recorded higher effect sizes than the standard channels, with the exception of the P300-component. Our results show that signal losses from scalp- to ear-EEG are about a third of the effect size. Yet, the ear-configuration still ranges from moderate to high effect sizes for all but the MMN-component. Therefore, even when effect sizes are low compared to individualized scalp-EEG, ERPs could still be satisfactorily recorded. That is, we recorded moderate to high effect sizes on the individual level.

The signal loss from scalp- to ear-EEG depends on the ERP-component

For all EEG-configurations, the magnitude of effect sizes differs between ERP-components. Importantly, the drop in effect size from scalp- to ear-EEG differs from ERP to ERP as well. So whether an ERP-component can be recorded well with ear-EEG cannot be inferred from scalp-EEG alone. Yet, our results show that while signal losses differ between ERP-components, the overall ranking of subjects remains constant, i.e. a subject whose effect size is high in the scalp-configuration also has a high effect size in the ear-configuration. This was true for all investigated ERPs. Therefore, on a single subject level, we find our results to be transferable from ear- to scalp-EEG.

Short screenings could improve studies using ear-EEG

For each ERP-component, we see considerable variability in the effect sizes between subjects. As mentioned above however, the ranking of a subject is stable over ERP-components, meaning that some subjects have low effect sizes across tasks. Thus, a reliably recorded ear-EEG cannot be guaranteed for every individual. To see if we can record satisfactory ERPs (or other brain signals) from a subject, it is recommendable to perform a short screening prior to long-term ear-EEG recordings (Hölle et al. (2021)). Our results show that the effect size recorded from a single ERP-component yields information about how well the other components can be recorded as well. Notably, a few subjects being difficult to record from is a problem common to all EEG-research and is not exclusive to recordings from around the ear.

Unilateral ear-EEG records ERP-components well

In the context of an unobtrusive EEG, it is desirable to use a unilateral ear-EEG, because it means less visible equipment and preparation time needed. While in our study, unilateral ear-EEG shows slightly higher signal losses than bilateral ear-EEG, those signal losses are in the range of only 10-15%. Therefore, the practical impact of these differences between uni- and bilateral ear-EEG is small. This is supported by the observation that there is no significant difference between left- and right unilateral ear-EEG, showcasing

that the unilateral measurement was of considerable stability. Maximum transparency, i.e. neither the users nor their social environment noticing the EEG-equipment (Bleichner and Debener (2017)), can be key in a study design. In these cases, we conclude that unilateral ear-EEG is still an option.

A multi-channel configuration yields the best results

For many applications of ear-EEG, a low number of channels, i.e. a simple setup, is necessary. In that case, it is important where the electrodes should be placed. It was our objective to find patterns and to make recommendations for this placement of ear-electrodes. When we looked for the largest amplitudes, unsurprisingly, they were recorded from the channels with the largest inter-electrode distance. But importantly, high amplitude does not imply a large effect size. Our data suggests two things in this regard: first, the pattern of high distance equalling high performance of the channel completely disappears when moving from amplitudes to effect sizes, i.e. the inter-electrode distance of a channel is not a reliable predictor for performance. That is because the effect size measure is dependent on the amount of signal of interest relative to the amount of noise. As the amplitude of the signal of interest (i.e. the N100) increases with distance between electrodes, so does the amount of background noise. Second, like distance, the orientation of the channels does not work to predict performance: between the four groups we defined (vertical, horizontal, left/right diagonal), no group shows a consistently higher effect size. It is possible that more subtle benefits of channel orientations can be observed with a higher density of electrodes around the ear. From such work, one might be able to derive a "single best ear-channel" per ERP-task that works sufficiently well for all participants. Still, the high amount of intra- and inter-individual differences in optimal channels clearly suggests that no single channel is suited to reliably record any ERP. We therefore highly recommend a multi-channel design and an individualized channel selection where-ever possible. Another observation supports this recommendation: within a subject, we find that there is a large discrepancy between the best and the worst channel in terms of effect size. Since there seems to be no clear pattern as to which channel records each ERP-component best, having only a single channel would highly inflate the inter-subject variability in effect sizes. In addition, ear-electrodes were contaminated by artefacts in a substantial number (~25%) of our recordings, which is a concern for single-channel recordings. For methods of artefact rejection (like the ICA used for the scalp data in our analysis), signal processing benefits from multi-channel recordings. Additional channels are used to estimate noise components that can subsequently be removed from the data. Given the low channel count in this study, we decided against using the ICA-approach for the ear-configuration, since it can be expected that the number of sources is substantially higher than the number of electrodes. Therefore, a satisfying separation of signal- and artefact sources cannot be expected. However, using ICA with a higher-density ear-EEG could result in higher signal-to-noise ratio than with fewer electrodes. Investigating this thought will be part of future research.

We conclude that ear-EEG is most reliably and effectively recorded from a multi-channel setup. We recommend to prioritize the quality of signal acquisition over potentially lower visibility in this case.

5 Conclusion

We quantified auditory ERP-components recorded with ear-EEG. These signals recorded from around the ear are highly similar to those from classical high-density EEG in morphology and effect sizes. The signal loss from scalp- to ear-EEG ranged from medium to low for both bilateral and unilateral ear-EEG. The inter-individual differences in effect size we found for ear-EEG are in the same range as those found in high-density EEG. We observed a significant inter- and intra-subject variability in the optimal channel location.

We therefore recommend the use of multi-channel ear-EEG and individualized channel selection to account for those differences. Our work can guide future researchers that transit from scalp-EEG to ear-EEG and provides an estimate of the to be expected effect sizes for different auditory paradigms.

Accepted Manuscript

6 Supplementary Information

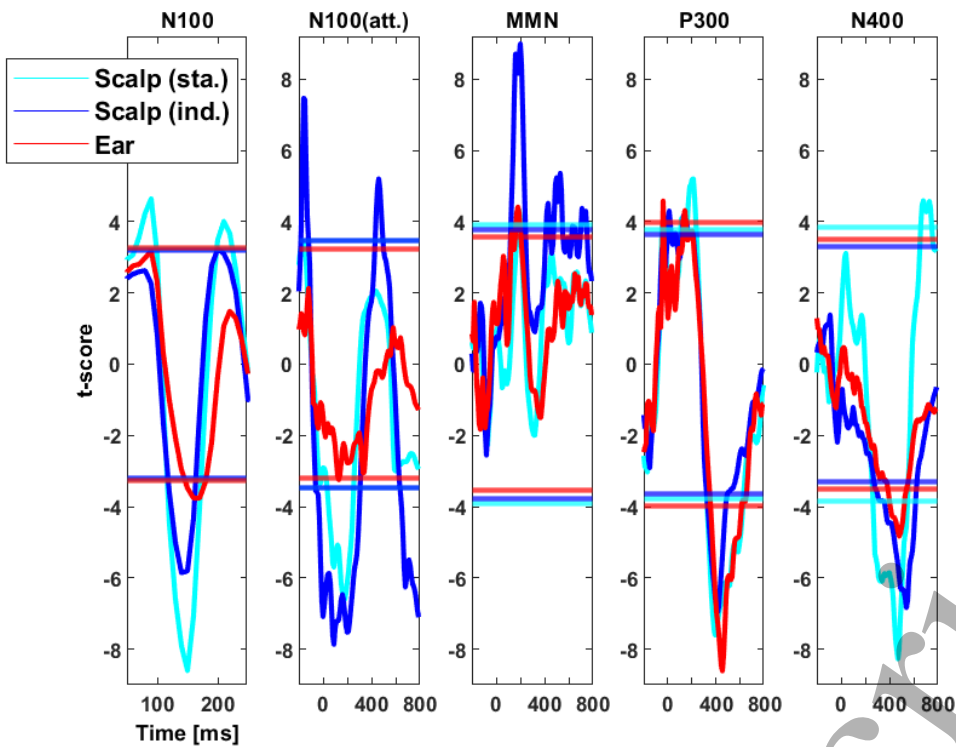


Figure 7: Statistical analysis of condition differences for five auditory ERPs. We show t-scores (y-axis) over time (x-axis) as a result of mass univariate analysis for five ERP-components. Group-level results are depicted for three EEG-configurations: scalp(sta.) in cyan, scalp(ind.) in blue and ear-EEG (bilateral) in red. Horizontal lines are color-coded to the respective EEG-configurations and represent the critical threshold. If t-scores exceed this value, the difference between target and control was considered significant at this time point.

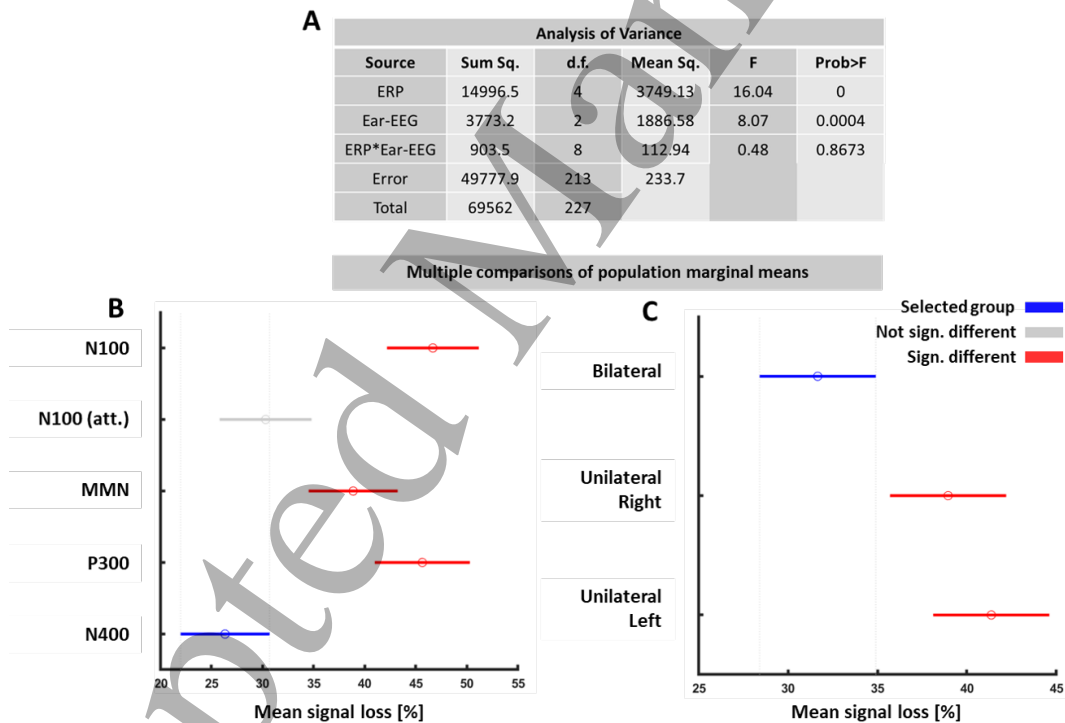


Figure 8: Results of a 3x5 two-way ANOVA with factors ear-EEG (levels are bilateral, left unilateral and right unilateral ear-EEG) and ERP-component (levels are N100(att.), N100, MMN, P300 and N400). In **A**, statistical details of the ANOVA are shown. In **B** we show the results of the post-hoc Tukey-Kramer test of the factor ERP. The five components are shown on the y-axis, the mean signal loss from scalp(ind.)- to ear-configuration on the x-axis. We take the N400 (in blue) as exemplary data: Since the lines of the N400 and the N100 do not overlap, they cannot be considered significantly different. The other three components, however, do not overlap with the N400, so these three components show a significantly higher mean signal loss than the N400. Likewise in **C**, the bilateral ear-EEG does not intersect with the unilateral ear-EEGs and therefore differs from them in signal loss. Left- and right unilateral ear-EEG do intersect; their means do not differ significantly.

7 Ethical statement

The conduction of this study was approved by the ethics committee of the University of Oldenburg.

8 Data and code availability statement

The code used for pre-processing and analysis of the data is available at DOI 10.17605/OSF.IO/FDZ3U

A general publication of the EEG-data is not possible, due to the uncertain legal situation regarding data sharing of EEG-data at our university. EEG can be considered as personal information under German legislation (DSGVO) and should not be generally published. We can share the data in anonymized form upon reasonable request.

9 Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

10 Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Emmy-Noether program - BL 1591/1-1 - Project ID 411333557.

11 Acknowledgements

The authors would like to thank Daniel Hölle and Marc Rosenkranz for their editorial comments on the manuscript. We thank Stefan Debener for all the fruitful discussions about the content of this study and the provision of the lab space. Finally, we thank Tatiana Goregliad Fjaellingsdal for the provision of the auditory material of the N400-task. We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Emmy-Noether program - BL 1591/1-1 - Project ID 411333557.

12 CRediT author statement

Arnd Meiser: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review and Editing, Visualization, Project administration.

Martin Bleichner: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review and Editing, Supervision, Project administration, Funding acquisition.

References

S. P. Ahlfors, J. Han, J. W. Belliveau, and M. S. Hämäläinen. Sensitivity of meg and eeg to source orientation. *Brain topography*, 23(3):227–232, 2010.

- C. Başar-Eroglu, T. Demiralp, M. Schürmann, and E. Başar. Topological distribution of oddball 'p300' responses. *International Journal of Psychophysiology*, 39(2-3):213–220, 2001.
- J. Y. Bennington and J. Polich. Comparison of p300 from passive and active tasks for auditory and visual stimuli. *International Journal of Psychophysiology*, 34(2):171–177, 1999.
- C. Bledowski, D. Prvulovic, R. Goebel, F. E. Zanella, and D. E. Linden. Attentional systems in target and distractor processing: a combined erp and fmri study. *Neuroimage*, 22(2):530–540, 2004.
- M. G. Bleichner and S. Debener. Concealed, unobtrusive ear-centered eeg acquisition: cee grids for transparent eeg. *Frontiers in human neuroscience*, 11:163, 2017.
- M. G. Bleichner and R. Emkes. Building an ear-eeg system by hacking a commercial neck speaker and a commercial eeg amplifier to record brain activity beyond the lab. *Journal of Open Hardware*, 4(1), 2020.
- M. G. Bleichner, M. Lundbeck, M. Selisky, F. Minow, M. Jäger, R. Emkes, S. Debener, and M. De Vos. Exploring miniaturized eeg electrodes for brain-computer interfaces. an eeg you do not see? *Physiological reports*, 3(4), 2015.
- M. G. Bleichner, B. Mirkovic, and S. Debener. Identifying auditory attention with ear-eeg: cee grid versus high-density cap-eeg comparison. *Journal of neural engineering*, 13(6):066004, 2016.
- I. Choi, S. Rajaram, L. A. Varghese, and B. G. Shinn-Cunningham. Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Frontiers in Human Neuroscience*, 7:115, 2013.
- D. J. Chwilla, C. M. Brown, and P. Hagoort. The n400 as a function of the level of processing. *Psychophysiology*, 32(3):274–285, 1995.
- S. Debener, R. Emkes, M. De Vos, and M. Bleichner. Unobtrusive ambulatory eeg using a smartphone and flexible printed electrodes around the ear. *Scientific reports*, 5:16743, 2015.
- A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.
- L. Fiedler, J. Obleser, T. Lunner, and C. Graversen. Ear-eeg allows extraction of neural responses in challenging listening scenarios—a future technology for hearing aids? In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5697–5700. IEEE, 2016.
- T. Goregliad Fjaellingsdal, E. Ruigendijk, S. Scherbaum, and M. G. Bleichner. The n400 effect during speaker-switch—towards a conversational approach of measuring neural correlates of language. *Frontiers in psychology*, 7:1854, 2016.
- D. M. Groppe, T. P. Urbach, and M. Kutas. Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review. *Psychophysiology*, 48(12):1711–1725, 2011.
- Y. Gu, E. Cleeren, J. Dan, K. Claes, W. Van Paesschen, S. Van Huffel, and B. Hunyadi. Comparison between scalp eeg and behind-the-ear eeg for development of a wearable seizure detection system for patients with focal epilepsy. *Sensors*, 18(1):29, 2018.
- E. Halgren, N. K. Squires, C. L. Wilson, J. W. Rohrbaugh, T. L. Babb, and P. H. Crandall. Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science*, 210(4471):803–805, 1980.

- D. Hölle, J. Meekes, and M. G. Bleichner. Mobile ear-eeeg to study auditory attention in everyday life. *Behavior Research Methods*, pages 1–12, 2021.
- L. Howard and J. Polich. P300 latency and memory span development. *Developmental Psychology*, 21(2): 283, 1985.
- R. Kanai and G. Rees. The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, 12(4):231–242, 2011.
- S. L. Kappel, D. Looney, D. P. Mandic, and P. Kidmose. Physiological artifacts in scalp eeg and ear-eeeg. *Biomedical engineering online*, 16(1):1–16, 2017.
- S. L. Kappel, S. Makeig, and P. Kidmose. Ear-eeeg forward models: Improved head-models for ear-eeeg. *Frontiers in neuroscience*, 13:943, 2019.
- E. S. Kappenman, J. L. Farrens, W. Zhang, A. X. Stewart, and S. J. Luck. Erp core: An open resource for human event-related potential research. *NeuroImage*, 225:117465, 2021.
- J. Katayama and J. Polich. Auditory and visual p300 topography from a 3 stimulus paradigm. *Clinical Neurophysiology*, 110(3):463–468, 1999.
- M. Koivisto and A. Revonsuo. Cognitive representations underlying the n400 priming effect. *Cognitive Brain Research*, 12(3):487–490, 2001.
- M. Kutas and S. A. Hillyard. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205, 1980.
- A. Meiser, F. Tadel, S. Debener, and M. G. Bleichner. The sensitivity of ear-eeeg: Evaluating the source-sensor relationship using forward modeling. *Brain Topography*, pages 1–12, 2020.
- K. B. Mikkelsen, S. L. Kappel, D. P. Mandic, and P. Kidmose. Eeg recorded from the ear: Characterizing the ear-eeeg method. *Frontiers in neuroscience*, 9:438, 2015.
- B. Mirkovic, M. G. Bleichner, M. De Vos, and S. Debener. Target speaker detection with concealed eeg around the ear. *Frontiers in neuroscience*, 10:349, 2016.
- R. Naatanen. Event-related brain potentials in research of cognitive processes: A classification of components. *Knowledge aided information processing*, pages 241–273, 1987.
- R. Näätänen and K. Alho. Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience*, 80(1-4):317–337, 1995.
- R. Näätänen, P. Paavilainen, H. Titinen, D. Jiang, and K. Alho. Attention and mismatch negativity. *Psychophysiology*, 30(5):436–450, 1993.
- R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, and I. Winkler. ‘primitive intelligence’ in the auditory cortex. *Trends in neurosciences*, 24(5):283–288, 2001.
- R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho. The mismatch negativity (mmn) in basic research of central auditory processing: a review. *Clinical neurophysiology*, 118(12):2544–2590, 2007.
- R. C. Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.

1
2
3 F. Perrin and L. Garcia-Larrea. Modulation of the n400 potential during auditory phonological/semantic
4 interaction. *Cognitive Brain Research*, 17(1):36–47, 2003.
5
6
7 J. Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, 118(10):2128–2148,
8 2007.
9
10
11 M. Stropahl, A.-K. R. Bauer, S. Debener, and M. G. Bleichner. Source-modeling auditory processes of eeg
12 data using eeglab and brainstorm. *Frontiers in neuroscience*, 12:309, 2018.
13
14
15 F. C. Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener. Semi-automatic identification
16 of independent components representing eeg artifact. *Clinical Neurophysiology*, 120(5):868–877, 2009.
17
18
19 S. Watkins, P. Dalton, N. Lavie, and G. Rees. Brain mechanisms mediating auditory attentional capture in
20 humans. *Cerebral Cortex*, 17(7):1694–1700, 2007.
21
22
23 I. Zibrandtsen, P. Kidmose, M. Otto, J. Ibsen, and T. Kjaer. Case comparison of sleep features from ear-eeg
24 and scalp-eeg. *Sleep Science*, 9(2):69–72, 2016.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60