



## The timecourse of multisensory speech processing in unilaterally stimulated cochlear implant users revealed by ERPs

Natalie Layer<sup>a,\*</sup>, Anna Weglage<sup>a</sup>, Verena Müller<sup>a</sup>, Hartmut Meister<sup>b</sup>, Ruth Lang-Roth<sup>a</sup>, Martin Walger<sup>a,b</sup>, Micah M. Murray<sup>c,d,e,f</sup>, Pascale Sandmann<sup>a</sup>

<sup>a</sup> University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Otorhinolaryngology, Head and Neck Surgery, Audiology and Pediatric Audiology, Cochlear Implant Center, Germany

<sup>b</sup> Jean-Uhrmacher-Institute for Clinical ENT Research, University of Cologne, Germany

<sup>c</sup> The Sense Innovation and Research Center, Lausanne and Sion, Switzerland

<sup>d</sup> The LINE (The Laboratory for Investigative Neurophysiology), Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

<sup>e</sup> CIBM Center for Biomedical Imaging of Lausanne and Geneva, Lausanne, Switzerland

<sup>f</sup> Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

### ARTICLE INFO

#### Keywords:

Cochlear implant  
Event-related potential  
Cortical plasticity  
Multisensory integration  
Audiovisual interaction  
Audiovisual speech perception

### ABSTRACT

A cochlear implant (CI) is an auditory prosthesis which can partially restore the auditory function in patients with severe to profound hearing loss. However, this bionic device provides only limited auditory information, and CI patients may compensate for this limitation by means of a stronger interaction between the auditory and visual system. To better understand the electrophysiological correlates of audiovisual speech perception, the present study used electroencephalography (EEG) and a redundant target paradigm. Postlingually deafened CI users and normal-hearing (NH) listeners were compared in auditory, visual and audiovisual speech conditions. The behavioural results revealed multisensory integration for both groups, as indicated by shortened response times for the audiovisual as compared to the two unisensory conditions. The analysis of the N1 and P2 event-related potentials (ERPs), including topographic and source analyses, confirmed a multisensory effect for both groups and showed a cortical auditory response which was modulated by the simultaneous processing of the visual stimulus. Nevertheless, the CI users in particular revealed a distinct pattern of N1 topography, pointing to a strong visual impact on auditory speech processing. Apart from these condition effects, the results revealed ERP differences between CI users and NH listeners, not only in N1/P2 ERP topographies, but also in the cortical source configuration. When compared to the NH listeners, the CI users showed an additional activation in the visual cortex at N1 latency, which was positively correlated with CI experience, and a delayed auditory-cortex activation with a reversed, rightward functional lateralisation. In sum, our behavioural and ERP findings demonstrate a clear audiovisual benefit for both groups, and a CI-specific alteration in cortical activation at N1 latency when auditory and visual input is combined. These cortical alterations may reflect a compensatory strategy to overcome the limited CI input, which allows the CI users to improve the lip-reading skills and to approximate the behavioural performance of NH listeners in audiovisual speech conditions. Our results are clinically relevant, as they highlight the importance of assessing the CI outcome not only in auditory-only, but also in audiovisual speech conditions.

### 1. Introduction

A cochlear implant (CI) is a neuroprosthesis that can help regaining hearing abilities and communication in patients with sensorineural hearing loss. However, the electrical CI signal transmits only limited

spectral and temporal information compared to natural acoustic hearing (Drennan and Rubinstein, 2008). Consequently, after implantation, the central auditory system has to learn to recognise the new, artificial input as meaningful sounds (Giraud et al., 2001a; Sandmann et al., 2015). This learning is an example of neural plasticity, wherein the nervous system

\* Corresponding author.

E-mail address: [natalie.layer@uk-koeln.de](mailto:natalie.layer@uk-koeln.de) (N. Layer).

<https://doi.org/10.1016/j.nicl.2022.102982>

Received 3 December 2021; Received in revised form 1 March 2022; Accepted 2 March 2022

Available online 4 March 2022

2213-1582/© 2022 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

adapts to the changing environment (Glennon et al., 2020; Merzenich et al., 2014).

Neural plasticity in CI users has been examined in several studies, reporting an increasing activation in the auditory cortex to auditory stimuli over the first months after implantation (Giraud et al., 2001c; Green et al., 2005; Sandmann et al., 2015). These experience-related cortical changes may even extend to the visual cortex, as indicated by the finding that CI patients engage not only the auditory, but also the visual cortex in purely auditory speech tasks (Chen et al., 2016; Giraud et al., 2001c). However, the latter results are based on positron emission tomography (PET) neuroimaging, which provides better spatial than temporal resolution. The temporal properties of the cross-modal response in the visual cortex of CI users have not yet been characterised.

As both, auditory and visual input (articulatory movements, gestures), contribute to speech perception, spoken language communication is an inherently audiovisual task (Drijvers and Özyürek, 2017; Grant et al., 1998; Sumbly and Pollack, 1954). Especially in difficult hearing situations (for instance speech in background noise) the supportive role of the visual modality can be crucial for understanding speech (Sumbly and Pollack, 1954). This has been characterised by the “principle of inverse effectiveness”, which states that if unisensory signals are poorly perceptible, there is a remarkable perception enhancement for multisensory signals (Stein and Meredith, 1993). The principle of inverse effectiveness has been assessed for audiovisual speech, in particular words, and has shown that if the visual or auditory presentation alone of a word was difficult, there was a higher benefit to understand audiovisual words (van de Rijjt et al., 2019). It is therefore not surprising that CI users show an enhanced visual influence on auditory perception (Desai et al., 2008) and stronger audiovisual interactions (Stevenson et al., 2017) due to the limited CI input, on both behavioural and cortical levels. CI users are not only better at integrating audiovisual speech information when compared to normal-hearing (NH) listeners (Bavelier et al., 2006; Kaiser et al., 2003; Mitchell and Maslin, 2007; Rouger et al., 2007). They also show a remarkably strong binding between the activation in the visual and auditory cortices during visual and audiovisual speech processing (Giraud et al., 2001c; Strelnikov et al., 2015). Given that previous findings on altered audiovisual speech processing in CI users are based on PET imaging, it is currently not well understood, whether these cortical alterations are already present at initial (sensory) or only at later cognitive processing stages.

Event-related potentials (ERPs) derived from continuous electroencephalography (EEG) represent a suitable tool to study cortical alterations in CI users (Sandmann et al., 2009, 2015; Sharma et al., 2002; Viola et al., 2012). ERPs have a high temporal resolution, which allows tracking single cortical processing steps (Biasucci et al., 2019; Michel and Murray, 2012). In particular, the auditory N1 (negative potential around 100 ms after stimulus onset) and the P2 (positive potential around 200 ms after stimulus onset) are elicited for auditory stimuli and are generated at least partly in the primary and secondary auditory cortex (Ahveninen et al., 2006; Bosnyak et al., 2004; Näätänen and Picton, 1987). Current frameworks suggest an automatic and sequential mode of cortical auditory processing, and a parallel processing of auditory information of distinct stimulus features within the supratemporal plane (De Santis et al., 2007; Inui et al., 2006). These auditory processes however can be influenced by top-down effects, in particular attention or predictions of incoming auditory events mediated by the frontal cortex (Dürschmid et al., 2019). Most of the previous ERP studies with CI users have focused on the N1 ERP to *unisensory auditory stimuli*, reporting that the N1 in CI users is reduced in amplitude and/or prolonged in latency compared to NH listeners (Finke et al., 2016a; Henkin et al., 2014; Sandmann et al., 2009; Timm et al., 2012). Recent studies, however, used more rudimentary, non-linguistic audiovisual stimuli (sinusoidal tones and white discs) to show an enhanced *visual* modulation of the *auditory* N1 ERP in CI users when compared to NH listeners (Schierholz et al., 2015, 2017). These findings in CI users suggest not only difficulties in the cortical sensory processing of auditory stimuli,

but also enhanced audiovisual interactions at sensory processing stages, particularly in conditions involving basic audiovisual stimuli.

The present EEG study extends previous research (Schierholz et al., 2015, 2017) by examining whether initial sensory cortical processing is also altered in CI patients in more complex audiovisual stimulus situations, especially in speech conditions. We used auditory syllables in combination with a computer animation of a talking head, providing the visual component of the stimuli (Fagel and Clemens, 2004; Schreitmüller et al., 2018). These types of speech stimuli are advantageous as they are highly controllable, reproducible, speaker-independent and perfectly timed. We compared the cortical processing of auditory-only and audiovisual syllables between CI users and NH listeners by means of electrical neuroimaging (Michel et al., 2009), including topographic and ERP source analysis to explore the timecourse of cortical processing. Compared to the traditional analysis of ERP data, which is based on waveform morphology at certain electrode positions, electrical neuroimaging is reference-independent and considers the spatial characteristics and the temporal dynamics of the global electric field to distinguish between effects of response strength, latency and distinct topographies (Michel et al., 2009).

Based on previous results with simple audiovisual stimuli (Schierholz et al., 2015, 2017), we predicted a multisensory effect in CI users and NH listeners for the more complex audiovisual speech stimuli as well, both on the behavioural and on the cortical level. However, due to the limited CI input and experience-related cortical changes after implantation, we expected group differences in voltage topographies and cortical source configurations, which would point to altered cortical activation in CI users during auditory and audiovisual speech processing.

## 2. Material and methods

### 2.1. Participants

In total, forty participants took part in the present study. Among these participants, six had to be excluded from further analyses due to the following reasons: bad EEG signal quality ( $n = 2$ ), low number of correct trials ( $n = 2$ ), one CI patient was hearing too well on the contralateral ear (rather single-sided deaf), and one NH participant was excluded due to the use of psychotropic drugs. Thus, for further analyses the total number of participants was thirty-four, with seventeen CI users (11 female, mean age: 59.3 years, range: 27–75 ± 12.1 years) and seventeen NH listeners (11 female, mean age: 59.9 years, range: 34–79 ± 11.1 years). NH controls were matched by gender, age, handedness, stimulated ear and years of education. The CI users were post-lingually deafened, implanted either unilaterally ( $n = 3$ , all left-implanted using a hearing-aid on the contralateral ear) or bilaterally ( $n = 14$ ). All CI users had been using their CI continuously for at least 12 months prior to the experiment. Therefore, in the case of bilateral implantation, either the ear which met that criterion or the ‘better’ ear (the ear showing the higher speech perception scores in the Freiburg Monosyllabic test) was used as stimulation side. Details on the implant system and the demographic variables can be found in Table 1.

To verify age-appropriate cognitive abilities, the DemTect Ear test battery was used (Brünecke et al., 2018). DemTect Ear is a version of the conventional DemTect (Kalbe et al., 2004) especially adjusted for patients with limited hearing abilities. This version enables to test cognitive skills independently of hearing and prevents disadvantages caused by hearing loss. It consists of various subtests including a word list, a number transcoding task, a word fluency task, digit span reverse, and delayed recall of the word list. These tests measure attention, memory and word fluency skills. All participants achieved total scores within the normal, age-appropriate range (13–18 points). Additionally, speech recognition abilities were measured by means of the German Freiburg monosyllabic speech test (Hahlbrock, 1970), using a sound intensity level of 65 dB SPL (see Table 4). None of the participants had a history of

**Table 1**  
Demographic information on the CI participants; HA = hearing aid.

ID	Sex	Age	Handedness	Fitting	Stimulated ear	Etiology	Age at onset of hearing impairment (years)	CI use of the stimulated ear (months)	CI manufacturer
1	m	61	right	CI + CI	left	unknown	41	15	MedEL
2	f	62	right	CI + CI	left	unknown	10	36	Cochlear
3	f	51	right	CI + CI	right	amalgam poisoning	20	54	Cochlear
4	m	65	right	CI + CI	left	hereditary	28	66	MedEL
5	m	53	right	CI + CI	left	Cogan syndrome	43	99	Cochlear
6	f	75	right	CI + CI	left	hereditary	57	30	Advanced Bionics
7	f	39	right	CI + CI	right	otosclerosis	24	17	Advanced Bionics
8	f	70	right	CI + CI	left	unknown	37	56	MedEL
9	f	70	right	CI + CI	left	meningitis	69	20	MedEL
10	m	59	right	CI + HA	left	unknown	49	33	Advanced Bionics
11	f	63	right	CI + CI	left	meningitis	20	106	Advanced Bionics
12	f	64	right	CI + CI	left	whooping cough	9	78	Cochlear
13	m	53	right	CI + CI	left	unknown	30	235	Cochlear
14	f	58	right	CI + HA	left	unknown	49	18	Advanced Bionics
15	f	71	right	CI + CI	left	otitis media	21	63	Cochlear
16	f	27	right	CI + HA	left	sudden hearing loss	6	236	MedEL
17	m	56	right	CI + CI	right	hereditary	19	63	MedEL

psychiatric illness and all participants were native German speakers. All participants had normal or corrected-to-normal vision, as measured by the Landolt test (Landolt C) according to the DIN-norm given by [Wesemann et al. \(2010\)](#) and all participants were right-handed (assessed by the Edinburgh inventory; range: 80–100%; [Oldfield, 1971](#)).

All participants gave written informed consent prior to data collection and were reimbursed. The study was conducted in accordance with the Declaration of Helsinki (2013) and was approved by the Ethics Committee of the medical faculty of the University of Cologne (application number: 18–257).

## 2.2. Stimuli

The stimuli used in this experiment were presented in three different conditions: visual-only (V), auditory-only (A) and audiovisual (AV). Additionally, there were trials with a black screen only ('nostim'), to which the participants were instructed to not respond. We used the Presentation software (Neurobehavioral Systems, version 21.1) and a computer in combination with a duplicated monitor (69 in.) to control stimulus delivery. The stimuli were the two syllables /ki/ and /ka/ and they were taken from the Oldenburg logatome speech corpus (OLLO; [Wesker et al., 2005](#)). The syllables were cut out of the available logatomes from one speaker (female speaker 1, V6 'normal spelling style', no dialect). These two syllables differed in their phonetic distinctive features (vowel place and height of articulation) in the vowel contrast (/a/ vs. /i/; [Micco et al., 1995](#)). These German vowels are different in terms of central frequencies of the first (F1) and second formant (F2). They represent the highest contrast between German vowels (e.g. [Obleser et al., 2003](#)), which makes them highly distinguishable for CI patients. These two syllables not only highly differ in terms of auditory (phoneme) realisation, but also in their visual articulatory (viseme) realisation. A viseme is the visual equivalent of the phoneme: a static image of a person articulating a phoneme ([Dong et al., 2003](#)). There are some phonemes that share identical visemes ([Cappelletta and Harte, 2012](#); [Lucey et al., 2004](#); [Mahavidyalaya, 2014](#)), but for the vowels of the syllables used in this study, the visemes are clearly distinguishable (see illustrations in [Jachimski et al., 2018](#)), which is of importance given that we present visual-only trials as well. The syllables were edited using Audacity (version 3.0.2) in order to be cut and adjusted to the same duration of 400 ms. The intensity level of the syllables was adjusted in

Adobe Audition CS6 (version 5.0.2) to ensure an equal intensity level of the stimuli (adjusted to the maximal amplitude).

The visualisation of the syllables was created with MASSY (Modular Audiovisual Speech SYNthesizer; [Fagel and Clemens, 2004](#)), which is a computer-based video animation of a talking head. This talking head has been previously validated for CI patients and is an objective tool to create visual and audiovisual speech stimuli ([Massaro and Light, 2004](#); [Meister et al., 2016](#); [Schreitmüller et al., 2018](#)). To be able to create articulatory movements corresponding to the auditory speech sounds, it is necessary to create files that transform the previously transcribed sounds into a probabilistic pronunciation model providing the segmentation and the timing of every single phoneme. This can be achieved by means of the web-based tool MAUS (Munich Automatic Segmentation; [Schiel, 1999](#)). To create a video file of the MASSY output, the screen recorder Bandicam (version 4.1.6) was used. The final editing of the stimuli was done with Pinnacle Studio 22 (version 22.3.0.377), creating video files of each stimulus in each condition: 1) Audiovisual (AV): lip-movements with corresponding speech sounds, 2) Auditory-only (A): black screen (video track turned off) combined with speech sounds, 3) Visual-only (V): lip-movements without speech sounds (audio track turned off). Each trial consisted of a static face (500 ms) and the video, which had a duration of 800 ms (20 ms initiation of lip movements + 400 ms auditory syllable + 380 ms completion of lip movements). For further analyses, we focused on the processing of the moving face (starting 500 ms post-stimulus onset/after the static face), as the responses to static faces comparing NH listeners and CI users have been investigated previously ([Stropahl et al., 2015](#)).

For CI users, the stimuli were presented via a loudspeaker (Audiometer-Box, type: LAB 501, Westra Electronic GmbH) placed in front of the participant. Since the CI users were supplied with a hearing aid or a second CI at the contralateral side, the device was removed during the recording. Additionally, the contralateral ear was closed with an ear-plug to ensure that only one ear was stimulated. For the NH participants, the stimuli were presented monaurally via insert earphones (3 M E-A-RTONE 3A) at the same side as in the matched CI users, and the contralateral ear was covered with an ear-plug as well. Stimulus presentation via insert earphones is advantageous when compared to free-field presentations because insert earphones avoid a confounding stimulation of the second ear. Similar to the CI users, three of the NH listeners were stimulated on the right ear, and fourteen on the left ear. The

syllables were presented at 65 dB SPL. Both the CI users and the NH listeners rated the perceived loudness of the speech sounds by means of a seven-point loudness rating scale (Sandmann et al., 2009, 2010), which allowed to adjust the sound intensity to a moderate level of 60–70 dB (Allen et al., 1990).

### 2.3. Procedure

The participants were seated comfortably in an electromagnetically shielded and dimly lit room. The distance from head to screen was approximately 175 cm. The task was to discriminate the syllables /ki/ and /ka/ regardless of their modality (AV, A, V) by pressing a corresponding button as fast as possible. Each syllable was assigned to one of the two buttons of a mouse. The sides of the assigned buttons were counterbalanced across participants to avoid a potential bias caused by the used finger.

In all conditions (AV, A, V, 'nostim'), 90 trials each were presented per syllable, resulting in a total number of 630 trials (90 repetitions  $\times$  3 conditions (AV, A, V)  $\times$  2 syllables (/ki/, /ka/) + 90 'nostim'-trials). Each trial started with a static face of the talking head (500 ms) followed by a visual-only, auditory-only, audiovisual syllable or 'nostim'. Right after, a fixation cross was shown until a response has been given by the participant. The trials were pseudo-randomised such that no trial of the same condition and syllable appeared twice in a row. In total, there were five blocks which lasted for five minutes each, resulting in a total experimental time of approximately 25 minutes. After each block, the participants were given a short break. Before the experiment, the participants were given a practice block consisting of five trials per condition to ensure that the task was understood. An overview of the paradigm is displayed in Fig. 1A.

### 2.4. Additional behavioural measures

Apart from hit rates and response times recorded during the EEG task, we obtained additional behavioural measures. After each experimental block in the EEG task, the participants were asked to indicate on a scale, how effortful it was to perform the task. For that purpose, we used the 'Borg Rating of Perceived Exertion'-scale (Borg RPE-scale; Williams, 2017). Further, we examined the lip-reading abilities of the participants by using a behavioural lip-reading test, which has been applied in previous studies (Stropahl et al., 2015; Stropahl and Debener,

2017). It consists of monosyllabic words taken from the German Freiburg monosyllabic speech test (Hahlbrock, 1970) and is video-taped with various professional speakers. The task of the participants was to watch the video of the visual performance of a word and to report, which word was lip-read. Additionally, the CI patients had to fill in the Nijmegen Cochlear Implant Questionnaire (NICQ; Hinderink et al., 2000) assessing the quality of life in CI users.

### 2.5. EEG recording

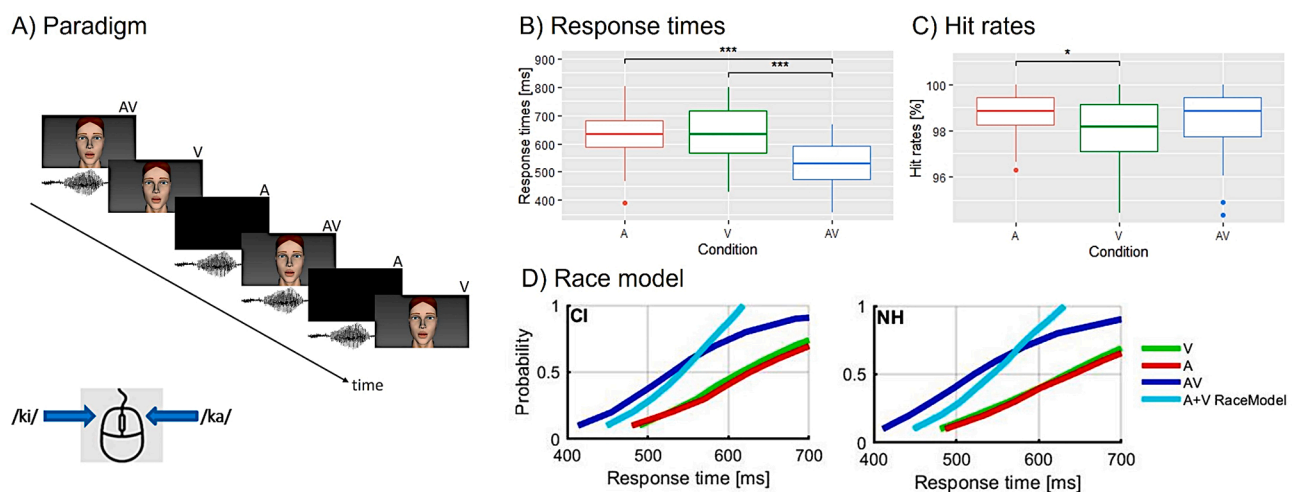
EEG data were recorded with 64 AG/AgCl ActiCap slim electrodes using a BrainAmp system (BrainProducts, Gilching, Germany) and a customised electrode cap with an electrode layout (Easycap, Herrsching, Germany) according to the 10–10 system. Two of the 64 electrodes were placed below (vertical eye movements) and beside (horizontal eye movements) the left eye to record an electrooculogram (EOG). All channels were recorded against a nose-tip reference, with a midline ground electrode, placed slightly anterior to Fz. Data were recorded with a sampling rate of 1000 Hz and were online analogically filtered between 0.02 and 250 Hz. Electrode impedances were kept below 10 k  $\Omega$  throughout the recording.

### 2.6. Data analysis

Data analysis was performed using MATLAB 9.8.0.1323502 (R2020a; Mathworks, Natick, MA) and R (version 3.6.3; R Core Team (2020), Vienna, Austria). Topographic analyses were conducted in CARTOOL (version 3.91; Brunet et al., 2011). For source analysis, Brainstorm (Tadel et al., 2011) was used. The following R packages have been utilised: ggplot2 (version 2.3.3) for creating plots; dplyr (version 1.0.4), tidyverse (version 1.3.0) and tidyr (version 1.1.3) for data formatting; ggpubr (version 0.4.0) and rstatix (version 0.7.0) for statistical computations.

#### 2.6.1. Behavioural data

First, trials with missing or false alarm responses were removed from the dataset. For each participant, outlier trials with reaction times (RTs) exceeding the individual mean by more than three standard deviations were removed for each condition separately. Afterwards, the mean hit rates and RTs were calculated for each condition (AV, A, V). We focused on the averages of hit rates and RTs computed across the two syllables



**Fig. 1.** Behavioural results. A) Simplified illustration of the paradigm. B) Mean response times for auditory (red), visual (green) and audiovisual (blue) syllables averaged over both groups, illustrating shorter response times for audiovisual syllables compared to auditory-only and visual-only RTs. C) Mean hit rates for auditory (red), visual (green) and audiovisual (blue) syllables averaged over both groups, exhibiting higher hit rates for auditory compared to visual syllables. D) Cumulative distribution functions for CI and NH. For both groups, the race model is violated, since both groups show that the probability of faster response times is higher for audiovisual stimuli (blue line) compared to the ones estimated by the race model (cyan line). Significant differences are indicated (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(/ki/, /ka/), given that supplementary analyses revealed no group differences and no group-specific effects with regards to the syllable conditions (Supplementary Fig. 1). To compare the performance for each group and for each condition, the hit rates and RTs were separately entered into a  $3 \times 2$  mixed ANOVA, with condition (AV, A, V) as the within-subjects factor and group (NH, CI) as the between-subjects factor. In case of violation of the sphericity assumption, a Greenhouse-Geisser correction was applied. Moreover, post-hoc t-tests were performed and corrected for multiple comparisons using a Bonferroni correction, if there were significant main effects or interactions ( $p \leq .05$ ).

The effect of facilitation in RTs for audiovisual input (i.e. faster RTs for AV) compared to unimodal stimuli (A, V) is known as the redundant signals effect (Miller, 1982). There are two models proposing an explanation for this issue: the race model and the coactivation model. The race model postulates that no neural integration is required to get a redundant signals effect (Raab, 1962). There is a competition of the independent unimodal stimuli (A, V) which are leading to a 'race' that determines whether one or the other unimodal stimulus determines the RTs. Therefore, the RTs of redundant signals (AV) can be assumed to be faster due to statistical facilitation, since the probability of either of the stimuli (A and V) to show a fast RT is higher than from one stimulus (A or V) alone. Whereas the coactivation model (Miller, 1982) claims that there is an interaction of the neural responses of the single sensory stimuli of a pair resulting in a combination and creation of a new product before a motor response is started. This process leads to faster RTs. To examine whether faster RTs for audiovisual syllables resulted from statistical facilitation or were based on multisensory processes (coactivation), the race model inequality (Miller, 1982) was applied. This approach is widely used in the area of multisensory research. The assumption is that violation of the race model is evidence for the presence of multisensory interactions (Ulrich et al., 2007) and evidence against independent processing of each stimulus of a pair. According to the race model inequality, the cumulative distribution function (CDF) of the RTs in the audiovisual condition can never be larger than the sum of the CDFs of the unisensory (A, V) conditions:

$$P(RT_{AV} \leq t) \leq P(RT_A \leq t) + P(RT_V \leq t), \text{ for all } t \leq 0,$$

where  $P(RT_x \leq t)$  is the probability of a condition  $x \in \{AV, A, V\}$  to be lower than an arbitrary value  $t$ . For any given value of  $t$ , violation of the race model points to the existence of multisensory interactions (see also Ulrich et al. (2007) for further detailed information). We used the RMITest software by Ulrich et al. (2007) to prove violation of the race model inequality. For each participant, the CDFs of the RT distributions for each condition (AV, A, V) and for the sum of the modality-specific conditions (A + V) were estimated. To determine percentile values, the individual RTs were rank ordered for each condition (Ratcliff, 1979). Next, the CDFs for the modality-specific sum (A + V) and for the redundant signals conditions (AV) were compared for the five fastest deciles (bin width: 10 %) for each group separately (NH, CI). We used one-tailed paired t-tests with subsequent Bonferroni correction to account for multiple comparisons. Significance at any decile bin pointed to a violation of the race model suggesting multisensory interactions to occur.

Concerning the other behavioural measures, in particular the lip-reading test and a subjective rating of the listening effort during the EEG task, we calculated two-sample t-tests to assess differences between CI users and NH listeners.

### 2.6.2. EEG preprocessing

EEG data were analysed with EEGLAB (version v2019.1; Delorme and Makeig, 2004) which is running in the MATLAB environment (Mathwork, Natick, MA). The raw datasets were downsampled to 500 Hz and filtered using a FIR-filter with a high pass cut-off frequency of 0.5 Hz and a maximum possible transition bandwidth of 1 Hz (cut-off frequency multiplied by two). Additionally, we applied a low pass cut-

off frequency of 40 Hz and a transition bandwidth of 2 Hz. The Kaiser-window (Kaiser- $\beta = 5.653$ , max. stopband attenuation = -60 dB, max. passband deviation = 0.001) was used for both filters (Widmann et al., 2015). This procedure maximises the energy concentration in the main lobe by averaging out noise in the spectrum and minimising information loss at the margins of the window. For CI users, the channels located around the region of the speech processor and transmitter coil were removed (mean: 3.2 electrodes; standard error of the mean: 1.0, range: 1–4). Next, the datasets were epoched into 2 s dummy epoch segments, and pruned of unique, non-stereotype artefacts using an amplitude threshold criterion of four standard deviations. An independent component analysis (ICA) was computed (Bell and Sejnowski, 1995) and the resulting ICA weights were applied to the epoched original data (1–40 Hz, -200 to 1220 ms relative to the stimulus onset (including the static and moving face). Independent components exhibiting horizontal and vertical eye movements, electrical heartbeat activity, as well as other sources of non-cerebral activity were removed (Jung et al., 2000). Similar to the procedures used in previous EEG studies with CI users (Debener et al., 2008; Sandmann et al., 2009; Viola et al., 2012), independent components accounting for the electrical CI artefact were identified and removed by means of ICA. The identification of these CI-related components was based on the stimulation side and the time course of the component activity, showing a pedestal artefact around 700 ms after the auditory stimulus onset (520 ms). Afterwards, the previously removed channels in CI users located around the speech processor and the transmitter coil were interpolated using a spherical spline interpolation (Perrin et al., 1989). This interpolation procedure enables a good dipole source localisation of auditory event-related potentials (ERPs) in CI participants (Debener et al., 2008; Sandmann et al., 2009). Only trials yielding correct behaviour (NH: 92.3 %  $\pm$  4.27 %; CI: 88.96 %  $\pm$  4.0 %) were retained for further analyses.

### 2.6.3. EEG data analysis

Event-related potentials (ERPs) of all conditions (AV, A, V) were compared between CI and NH participants. Similar to the behavioural data analysis, we focused on the ERP averages computed across the two syllables (/ki/, /ka/), given that supplementary analyses revealed no group differences and no group-specific effects with regards to the syllable conditions (Supplementary Fig. 1). To examine multisensory interactions, we used the additive model which is expressed by the equation  $AV = A + V$  (Barth et al., 1995). If the processing of multisensory (AV) stimuli is the sum of the processing of unisensory (A, V) stimuli, the model is satisfied and suggests independent processing steps or interactions that are fully linear. By contrast, if the model is not satisfied, it is assumed that there are non-linear interactions between the sensory modalities (Barth et al., 1995). As a next step, the equation was rearranged to  $A = AV - V$ . This allowed to compare the recorded ERP response to auditory-only stimuli (A) and the term  $[AV - V]$ , reflecting the ERP difference wave that is calculated by subtracting the visual ERP from the audiovisual ERP response. The term  $[AV - V]$  represents a visually-modulated auditory ERP response, which is the estimation of the auditory response in a multisensory context. If there is no interaction between the auditory and the visual modalities, the auditory (A) and the modulated auditory (AV-V) ERPs should be the same. By contrast, if the ERPs for A and AV-V are not equal, this can be interpreted as indication for non-linear multisensory interactions (Besle et al., 2004; Stekelenburg and Vroomen, 2007; Stevenson et al., 2014; Feder-Säljörvi et al., 2005; Vroomen and Stekelenburg, 2010). The interpretation of such non-linear effects as either sub-additive ( $A > AV - V$ ) or super-additive ( $A < AV - V$ ) is not straightforward and requires measurements that are reference-independent, measurements of power, or measurements of source estimates (e.g. Cappe et al., 2010). To ensure that an equal number of epochs for each condition contributed to the difference wave (AV-V), we reduced the number of epochs according to the condition with the lowest number of epochs for each subject individually. This procedure reduced the number of the initial epochs to 88.04 % ( $\pm$  4.63

%). After reducing the epochs to the same amount for each condition, we computed the difference waves (AV-V).

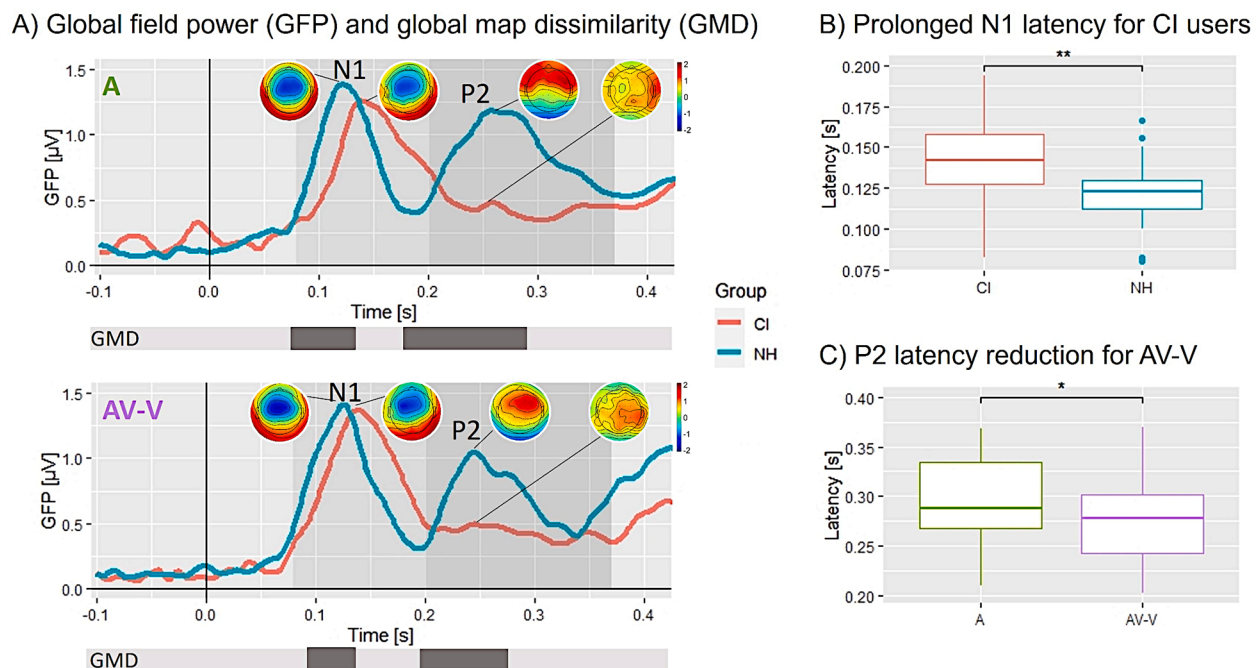
In this study, we used an electrical neuroimaging (Michel et al., 2009) analysis framework, including topographic and ERP source analysis to compare auditory (A) and modulated (AV-V) ERPs between and within groups (NH, CI). We explored these ERP differences by analysing the global field power (GFP) and the global map dissimilarity (GMD) to quantify ERP differences in response strength and response topography, respectively (Murray et al., 2008). In a first step, we analysed the GFP, at the time window of the N1 and the P2 (N1: 80–200 ms; P2: 200–370 ms). The time windows were chosen based on visual inspection of the GFP computed for the grand average ERP. The GFP, which was first introduced by Lehmann and Skrandies (1980), equals the root mean square (RMS) across the average-referenced electrode values at a given instant in time or, more simply put, the spatial standard deviation of all electrodes at a given time (Murray et al., 2008). We deliberately chose the GFP in this study to analyse the cortical response strength without having the disadvantage that arbitrary channel selection may bias the results of the peak detection. Fig. 2A displays the ERPs for A and AV-V for CI users and NH listeners. The individual GFP peak mean amplitudes and latencies were detected (for each condition (A, AV-V) and time window (N1 and P2)) and were statistically analysed by using a  $2 \times 2$  mixed ANOVA with group (NH, CI) as the between-subjects factor and condition (A, AV-V) as the within-subjects factor. This was done separately for each peak (N1, P2).

In a second step, we looked at the GMD (Lehmann and Skrandies, 1980) which quantifies topographic differences (and by extension, distinct configurations of neural sources; Vaughan Jr, 1982) between groups or experimental conditions, independently of the signal strength (Murray et al., 2008). The GMD was analysed in CARTOOL (Brunet et al., 2011) by computing a so-called 'topographic ANOVA' (TANOVA; Murray et al. 2008) to explore topographic differences between groups for each condition (CI(A) vs. NH(A) and CI(AV-V) vs. NH(AV-V)). Importantly, this is no analysis of variance, but a non-parametric

randomisation test. This randomisation test was conducted by using 5,000 permutations and by computing sample-by-sample p-values. An FDR-correction was applied to control for multiple comparisons (FDR = false discovery rate; Benjamini and Hochberg, 1995). The minimal significant duration was set to 15 consecutive time frames, given that previous observations suggest that ERP topographies do not continuously change as time elapses, but stay stable for a certain period of time before changing to another topography (Michel and Koenig, 2018) and to account for temporal autocorrelation.

#### 2.6.4. Hierarchical clustering and single-subject fitting analysis

Our results revealed differences across groups for the GMD for each condition (see section 'ERP results on the sensor level: GMD' for more details), which is an indication for different underlying neural generators (e.g. Vaughan Jr, 1982). But, if this is the case, it is interesting to know, how these differences are caused. On the one hand, a GMD may be explained by the fact that CI users and NH listeners reveal a fundamentally different configuration of neural generators during audiovisual speech processing. On the other hand, a difference in topography can also result from a latency shift which causes similar topographic maps to be shifted in time (Murray et al., 2008). To distinguish between the two exploratory approaches, we performed a hierarchical topographic clustering analysis with group-averaged data (NH(A), NH(AV-A), CI(A), CI(AV-V)) to identify template topographies in the time windows of interest (N1, P2). This analysis was done in CARTOOL (Brunet et al., 2011). Specifically, we applied the atomize and agglomerate hierarchical clustering (AAHC) which has been devised for EEG-data by Murray et al. (2008). It includes the global explained variance of a cluster and hinders blind combinations (or agglomerations) of clusters with short durations. Hence, the topographic clustering finds the minimal number of topographies explaining the greatest variance in a given dataset (here CI(A), CI(AV-V), NH(A), NH(AV-V)). An ERP topography does not show a random variation across time, but rather stays stable for some time before changing to another topography. This empirical



**Fig. 2.** Sensor ERP results. A) GFP of the conditions A and AV-V for NH listeners (blue) and CI users (red). Note that the GFP provides only positive values, given that it represents the standard deviation of all electrodes separately for each time point. The ERP topographies at the GFP peaks (N1, P2) are given separately for each group. Grey-shaded areas indicate the N1 and P2 time windows for peak and latency detection. Grey bars below mark the time window of significant GMDs between the two groups. B) N1 group effect (independent of the condition): the N1 latency is prolonged for CI users (red) compared to NH listeners (blue). C) P2 condition effect (independent of the group): the P2 latency is shorter for AV-V compared to A. Significant differences are indicated (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observation has been referred to as microstates (Michel and Koenig, 2018). A more detailed description of this approach can be found in Murray et al. (2008).

Next, the template maps identified by the AAHC were submitted to a single-subject fitting (Murray et al., 2008) to see, how specific templates are distributed on a single-subject level. This was achieved by computing sample-wise correlations for each subject and condition between each template topography and the observed voltage topographies. Each sample was matched to the template map with the highest spatial correlation. We conducted a statistical analysis on the output, which was the *first onset of maps* (latency) and the *map presence* (number of samples in time frames) being assigned to a specific template topography. Specifically, we performed a mixed ANOVA with group (NH, CI) as the between-subjects factor and condition (A, AV-V) and template map as within-subject factors, separately for each time window (N1, P2). In the case of significant three-way interactions, group-wise mixed ANOVAs (condition  $\times$  template map) were computed. In case of violation of the sphericity assumption, a Greenhouse-Geisser correction was applied. Follow-up t-tests were computed and corrected for multiple comparisons using a Bonferroni correction.

### 2.6.5. Source analysis

We computed an ERP source analysis for each group and condition by using Brainstorm (Tadel et al., 2011). This was done to assess the question, whether the observed topographic group differences at N1 latency can be explained by a fundamentally different configuration of neural generators during audiovisual speech processing. The source analysis was conducted according to the step-by-step tutorial provided by Stropahl et al. (2018). There are various proposed options to estimate source activities, but we chose the method of dynamic statistical parametric mapping (dSPM; Dale et al. 2000). This method has been successfully applied in previous studies with CI patients (Bottari et al., 2020; Stropahl et al., 2015; Stropahl and Debener, 2017). The dSPM method is able to localise deeper sources more precisely than standard norm methods, but the spatial resolution tends to remain low (Lin et al., 2006). dSPM uses the minimum-norm inverse maps with constrained dipole orientations to estimate the locations of the scalp-recorded electrical activity. It works well for localising auditory cortex sources, even though it is a relatively small cortical area (Stropahl et al., 2018). Single-trial pre-stimulus onset baseline intervals (-50 to 0 ms) were taken to calculate individual noise covariances to get an estimation of single-subject based noise standard deviations at each location (Hansen et al., 2010). As forward solution, the boundary element method (BEM) which is implemented in OpenMEEG was used as head model. The BEM gives three realistic layers and representative anatomical information (Gramfort et al., 2010). The activity data is shown as absolute values with arbitrary units based on the normalisation within the dSPM algorithm.

The inbuilt function to perform t-tests against zero (Bonferroni corrected,  $p = .001$ ) was applied to the averaged dataset (averaged over conditions and groups) to obtain regions of interest (ROIs) based on the maximal source activity around the visual and auditory cortices in the time windows of interest (N1, P2). The predefined regions implemented in Brainstorm of the Destrieux-atlas (Destrieux et al., 2010; Tadel et al., 2011) were chosen based on the t-test as they were lying within this area. The *auditory* ROI was defined as a combination of four smaller regions from the Destrieux-atlas (Destrieux: G\_temp\_sup-G\_T\_transv, S\_temporal\_transverse, G\_temp\_sup-Plan\_tempo and Lat\_Fis-post) of which the first three regions have been used in e.g. Stropahl and Debener (2017) or Stropahl et al. (2015). Similarly, previous studies have reported activity in the auditory cortex during unisensory auditory processing in CI patients in Brodmann area 41 and 42 (Sandmann et al., 2009; Stropahl et al., 2018) which can be approximated by the chosen ROIs. As *visual* ROIs, three regions again from the Destrieux-atlas were chosen (Destrieux: G\_cuneus, S\_calcarine, S\_parieto\_occipital) that matched the results of the t-test against zero and which comprise the

secondary visual cortex. Similar visual regions have been identified in previous studies with CI users (Giraud et al., 2001a; Giraud et al., 2001b; Prince et al., 2021). The selected ROIs can be inspected in the boxes of Fig. 4A.

Single-subject source activities for each ROI, condition and group were exported from Brainstorm, and they were statistically analysed in the software R. First, the peak means and latencies of the peaks were calculated. We performed a mixed ANOVA separately for each time window of interest (N1: 80–200 ms, P2: 200–370 ms), with group (NH, CI) as between-subject factor and condition (A, AV-V), ROI (auditory, visual) and hemisphere (left, right) as within-subjects factors. If the sphericity assumption was violated, a Greenhouse-Geisser correction was applied. In the case of significant interactions or main effects, follow-up t-tests were computed and corrected for multiple comparisons using a Bonferroni correction.

## 3. Results

### 3.1. Behavioural results

In general, the participants achieved a very high-performance level (hit rates  $\geq 97.9\%$ ) in all conditions, and the mean of the RT was between 522 ms and 640 ms in both groups (Table 2). Importantly, the CI users with a hearing-aid on the contralateral ear (implanted unilaterally) were no behavioural outliers in comparison to the bilaterally implanted CI users (mean of response times  $\pm$  one standard deviation: unilaterally implanted: 589 ms  $\pm$  152 ms, bilaterally implanted: 597 ms  $\pm$  96.2 ms). The  $3 \times 2$  mixed ANOVA with condition (AV, A, V) as within-subject factor and group (CI, NH) as between-subject factor showed no main effect of group ( $F_{1,32} = 0.00039$ ;  $p = .98$ ,  $\eta^2 = 0.00001$ ), but a main effect of condition for RTs ( $F_{1,41} = 69.18$ ;  $p \leq .001$ ,  $\eta^2 = 0.22$ ). Subsequent post-hoc t-tests revealed that RTs to redundant signals (AV) were significantly faster when compared to V ( $t(33) = 15.4$ ;  $p \leq .001$ ) or A ( $t(33) = 10.9$ ;  $p \leq .001$ ). There was no difference in RTs between the unisensory stimuli A and V ( $t(33) = -0.2$ ;  $p = 0.87$ ). These results are illustrated in Fig. 1B.

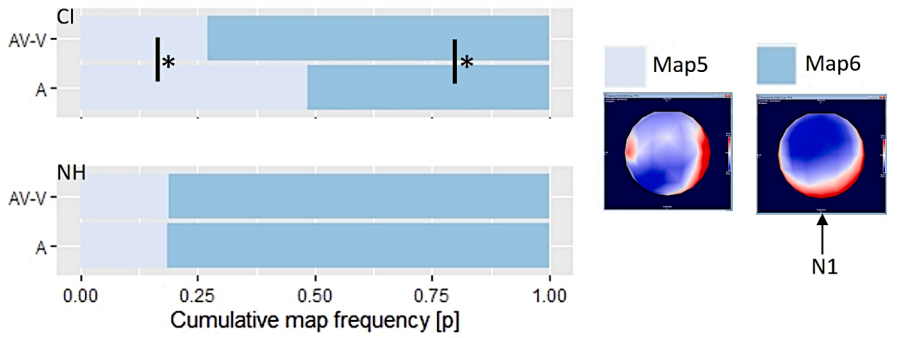
Concerning the hit rates, the  $3 \times 2$  mixed ANOVA with condition (AV, A, V) as within-subject factor and group (CI, NH) as between-subject factor revealed a significant main effect of condition ( $F_{2,64} = 3.8$ ;  $p \leq .05$ ,  $\eta^2 = 0.05$ ). Follow-up t-tests showed a difference in hit rates between A and V ( $t(33) = 2.74$ ;  $p \leq .05$ ), with A having significantly more correct hits than V. The other comparisons (AV vs. A and AV vs. V) were not significant (AV vs. A ( $t(33) = 0.29$ ;  $p = 0.88$ ) and AV vs. V ( $t(33) = -1.89$ ;  $p = 0.2$ )). These results are illustrated in Fig. 1C.

To examine the violation of the race model, the race model inequality was used. The one sample t-tests were significant in at least one decile for each group (Table 3). This means that the probability of faster response times is higher for audiovisual conditions compared to the ones estimated by the race model. This can also be observed in Fig. 1D. In sum, the violation of the race model in CI users and NH listeners points out the presence of multisensory integration for both groups.

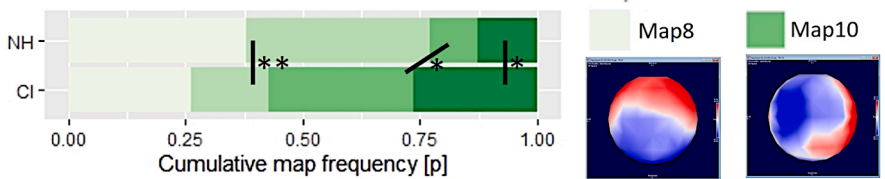
### 3.2. Other behavioural results

We calculated a two-sample t-test to assess differences in *auditory word recognition ability* and *(visual) lip-reading abilities* between CI users and NH listeners. The results revealed poorer auditory performance ( $p \leq .001$ , *Cohen's* $d = 2.48$ ), but better lip-reading skills for CI users compared to NH listeners ( $p \leq .001$ , *Cohen's* $d = 1.42$ ). Concerning the *subjective listening effort* measured during the EEG task, the t-test did not show a difference between CI users and NH listeners ( $p \geq .05$ , *Cohen's* $d = 0.19$ ). Therefore, none of the two groups found the task more difficult than the other group. The scores for these measures are displayed in Table 4.

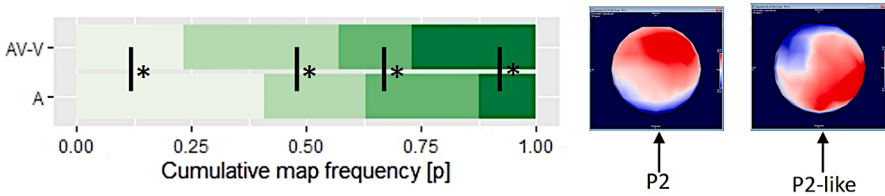
A) 80-200 ms: Condition effect for CI users



B) 200-370 ms: Group effect



C) 200-370 ms: Condition effect



**Fig. 3.** Hierarchical clustering and single-subject fitting results. A) Cumulative map frequency of the N1 maps: the CI users, but not the NH listeners show a conditions effect, with more frequent N1 map presences for AV-V compared to A. The corresponding map topographies are displayed on the right side, with map 6 being referred to as the N1 topography. B) Cumulative map frequency of the P2 group effect: NH listeners reveal a more frequent presence of a P2 topography (map 9) compared to CI users. C) Cumulative map frequency of the P2 condition effect: there is an increase in the presence for P2 (map 9) and P2-like (map 11) topographies for AV-V compared to A. Significant differences are indicated (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

3.3. ERP results on the sensor level: GFP

Fig. 2A displays the GFP of the grand averaged auditory ERPs for each group, in particular the unisensory auditory (A) and the visually modulated auditory (AV-V) ERP responses. The first prominent peak is visible for both groups approximately around 120 ms, which corresponds to the time window of the N1. The second peak around 220 ms is less prominent for CI users than for NH listeners and is referred to as the P2 ERP. For the N1, a  $2 \times 2$  mixed ANOVA with group (NH, CI) as between-subject variable and condition (A, AV-V) as within-subject factor was calculated for the GFP amplitude peak mean and the GFP peak latency. Concerning the peak mean, there were no statistically significant interactions or main effects. But, for peak latency, there was a significant main effect of group ( $F_{1,32} = 5.07$ ;  $p \leq .05$ ,  $\eta^2 = 0.13$ ) for both A and AV-V. Subsequent t-tests revealed a prolonged N1 latency in CI users than NH listeners ( $t(33) = 2.78$ ;  $p \leq .05$ ; Fig. 2B).

Similar to the N1 ERP analysis, a  $2 \times 2$  mixed ANOVA has been applied for the P2 GFP peak mean and latency. Regarding the P2 peak mean, the mixed ANOVA showed no significant interactions or main effects. For the P2 latency, the mixed ANOVA revealed a significant condition main effect ( $F_{1,32} = 8.75$ ;  $p \leq .05$ ,  $\eta^2 = 0.04$ ). Follow-up t-tests showed a significant difference between A and AV-V ( $t(33) = 3.0$ ;  $p \leq .05$ ) with shorter latencies for AV-V, indicating multisensory interaction effects to occur, independent of the factor group (Fig. 2C).

3.4. ERP results on the sensor level: GMD

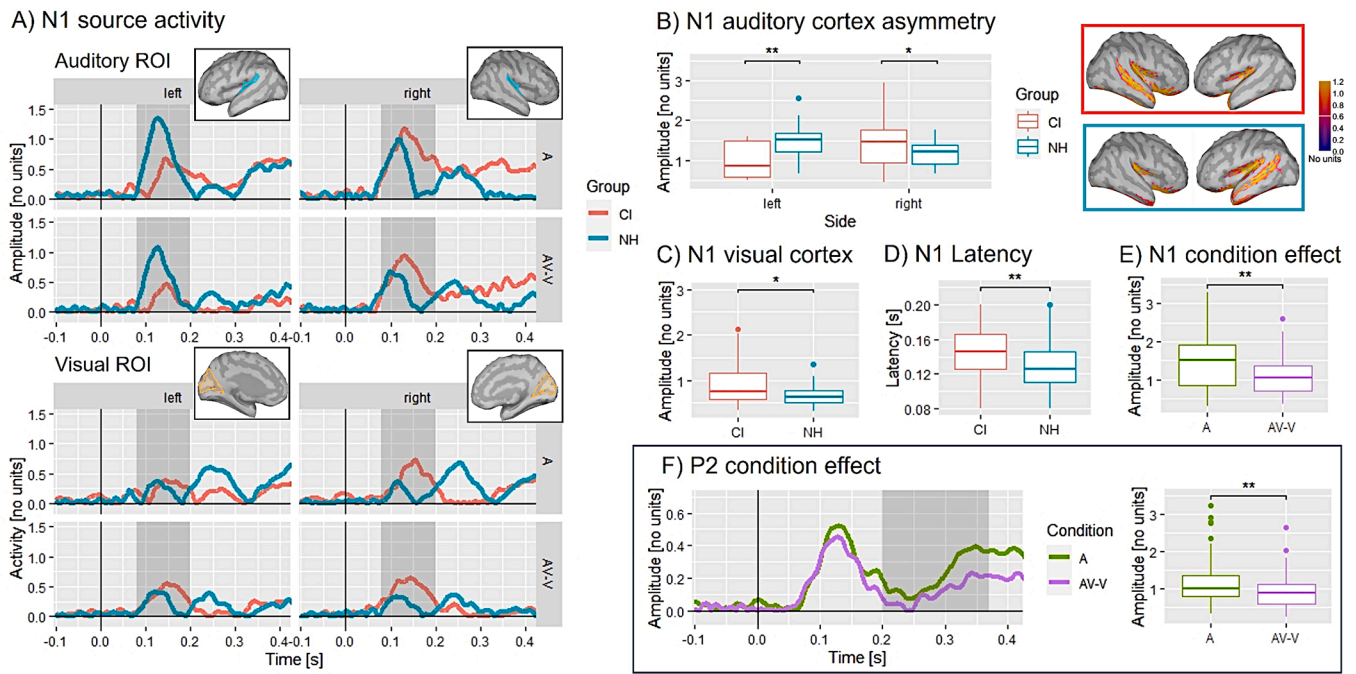
Regarding the GMD, we computed sample-by-sample p-values to quantify differences in ERP topographies between groups and conditions. Comparing the CI users and NH listeners *separately for each condition* (CI(A) vs. NH(A) and CI(AV-V) vs. NH(AV-V)), the results revealed

significant group differences for each condition (A and AV-V), both at the N1 (A = 80–136 ms; AV-V = 92–136 ms) and the P2 (A = 184–290 ms; AV-V = 198–278 ms) latency range (Fig. 2A, grey bars beneath the GFP plots). Additionally, we compared the ERP topographies of the groups with regards to the *condition difference*, which was obtained by subtracting A from AV-V (CI(AV-V)-A) vs. NH(AV-V)-A). This analysis revealed a significant group difference at the time window 156–174 ms, pointing towards group-specific audiovisual interaction processes.

3.5. ERP results on the sensor level: Hierarchical clustering and single-subject fitting results

To explore the origin of the underlying topographic difference between the two groups, we performed a hierarchical topographic clustering analysis with group-averaged data (CI(A); CI(AV-V); NH(A); NH(AV-V)) to identify template topographies in the time windows of interest (N1, P2). We applied the atomize and agglomerate hierarchical clustering (AAHC) to find the minimal number of topographies explaining the greatest variance in our dataset (i.e. the group-averaged ERPs). This procedure identified 12 template maps in 8 clusters that collectively explained 87.8% of these concatenated data. In particular, we observed two prominent maps for the N1 (map 5 and map 6) and four for the P2 (map 8, 9, 10, 11). These template maps were submitted to a single-subject fitting (Murray et al., 2008) to see, how the specific template maps are distributed on a single-subject level. In the following, *template map 6* will be referred to as N1 topography, since it matches the topography of the N1 peak from previous studies (Finke et al., 2016; Sandmann et al., 2015) and from our ERP results in both groups (Fig. 2A). The *template map 9* matches the P2 topography from previous studies (Finke et al., 2016a; Schierholz et al., 2021) and from NH





**Fig. 4.** Source analysis results. A) N1 source activity for CI (red) and NH (blue) for each ROI separated by hemisphere. The source activity is shown as absolute values and has arbitrary units based on the normalisation within the dSPM algorithm in Brainstorm. The grey shaded areas mark the N1 time window. The location of the defined ROIs is illustrated in the boxes with auditory ROIs in blue and visual ROIs in yellow. B) Hemisphere effect of the N1 peak mean: NH listeners (blue) show more activity in the left auditory cortex and CI users (red) show more activity in the right auditory cortex regardless of the condition. C) Group effect of the N1 peak mean: CI users show more activity in the visual cortex compared to NH listeners, regardless of the condition. D) N1 latency effect: CI users show a prolonged N1 latency compared to NH listeners, regardless of the condition. E) Condition effect of the N1 peak mean: there is a difference in activity between A and AV-V in the auditory cortex, indicating multisensory processes in both groups. F) P2 condition effect for the P2 peak mean: there is a significantly reduced auditory-cortex activation in AV-V compared to A, pointing towards cortical audiovisual interactions in both groups. Significant differences are indicated (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Mean hit rates (in %) and mean response times (in ms).

Condition	Hit rates		Response times	
	NH	CI	NH	CI
A	99 ± 0.8	98.5 ± 1	627 ± 81.1	632 ± 99.1
V	98.1 ± 1.3	97.9 ± 1.5	640 ± 102	623 ± 108
AV	98.5 ± 1.5	98.3 ± 1.4	522 ± 86.2	532 ± 86.3

**Table 3**

Redundant signals and modality-specific sum in each decile. AV corresponds to the redundant signals condition. A + V corresponds to the modality-specific sum. Paired-samples one-tailed t-tests with Bonferroni correction for multiple comparisons were conducted for each group. An asterisk indicates a significant result ( $p \leq .05/5$ ).

Decile	NH			CI		
	AV	A + V	<i>p</i>	AV	A + V	<i>p</i>
0.10	411	448	0.000*	413	449	0.000*
0.20	443	481	0.000*	455	481	0.003*
0.30	471	505	0.000*	480	505	0.007*
0.40	497	524	0.002*	506	525	0.030
0.50	521	543	0.011	530	542	0.108

listeners in the current results (Fig. 2A), which is why this template map is henceforth referred to as the P2 topography. Accordingly, the *template map 11* is comparable to the P2 topography observed in our CI users (Fig. 2A below). Given that the CI users in the present study showed only a weakly pronounced P2 ERP, the template map 11 is called P2-like topography in the next section.

The observed topographic differences between groups and conditions (as described in section 'ERP results on the sensor level: GMD') could be due to a latency shift of ERPs and/or due to a fundamentally

**Table 4**

Other behavioural measures for CI users and NH listeners. A score of 100% means that all words have been repeated correctly both in the Freiburg monosyllabic test and the lip-reading test. Concerning the Nijmegen Cochlear Implant Questionnaire (NCIQ), a higher percentage value corresponds to a better quality of life. A higher value for the listening effort corresponds to more effort to perform the task.

Group	Freiburg test (%)	Lip-reading test (%)	NCIQ total (%)	Listening effort
NH	96.2 ± 5.7	15.8 ± 10.2	not applicable	11.8 ± 2
CI	73.8 ± 11.4	35.5 ± 16.4	69.7 ± 15.1	12.2 ± 1.9

different configuration of neural generators during speech processing. To distinguish between these two explanatory approaches, we focused on two output parameters in the hierarchical clustering analysis, in particular the *first onset of maps* as well as the *map presence*.

In a first step, we analysed the *first onset of maps* to address the question whether the observed differences in GMD can be explained by a latency shift. For this purpose, we performed a mixed-model ANOVA with group (NH, CI) as the between-subjects factor and condition (A, AV-V) and template map as the within-subject factors, separately for the N1 and P2 time windows. For the N1 maps, the three-way mixed ANOVA showed a significant group × map interaction ( $F_{1,21} = 6.37$ ;  $p \leq .05$ ,  $\eta^2 = 0.1$ ). Post-hoc t-tests exhibited a difference between CI and NH for map 6 ( $t(28) = 3.79$ ;  $p \leq .001$ ). Given that template map 6 corresponds to a conventional N1 topography according to the distribution of the potentials on the scalp, these results confirm that the N1 of the CI users is generated later than the one of the NH listeners. For the first onset of the P2 maps, there were no significant main effects or interactions.

In a second step, we analysed the *map presence* (the number of time frames a corresponding template map is best-correlated to the single-subject data) to explore whether the topographic differences can be explained by group-specific patterns of ERP maps, which would point to distinct underlying neural generators between the two groups. Again, we performed a mixed-model ANOVA with group (NH, CI) as the between-subjects factor and condition (A, AV-V) and template map as the within-subjects factors, separately for the N1 and P2 time window.

The results revealed a group  $\times$  map  $\times$  condition interaction ( $F_{1,32} = 9.1$ ;  $p \leq .005$ ,  $\eta p^2 = 0.05$ ). Post-hoc t-tests revealed in particular for the group of CI users a significant difference of map presence between A and AV-V for the template map 5 ( $t(16) = 3.42$ ;  $p \leq .01$ ) and the template map 6 ( $t(16) = -3.42$ ;  $p \leq .01$ ). In CI users, the map 6 was similarly frequent to the map 5 in the auditory-only condition (A; number of samples  $28.9 \pm 20.6$  vs.  $27.1 \pm 20.6$ ), whereas the map 6 was more frequent compared to map 5 in the modulated condition (AV-V; number of samples  $40.8 \pm 15.6$  vs.  $15.2 \pm 15.8$ ). Thus, in CI users, the template map 6 better characterised the modulated response (AV-V) than the unisensory response (A). By contrast, the NH listeners in general showed a greater presence of map 6, irrespective of the condition (A: number of samples  $45.6 \pm 7.81$ ; AV-V: number of samples  $45.4 \pm 8.81$ ), and these individuals did not show a difference in the N1 ERP map presence between the modulated (AV-V) and the unisensory (A) responses ( $t(16) = 0.12$ ;  $p = .9$ ). Given that the template map 6 depicts a conventional N1 topography, the results suggest that specifically the CI users generate a N1 ERP map for the modulated response (AV-V) more frequently compared to the unisensory (A) condition. The comparison of groups by means of post-hoc t-tests revealed only a difference between CI users and NH listeners for the unisensory (A) ( $t(16) = -2.79$ ;  $p \leq .05$ ), but not for the modulated response ( $t(16) = -1.42$ ;  $p = .16$ ).

In sum, our results about the *first onset of maps* and the *map presence* at N1 latency suggest that the group differences in N1 topography originate from two reasons, in particular 1) generally slowed cortical N1 ERPs in CI users (regardless of condition), and 2) the distinct pattern of ERP maps in CI users as compared to NH listeners. The observation that the N1 topographies of the CI users differ from NH listeners especially in the auditory-only condition and that the two groups approximate in conditions with additional visual information, suggests that the CI users have a particularly strong visual impact on auditory speech processing. This visual impact on auditory speech processing is a manifestation of multisensory processes that remain intact in CI users despite the limited auditory signal provided by the CI.

Regarding the P2, the three-way mixed ANOVA revealed a significant group  $\times$  map ( $F_{2,77} = 4.35$ ;  $p \leq .05$ ,  $\eta p^2 = 0.09$ ) and a condition  $\times$  map ( $F_{3,96} = 6.79$ ;  $p \leq .001$ ,  $\eta p^2 = 0.06$ ) interaction. For the group  $\times$  condition interaction, post-hoc t-tests revealed significant differences between NH and CI for three maps, in particular map 9 ( $t(33) = -3.48$ ;  $p \leq .001$ ), map 10 ( $t(33) = 3.2$ ;  $p \leq .01$ ) and map 11 ( $t(33) = 2.34$ ;  $p \leq .05$ ). NH listeners showed a conventional P2 topography (map 9) according to the distribution of the potentials on the scalp more often than CI users regardless of the condition (see Fig. 3B).

Following the condition  $\times$  map interaction, post-hoc t-tests revealed significant differences between A and AV-V for all maps (map 8 ( $t(33) = 3.11$ ;  $p \leq .01$ ), map 9 ( $t(33) = -2.1$ ;  $p \leq .05$ ), map 10 ( $t(33) = 2.27$ ;  $p \leq .05$ ), map 11 ( $t(33) = -2.69$ ;  $p \leq .05$ )). This result shows that P2 topographies (map 9) and P2-like topographies (map 11) are generated more often for modulated responses (AV-V) compared to unmodulated responses (A), which is illustrated in Fig. 3C.

In sum, our results about the *first onset of maps* and the *map presence* at P2 latency suggest group-specific topographic differences at P2 latency, with a stronger presence of a conventional P2 topography (map 9) in NH listeners, and a stronger presence of the slightly different P2-like topography (map 11) in CI users. Both of these maps 9 and 11 are more frequent in the modulated than in the auditory-only condition, pointing to changes in the cortical activation due to the presence of additional

visual information in the speech signal.

### 3.6. Results from ERP source analysis

A source analysis was conducted to further explore the observed group differences and to evaluate the visual modulation of auditory speech processing in different cortical areas. Based on previous findings of experience-related cortical changes in CI users (Chen et al., 2016; Giraud et al., 2001b; Giraud et al., 2001c; Green et al., 2005; Sandmann et al., 2015), we focused on the auditory and visual cortex activity in both hemispheres. Single-subject source activities for each ROI, condition and group were exported from Brainstorm and were statistically analysed. The source waveforms for the N1 are displayed in Fig. 4A, showing the response in the auditory cortex (CI mean =  $141 \text{ ms} \pm 29 \text{ ms}$ ; NH mean =  $125 \text{ ms} \pm 25 \text{ ms}$ ) and the visual cortex (CI mean =  $147 \text{ ms} \pm 30 \text{ ms}$ ; NH mean =  $137 \text{ ms} \pm 34 \text{ ms}$ ) for both groups. First, the peak means and latencies of the peaks were calculated, which were the dependent variables for the subsequent ANOVA. We performed a mixed-model ANOVA with group (NH, CI) as the between-subjects factor and condition (A, AV-V), ROI (auditory, visual) and hemisphere (left, right) as the within-subject factors for each time window of interest (N1, P2) separately.

For the N1 *peak mean*, the mixed ANOVA revealed a group  $\times$  ROI  $\times$  hemisphere interaction ( $F_{1,32} = 8.44$ ;  $p \leq .01$ ,  $\eta p^2 = 0.03$ ) and a ROI  $\times$  condition interaction ( $F_{1,32} = 16.81$ ;  $p \leq .001$ ,  $\eta p^2 = 0.02$ ). For the group  $\times$  ROI  $\times$  hemisphere interaction, post-hoc t-tests confirmed a significant difference in the *left* auditory cortex activity between NH listeners and CI users ( $t(28) = -4.04$ ;  $p \leq .001$ ), with NH listeners showing stronger activation in the *left* auditory cortex than the CI users, regardless of condition. Whereas, post-hoc t-tests confirmed a significant difference in *right* auditory cortex activity between NH and CI ( $t(28) = 2.03$ ;  $p \leq .05$ ), with CI users showing stronger activation in the *right* auditory cortex than the NH listeners (Fig. 4B), regardless of the condition.

To resolve the ROI  $\times$  condition interaction from the initial ANOVA, post-hoc t-tests confirmed a difference between A and AV-V for the *auditory cortex* ( $t(67) = 5.71$ ;  $p \leq .001$ ), with more activity for A compared to AV-V, pointing towards multisensory interaction processes in both groups (Fig. 4E), regardless of the hemisphere.

For the *visual cortex*, post-hoc t-tests showed a significant difference in peak mean between NH listeners and CI users ( $t(67) = 3.6$ ;  $p \leq .001$ ), with CI users exhibiting a stronger activation in the visual cortex than NH listeners (Fig. 4C), regardless of the condition and hemisphere.

Concerning the N1 latency, the four-way mixed ANOVA showed a main effect of group ( $F_{1,32} = 7.39$ ;  $p \leq .05$ ,  $\eta p^2 = 0.05$ ) and ROI ( $F_{1,32} = 4.89$ ;  $p \leq .05$ ,  $\eta p^2 = 0.02$ ). Post-hoc t-tests revealed a difference in N1 latency between NH listeners and CI users ( $t(135) = 3.82$ ;  $p \leq .001$ ) and a difference between the auditory and visual cortex ( $t(135) = -2.53$ ;  $p \leq .05$ ). Again, the results confirmed that the N1 is generated later in CI users compared to NH listeners regardless of the condition and hemisphere (Fig. 4D).

Similar to the N1 ERP, a group (NH, CI)  $\times$  condition (A, AV-V)  $\times$  ROI (auditory, visual)  $\times$  hemisphere (left, right) ANOVA was computed on the peak mean and latency of the P2. Regarding the P2 *peak mean*, the mixed ANOVA revealed a main effect of ROI ( $F_{1,32} = 18.31$ ;  $p \leq .001$ ,  $\eta p^2 = 0.08$ ) and condition ( $F_{1,32} = 18.97$ ;  $p \leq .001$ ,  $\eta p^2 = 0.04$ ). Post-hoc t-tests showed a difference between the auditory and visual cortex ( $t(136) = 5.98$ ;  $p \leq .001$ ), with more activity for the auditory cortex compared to the visual cortex, and a significant difference in peak mean between A and AV-V ( $t(136) = 5.25$ ;  $p \leq .001$ ), indicating multisensory interaction in both groups. This multisensory effect at P2 latency is illustrated in Fig. 4F, providing the source waveforms (and the corresponding box plot) separately for the auditory-only (A) and the modulated (AV-V) response, averaged across the two groups.

Regarding the P2 latency, we found no statistically significant

interactions or main effects for the P2 latency.

### 3.7. Correlations

We performed targeted correlations matching findings from previous studies using the Pearson's correlation. The Benjamini-Hochberg (BH) procedure was applied to account for multiple comparisons (Benjamini and Hochberg, 1995). First, we verified that lip-reading abilities are related to the CI experience and the age at onset of hearing loss (Stropahl et al., 2015; Stropahl and Debener, 2017). The results revealed a positive relationship between lip-reading abilities and CI experience ( $r = .51$ ;  $p = 3.8e - 2$ ; significant according to BH), as well as a negative relationship between lip-reading abilities and the age at onset of hearing loss ( $r = -.73$ ;  $p = 9.5e - 4$ ; significant according to BH). This means that the longer the CI experience and the earlier the onset of hearing impairments, the more pronounced are the lip-reading abilities in CI users. Second, we explored whether there is a relationship between CI experience and the activation in the visual cortex (Giraud et al., 2001c). Indeed, there was a positive correlation between CI experience and visual cortex activation at N1 latency ( $r = .68$ ;  $p = 2.9e - 3$ ; significant according to BH), indicating that a longer experience with the CI results in enhanced recruitment of the visual cortex during auditory speech processing. Moreover, previous studies reported a positive relationship between left-auditory cortex activation and speech comprehension, which however was not confirmed by the present results (Freiburg monosyllabic test:  $r = .27$ ;  $p = .29$ ; not significant according to BH). Finally, we correlated the NCIQ with the Freiburg monosyllabic test (Vasil et al., 2020) and found a strong positive relationship ( $r = .75$ ;  $p = 4.9e - 4$ ; significant according to BH), demonstrating that the quality of life is rising with better hearing performance.

## 4. Discussion

The present study investigated audiovisual interactions in CI users and NH listeners by means of EEG and behavioural measures. On the behavioural level, the results showed multisensory integration for both groups, as indicated by shortened response times for the audiovisual as compared to the two unisensory conditions. A multisensory effect was confirmed by the ERP analyses for both groups, showing a reduced activation in the auditory cortex for the modulated (AV-V) compared to the auditory-only (A) response, both at the N1 and the P2 latency. Nevertheless, we found specifically in the group of CI users a change of N1 voltage topographies when visual information was presented in addition to auditory information, which resulted in an approximation of the N1 topographies of the two groups. Thus, our behavioural and ERP findings demonstrate a clear audiovisual benefit in both groups, with a particularly strong visual impact on auditory speech processing in the CI users.

Apart from these condition effects, we found differences between CI users and NH listeners, not only in ERP topographies (N1 and P2 latency) but also in auditory- and visual-cortex activation (N1 latency). Regarding the latter, the ERP source analyses revealed that the CI users, when compared to the NH listeners, generally have an enhanced recruitment of the *visual cortex* at N1 latency, which is positively correlated with the duration of CI experience. Further, the ERP source analyses revealed a delayed N1 response in CI users for the *auditory cortex*, and a group-specific pattern of activation, with a leftward hemispheric asymmetry in the NH listeners and a rightward hemispheric asymmetry in the CI users.

Taken together, these results suggest that the topographic group differences observed at N1 latency are caused by two different factors, notably the generally slowed cortical N1 response in CI users, and the group-specific differences in cortical source configuration.

### 4.1. Multisensory integration in both groups – behavioural level

Behavioural measures revealed that both the NH listeners and the CI users showed shorter reaction times for audiovisual syllables compared to unisensory (auditory-alone, visual-alone) syllables. There was no difference between auditory and visual conditions. Therefore, both groups show a clear redundancy effect for audiovisual syllables, suggesting that on the behavioural level, the benefit for cross-modal input is comparable between the CI users and the NH listeners (Laurienti et al., 2004; Schierholz et al., 2015), at least in conditions with short syllables that are combined with a computer animation of a talking head. Even though the CI provides only a limited input, the CI users' responses were not delayed on the behavioural level. In line with this, there was no difference in subjective ratings of listening effort between the two groups, indicating that both groups were able to perform the task with comparable listening effort.

Concerning purely visual syllables, we expected to find differences between CI users and NH listeners with shorter reaction times for CI users compared to NH listeners. This expectation is based on results from studies showing shorter reaction times for visual stimulation in congenitally deaf individuals (Bottari et al., 2014; Finney et al., 2003; Hauthal et al., 2014) and visually-induced activation in the auditory cortex, both in deaf individuals and CI users (Bavelier and Neville, 2002; Bottari et al., 2014; Heimler et al., 2014; Sandmann et al., 2012). This cross-modal reorganisation seems to be induced by auditory deprivation and might provide the neural substrate for enhanced superior visual abilities. CI users experience not only a time of deprivation before implantation, but they also perceive a limited input provided by the implant, which may force these individuals to develop compensatory strategies. Indeed, previous studies have revealed that CI users show specific enhanced behavioural visual abilities, in particular improvements in (visual) lip-reading skills when compared to NH listeners (Rouger et al., 2007; Schreitmüller et al., 2018; Stropahl et al., 2015; Stropahl and Debener, 2017).

Nevertheless, the behavioural results on *syllables* in the present study show comparable results between CI users and the NH listeners, which resembles previous results from a speeded response task with basic stimuli, presenting tones and white discs as auditory and visual stimuli, respectively (Schierholz et al., 2015). The authors argued that CI users do not have better reactivity to visual stimuli compared to NH listeners when performing a simple detection task. However, the present task required detection *and* discrimination of the syllables. But, since there was only a choice of two syllables that had to be discriminated, the task might have been too easy to detect differences between the two groups, leading to ceiling effects (hit rates  $\geq 97.9\%$ ). This argument is supported by our finding that CI users did *not* report an enhanced listening effort when compared to NH listeners, which contradicts previous studies that used more complex speech stimuli and which reported an enhanced listening effort in CI users (Finke et al., 2016a; Finke et al., 2016b). Nevertheless, similar to our study, Stropahl and Debener (2017) used a syllable identification task to compare CI users and NH listeners. They did not find behavioural group differences for visual syllables as well, although CI users in general showed enhanced lip-reading abilities, as in the present study. It seems that behavioural improvements in CI users are stimulus- and task-selective, and may be observed only under specific conditions and might be more pronounced if speech stimuli include semantic information (Moody-Antonio et al., 2005; Rouger et al., 2008; Tremblay et al., 2010).

A benefit for cross-modal input, which is represented by a redundant signals effect for audiovisual stimuli, does not necessarily confirm multisensory integration. Faster reaction times for audiovisual stimuli might also result from statistical facilitation due to a competition between the unimodal auditory or visual stimuli, leading to a 'race' that determines the reaction times (Miller, 1982). To test whether statistical facilitation or coactivation were the cause for the redundant signals effect, we applied the race model inequality (Miller, 1982). Given that

the race model was violated for both the NH listeners and the CI users over the faster deciles of the RT distribution, we can conclude that audiovisual integration was contributing to the observed redundant signals effect in both groups.

Our results on syllable perception revealed that the audiovisual gain in CI users and NH listeners was comparable. However, Rouger et al. (2007) used a word recognition task and showed for the CI users an enhanced gain in word recognition performance when the visual and auditory input was combined. The authors concluded that CI users are ‘better audiovisual integrators’ when compared to NH listeners. The inconsistency to our results might originate from the difference of demands in the task. As argued before, the task in our experiment and in the previous one with tones (Schierholz et al., 2015) was probably too easy to detect differences between the groups on the behavioural level. Schierholz et al. (2015) suggested that audiovisual benefits in CI users might be more pronounced in more ecologically valid stimulus conditions, since CI users might need their compensatory strategies rather for more complex speech stimuli.

However, it seems that syllables are not yet complex enough to show a group difference in multisensory enhancement between CI users and NH listeners. As both groups performed with a very high hit rate, a difference was not detected due to ceiling effects. We suppose that experimental designs with even more complex speech stimuli, such as words or sentences, are necessary to reveal behavioural differences in audiovisual speech processing between CI users and NH listeners.

#### 4.1.1. Multisensory interactions in both groups - cortical level

To analyse the influence of the visual input on the auditory cortical response, we compared the ERPs to the auditory condition (A) with the modulated ERP response (AV-V). If there is a difference between the auditory and the modulated response, this can be interpreted as indication for multisensory interactions (Besle et al., 2004; Stekelenburg and Vroomen, 2007; Teder-Sälejärvi et al., 2005; Vroomen and Stekelenburg, 2010). In that case, the difference can be either subadditive ( $A > AV - V$ ) or superadditive ( $A < AV - V$ ) (Stevenson et al., 2014). In multisensory research, both subadditive results (e.g. Cappe et al., 2010) and superadditive results (e.g. Schierholz et al., 2015) have been reported for ERPs. Regarding the current study, the source analyses of the N1 and P2 ERPs revealed a reduced cortical activation for the modulated (AV-V) compared to the auditory-only (A) ERPs. Thus, our results point to a subadditive affect in both the CI users and the NH listeners, which is present at different processing stages, specifically at N1 and P2 latency.

Further evidence of cortical multisensory interaction in both the CI users and the NH listeners comes from our analysis of the global field power (GFP) and the topographic clustering analysis. Regarding the GFP, we found that the peak in the P2 time window was shortened for the modulated (AV-V) compared to the auditory-alone (A) responses. This observation confirms previous conclusions that a reduction in latency can be a sign for multisensory integration processes (van Wassenhove et al., 2005). Regarding the topographic clustering analysis, we found an increased presence of the P2 and P2-like topographies for the modulated responses (AV-V) compared to auditory responses (A). This effect was not specific to one group but was observed for both the CI and the NH individuals. Thus, our findings point to a visual modulation of auditory ERPs in both the CI users and the NH listeners, which confirms our behavioural results of multisensory integration in both groups.

Similar to the present study, Schierholz et al. (2015) investigated audiovisual interactions by applying the additive model to ERP data. They showed a cortical audiovisual interaction as well, as indicated by a significant visual modulation of auditory cortex activation at N1 latency, specifically for elderly CI users as opposed to young CI users and elderly NH individuals. In our study however, the visual modulation in cortical activation was present in both the NH and the CI group. This discrepancy in results may result from the fact that we did not include two age groups in our study, as it has been done in Schierholz et al. (2015).

## 4.2. Group differences in auditory and audiovisual processing

There was no difference in global field power (GFP) between CI users and NH listeners, indicating that the strength of cortical responses was comparable between the two groups. However, there were group differences in ERP topographies as shown by the global map dissimilarity (GMD). In particular, GMD group differences were found not only for the auditory-only condition (A) and the modulated response (AV-V), but also for the difference wave ((AV-V)-A).

In sum, these results suggest that differences between groups did not result from differences in signal strength, but rather from differences in topography. A distinct electric field topography however can be caused by a latency shift of ERPs and/or by distinct configurations of the neural sources. We performed different analyses, including hierarchical clustering, single-subject fitting and source analysis, to distinguish between these two causes for topographic group differences. The following sections discuss the results of these analyses.

### 4.2.1. P2 time window

Hierarchical clustering and single-subject fitting analyses for the P2 time window revealed a condition effect (independent of group) and a group effect (independent of condition). The *condition* effect, showing a more frequent map presence for P2 and P2-like topographies has been discussed already in the section 4.1.1. Concerning the *group* effect, we found a more frequent presence for P2 topographies for NH listeners compared to CI users and more frequent P2-like topographies for CI users compared to NH listeners. But, as a whole, four template maps have been identified by the clustering analysis, showing variable distributions across groups and conditions. This leads to the conclusion that group differences are relatively probable but are not easy to depict due to the high variability in the data. Therefore, it is unlikely that a source analysis will find clear results to reveal group-specific cortical activation patterns. Indeed, the source analysis did not reveal group differences, neither in the auditory, nor the visual cortex activation. However, there was an amplitude reduction for AV-V compared to A over both groups, confirming audiovisual interactions but no differences between CI users and NH listeners (see section 4.1.1).

It may be argued that our selected auditory ROI was not appropriate enough for P2 responses, which could have prevented finding a group difference in auditory-cortex activation. To verify the ROI selection, we conducted a supplementary source analysis for the P2 ERP with an adjusted, smaller ROI that was selected by considering previous studies. Specifically, previous studies using MEG (magnetoencephalography) or fMRI (functional magnetic resonance imaging) have reported that the auditory P2 is generated in the anterior part of the auditory cortex, in particular the lateral part of the Heschel’s gyrus (Bosnyak et al., 2004; Lütkenhöner and Steinsträter, 1998; Ross and Tremblay, 2009; Tremblay et al., 2014). This region is also implemented in the Destrieux-atlas in Brainstorm (G\_temp\_sup-G\_T\_transv), a subregion of our originally selected auditory ROI. Statistical analyses did not show any different results between the adjusted (i.e. smaller) and the original ROI. Therefore, we conclude that CI users and NH listeners show a comparable recruitment of the auditory cortex during auditory and audiovisual speech processing, at least at the P2 latency range.

### 4.2.2. N1 time window

Hierarchical clustering and single-subject fitting results for the N1 time window demonstrated a more frequent presence of N1 topographies for the modulated responses (AV-V) compared to purely auditory responses (A) for CI users, but not for NH listeners. Specifically for CI users, there was a difference between A and AV-V, showing a more frequent map 6 (N1 topography) and a less frequent map 5 presence (blurred N1 topography) for modulated responses and an equal map presence of map 5 and map 6 for purely auditory responses. By contrast, the NH listeners showed no differences in map presence between auditory and modulated responses, hinting towards similar neural processes

to evaluate these stimuli.

Regarding the CI users, the increased presence of a conventional N1 topography (map 6) specifically for the modulated response indicates alterations in audiovisual processing and a multisensory benefit for CI users, if additional visual information is provided. It is likely that these cortical alterations in CI users reflect compensatory strategies, which CI users develop to overcome the limited CI input. This compensatory processing strategy may allow the CI users to better use the visual input to reach a processing level closer to the one of the NH listeners. This is confirmed by direct group comparisons showing differences between NH listeners and CI users only for the auditory, but not for the modulated response. Overall, as suggested by [Stevenson et al. \(2014\)](#), hierarchical clustering with subsequent single-subject fitting is a powerful approach to study multisensory interactions. This was confirmed in this study, since this method allowed us to demonstrate a clear audiovisual benefit for CI users when they were provided with additional visual input. We suggest that further studies using more complex speech stimuli should apply this approach as well to study multisensory interactions in CI users. As with behavioural results ([Moody-Antonio et al., 2005](#); [Rouger et al., 2008](#); [Tremblay et al., 2010](#)), it may be speculated that cortical audiovisual interactions in CI users are more pronounced if speech stimuli with semantic information are used.

On the sensor level, ERP data showed a prolonged N1 latency for CI users compared to NH listeners, for both the modulated and the auditory-only responses. Hierarchical clustering and single-subject fitting analyses confirmed the delayed N1 topography in CI users compared to NH listeners by assessing the first onset of map 6 (N1 topography). On the source level, again, the source activity results confirmed a delayed N1 response for CI users compared to NH listeners for both auditory and visual cortices. In fact, one might assume that the time until the electrical signal reaches the auditory nerve with a CI is faster than in NH listeners, since the signal does not have to undergo the processing through the middle and inner ear due to the direct stimulation of auditory nerve fibres. Interestingly, in the study of [Schierholz et al. \(2017\)](#), delayed cortical responses have been found not only in patients with a CI but also in patients with an auditory brainstem (ABI; direct stimulation of the cochlear nucleus) and an auditory midbrain implant (AMI; direct stimulation of the inferior colliculus). Although central auditory implants bypass more structures than the CI, ABI and AMI patients revealed even more delayed cortical responses and poorer behavioral abilities when compared to CI patients ([Schierholz et al., 2017](#)). Given that ABI and AMI are not yet optimised in electrode array designs and processing strategies ([Lim and Lenarz, 2015](#); [McKay et al., 2013](#)), the remarkable delay of the cortical response seems to be particularly related to the insufficient input provided by the central auditory implants ([Schierholz et al., 2017](#)). Similarly, the auditory input provided by the CI is limited, and the delayed N1 ERP responses in CI users, when compared to NH listeners, may reflect enhanced difficulties to process speech sounds, resulting in more time to encode the stimuli. Indeed, previous studies with NH listeners have shown that difficult acoustic listening conditions, for instance speech in background noise, cause a delay in the N1 response ([Billings et al., 2011](#); [Finke et al., 2016a](#)). Moreover, our results of a slowed cortical N1 ERP in CI users are consistent with several other studies, which used purely auditory stimuli of different acoustic complexity in unilaterally stimulated CI users ([Beynon et al., 2005](#); [Finke et al., 2016a](#); [Sandmann et al., 2009, 2015](#)) and in bilaterally stimulated CI users ([Senkowski et al., 2014](#)).

Taken together, our results revealed group differences for the N1 ERP. First, the hierarchical clustering and single-subject fitting analyses revealed a *condition effect* specifically for the group of CI users, showing distinct patterns of N1 topography between the auditory-only (A) and the modulated auditory response (AV-V; [Fig. 3A](#)). By contrast, the NH listeners showed no differences in map presence between auditory and modulated responses, hinting towards similar neural processes to evaluate these stimuli ([Fig. 3A](#)). However, the CI-specific condition effect points to a particularly strong visual impact in CI users on auditory

speech processing, which may allow these individuals to compensate for the degraded CI input. Apart from this CI-specific condition effect, our results revealed a *map dissimilarity at N1 latency between the two groups*, independent of the condition. The topographic and source analyses revealed that these group differences at N1 latency can be explained by a delayed cortical N1 response in CI users when compared to NH listeners. However, as discussed in the next two sections (4.2.3 and 4.2.4), the source analyses also revealed that this map dissimilarity at N1 latency is additionally caused by a distinct configuration of neural sources. When compared to the NH listeners, the CI users showed an additional activation in the visual cortex and a delayed cortical activation with a reversed, rightward functional lateralisation in the auditory cortex (see sections 4.2.3 and 4.2.4 for more details).

#### 4.2.3. Intra-modal plasticity in the auditory cortex

The source analysis of cortical activity revealed hemispheric differences between NH listeners and CI users, with NH listeners showing more activity in the left auditory cortex and CI users showing more activity in the right auditory cortex. There are two possible explanations for these results.

The first explanation is associated with the CI-related stimulus degradation, which affects the basic acoustic properties of the presented syllables. Indeed, many CI users reported that they did not perceive the syllables as syllables, but rather as distinct sounds. Therefore, it is conceivable that the CI users were able to discriminate the different stimuli, although the sounds were not identified as speech sounds. The two hemispheres of the brain have been found to be specialised for the processing of distinct basic acoustic properties ([Belin et al., 2000](#); [Poeppl, 2003](#); [Zatorre and Belin, 2001](#)). The left auditory cortex is involved in speech processing, as it preferentially processes fast temporal cues, which are largely contained in speech signals ([Boemio et al., 2005](#)). Whereas the right auditory cortex has been found to be more involved in the processing of spectral aspects of sounds and slowly modulated temporal acoustic cues, which is a property of musical stimuli ([Belin et al., 2000](#); [Poeppl, 2003](#); [Zatorre and Belin, 2001](#)). Thus, the hemispheric differences during speech and music processing can be attributed to the relative specialisation of the two hemispheres for basic acoustic stimulus properties, in particular fast temporal versus slow spectrotemporal acoustic cues. For this reason, one can assume that there must be more involvement of the left auditory cortex for processing syllables, which is the case for NH listeners. For CI users however, the present results show more recruitment of the right auditory cortex. It is likely that this hemispheric difference was caused by the CI-related syllable degradation, resulting in alterations of basic acoustic properties of the CI signal and leading to a relatively enhanced activation in the right auditory cortex. This interpretation is confirmed by our finding that CI users rather perceived the syllables more as meaningless sounds and not as linguistic input.

The second explanation for group differences in hemispheric asymmetries is related to cortical reorganisation, which might be caused by deafness before implantation and subsequent fitting with a CI. In the study of [Sandmann et al. \(2009\)](#) this claim has been investigated by comparing processing of NH listeners for normal and noise-vocoded stimuli to simulate the hearing experience of a CI patient. The results revealed no difference between the normal and vocoded responses in NH listeners, providing evidence against pure stimulus degradation effects. The authors rather suggest that functional differences arose due to a distinct hearing experience as a consequence of cortical plasticity. Our CI users experienced profound deafness, and through the supply of a CI, the hearing was restored but was still artificial and highly different compared to normal acoustic hearing. This process leads to experience-related changes in the normal pattern of response asymmetry. A previous study with unilaterally-implanted CI users ([Mortensen et al., 2006](#)) suggests that it is important to maintain the left auditory cortex for auditory processing to show more success with a CI. This study found associations for speech-induced left superior temporal gyrus activation

with high levels of speech comprehension in CI users. Accordingly, [Giraud et al., \(2001a\)](#) demonstrated enhanced activation in the left auditory cortex for phonology-specific stimuli for NH listeners compared to unilaterally-implanted CI users, which was interpreted as decreased left hemispheric specialisation for the processing of speech in CI users. This is in line with other studies on patients with profound hearing loss ([Fujiki et al., 1998](#); [Ponton et al., 2001](#)) and studies with CI patients showing cortical reorganisation after cochlear implantation ([Green et al., 2005](#); [Pantev et al., 2006](#)).

In the present study, we can report group differences in auditory cortex asymmetries as well, as our groups were carefully matched in stimulation side. Since all participants were right-handed, we can exclude lateralisation effects caused by handedness. However, we cannot exclude that differences happened due to the side of implantation, as we could not split up our groups into right-implanted and left-implanted patients due to the limited sample size. Further studies are required to systematically disentangle the effects of implantation side and cortical asymmetries in CI users for different kinds of stimulation (speech and non-speech).

#### 4.2.4. Cross-modal plasticity in the visual cortex

The source analysis showed a stronger activation in the visual cortex for CI users compared to NH listeners, hinting towards cross-modal activation of the visual cortex. This confirms the results from previous studies, reporting that cortical changes in CI users are not restricted to the auditory cortex ([Campbell and Sharma, 2013, 2016](#); [Chen et al., 2016](#); [Giraud et al., 2001a](#); [Giraud et al., 2001b](#)), but are also present in visual and even in inferior frontal areas ([Rouger et al., 2012](#)).

Previous results have suggested that CI users compensate the limited auditory input provided by the CI by additionally recruiting the visual cortex ([Doucet et al., 2006](#); [Giraud et al., 2001c](#); [Strelnikov et al., 2010, 2013](#)). [Giraud et al. \(2001c\)](#) reported in their PET (positron emission tomography) study greater auditory activation in the visual cortex for meaningful speech sounds in unilaterally-implanted CI users compared to NH listeners. This additional visual activity was associated with enhanced speech-reading abilities and longer CI experience ([Giraud et al., 2001c](#)). The authors suggest that the activation of the visual cortex for auditory processing is present due to an enhanced synergy of the auditory and visual modalities in CI users. It is conceivable that this audiovisual synergy helps recovering speech comprehension after cochlear implantation, when CI users need to learn to match the novel auditory speech inputs with the corresponding visual speech cues ([Strelnikov et al., 2013](#)). Indeed, the finding of a positive correlation between visual-cortex activation and the duration of CI experience indicates that the visual modality helps to restore auditory speech perception. This is in line with our results, demonstrating a positive relationship between CI experience and visual cortex activation, supporting experience-related plastic changes after implantation.

Up to now, the cross-modal recruitment of the visual cortex for processing auditory stimuli in CI users has been observed for syllables, words, environmental sounds, pure tones and reversed words ([Chen et al., 2016](#); [Giraud et al., 2001a](#); [Giraud et al., 2001c](#)). Thus, cross-modal activation of the visual cortex is not restricted to a semantic content of the auditory stimulus ([Chen et al., 2016](#)), but seems to be highly beneficial with regards to behavioural abilities, such as lip-reading ([Rouger et al., 2007](#); [Stropahl et al., 2015](#); [Stropahl and Debener, 2017](#)) and audiovisual word integration ([Rouger et al., 2007](#)).

Our study is the first EEG study showing visual recruitment for auditory (and audiovisual) stimuli and a positive correlation of CI experience with visual cortex activation in CI users, extending existing evidence in this direction. This confirms that EEG is a suitable tool to study cortical reorganisation in CI users with the advantage of the highly precise temporal resolution allowing to track single processing steps. Here, we found differences in cortical activation between CI users and NH listeners already at the initial cortical processing stage of the N1 time window (approximately 100 ms post stimulus onset).

#### 4.3. Clinical implications

Given that life expectancy and the exposure to noise are rising, hearing problems will become increasingly prevalent in the ageing population ([Sun et al., 2021](#)). Therefore, treatment methods have to be optimised, as hearing loss is an important factor influencing the quality of life (confirmed here with the NCIQ). As both auditory and visual input (lip movements, gestures) contribute to speech perception, spoken language communication is an audiovisual task ([Drijvers and Özyürek, 2017](#); [Grant et al., 1998](#); [Sumbly and Pollack, 1954](#)). Especially in difficult hearing situations (for instance speech in background noise) the supportive role of the visual modality can be crucial for understanding speech ([Sumbly and Pollack, 1954](#)). Enhanced lip-reading and audiovisual integration abilities may support advantages in perceiving speech in ecologically valid listening situations comparable to real-life situations including both modalities ([Rouger et al., 2007, 2012](#); [Stropahl et al., 2015, 2017](#)). Therefore, it is essential to not only investigate auditory or visual processing in CI users separately, but to focus on the interaction of the two modalities. Currently, the clinical assessment of speech perception in CI users is typically restricted to purely auditory methods. Given the importance of visual input in real-life situations, this approach is insufficient for gauging the actual capabilities of CI patients ([Stevenson et al., 2017](#)). To further optimise treatment options for CI patients, it is important to develop audiovisual, ecologically valid assessment methods, which requires an improved understanding of the relationship between multisensory interactions and hearing performance with a CI. The present study confirmed that CI users highly benefit from additional visual input and seem to use distinct processing strategies compared to NH listeners. In particular, their multisensory processes remain intact despite the insufficient auditory CI input. Our results are of clinical relevance, because strong audiovisual interactions could help optimising auditory rehabilitation after cochlear implantation. In the first weeks and months after implantation, the CI users need to learn to match the novel auditory inputs from the CI with the corresponding visual cues ([Strelnikov et al., 2009](#)). It may be speculated that audiovisual trainings can fasten and improve the auditory rehabilitation with a CI. Indeed, first results from NH listeners (with vocoded speech sounds) as well as results from late-implanted deafened ferrets have indicated that audiovisual training can enhance auditory perceptual learning ([Bernstein et al., 2013](#); [Bernstein et al., 2014](#); [Isaiah et al., 2014](#); [Kawase et al., 2009](#)). Thus, future studies should examine whether these positive effects of audiovisual training on auditory-only perceptual learning can also be confirmed in CI users by including different groups (control group, training group) and by differentiating patients with short and long CI experience. This would allow to find out whether an audiovisual-based rehabilitation strategy can optimise rehabilitation and whether this approach is applicable even several months after CI switch-on.

#### 5. Limitations

One limitation of this study is that we chose to stimulate only one ear, which might be unnatural for both the CI users and the NH listeners. This limits the conclusions regarding cross-modal reorganisation in CI users and makes it difficult to directly connect the experimental setting to the real world, as usually both ears are used. However, unilateral stimulation enabled us to control for the possible confounds caused by the contralateral ear (different hearing systems and hearing thresholds). Another limitation is that we used headphones for NH controls and loudspeakers for the CI users to present the stimuli. As we chose to stimulate one ear only, one cannot use loudspeakers for NH controls, as it is not possible to appropriately mute the second ear. Nevertheless, both the loudspeakers and the headphones were equally calibrated to an intensity level of 65 dB SPL, and a loudness rating scale was used to verify that the stimulus situation was as similar as possible across the two groups of participants. Moreover, one could argue that the different

approach of auditory presentation could have influenced audiovisual integration. We are not aware about studies that systematically examined differences in audiovisual speech, comparing the stimulation via headphones with the stimulation via loudspeakers. However, precise time-alignment of the auditory and visual stimulus – an important factor in multimodal processing – was given in both presentation schemes. Still, in order to better understand alterations in audiovisual processing after cochlear implantation, future studies are required to systematically compare the stimulation of the CI ear alone, the contralateral ear alone and both ears together. This should be done with ecologically valid stimuli, such as complex speech stimuli, and different task conditions (e. g. passive vs. active). Further, we suggest that future studies should compare different subgroups of CI patients (bimodal (CI + hearing aid), bilateral (CI + CI), single-sided deaf (CI + NH)) to explore potential differences in multisensory processing and cross-modal reorganisation across the different CI groups. Specifically, these studies need to assess whether the observed alterations in audiovisual speech processing in different groups of CI users can be used to fasten and improve the auditory rehabilitation after cochlear implantation.

## 6. Conclusion

This EEG study used electrical neuroimaging, including topographic and ERP source analysis, to examine cortical interactions during audiovisual syllable discrimination in CI users and NH listeners. Our results revealed a clear multisensory effect in both the CI users and the NH listeners, as indicated by shorter behavioural response times and reduced auditory-cortex activation at N1 and P2 latency. In addition, for the N1 ERP topography we found a condition effect specifically for the group of CI users, showing a distinct pattern of N1 topography between the auditory-only (A) and the modulated auditory response (AV-V). This observation points to a strong visual impact on auditory speech processing in CI users, which allows these individuals to compensate for the limited CI input. This visual impact on auditory speech processing is a manifestation of multisensory processes that remain intact in CI users despite the limited auditory signal provided by the CI. Finally, for the N1 ERP topography we found a group effect, which was independent of the condition (A, AV-V). Our source analyses suggest that this group difference is caused by two different reasons. First, the CI users show a delay in the cortical N1 response, pointing to difficulties in the cortical processing of the degraded signal. Second, the groups show distinct configurations of cortical sources at N1 latency, including not only a distinct directionality of auditory-cortex asymmetry (left-ward lateralisation in NH listeners, right-ward lateralisation in CI users), but also the additional recruitment of the visual cortex in the group of CI users, which was positively related to the duration of CI experience. These results confirm distinct multisensory processing between CI users and NH listeners, highlighting the need for audiovisual testing in clinical practice and for developing audiovisual methods in order to improve auditory rehabilitation.

## Funding

This work was supported by the German Research Foundation (DFG) [SA 3615/1-1 (project number: 415896102) and SA 3615/2-1 (project number: 416867313)] and by the Köln Fortune Program/Faculty of Medicine, University of Cologne [278/2020] to PS. MMM is supported by the Swiss National Science Foundation (169206).

## CRedit authorship contribution statement

**Natalie Layer:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration. **Anna Weglage:** Software, Validation, Formal analysis, Investigation, Writing – review & editing. **Verena Müller:** Conceptualization, Writing – review & editing.

**Hartmut Meister:** Conceptualization, Software, Resources, Writing – review & editing. **Ruth Lang-Roth:** Conceptualization, Writing – review & editing. **Martin Walger:** Conceptualization, Writing – review & editing. **Micah M. Murray:** Software, Formal analysis, Validation, Writing – review & editing. **Pascale Sandmann:** Conceptualization, Methodology, Validation, Resources, Writing – original draft, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

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.

## Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG) [SA 3615/1-1 (project number: 415896102) and SA 3615/2-1 (project number: 416867313)] and Köln Fortune [278/2020] for providing funding for this research. Additionally, we thank Pauline Burkhardt and Lina Wiesel for their help in data collection and Moritz Layer for his support with the implementation of the analysis. Finally, we thank all participants for their time and effort.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2022.102982>.

## References

- Ahveninen, J., Jääskeläinen, I.P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B.G., Witzel, T., Belliveau, J. W., 2006. Task-modulated “what” and “where” pathways in human auditory cortex. *Proc. Natl. Acad. Sci.* 103 (39), 14608–14613.
- Allen, J.B., Hall, J.L., Jeng, P.S., 1990. Loudness growth in 1/2-octave bands (LGOB)—a procedure for the assessment of loudness. *J. Acoust. Soc. Am.* 88 (2), 745–753.
- Barth, D.S., Goldberg, N., Brett, B., Di, S., 1995. The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Res.* 678 (1–2), 177–190.
- Bavelier, D., Dye, M.W.G., Hauser, P.C., 2006. Do deaf individuals see better? *Trends Cognitive Sci.* 10 (11), 512–518.
- Bavelier, D., Neville, H.J., 2002. Cross-modal plasticity: Where and how? *Nat. Rev. Neurosci.* 3 (6), 443–452.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403 (6767), 309–312.
- Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7 (6), 1129–1159.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 57 (1), 289–300.
- Bernstein, L.E., Auer, E.T., Eberhardt, S.P., Jiang, J., 2013. Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Front. Neurosci.* 7, 1–16.
- Bernstein, L.E., Eberhardt, S.P., Auer, E.T., 2014. Audiovisual spoken word training can promote or impede auditory-only perceptual learning: results from prelingually deafened adults with late-acquired cochlear implants and normal-hearing adults. *Front. Psychol.* 5, 1–20.
- Besle, J., Fort, A., Delpuech, C., Giard, M.-H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20 (8), 2225–2234.
- Beynon, A.J., Snik, A.F.M., Stegeman, D.F., van den Broek, P., 2005. Discrimination of speech sound contrasts determined with behavioral tests and event-related potentials in cochlear implant recipients. *J. Am. Acad. Audiol.* 16 (01), 042–053.
- Biasiucci, A., Franceschiello, B., Murray, M.M., 2019. Electroencephalography. *Curr. Biol.* 29 (3), R80–R85.
- Billings, C.J., Bennett, K.O., Molis, M.R., Leek, M.R., 2011. Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear.* 32 (1), 53.
- Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8 (3), 389–395.
- Bosnyak, D.J., Eaton, R.A., Roberts, L.E., 2004. Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb. Cortex* 14 (10), 1088–1099.
- Bottari, D., Bednaya, E., Dormal, G., Villwock, A., Dzhelevova, M., Grin, K., Pietrini, P., Ricciardi, E., Rossion, B., Röder, B., 2020. EEG frequency-tagging demonstrates increased left hemispheric involvement and crossmodal plasticity for face processing in congenitally deaf signers. *NeuroImage* 223, 117315.

- Bottari, D., Heimler, B., Caclin, A., Dalmolin, A., Giard, M.-H., Pavani, F., 2014. Visual change detection recruits auditory cortices in early deafness. *Neuroimage* 94, 172–184.
- Brüneck, I., Hölsken, S., Kessler, J., 2018. Der DemTect Eye+ Ear-Neues kognitives Screeningverfahren bei schwerhörigen Menschen mit Demenzverdacht. *Zeitschrift für Audiologie* 57 (3), 121.
- Brunet, D., Murray, M. M., Ichel, C. M. (2011). *Spatiotemporal analysis of multichannel EEG: CARTOOL. Computational Intelligence and Neuroscience, 2011.*
- Campbell, J., Sharma, A., 2013. Compensatory changes in cortical resource allocation in adults with hearing loss. *Front. Syst. Neurosci.* 7, 71.
- Campbell, J., Sharma, A., 2016. Visual cross-modal re-organization in children with cochlear implants. *PLoS ONE* 11 (1), e0147793.
- Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30 (38), 12572–12580.
- Cappelletta, L., Harte, N., 2012. Phoneme-to-viseme mapping for visual speech recognition. *ICPRAM* 2, 322–329.
- Chen, L.-C., Sandmann, P., Thorne, J.D., Bleichner, M.G., Debener, S., 2016. Cross-modal functional reorganization of visual and auditory cortex in adult cochlear implant users identified with fNIRS. *Neural Plasticity* 2016, 1–13.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26 (1), 55–67.
- De Santis, L., Clarke, S., Murray, M.M., 2007. Automatic and intrinsic auditory “what” and “where” processing in humans revealed by electrical neuroimaging. *Cereb. Cortex* 17 (1), 9–17.
- Debener, S., Hine, J., Bleek, S., Eyles, J., 2008. Source localization of auditory evoked potentials after cochlear implantation. *Psychophysiology* 45 (1), 20–24.
- Delorme, A., Makeig, S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21.
- Desai, S., Stickney, G., Zeng, F.-G., 2008. Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *J. Acoust. Soc. Am.* 123 (1), 428–440.
- Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53 (1), 1–15.
- Dong, L., Foo, S. W., Lian, Y. (2003). Modeling continuous visual speech using boosted viseme models. *Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, 3*, 1394–1398.
- Doucet, M., Bergeron, F., Lassonde, M., Ferron, P., Lepore, F., 2006. Cross-modal reorganization and speech perception in cochlear implant users. *Brain* 129 (12), 3376–3383.
- Drennan, W. R., & Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *J. Rehabilitation Res. Develop.*, 45(5), 779.
- Drijvers, L., Özyürek, A., 2017. Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J. Speech Language Hearing Res.* 60 (1), 212–222.
- Dürschmid, S., Reichert, C., Hinrichs, H., Heinze, H.-J., Kirsch, H.E., Knight, R.T., Deouell, L.Y., 2019. Direct evidence for prediction signals in frontal cortex independent of prediction error. *Cereb. Cortex* 29 (11), 4530–4538.
- Fagel, S., Clemens, C., 2004. An articulation model for audiovisual speech synthesis—determination, adjustment, evaluation. *Speech Commun.* 44 (1–4), 141–154.
- Finke, M., Büchner, A., Ruigendijk, E., Meyer, M., Sandmann, P., 2016a. On the relationship between auditory cognition and speech intelligibility in cochlear implant users: an ERP study. *Neuropsychologia* 87, 169–181.
- Finke, M., Sandmann, P., Bönitz, H., Kral, A., Büchner, A., 2016b. Consequences of stimulus type on higher-order processing in single-sided deaf cochlear implant users. *Audiol. Neurotol.* 21 (5), 305–315.
- Finney, E.M., Clementz, B.A., Hickok, G., Dobkins, K.R., 2003. Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG. *NeuroReport* 14 (11), 1425–1427.
- Fujiki, N., Naito, Y., Nagamine, T., Shiomi, Y., Hirano, S., Honjo, I., Shibasaki, H., 1998. Influence of unilateral deafness on auditory evoked magnetic field. *NeuroReport* 9 (14), 3129–3133.
- Giraud, A.L., Price, C.J., Graham, J.M., Frackowiak, R.S., 2001a. Functional plasticity of language-related brain areas after cochlear implantation. *Brain* 124 (7), 1307–1316.
- Giraud, A.L., Truy, E., Frackowiak, R., 2001b. Imaging plasticity in cochlear implant patients. *Audiol. Neurotol.* 6 (6), 381–393.
- Giraud, A.-L., Price, C.J., Graham, J.M., Truy, E., Frackowiak, R.S.J., 2001c. Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron* 30 (3), 657–664.
- Glennon, E., Svirsky, M.A., Froemke, R.C., 2020. Auditory cortical plasticity in cochlear implant users. *Curr. Opin. Neurobiol.* 60, 108–114.
- Gramfort, A., Papadopoulos, T., Olivi, E., Clerc, M., 2010. OpenMEEG: Opensource software for quasistatic bioelectromagnetics. *Biomed. Eng. Online* 9 (1), 1–20.
- Grant, K.W., Walden, B.E., Seitz, P.F., 1998. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *J. Acoust. Soc. Am.* 103 (5), 2677–2690.
- Green, K.M.J., Julian, P.J., Hastings, D.L., Ramsden, R.T., 2005. Auditory cortical activation and speech perception in cochlear implant users: effects of implant experience and duration of deafness. *Hear. Res.* 205 (1–2), 184–192.
- Hahlbrock, K.H., 1970. *Sprachaudiometrie: Grundlagen und praktische Anwendung einer Sprachaudiometrie für das deutsche Sprachgebiet.* Georg Thieme Verlag.
- Hansen, P., Kringsbach, M., Salmelin, R., 2010. *MEG: An introduction to methods.* Oxford University Press.
- Hauthal, N., Thorne, J.D., Debener, S., Sandmann, P., 2014. Source localisation of visual evoked potentials in congenitally deaf individuals. *Brain Topogr.* 27 (3), 412–424.
- Heimler, B., Weisz, N., Collignon, O., 2014. Revisiting the adaptive and maladaptive effects of crossmodal plasticity. *Neuroscience* 283, 44–63.
- Henkin, Y., Yaar-Soffer, Y., Steinberg, M., Muchnik, C., 2014. Neural correlates of auditory-cognitive processing in older adult cochlear implant recipients. *Audiol. Neurotol.* 19 (1), 21–26.
- Hinderink, J.B., Krabbe, P.F.M., Van Den Broek, P., 2000. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the nijmegen cochlear implant questionnaire. *Otolaryngol.-Head Neck Surg.* 123 (6), 756–765.
- Inui, K., Okamoto, H., Miki, K., Gunji, A., Kakigi, R., 2006. Serial and parallel processing in the human auditory cortex: a magnetoencephalographic study. *Cereb. Cortex* 16 (1), 18–30.
- Isaiah, A., Vongpaisal, T., King, A.J., Hartley, D.E.H., 2014. Multisensory training improves auditory spatial processing following bilateral cochlear implantation. *J. Neurosci.* 34 (33), 11119–11130.
- Jachimski, D., Czyzewski, A., Ciszewski, T., 2018. A comparative study of english viseme recognition methods and algorithms. *Multimedia Tools Appl.* 77 (13), 16495–16532.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M.J., Iragui, V., Sejnowski, T.J., 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37 (2), 163–178.
- Kaiser, A.R., Kirk, K.I., Lachs, L., Pisoni, D.B., 2003. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J. Speech Language Hear. Res.* 46 (2), 390–404.
- Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A.P., Brand, M., Bullock, R., 2004. DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int. J. Geriatric Psych.* 19 (2), 136–143.
- Kawase, T., Sakamoto, S., Hori, Y., Maki, A., Suzuki, Y., Kobayashi, T., 2009. Bimodal audio–visual training enhances auditory adaptation process. *NeuroReport* 20, 1231–1234.
- Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T., 2004. Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158 (4), 405–414.
- Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.* 48 (6), 609–621.
- Lim, H.H., Lenarz, T., 2015. Auditory midbrain implant: research and development towards a second clinical trial. *Hear. Res.* 322, 212–223.
- Lin, F.-H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31 (1), 160–171.
- Lucey, P., Martin, T., Sridharan, S. (2004). Confusability of phonemes grouped according to their viseme classes in noisy environments. *Proc. Of Australian Int. Conf. On Speech Science & Tech*, 265–270.
- Lütkenhöner, B., Steinsträter, O., 1998. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol. Neurotol.* 3 (2–3), 191–213.
- Mahavidyalaya, B. V. (2014). Phoneme and viseme based approach for lip synchronization. *Int. J. Signal Process., Image Process. Pattern Recogn.*, 7(3), 385–394.
- Massaro, D.W., Light, J., 2004. Using visible speech to train perception and production of speech for individuals with hearing loss. *J. Speech Lang Hear. Res.* 47 (2), 304–320.
- McKay, C.M., Lim, H.H., Lenarz, T., 2013. Temporal processing in the auditory system: insights from cochlear and auditory midbrain implantees. *J. Assoc. Res. Otolaryngol.* 14 (1), 103–124.
- Meister, H., Fuersen, K., Schreitmueller, S., Walger, M., 2016. Effect of acoustic fine structure cues on the recognition of auditory-only and audiovisual speech. *J. Acoust. Soc. Am.* 139 (6), 3116–3120.
- Merzenich, M.M., Van Vleet, T.M., Nahum, M., 2014. Brain plasticity-based therapeutics. *Front. Hum. Neurosci.* 8, 385.
- Micco, A.G., Kraus, N., Koch, D.B., McGee, T.J., Carrell, T.D., Sharma, A., Nicol, T., Wiet, R.J., 1995. Speech-evoked cognitive P300 potentials in cochlear implant recipients. *Am. J. Otol.* 16 (4), 514–520.
- Michel, C.M., Koenig, T., 2018. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage* 180, 577–593.
- Michel, C.M., Koenig, T., Brandeis, D., Gianotti, L.R., Wackermann, J., 2009. *Electrical Neuroimaging.* Cambridge University Press.
- Michel, C.M., Murray, M.M., 2012. Towards the utilization of EEG as a brain imaging tool. *Neuroimage* 61 (2), 371–385.
- Miller, J., 1982. Divided attention: Evidence for coactivation with redundant signals. *Cogn. Psychol.* 14 (2), 247–279.
- Mitchell, T.V., Maslin, M.T., 2007. How vision matters for individuals with hearing loss. *Int. J. Audiol.* 46 (9), 500–511.
- Moody-Antonio, S., Takayangi, S., Masuda, A., Auer Jr, E.T., Fisher, L., Bernstein, L.E., 2005. Improved speech perception in adult congenitally deafened cochlear implant recipients. *Otol. Neurotol.* 26 (4), 649–654.
- Mortenson, M.V., Mirz, F., Gjedde, A., 2006. Restored speech comprehension linked to activity in left inferior prefrontal and right temporal cortices in postlingual deafness. *Neuroimage* 31 (2), 842–852.
- Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20 (4), 249–264.



- Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24 (4), 375–425.
- Obleser, J., Elbert, T., Lahiri, A., Eulitz, C., 2003. Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Res.* 15 (3), 207–213.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9 (1), 97–113.
- Pantev, C., Dinnesen, A., Ross, B., Wollbrink, A., Knief, A., 2006. Dynamics of auditory plasticity after cochlear implantation: a longitudinal study. *Cereb. Cortex* 16 (1), 31–36.
- Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72 (2), 184–187.
- Poeppl, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time”. *Speech Commun.* 41 (1), 245–255.
- Ponton, C.W., Vasama, J.-P., Tremblay, K., Khosla, D., Kwong, B., Don, M., 2001. Plasticity in the adult human central auditory system: Evidence from late-onset profound unilateral deafness. *Hear. Res.* 154 (1-2), 32–44.
- Prince, P., Paul, B.T., Chen, J., Le, T., Lin, V., Dimitrijevic, A., 2021. Neural correlates of visual stimulus encoding and verbal working memory differ between cochlear implant users and normal-hearing controls. *Eur. J. Neurosci.*
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Raab, D. H. (1962). Division of psychology: Statistical facilitation of simple reaction times. *Trans. New York Acad. Sci.*, 24(5 Series II), 574–590.
- Ratcliff, R., 1979. Group reaction time distributions and an analysis of distribution statistics. *Psychol. Bull.* 86 (3), 446–461.
- van de Rijt, L.P., Roye, A., Mylanus, E.A., van Opstal, A.J., van Wanrooij, M.M., 2019. The principle of inverse effectiveness in audiovisual speech perception. *Front. Hum. Neurosci.* 13, 335.
- Ross, B., Tremblay, K., 2009. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hear. Res.* 248 (1-2), 48–59.
- Rouger, J., Fraysse, B., Deguine, O., Barone, P., 2008. McGurk effects in cochlear-implanted deaf subjects. *Brain Res.* 1188, 87–99.
- Rouger, J., Lagleyre, S., Démonet, J.-F., Fraysse, B., Deguine, O., Barone, P., 2012. Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Hum. Brain Mapp.* 33 (8), 1929–1940.
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Barone, P., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc. Natl. Acad. Sci.* 104 (17), 7295–7300.
- Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R.D., Marcar, V.L., Jäncke, L., Debener, S., 2012. Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain* 135 (2), 555–568.
- Sandmann, P., Eichele, T., Buechler, M., Debener, S., Jäncke, L., Dillier, N., Hugdahl, K., Meyer, M., 2009. Evaluation of evoked potentials to dyadic tones after cochlear implantation. *Brain* 132 (7), 1967–1979.
- Sandmann, P., Kegel, A., Eichele, T., Dillier, N., Lai, W., Bendixen, A., Debener, S., Jancke, L., Meyer, M., 2010. Neurophysiological evidence of impaired musical sound perception in cochlear-implant users. *Clin. Neurophysiol.* 121 (12), 2070–2082.
- Sandmann, P., Plotz, K., Hauthal, N., de Vos, M., Schönfeld, R., Debener, S., 2015. Rapid bilateral improvement in auditory cortex activity in postlingually deafened adults following cochlear implantation. *Clin. Neurophysiol.* 126 (3), 594–607.
- Schiel, F., 1999. Automatic phonetic transcription of non-prompted speech. *Open Access LMU*.
- Schierholz, I., Finke, M., Kral, A., Büchner, A., Rach, S., Lenarz, T., Dengler, R., Sandmann, P., 2017. Auditory and audio-visual processing in patients with cochlear, auditory brainstem, and auditory midbrain implants: an EEG study. *Hum. Brain Mapp.* 38 (4), 2206–2225.
- Schierholz, I., Finke, M., Schulte, S., Hauthal, N., Kantzke, C., Rach, S., Büchner, A., Dengler, R., Sandmann, P., 2015. Enhanced audio-visual interactions in the auditory cortex of elderly cochlear-implant users. *Hear. Res.* 328, 133–147.
- Schierholz, I., Schönemark, C., Ruigendijk, E., Kral, A., Kopp, B., Büchner, A., 2021. An event-related brain potential study of auditory attention in cochlear implant users. *Clinical* 132 (9), 2290–2305.
- Schreitmüller, S., Frenken, M., Bentz, L., Ortmann, M., Walger, M., Meister, H., 2018. Validating a method to assess lipreading, audiovisual gain, and integration during speech reception with cochlear-implanted and normal-hearing subjects using a talking head. *Ear Hear.* 39 (3), 503–516.
- Senkowski, D., Pomper, U., Fitzner, I., Engel, A.K., Kral, A., 2014. Beta-band activity in auditory pathways reflects speech localization and recognition in bilateral cochlear implant users. *Hum. Brain Mapp.* 35 (7), 3107–3121.
- Sharma, A., Dorman, M.F., Spahr, A.J., 2002. A sensitive period for the development of the central auditory system in children with cochlear implants: Implications for age of implantation. *Ear Hear.* 23 (6), 532–539.
- Stein, B., Meredith, M., 1993. *The merging of the senses*. The MIT Press, Cambridge, MA.
- Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cognit. Neurosci.* 19 (12), 1964–1973.
- Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L. R., Siemann, J.K., James, T.W., Wallace, M.T., 2014. Identifying and quantifying multisensory integration: a tutorial review. *Brain Topogr.* 27 (6), 707–730.
- Stevenson, R., Sheffield, S.W., Butera, I.M., Gifford, R.H., Wallace, M., 2017. Multisensory integration in cochlear implant recipients. *Ear Hear.* 38 (5), 521.
- Strelnikov, K., Rouger, J., Barone, P., Deguine, O., 2009. Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scand. J. Psychol.* 50 (5), 437–444.
- Strelnikov, K., Rouger, J., Démonet, J.-F., Lagleyre, S., Fraysse, B., Deguine, O., Barone, P., 2010. Does brain activity at rest reflect adaptive strategies? Evidence from speech processing after cochlear implantation. *Cereb. Cortex* 20 (5), 1217–1222.
- Strelnikov, K., Rouger, J., Démonet, J.-F., Lagleyre, S., Fraysse, B., Deguine, O., Barone, P., 2013. Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain* 136 (12), 3682–3695.
- Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Démonet, J.-F., Déguine, O., Barone, P., 2015. Increased audiovisual integration in cochlear-implanted deaf patients: independent components analysis of longitudinal positron emission tomography data. *Eur. J. Neurosci.* 41 (5), 677–685.
- Stropahl, M., Bauer, A.-K.-R., Debener, S., Bleichner, M.G., 2018. Source-modeling auditory processes of EEG data using EEGLAB and brainstorm. *Front. Neurosci.* 12, 309.
- Stropahl, M., Chen, L.-C., Debener, S., 2017. Cortical reorganization in postlingually deaf cochlear implant users: intra-modal and cross-modal considerations. *Hear. Res.* 343, 128–137.
- Stropahl, M., Debener, S., 2017. Auditory cross-modal reorganization in cochlear implant users indicates audio-visual integration. *NeuroImage: Clinical* 16, 514–523.
- Stropahl, M., Plotz, K., Schönfeld, R., Lenarz, T., Sandmann, P., Yovel, G., De Vos, M., Debener, S., 2015. Cross-modal reorganization in cochlear implant users: auditory cortex contributes to visual face processing. *NeuroImage* 121, 159–170.
- Summy, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26 (2), 212–215.
- Sun, Z., Seo, J.W., Park, H.J., Lee, J.Y., Kwak, M.Y., Kim, Y., Lee, J.Y., Park, J.W., Kang, W.S., Ahn, J.H., Chung, J.W., Kim, H., 2021. Cortical reorganization following auditory deprivation predicts cochlear implant performance in postlingually deaf adults. *Hum. Brain Mapp.* 42 (1), 233–244.
- Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: A user-friendly application for MEG/EEG analysis. *Comput. Intel. Neurosci.* 2011, 1–13.
- Teder-Sälejärvi, W.A., Di Russo, F., McDonald, J.J., Hillyard, S.A., 2005. Effects of spatial congruity on audio-visual multimodal integration. *J. Cognit. Neurosci.* 17 (9), 1396–1409.
- Timm, L., Agrawal, D., Viola, F.C., Sandmann, P., Debener, S., Büchner, A., Dengler, R., Wittfoth, M., Malmierca, M.S., 2012. Temporal feature perception in cochlear implant users. *PLoS ONE* 7 (9), e45375.
- Tremblay, C., Champoux, F., Lepore, F., Théoret, H., 2010. Audiovisual fusion and cochlear implant proficiency. *Restor. Neurol. Neurosci.* 28 (2), 283–291.
- Tremblay, K., Ross, B., Inoue, K., McClannahan, K., Collet, G., 2014. Is the auditory evoked P2 response a biomarker of learning? *Front. Syst. Neurosci.* 8, 28.
- Ulrich, R., Miller, J., Schröter, H., 2007. Testing the race model inequality: an algorithm and computer programs. *Behav. Res. Methods* 39 (2), 291–302.
- van Wassenhove, V., Grant, K.W., Poeppl, D., 2005. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci.* 102 (4), 1181–1186.
- Vasil, K.J., Lewis, J., Tamati, T., Ray, C., Moberly, A.C., 2020. How does quality of life relate to auditory abilities? A subitem analysis of the nijmegen cochlear implant questionnaire. *J. Am. Acad. Audiol.* 31 (04), 292–301.
- Vaughan, H.G., 1982. The neural origins of human event-related potentials. *Ann. N. Y. Acad. Sci.* 388 (1), 125–138.
- Viola, F.C., De Vos, M., Hine, J., Sandmann, P., Bleeck, S., Eyles, J., Debener, S., 2012. Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials. *Hear. Res.* 284 (1-2), 6–15.
- Vroomen, J., Stekelenburg, J.J., 2010. Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cognit. Neurosci.* 22 (7), 1583–1596.
- Wesemann, W., Schiefer, U., Bach, M., 2010. Neue DIN-Normen zur Sehschärfebestimmung New DIN norms for determination of visual acuity. *Der Ophthalmologe* 107 (9), 821–826.
- Wesker, T., Meyer, B., Wagener, K., Anemüller, J., Mertins, A., Kollmeier, B., 2005. Oldenburg logatome speech corpus (OLLO) for speech recognition experiments with humans and machines. Ninth European Conference on Speech Communication and Technology.
- Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data—a practical approach. *J. Neurosci. Methods* 250, 34–46.
- Williams, N., 2017. The borg rating of perceived exertion (RPE) scale. *Occup. Med.* 67 (5), 404–405.
- Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11 (10), 946–953.