Journal of Neural Engineering

PURPOSE-LED PUBLISHING™

PAPER

Enhanced detection of envelope-following responses for objective fitting of cochlear-implant users

To cite this article: Julian Schott *et al* 2025 *J. Neural Eng.* **22** 026057

View the article online for updates and enhancements.

You may also like

- Evaluation of focused multipolar stimulation for cochlear implants: a preclinical safety study
Robert K Shepherd, Andrew K Wise, Ya Lang Enke et al.
- Optimizing deep brain stimulation based on isostable amplitude in essential tremor patient models
Benoit Duchet, Gihan Weerasinghe, Christian Bick et al.
- Development of the Electric Device for Skin Barrier Healing
Mayu Nakabayashi, Yuina Abe, Kuniaki Nagamine et al.

This content was downloaded from IP address 134.58.253.56 on 22/10/2025 at 12:15

Journal of Neural Engineering

PAPER

Enhanced detection of envelope-following responses for objective fitting of cochlear-implant users

Julian Schott[1,2,*], Robin Gransier[2], Marc Moonen[1] and Jan Wouters[2]

[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, 3001 Leuven, Belgium
[2] KU Leuven, Department of Neurosciences, ExpORL 3001 Leuven, Belgium
[*] Author to whom any correspondence should be addressed.

E-mail: julian.schott@kuleuven.be

RECEIVED
9 December 2024

REVISED
7 March 2025

ACCEPTED FOR PUBLICATION
28 March 2025

PUBLISHED
11 April 2025

Keywords: cochlear implant, objective measures, neural tracking, EEG, artifact, system identification, objective fitting

Abstract

Objective. Electrically evoked auditory steady-state responses (EASSRs) are potential neural responses for objectively determining stimulation parameters of cochlear implants (CIs). Unfortunately, they are difficult to detect in electroencephalography (EEG) recordings due to the electrical stimulation artifacts of the CI. This study investigates a novel stimulation paradigm hypothesized to improve artifact removal efficacy via system identification (SI), and therefore to improve response detection and clinical applicability. *Approach*. An amplitude-modulated (AM) CI stimulation pulse train with a step-wise increase in modulation frequency is created (referred to as SWEEP stimulation). Another stimulation is created by randomly shuffling modulation frequencies of the SWEEP stimulation (referred to as Shuffled-SWEEP stimulation). AM pulse trains with fixed modulation frequency (referred to as conventional AM stimulation), which elicit EASSRs, are also created for comparison. EEG data is collected from four CI users. A supra-threshold stimulation condition is used to investigate whether the SWEEP and Shuffled-SWEEP stimulation can elicit envelope-following responses (EFRs). A sub-threshold stimulation condition allows the collection of artifact-only EEG data, which is used to compare the SI accuracy on recordings from the SWEEP and the conventional AM stimulation. *Main results*. In all CI users, neural responses, following the SWEEP, Shuffled-SWEEP, and conventional AM stimulation are detected after artifact removal with SI. The validation with artifact-only EEG data shows higher $F1$ scores when comparing recordings with SWEEP stimulation ($F1 = 0.9$) to recordings with conventional AM stimulation ($F1 = 0.82$). *Significance*. Being able to accurately identify the response within one EEG recording enables the development of effective, online, objective fitting protocols. The increased neural response detection sensitivity with SWEEP stimulation reduces clinical recording time on average by a factor of 2.07. Detecting EFRs following complex stimulation paradigms offers a potential advancement in the systematic assessment of the temporal envelope processing in CI users.

1. Introduction

Cochlear implants (CIs) are considered the most successful neuroprostheses that are currently in clinical use. They have revolutionized hearing rehabilitation for individuals with severe to profound sensorineural hearing loss. By bypassing the damaged hair cells in the cochlea and electrically stimulating the auditory nerve, they enable hearing. One challenge in the hearing restoration of CI users is the recurring adjustment of the stimulation parameters. This so-called CI fitting is a time-consuming interaction between the CI user and a trained clinician. The quality therefore highly depends on the communication between both parties, the expertise of the clinician, and user dependent factors. Fitting results are

© 2025 IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

especially subjective and variable in challenging CI user populations such as infants, and can vary greatly over fitting sessions. Consequently, in the past years, an increasing interest in optimal and personalized fitting using objective measures has emerged in hearing rehabilitation research. Here, neural responses following CI stimulation are recorded, for instance with electroencephalography (EEG), and utilized to adjust the stimulation parameters.

Commonly used neural responses include the electrically evoked stapedius reflex (Gordon *et al* 2004a, 2004b, Caner *et al* 2007), the electrically evoked auditory brainstem response (EABR, Brown *et al* 2000, Miller *et al* 2008), electrically evoked compound action potential (ECAP Hughes *et al* 2000, Thai-Van *et al* 2004, Gordon *et al* 2004a, 2004b, Miller *et al* 2008), cortical auditory evoked potential (CAEP, Tavora-Vieira *et al* 2018), and electrically evoked auditory steady-state response (EASSR, Hofmann and Wouters 2010, 2012).

EABRs and ECAPs are transient auditory responses following single- or short bursts of CI stimulation pulses. They are relatively simple to detect and can even be recorded with the CI itself (Dillier *et al* 2002, Dijk *et al* 2007), making them popular in clinical practice. However, stimulation levels determined from these responses are suboptimal, as the CI stimulation to elicit these responses and the clinically used CI stimulation differ too much in stimulation rate (McKay *et al* 2013, McKay and Smale 2017). Furthermore, CI fitting based on ECAP recordings has lead to poorer speech comprehension scores than conventional subjective fitting (Seyle and Brown 2002), or to inconclusive results about the relevance of ECAPs (Smoorenburg *et al* 2002).

The CAEP is a cortical response following CI stimulation pulse bursts, that represent speech tokens such as /m/, /g/, /t/, or speech syllables such as /ba/, /da/, /ga/. The presence of CAEPs in EEG recordings may indicate that speech (acoustic cues in speech) is perceived and audible to the CI user (Carter *et al* 2013). CI stimulation levels determined from CAEP recordings have shown good correlation with behaviorally determined stimulation levels (Visram *et al* 2015, Mao *et al* 2019) and adjustments to stimulation parameters based on the presence of CAEPs can lead to improvements in speech comprehension scores (Tavora-Vieira *et al* 2022). However, CAEPs are often elicited with short stimuli, that cannot fully represent the temporal aspects of speech in daily-life communication.

More appropriate neural responses as objective measure are EASSRs, as the main signal feature of an acoustic input signal that is transmitted via CI stimulation, is the envelope, i.e. the slow amplitude fluctuations of the acoustic signal (Wouters *et al* 2015). EASSRs belong to the envelope-following responses (EFRs), where the brain activity phase-locks to the envelope of the stimulation pulse train. CI stimulation patterns that recreate the amplitude fluctuations of speech lead to good speech understanding performance of CI users in quiet listening conditions (Drullman *et al* 1994, Shannon *et al* 1995). In the case of EASSRs, the stimulation envelope is a sinusoidal signal with fixed modulation frequency and depth, i.e. an amplitude-modulated (AM) stimulation pulse train is used (further referred to as conventional AM stimulation). Recent research using EASSRs has shown that they can be used to assess the temporal envelope encoding in CI users (Luke *et al* 2015, Gransier *et al* 2020) and that the encoding ability correlates with speech-in-noise understanding performance (Gransier *et al* 2020). Furthermore, amplitude growth functions of auditory steady-state responses in normal hearing subjects show a correlation with perceived loudness growth with increasing stimulation levels Eeckhoutte *et al* (2016). The same has been confirmed in CI users with EASSRs in Eeckhoutte *et al* (2018).

It is therefore hypothesized that using EASSRs to objectively determine CI stimulation parameters is favorable, as the stimulation levels, that elicit EASSRs, ensure information transmission of the temporal modulations of the stimulus, which is essential for speech understanding. This is especially important in the early development stages of pediatric CI users, where fitting is challenging, and modulation perception is crucial for speech development (Petley *et al* 2024).

The main obstacle for using EASSRs clinically is the presence of stimulation artifacts of the CI, which contaminate the EEG recording. Like the EASSR, the artifact signal follows the envelope of the stimulation and is therefore impossible to distinguish from the neural response without sophisticated artifact removal methods. However, this is difficult due to the continuous, periodic nature of the stimulation and neural response, which intrinsically leads to a timely overlap of both signals. Conventional filtering is also not possible as the neural response and artifact also show overlapping frequency content, i.e. a component at the modulation frequency of the conventional AM stimulation (Deprez *et al* 2017). Many artifact removal methods for EASSR recordings have been introduced. They include model based approaches (Luke and Wouters 2017), blind source separation with independent component analysis (Deprez *et al* 2018, Alemi *et al* 2021), denoising source separation (DSS, Undurraga *et al* (2021)), template subtraction (Deprez *et al* 2017), selective downsampling of the EEG recording (Hu and Ewert 2021) and linear interpolation between artifact-free EEG recording samples (LI, Hofmann and Wouters (2010)). However, only the recently introduced system identification (SI) approach has shown good results for artifact removal and EASSR detection with clinical stimulation settings of the CI, i.e. high pulse rates at 900 pulses per second (pps) and monopolar stimulation (MP), and

with every individual EEG recording channel (ipsi- and contralateral to the CI) (Schott *et al* 2022, 2024). Good results have been achieved for a large EASSR dataset from different CI users ($N = 16$) at four different stimulation levels (Schott *et al* 2024). These findings demonstrate that EASSRs, elicited with clinically relevant stimulation parameters, can be used as neural responses for objective fitting, when SI is used for artifact removal and neural response detection.

The main shortcoming of the SI-based artifact removal as introduced by Schott *et al* (2022), is the need for at least two EASSR recordings, each at a different modulation frequency, to create a model that enables a group delay identification, which serves as validation of the presence of a neural response. When providing more input–output information to the SI, i.e. more recordings at different modulation frequencies, a better identification of the systems is achieved (Ljung 1999, Nelles 2001), and it is therefore hypothesized that this leads to a more reliable separation of stimulation artifact and neural response. However, currently, EASSR EEG recordings of approximately 5 min duration (to guarantee sufficiently low background noise levels) are acquired at each individual modulation frequency (Gransier *et al* 2020, Schott *et al* 2024), which leads to a high accumulated total recording time, which is undesired in clinical practice. Finally, the current way of recording and analyzing data does not allow the implementation of an online detection algorithm, where the presence of a neural response can continuously be checked while EEG is further recorded.

To overcome the described issues, a novel stimulation paradigm for CI users is introduced, which is based on well-established SI methods from other research or engineering fields (Farina 2000, Muller and Massarani 2001) and which has also already been used in normal-hearing subjects to determine the modulation frequency transfer function (MFTF) (Picton *et al* 2003, Purcell *et al* 2004, Purcell and John 2010). This new paradigm uses a sweep over modulation frequencies and can therefore be used to determine the group delay within a single EEG recording, which is assumed to improve the SI accuracy, time efficiency, and allows online detection of the neural response.

In this paper, the potential benefits of incorporating more input–output information, i.e. more modulation frequencies in the stimulation (further referred to as SWEEP stimulation), are analyzed for SI accuracy and neural response detection. This is achieved by creating hybrid EEG data that combines real artifact-only recordings with simulated neural response signals. Further, it is investigated whether EFRs can be elicited in CI users with the SWEEP stimulation and whether they relate to EASSRs. Additionally, the complexity of the SWEEP stimulation is further increased by randomly shuffling modulation frequencies during the stimulation epoch, and it is investigated whether EFRs can also be elicited and detected with this 'random' stimulation envelope.

2. Materials and methods

2.1. SWEEP stimulation

The idea of the SWEEP stimulation is based on general SI procedures, where sinusoidal sweeps are used to efficiently identify the transfer function of unknown systems (Ljung 1999, Muller and Massarani 2001, Nelles 2001). Applications of sinusoidal sweep recordings can be found in different fields such as biomedical engineering (Bernardi *et al* 2017), acoustics (Farina 2000), and others. In audiology, sinusoidal sweeps have been used, among other things, to evaluate temporal auditory acuity in normal hearing listeners (Purcell *et al* 2004, Purcell and John 2010). In Purcell *et al* (2004), an acoustic stimulus is generated by modulating a noise or pure-tone carrier signal with a sinusoid, where the modulation frequency first increases linearly and then decreases linearly. The results show that by using this stimulus, the MFTF can efficiently be evaluated and that the response at distinct frequencies is equal to ASSRs at the same frequency.

Two recent studies tested CI stimulation paradigms which involve AM pulse trains with alterations in modulation frequencies. In Undurraga *et al* (2021), a CI stimulation paradigm is proposed to assess temporal processing in CI users based on (electrically evoked) auditory change complexes (EACCs). The paradigm consists of alterations between two AM pulse trains with different modulation frequencies (19.2 Hz and 32 Hz), which elicits EACCs at every transition between the pulse trains. Additionally, EASSRs, following the modulation of each individual AM pulse train, are also elicited and detected. As the main focus of the study is the detection of EACCs, the difference between the two modulation frequencies is rather big. The elicited EASSRs therefore likely originate from different generators along the auditory pathway (Gransier *et al* 2016). The stimulation artifacts in the EEG recordings are removed with a combination of LI and DSS (Undurraga *et al* 2021), which requires a multichannel EEG system. LI alone may not be sufficient to remove all stimulation artifacts, as discussed in Undurraga *et al* (2021). Nonetheless, the study is interesting, as it shows that EASSRs can be elicited, even when alternating between the AM pulse trains at rates up to 6.9 Hz. In Aldag and Nogueira (2024), another CI stimulation paradigm with alterations of AM stimulation pulse trains with different modulation frequencies is proposed. Here, a constant pulse train is followed by two subsequent AM pulse trains with different modulation frequencies, and concluded with a silent period. Again, the focus of this stimulation paradigm is the elicitation of CAEPs, i.e.

onset-, two EACCs, and an offset response. EASSRs are additionally detected, but not the primary focus of the investigation. Interestingly, the discussion of the EASSR detection is rather brief, possibly due to difficulties in reliable response detection in all subjects. Furthermore, in the study, a CI pulse rate of only 500 pps was used to enable artifact removal via LI. Even though both studies involve CI stimulation paradigms with AM pulse trains with alterations in modulation frequencies, the stimuli are not intended and therefore not optimized to detect EASSRs. The results from Undurraga *et al* (2021), Aldag and Nogueira (2024), and Gransier *et al* (2020) suggest that EACCs may not be robust enough to serve as neural correlates of temporal envelope processing in CI users. As previously motivated, EASSRs/ EFRs may be the more relevant neural response, emphasizing the importance of developing and validating the SWEEP stimulation with a focus on reliable EFR detection, as undertaken in this work.

In Undurraga *et al* (2021) and Aldag and Nogueira (2024), the two AM pulse trains, each with a different modulation frequency, maintained a constant modulation frequency within a specified time segment. This may facilitate phase locking of the auditory nerve fibers to the modulation of the stimulus. We therefore hypothesize that a sweep with step-wise modulation frequency increase, i.e. where the modulation frequency is kept constant over a defined time segment T_{seg} after each increase, increases the chance to reliably detect responses. In Undurraga *et al* (2021), EASSRs are detected for durations of $T_{\text{seg}} = 2$ s and $T_{\text{seg}} \approx 0.14$ s, and in Aldag and Nogueira (2024) EASSRs are detected for $T_{\text{seg}} = 1$ s. In David *et al* (2023) and Gransier *et al* (2020), time segment durations of $T_{\text{seg}} = 1.024$ s have shown to elicit EASSRs, even if the continuous stimulus is interrupted by a silence period (David *et al* 2023) or a change in modulation depth (Gransier *et al* 2020).

For the SWEEP stimulation, we therefore choose to keep the modulation frequency constant over a time segment $T_{\text{seg}} = 1.024$ s for each modulation frequency step. The modulation frequencies are rounded so that only full cycles of the modulation are contained within the time segments of T_{seg}. This way, the instantaneous phase $\phi(t)$ of the modulating signal is continuous, also when the modulation frequency is increased. This is done to avoid jumps in the stimulation levels when increasing the modulation frequency or repeating the sweep, which would otherwise lead to undesirable 'clicks' in the stimulation.

To detect the neural response and assess its statistical significance, i.e. test whether the determined neural response differs from the neural background noise, a one-sample Hotelling T^2 (HT^2) test is applied to the variance of the neural response across multiple recordings of repetitions of the stimulation (epochs). The number of used epochs, therefore, determines the duration of one SWEEP stimulation epoch when aiming for a total recording duration of 5.12 minutes, in line with EASSR recording times (Gransier *et al* 2020, Schott *et al* 2024). This also means, that the number of modulation frequency steps is limited, when $T_{\text{seg}} = 1.024$ s. For convenience, a sample size of 60 epochs is chosen, which results in a SWEEP stimulation duration of 5.12 s, which means 5 modulation frequencies can be used within one SWEEP, each corresponding to a time segment of $T_{\text{seg}} = 1.024$ s. In the frequency range of interest from 36 Hz to 43 Hz, the following modulation frequencies are used:
$f_m^1 = 36.1328125$ Hz, $f_m^2 = 38.0859375$ Hz,
$f_m^3 = 40.0390625$ Hz, $f_m^4 = 41.015625$ Hz
and $f_m^5 = 42.96875$ Hz.

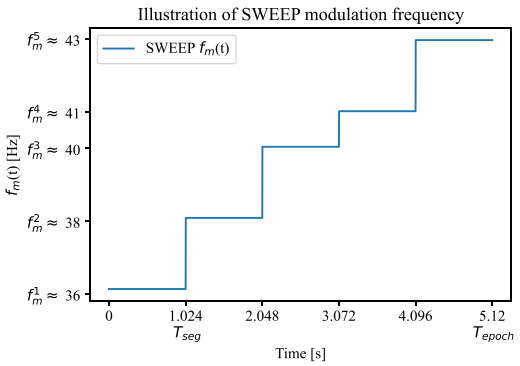


Figure 1. Illustration of time dependent modulation frequency $f_m(t)$ of the SWEEP stimulation. The modulation frequency $f_m(t)$ increases in discrete steps every T_{seg} seconds over the duration of one SWEEP stimulation epoch T_{epoch}.

The resulting instantaneous frequency $f_m(t)$ of the modulating signal $m(t)$ is shown in figure 1.

The modulating signal, i.e. the envelope of the CI stimulation pulse train, is then defined as

$$m(t) = \sin\left(2\pi f_m(t) \cdot t + \frac{\pi}{2}\right). \tag{1}$$

2.2. Shuffled-SWEEP stimulation

In Gransier and Wouters (2021), the *Temporal Envelope Speech Tracking* (TEMPEST) framework is introduced. This framework can be used to create stimuli to assess the temporal envelope processing in the auditory pathway in a systematic and standardized way. The complexity of the envelope of the stimulus can be varied, from a simple single-frequency envelope where the modulation is kept constant in frequency and depth (e.g. as in equ. (7)), to a speech-like envelope, where modulation depth and frequency continuously vary. The SWEEP stimulation of section 2.1 can be seen as a stimulus from Gransier and Wouters (2021), where the modulation depth stays constant, but the frequencies change over time.

To investigate the effect of a more complex stimulation paradigm than conventional-AM- or the

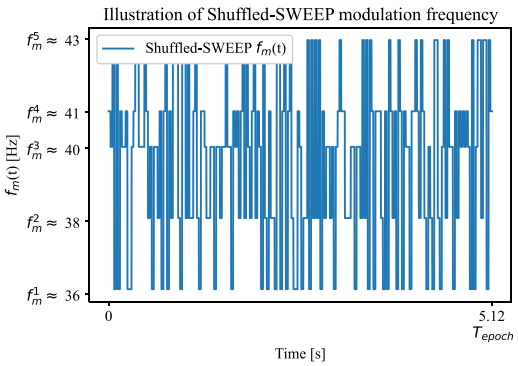


Figure 2. Illustration of time dependent modulation frequency $f_m(t)$ of the Shuffled-SWEEP stimulation. The modulation frequency $f_m(t)$ changes randomly between frequencies $f^1{}_m, f^2{}_m, f^3{}_m, f^4{}_m$, and $f^5{}_m$. Every random frequency is kept constant for a duration of $T_m = \frac{1}{f_m}$. The stimulation epoch has a duration of $T_{\text{epoch}} = 5.12$ s.

SWEEP stimulation, a TEMPEST-like stimulation for CI users, further referred to as Shuffled-SWEEP, is created. Here, the SWEEP stimulation of section 2.1 is used and the sequence of the instantaneous modulation frequency $f_m(t)$ is randomized. However, it is ensured, that for every modulation frequency in the random sequence, a full period of duration $T_m = \frac{1}{f_m}$ is created,

$$\text{mod}\,(t) = \sin\left(2\pi f_m \cdot t + \frac{\pi}{2}\right) \text{for t} = 0, \ldots T_m. \quad (2)$$

The complete modulating signal m_{SSW}, that is used for modulating the stimulation pulse train, is defined as the concatenation of all full periods

$$m_{\text{SSW}} = \left(\text{mod}^1, \text{mod}^2, \ldots \text{mod}^I\right), \quad (3)$$

with $I = T_{\text{seg}} \cdot f_m^1 + T_{\text{seg}} \cdot f_m^2 + T_{\text{seg}} \cdot f_m^3 + T_{\text{seg}} \cdot f_m^4 + T_{\text{seg}} \cdot f_m^5$, i.e. the total number of periods in the epoch.

By concatenating full periods, a continuous modulation m_{SSW} is again guaranteed. The resulting m_{SSW} is closer related to a speech envelope, where the modulation 'randomly' changes over time, without having confounding effects such as speech understanding influencing the strength of the elicited neural response.

The SWEEP and Shuffled-SWEEP provide equal input information for the SI, as the same modulation frequencies and equal number of full periods per frequency are used, only the sequence of the modulation frequencies randomly changes during one epoch of the Shuffled-SWEEP stimulation as seen in figure 2.

2.3. Artifact removal and neural response detection

The SI approach for detecting EASSRs, as introduced in Schott *et al* (2022) and further refined in Schott *et al* (2024), can also be applied to EEG recordings following SWEEP or Shuffled-SWEEP stimulation. The elicited neural responses will hereafter be referred to as EFRs or Shuffled-EFRs, respectively.

For completeness, we briefly review the model formulation as defined in Schott *et al* (2024).

For the EASSR recording following conventional AM stimulation, the recording model remains

$$\underbrace{\begin{bmatrix}\mathbf{z}^{37}\\ \mathbf{z}^{40}\\ \mathbf{z}^{43}\end{bmatrix}}_{\mathbf{z}} = \underbrace{\begin{bmatrix}\mathbf{M}^{37} & \mathbf{U}^{37}\\ \mathbf{M}^{40} & \mathbf{U}^{40}\\ \mathbf{M}^{43} & \mathbf{U}^{43}\end{bmatrix}}_{\mathbf{H}} \cdot \underbrace{\begin{bmatrix}\mathbf{b}_{\text{resp}}\\ \mathbf{b}_{\text{arti}}\end{bmatrix}}_{\mathbf{x}} + \mathbf{v}, \quad (4)$$

with recording $\mathbf{z}^{f_m} = \left[z^{f_m}[0] \quad z^{f_m}[1] \quad \ldots \quad z^{f_m}[K]\right]^{\mathsf{T}}$ being the mean epoch of the pre-processed EEG recording with modulation frequency f_m in the conventional AM stimulation, with a total number of $K+1$ samples. $\mathbf{v}$ is the additive noise. The input matrix $\mathbf{M}^{f_m}$ for modulation frequency f_m is defined as

$$\mathbf{M}^{f_m} = \begin{bmatrix}\mathbf{m}^{f_m}[0]\\ \mathbf{m}^{f_m}[1]\\ \vdots\\ \mathbf{m}^{f_m}[K]\end{bmatrix}, \quad (5)$$

with

$$\mathbf{m}^{f_m}[k] = \left[m[k] \quad m[k-1q] \quad \ldots \quad m[k-N_1 q]\right], \quad (6)$$

where $m[k]$ is the modulation of the conventional AM stimulation at discrete time k, with sampling interval T_s of the EEG recording system

$$m[k] = \sin\left(2\pi f_m k T_s + \frac{\pi}{2}\right). \quad (7)$$

The time-shift parameter q is set to $q = 4194$, as described in Schott *et al* (2024). N_1 is the order of the finite-impulse response (FIR) system, which models the EASSR in the EEG recording, and is set to $N_1 = 5$.

Matrix $\mathbf{U}^{f_m}$ is similarly defined as $\mathbf{M}^{f_m}$, i.e.

$$\mathbf{U}^{f_m} = \begin{bmatrix}\mathbf{u}^{f_m}[0]\\ \mathbf{u}^{f_m}[1]\\ \vdots\\ \mathbf{u}^{f_m}[K]\end{bmatrix}, \quad (8)$$

with

$$\mathbf{u}^{f_m}[k] = \left[u[k] \quad u[k-1] \quad \ldots \quad u[k-N_2]\right], \quad (9)$$

using

$$u[k] = A_c \cdot (1 + D \cdot m[k]) \cdot s[k]. \quad (10)$$

N_2 is the order of the FIR system, which models the stimulation artifact in the EEG recording, and is set to $N_2 = 364$ (Schott *et al* 2024).

Furthermore, $s[k]$ is the stimulation pulse train with pulse rate f_c, which contains 1's at the sample positions k, which are closest to the stimulation pulse positions and 0's elsewhere, i.e.

$$s[k] = \begin{cases} 1, & \text{for mod}\left(kT_s, \frac{1}{f_c}\right) < T_s \\ 0, & \text{otherwise} \end{cases}. \quad (11)$$

Finally, A_c and D are defined by the used stimulation level CL

$$A_c = \frac{\text{CL} + (\text{CL} - 50)}{2} \quad (12)$$

and

$$D = \frac{C - (C - 50)}{C + (C - 50)}. \quad (13)$$

For EEG recordings following SWEEP or Shuffled-SWEEP stimulation, the recording model $\mathbf{z}$ reduces to an unstacked model, as no combination of different recordings is required anymore,

$$\mathbf{z} = \underbrace{\begin{bmatrix}\mathbf{M} & \mathbf{U}\end{bmatrix}}_{\mathbf{H}} \cdot \underbrace{\begin{bmatrix}\mathbf{b_{resp}} \\ \mathbf{b_{arti}}\end{bmatrix}}_{\mathbf{x}} + \mathbf{v}. \quad (14)$$

Matrices $\mathbf{M}^{f_m}$ and $\mathbf{U}^{f_m}$ are similarly defined as in (5) and (8), now without the f_m superscripts. For the EFR model, the modulation $m[k]$ is adapted to incorporate the time varying modulation frequency

$$m[k] = \sin\left(2\pi f_m(kT_s)\, kT_s + \frac{\pi}{2}\right), \quad (15)$$

as defined by the SWEEP stimulation. For the Shuffled-EFR model, $m[k]$ uses m_{SSW} as defined in (3), sampled at discrete time intervals kT_s,

$$m[k] = m_{\text{SSW}}(kT_s). \quad (16)$$

As the SWEEP or Shuffled-SWEEP stimulation uses 5 distinct modulation frequencies in $m[k]$, N_1 is set to $N_1 = 9$ for the EFR and Shuffled-EFR model, and $q = 2330$.

The unknown model coefficients are estimated using the same Tikhnonov regularization as in Schott *et al* (2024), where regularization constant α is determined as in Schott *et al* (2024) and set to the same value $\alpha = 5000$,

$$\hat{\mathbf{x}} = \begin{bmatrix}\hat{\mathbf{b}}_{resp} \\ \hat{\mathbf{b}}_{arti}\end{bmatrix} = \left(\mathbf{H}^{\intercal}\mathbf{H} + \alpha\boldsymbol{\Gamma}^{\intercal}\boldsymbol{\Gamma}\right)^{-1}\mathbf{H}^{\intercal}\mathbf{z}. \quad (17)$$

The response FIR system, used to determine the presence of a neural response in the EEG recordings, is defined as

$$\text{FIR1}(z) = \begin{bmatrix}1 & z^{-q} & \dots & z^{-N_1 q}\end{bmatrix} \cdot \hat{\mathbf{b}}_{resp}. \quad (18)$$

The amplitude of the neural response is determined as

$$A(f_m) = \left|\text{FIR1}\left(z = e^{j\omega_m}\right)\right| = \left|\begin{bmatrix}1 & e^{-jq\omega_m} & \dots & e^{-jN_1 q\omega_m}\end{bmatrix} \cdot \hat{\mathbf{b}}_{resp}\right|, \quad (19)$$

with normalized angular modulation frequency $\omega_m = 2\pi f_m / F_s$. The phase of the neural response is similarly determined as

$$\theta(f_m) = \angle\text{FIR1}\left(z = e^{j\omega_m}\right). \quad (20)$$

The group delay of FIR1 represents the latency of the neural response and is determined from the slope of a linear regression fitted to the phases of the neural responses at different f_m, as described in Schott *et al* (2024).

2.4. Dataset

Four adult, unilateral CI users, all with Cochlear Ltd devices (CIC4 implants), and at least three years of CI experience were recruited for the EEG data collection. This study was approved by the Medical Ethical Committee of the UZ Leuven Hospital (study number: S65454). All participants gave written informed consent before participation, and the study was carried out in accordance with the Declaration of Helsinki principles.

The same recording setup as in Schott *et al* (2024) (Biosemi ActiveTwo Hyper-Rate EEG system) was used to collect EEG data at 8 recording channels. Six recording channels were placed according to the international 10–20 system at *P*9, *P*10, *Iz*, *Cz*, *Fz*, *Fpz*, and two extra channels at the left and right mastoids respectively (*MaL*, *MaR*). If an EEG recording channel overlapped with the CI coil position, an alternative EEG recording channel in the nearest proximity was used instead. The high sampling rate of 262 144 kHz of the EEG system enables more signal processing options of the recordings, but it is not a prerequisite for the presented results here. For instance, it allows using other artifact removal methods such as LI, mainly because the cut-off frequency of the front-end low-pass filter of the EEG system is much higher than for common systems. Here, the EEG signals are low-pass filtered with the system's built-in analog third-order antialiasing filter having a −3 dB point at 50 kHz, which reduces the temporal 'smearing' of the stimulation pulses in the recording. Consequently, the time duration of the stimulation artifact is shorter than an inter-pulse interval in recordings from contralateral EEG channels, which enables artifact removal with LI in these channels as seen in Gransier *et al* (2020) and Schott *et al* (2024).

For each condition of conventional AM stimulation, 300 epochs were recorded, while for each condition of SWEEP and Shuffled-SWEEP stimulation, 60 epochs were recorded, resulting in a total recording time of 5.12 min per condition.

Like the recording setup, the stimuli were created and presented using the same equipment as in Schott *et al* (2024). The stimuli were generated using custom Matlab scripts (R2020a) and directly delivered to the implant using the NIC4 research interface platform (Cochlear Ltd).

The CI stimulation pulse trains consisted of symmetric biphasic cathodic-first pulses, with user-dependent phase width of 25 μs or 37 μs, and an interphase gap of 8 μs. A pulse rate of 900 pps was used, and the stimuli were presented between the intracochlear electrode 15 and the extracochlear plate electrode (MP1 mode). These two settings are particularly relevant because the high pulse rate and MP stimulation reflect the clinical settings used in daily CI operation. Neural responses obtained with these settings are therefore presumed to be more relevant for objective fitting, as previously stated in the introduction.

For the conventional AM stimulation, used to elicit EASSRs, three distinct AM pulse trains with respective modulation frequencies of 37 Hz, 40 Hz, and 43 Hz were used and repeated at two different stimulation levels. A fixed modulation depth of 50 current levels (CL) between each defined stimulation level and the respective level minus 50 CL was used for the AM pulse trains. The two stimulation levels were defined as the user's C and user's $T-10$ CL ('Tm10') level, resulting in two CI stimulation pulse trains, with CLs between C and $C-50$ CL, and $T-10$ CL and $T-60$ CL respectively. For each user, the necessary T- and C-level were behaviorally determined before the EEG recordings. Here, the same fitting procedure, using a 7-point categorical loudness scale as described in Schott *et al* (2024), was applied.

SWEEP and Shuffled-SWEEP stimulation were generated using the same settings as for the conventional AM stimulation. However, instead of creating three distinct stimuli, only one stimulus with a time-varying modulation frequency, as described in 2.1 and 2.2, was created for each stimulation level in each paradigm.

Before applying artifact removal and neural response detection, the recorded EEG signals were pre-processed using Matlab R2020a. Using Matlab's built-in *detrend* function, the raw EEG signals were detrended with a second-order polynomial. The signals were then split into epochs, based on the stimulation triggers at the beginning of each epoch. Five percent of the epochs with the highest peak-to-peak amplitudes were rejected in each channel to reduce the effect of transient movement artifacts. A mean epoch for each channel was created and then referenced in the time domain with the mean epoch of channel *Fz*. The channels *MaR*, *MaL*, *P*9, *P*10, and *Iz* were used for the further analysis.

2.5. Validation with hybrid EEG data

To investigate the hypothesized benefit of using more modulation frequencies in the SWEEP stimulation on the SI accuracy and hence response detection, hybrid EEG data is created as a combination of real artifact-only EEG recordings and simulated responses. By using this hybrid data, various response systems, i.e. different response latencies and amplitudes, can be simulated, allowing for validation of the detection since the underlying ground truth is simulated and therefore known. Using artifact-only EEG recordings allows for the creation of realistic testing conditions.

The hybrid data is created by using EEG recordings, where the CLs are below the user's T-level ('Tm10' condition). This way, no response is elicited, and only stimulation artifact is present at the modulation frequencies in the EEG recording. The simulated responses are created by using the envelope of the stimulation, i.e. (7) for the conventional AM and (15) for SWEEP stimulation respectively, and digitally filtering it with a simulated FIR system defined as

$$H(z) = \sigma \cdot z^{-D} \tag{21}$$

with a simulated group delay in $D = \lfloor \text{delay} \cdot F_s \rceil$ samples at sampling frequency F_s, and simulated scaling σ. For instance an EASSR with 40 ms latency and amplitude of 0.3 μV can be simulated by setting $D = \lfloor 0.04\,\text{s} \cdot 262144\,\text{Hz} \rceil = 10485$ and $\sigma = 0.3$.

The simulated group delays, further referred to as response latencies, range from 0 to 100 ms in steps of 1 ms, and per simulated response latency, different scalings σ are used. The σ-s are determined from a range of simulated response SNRs, with a fixed hypothetical background noise level ν of $\nu = 50$ nV,

$$\sigma = 10^{\frac{\text{SNR}}{20}} \cdot 50 \cdot 10^{-9}. \tag{22}$$

The simulated SNRs range from -10 dB to 20 dB in steps of 1 dB.

All together, responses from 31×101 simulated response systems are created and added to artifact-only EEG recordings from ipsilateral EEG channels, to ensure sufficient artifact contamination and hence challenging testing conditions.

The conventional AM stimulation hybrid data is further differentiated and validated by either using only one recording of one modulation frequency (40 Hz), or all three recordings (37 Hz, 40 Hz and 43 Hz). For a fair comparison of the simulated conditions, the number of epochs of the recordings of the conventional AM stimulation are reduced, so that equal total recording lengths between conditions are compared. For instance, for the stacked recording model of 3 f_ms, only 1/3 of the epochs per 5.12 min recording are used to calculate the mean epoch, which is being analyzed. This way, the stacked recording uses the same amount of data ($3 \cdot \frac{5.12\,\text{min}}{3}$) as the

SWEEP recording. The included epochs are randomly selected.

Hybrid data is created for only three of the four subjects, as for one subject technical issues were experienced during the acquisition of the artifact-only EEG recording. Nonetheless, the 3131 conditions with three distinct artifact morphologies (originating from three different CI users) ensure a robust effect size for the analysis.

SI is used to perform artifact removal and neural response detection, and the accuracy is validated by checking the true positive rate (TPR) and $F1$ score. The TPR is the number of detected responses (true positives) over detectable responses, and is used as a measure of sensitivity. The $F1$ score is the harmonic mean of TPR and precision, of which precision is the ratio of true positives and predicted positives (true and false). $F1$ is a measure of predictive performance.

Furthermore, the neural response detection is also validated by analyzing the accuracy of the estimated response latencies and amplitudes.

3. Results

3.1. Hybrid EEG data results

The TPR and $F1$ score as functions of the simulated response latencies are shown in figure 3 and figure 4 respectively. For each simulated response latency, SI is applied on the hybrid data of all simulated scalings σ, and HT^2 is used to determine whether the identified response is significant, i.e. detected. Simulated responses with SNRs of 6 dB or higher, with recording background noise level v, should be, after successful artifact removal, detectable above the noise floor and therefore be statistically significant when applying the HT^2 test.

Each plot-line represents the results for a particular stimulation paradigm, i.e. blue for one recording with conventional AM stimulation, green for the combination of three recordings of conventional AM stimulation, and red for the recording of SWEEP stimulation. Solid lines represent the average value over subjects. Individual results per subject are plotted in shaded lines with different line-styles, as indicated in the figure legend. The following observations can be made from the figures 3 and 4.

First, when using only one recording of conventional AM stimulation, the TPR repeatedly drops to values close to 0%, meaning that no response is detected, even when the simulated SNRs are above 6 dB. When using three recordings of conventional AM stimulation or one recording of SWEEP stimulation, the steep drops in TPR disappear, however, fluctuations in TPR remain over different simulated response latencies.

Second, the more modulation frequencies are used in the analyzed recording model, the better the balance between detection accuracy and detection precision, which can be seen from the $F1$ scores. For one f_m, the mean value is $\bar{F1} = 0.75$, for three f_m $\bar{F1} = 0.82$, and for the sweep $\bar{F1} = 0.90$, indicating the best response detection with stimulation while avoiding false positive responses.

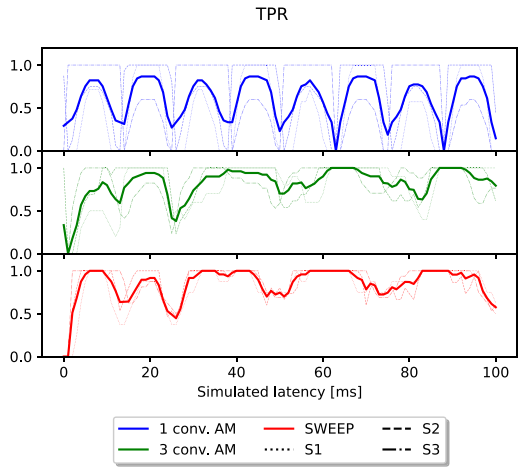


Figure 3. True positive rate (TPR) as function of simulated response latency. A TPR of 1 corresponds to a 100% response detection rate. Each sub-figure presents results from a distinct stimulation paradigm: the blue plot line represents a single recording with conventional AM stimulation, the green plot line represents results of the combination of three recordings of conventional AM stimulation, and the red plot line corresponds to a recording with SWEEP stimulation. Shaded plot lines indicate individual subject data, while the solid lines represent the mean TPR across subjects.

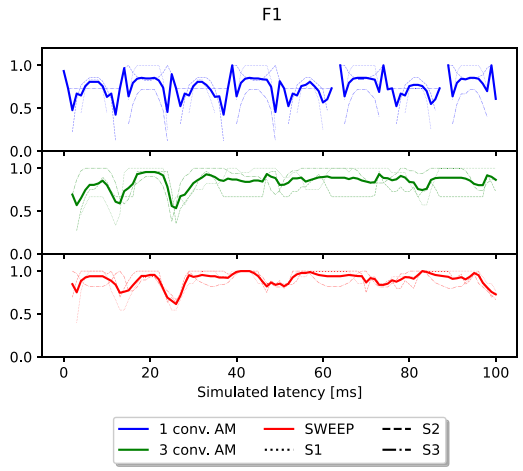


Figure 4. $F1$ score as function of simulated response latency. An $F1$ score of 1 corresponds to perfect precision and sensitivity. Each sub-figure presents results from a distinct stimulation paradigm: the blue plot line represents a single recording with conventional AM stimulation, the green plot line represents results of the combination of three recordings of conventional AM stimulation, and the red plot line corresponds to a recording with SWEEP stimulation. Shaded plot lines indicate individual subject data, while the solid lines represent the mean $F1$ scores across subjects.

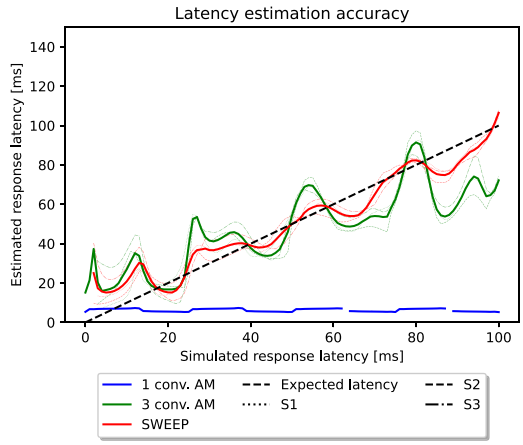


Figure 5. Estimated response latency as a function of simulated response latency. Each plot line represents a different stimulation paradigm: the blue line corresponds to a single recording with conventional AM stimulation, the green line shows the combined results from three recordings of conventional AM stimulation, and the red line represents a recording with SWEEP stimulation. Shaded plot lines indicate individual subject data, with different line-styles distinguishing subjects, while solid lines represent the mean estimated response latency across subjects. Only responses that are identified to be significantly different from neural background noise are included in the analysis.

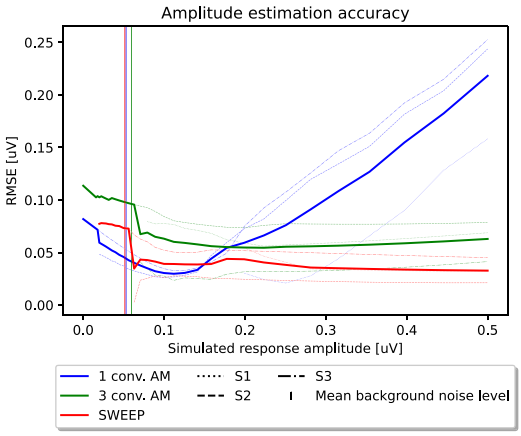


Figure 6. Root mean square error (RMSE) between estimated and simulated response amplitude at 40 Hz as function of simulated response amplitude. Each plot line represents a different stimulation paradigm: the blue line corresponds to a single recording with conventional AM stimulation, the green line shows the combined results from three recordings of conventional AM stimulation, and the red line represents a recording with SWEEP stimulation. Shaded plot lines indicate individual subject data, with different line-styles distinguishing subjects, while solid lines represent the mean estimated response latency across subjects. Vertical lines indicate the mean background noise level of the different stimulation paradigms. Only responses that are identified to be significantly different from neural background noise are included in the analysis.

Furthermore, the SI accuracy can also be analyzed based on the determined response latency compared to the simulated response latency. Here, only simulation conditions of different scaling factors σ, where the identified response is determined to be significantly different from background noise (p-value $<$ 0.05 after HT^2), are included in the latency analysis. The detected response latencies for these are averaged and plotted for each stimulation paradigm as a function of the simulated response latency in figure 5. Again, solid lines represent the average result over subjects. Individual results per subject are plotted in shaded lines with different line-styles, as indicated in the figure legend. The following observations can be made from the figure.

First, when using only one recording of conventional AM stimulation, no group delay can be determined with the response model, and hence the response latency cannot be detected.

Second, when using three recordings of conventional AM stimulation or one recording of SWEEP stimulation, the detected latency follows the simulated latency. The accuracy for hybrid data with SWEEP stimulation is the best, as seen from the smallest variation and smallest deviation from the simulated response latencies. Third, responses with lower simulated latencies are less accurately detected.

Similar to the accuracy validation with determined response latency, the determined response amplitude can also be used. Here, the determined response amplitude at modulation frequency $f_m = 40$ Hz for a subset of the simulated latency conditions is used. As seen in figure 5, the identifiability varies greatly with simulated response latency. Therefore, only hybrid data with simulated response latencies between 35 ms to 55 ms is validated, as these represent the latency range of interest for real EEG recordings with modulation frequencies around 40 Hz (Gransier *et al* 2016, 2017, Schott *et al* 2024).

In figure 6 the root-mean-square error (RMSE) between detected and simulated response amplitude for the range of simulated latencies is plotted for each stimulation paradigm. Again, only statistically significant responses are included in the analysis. In addition to the average results across subjects and individual subject results for each stimulation paradigm, the mean recording background noise levels across subjects are shown as vertical lines, following the same color conventions. The following observations can be made from the figure.

Using one recording of conventional AM stimulation results quite inaccurate amplitude estimations, as seen from the relatively high RMSE. When using three recordings of conventional AM stimulation, the RMSE significantly drops to a level that is very close to the noise level of the EEG recordings. When using the SWEEP stimulation recording, the RMSE is the lowest and independent of the simulated response amplitude. For all stimulation paradigms, the RMSE increases, when the simulated response amplitude decreases and approaches the noise level.

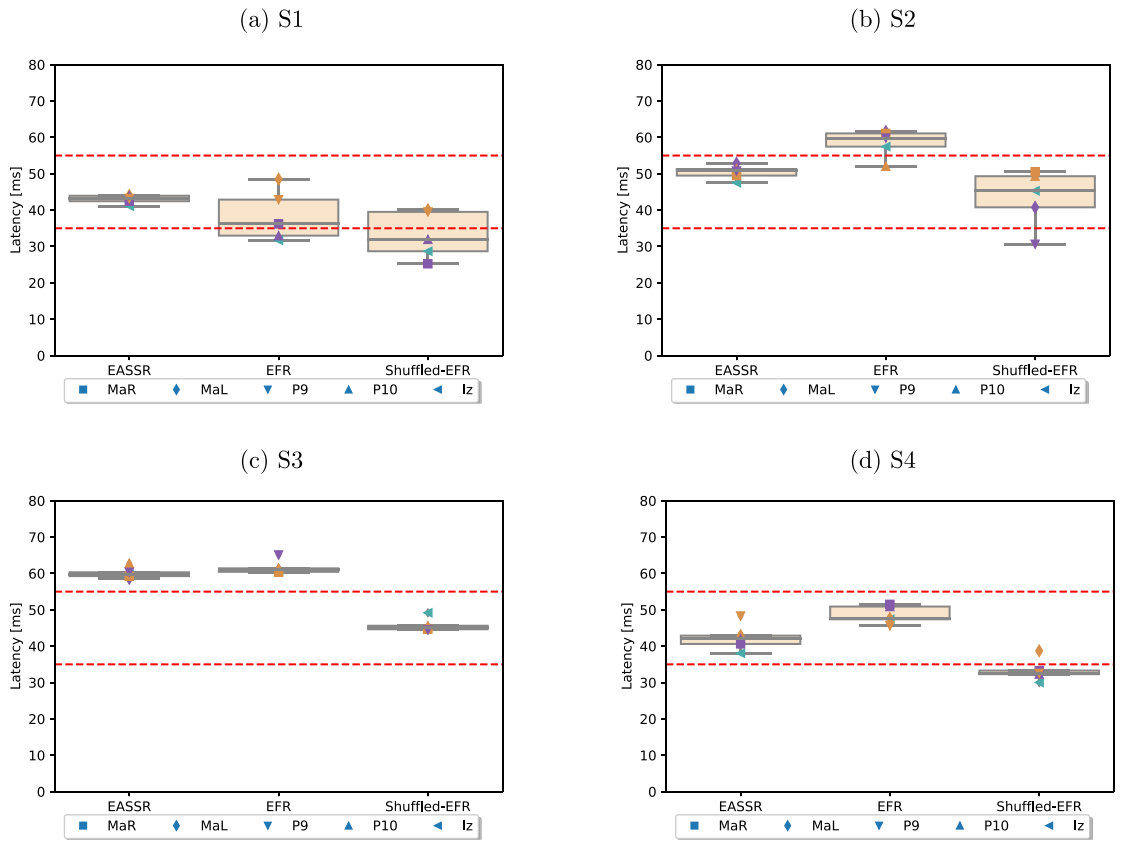


Figure 7. The determined response latencies following stimulation of the different paradigms for all subjects. Results for subject 1 are shown in sub-figure (a), while results for subjects 2, 3, and 4 are presented in sub-figures (b), (c), and (d), respectively. The expected latency range between 35 ms and 55 ms is indicated with dashed lines. The recording channel name is indicated by the marker, and the location with respect to the CI is indicated by the color (contralateral: ●, ipsilateral : ●, and central: ●.

3.2. EEG data results

3.2.1. Response latency

The determined response latencies of the conventional AM, SWEEP and Shuffled-SWEEP stimulation recordings of all subjects are shown in figure 7 ((a)–(d)). The expected latency for the used modulation frequency range lies between 35 ms and 55 ms ((Gransier *et al* 2016, 2017, Schott *et al* 2024)), as indicated by the red dashed line. The results from each stimulation paradigm include all analyzed EEG recording channels. The recording sites of the channels are indicated by color, where purple ● stands for contralateral, orange ● for ipsilateral and turquoise ● for central channels. The exact channel is indicated by the marker, which can be found in the legend.

For all subjects and all stimulation paradigms, all determined responses are statistically significant (as determined from HT^2), meaning that the determined latencies originate from the elicited response and not from background noise. The response latencies are all within or slightly outside the expected range. Due to the small number of samples, no statistical test is performed to test for a difference between stimulation paradigms. Based on visual inspection, responses detected in recordings with Shuffled-SWEEP stimulation appear to have lower response latencies.

The phase delays of the individual modulation frequencies are shown in figure 8. For each subject, one sub-figure is presented ((a)–(d)). The phase delays are offset-corrected so that for each stimulation paradigm, the phase delay at $f_m = 43$ Hz is $-200°$.

3.2.2. Response amplitude

The response amplitudes of the individual modulation frequencies are shown in figure 9. Again, each sub-figure ((a)–(d)) contains the response amplitudes of the responses elicited by the three stimulation paradigms of the same subject. Based on visual inspection, the response amplitudes are generally consistent across different stimulation paradigms within subjects, except for the ipsilateral recordings of the sweep paradigm in subject 2.

4. Discussion

Recent advances in CI stimulation artifact removal have enabled the use of EASSRs, elicited with clinically relevant stimulation parameters, as neural responses in CI users (Schott *et al* 2024). However, using EASSRs can still be impractical for given reasons. In this paper, an improved SWEEP stimulation paradigm has been proposed, which is assumed to

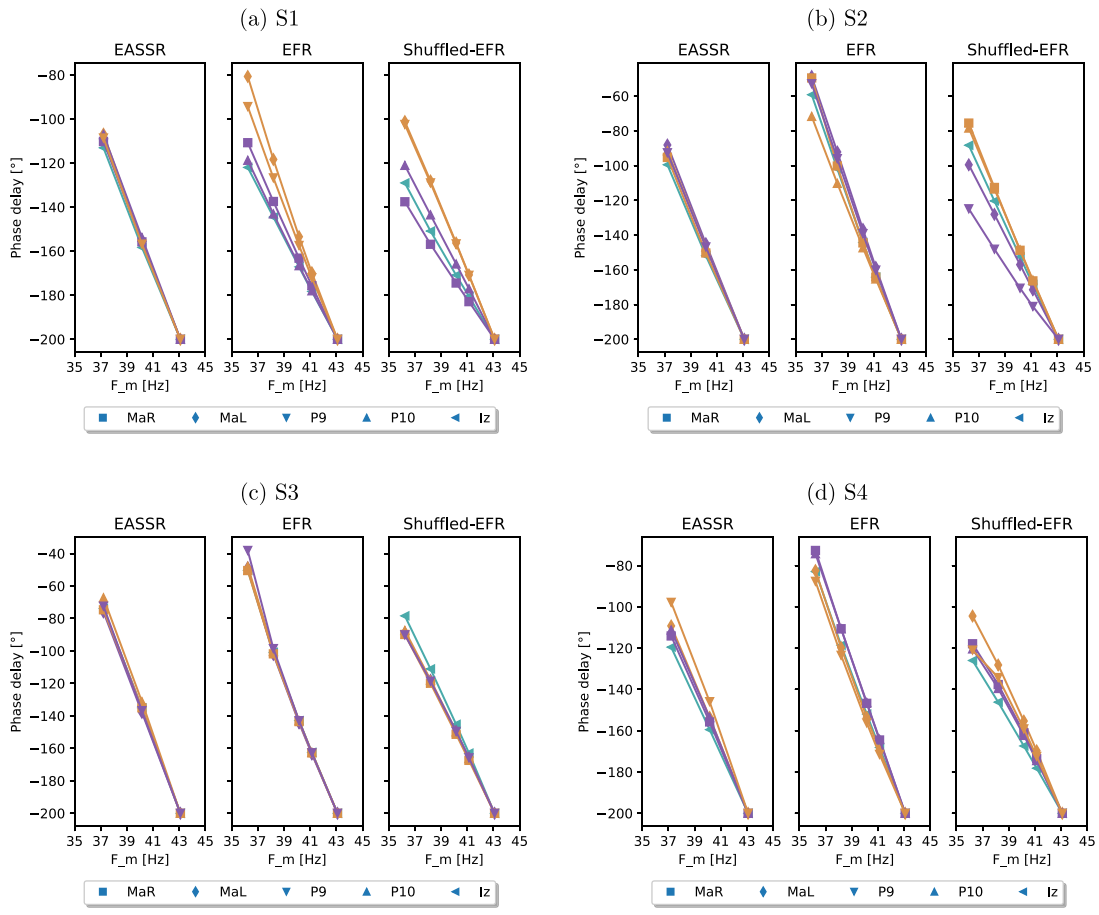


Figure 8. The determined response phases across the used modulation frequencies following stimulation of the different paradigms for all subjects. Results for subject 1 are shown in sub-figure (a), while results for subjects 2, 3, and 4 are presented in sub-figures (b), (c), and (d), respectively. The phase delays are offset-corrected so that for each stimulation paradigm, the phase delay at $f_m = 43$ Hz is $-200°$. The recording channel name is indicated by the marker, and the location with respect to the CI is indicated by the color (contralateral: ●, ipsilateral: ● and central: ●).

elicit another EFR. The SWEEP stimulation has several advantages over conventional AM stimulation as used to elicit EASSRs, and it has been shown that the SWEEP stimulation can be further adapted to create more complex stimulation envelopes, closer related to speech envelopes, as demonstrated by Gransier and Wouters (2021), referred to as Shuffled-SWEEP in this paper. EEG data from four CI users has been recorded at high stimulation levels for conventional AM, SWEEP and Shuffled-SWEEP stimulation, and EFRs have been detected. Furthermore, hybrid EEG datasets have been created by combining artifact-only EEG recordings with simulated responses, to show the benefits of the proposed SWEEP stimulation on artifact removal and response detection with SI.

4.1. Hybrid EEG data

The validation metrics on the hybrid EEG data show that using more input–output information, i.e. more modulation frequencies in the envelope of the stimulation, is beneficial for the SI and hence response detection. Using the SWEEP stimulation with 5 modulation frequencies provides the most accurate results in response detection. Although using three recordings of conventional AM stimulation is good enough to separate artifact and response (as seen from the determined latencies in figure 5) and achieve very good detection accuracies as well (figure 3), when looking at the $F1$ score as a function of the simulated response SNR (figure 10) the following can be concluded. In all hybrid datasets per individual subject, the recordings with SWEEP stimulation reach the maximum $F1$ score, which indicates correct identification of response, while avoiding false response detection (from stimulation artifacts), at lower SNRs than the combination of 3 recordings of AM stimulation. The differences in SNRs are around 2.12 dB, 2.5 dB and 4.46 dB, for the results of subject 1, 2 and 3 respectively, as indicated by the differences between vertical tick-marks in figure 10. This is due to better response and artifact separation, as seen in the superior latency (figure 5) or amplitude detection (figure 6), which consequently does not lead to false removal of the response when removing the estimated artifact part from the recording. This leads to an increase in the estimated response amplitude, which is why at lower SNRs the response is better detected. Using the SWEEP stimulation may therefore provide

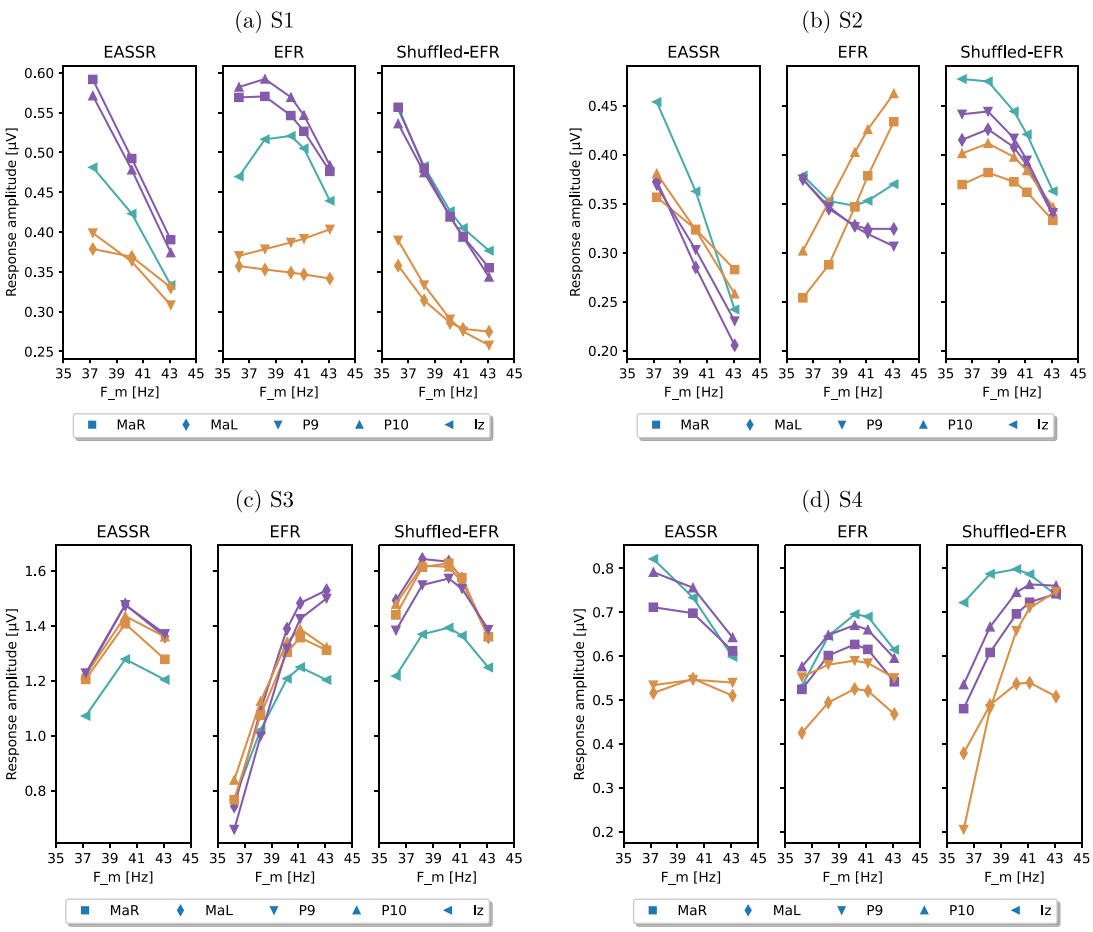


Figure 9. The determined response amplitudes across the used modulation frequencies following stimulation of the different paradigms for all subjects. Results for subject 1 are shown in sub-figure (a), while results for subjects 2, 3, and 4 are presented in sub-figures (b), (c) and (d), respectively. The recording channel name is indicated by the marker, and the location with respect to the CI is indicated by the color (contralateral: ●, ipsilateral: ●, and central: ●).

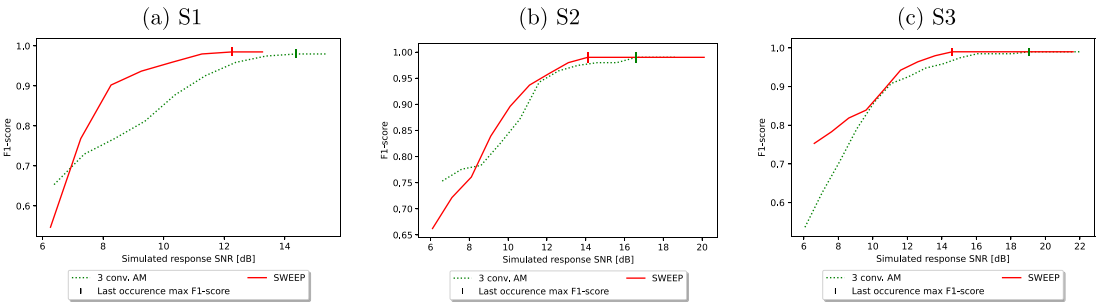


Figure 10. The $F1$ score over the range of simulated response SNRs per individual subject of the hybrid dataset. The results for conventional AM stimulation with 3 recordings (green, dotted line) are compared to the results of the SWEEP stimulation (red, solid line). The lowest SNR value, at which the highest $F1$ score is achieved, is indicated by vertical tick-mark. This marks the beginning decrease of $F1$ score, with decreasing SNR, and is used to compare the response detection quality of both stimulation paradigms.

better sensitivity in detecting neural responses in clinical recordings with stimulation levels close to the threshold level of the CI user. Furthermore, under the assumption that the background noise is Gaussian, the SNR is expected to increase with factor of the square root of the number of epochs during epoch averaging (Dobie and Wilson 1996). The SNR differences of 2.12 dB, 2.5 dB and 4.46 dB, between SWEEP and conventional AM stimulation correspond to factors of 1.28, 1.33, 1.67, meaning that respectively 1.63, 1.78, and 2.79 times more averages are needed to achieve equivalent results with three conventional

AM stimulation recordings. Consequently, using a SWEEP stimulation with 5 f_ms could reduce the clinical recording time on average by a factor of 2.07.

4.2. EEG data

The determined response latencies as seen in figure 7 show that for all three stimulation paradigms, neural responses were successfully elicited and detected in the EEG recordings of all subjects and at every recording channel position. Based on visual inspection, Shuffled-SWEEP stimulation appears to elicit neural responses with lower response latencies, which may have two origins.

One, the stimuli might activate different neural generators, leading to higher latencies when more cortical generators are activated and lower latencies when more brainstem generators are activated (Herdman *et al* 2002, Gransier *et al* 2016).

Two, as seen from the validation with hybrid EEG data, a difference in SI accuracy impacts the response detection and response latency estimation. The lower response latencies of the Shuffled-EFRs may therefore originate from the analysis of the EEG data. To gain insight into this possibility, we also used LI as one artifact removal method on the contralateral and central recording channels. These channels are known to be less contaminated by the stimulation artifact (Deprez *et al* 2017), hence LI can indeed successfully be applied. Note that this is only possible with recordings from contralateral channels and with the extremely high sampling rate as used here and seen in Gransier *et al* (2020) and Schott *et al* (2024). With LI, the artifact removal does not depend on the SI accuracy, i.e. is independent of the complexity of the envelope of the stimulation. If the determined response latencies of the different recordings are equal after artifact removal with LI, the differences in figure 7 may be attributed to differences in SI rather than to differences in elicited neural responses. However, as can be seen in figure 11, similar latency results are determined with LI. Again, Shuffled-EFRs appear to have lower response latencies. Furthermore, when comparing response latencies determined after LI with response latencies determined after SI, no obvious difference is observable. Together, this may indicate that the lower latencies in the Shuffled-EFRs effectively originate from a difference in elicited neural response, i.e. a difference in active neural generators, and not from a difference in artifact removal quality. Interestingly, in Undurraga *et al* (2021), it has been observed, that a higher alteration rate between AM pulse trains with different modulation frequencies (i.e., alterations at 0.5 Hz vs 6.9 Hz), possibly leads to higher cortical adaptation. This means that responses may be dominated by sub-cortical generators, which is in line with the lower response latencies, as seen for the Shuffled-EFRs. However, due to the small sample size (only 4 subjects), these findings may not be conclusive. Only an investigation on a larger scale can give a clear insight into the differences in elicited EFRs.

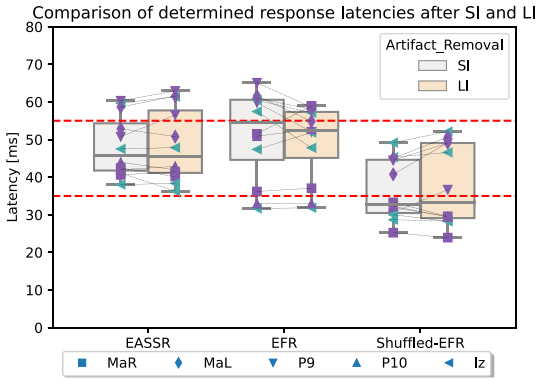


Figure 11. Comparison of determined response latencies after SI (left, grey box) and LI (right, bisque box), per different type of envelope-following response. Only EEG recordings from contralateral and central recording channels are compared. Results from the same subject and same EEG recording channel are connected via dashed lines. No apparent differences in determined response latencies between artifact removal methods, may indicate a possible difference in active neural generators for Shuffled-EFRs.

4.3. Limited sample size

Despite the limited sample size of the study, the findings demonstrate the feasibility and potential benefits of the SWEEP stimulation paradigm for objective CI fitting. The hybrid results show that SWEEP stimulation improves artifact and response separation compared to the conventional AM stimulation, which could enhance clinical efficiency. A total of 3131 simulation conditions, each repeated with three distinct artifact morphologies, provide sufficient data for a reliable comparison of the different stimulation paradigms. Clear advantages when using the SWEEP stimulation can be seen in figures 5, 6, and 10. Beyond these technical advantages, the presented results also confirm, that the proposed CI stimulation paradigms with time-varying temporal envelope modulations of different levels of complexity (SWEEP and Shuffled-SWEEP) can be used to reliably elicit envelope-following responses in CI users. As this is the case for all tested CI users (as seen in figures 7–9), a sample size of 4 subjects is sufficient to show that the novel stimulation paradigms can be used in clinical applications.

Since the primary goal of the presented study was the development of a novel stimulation paradigm to enhance the response detection, rather than to analyze the differences in neural responses per se, no statistical tests were applied to compare different EFRs. A sample size of 4 subjects may result in insufficient effect size, making the statistical comparison unreliable. Nonetheless, a preliminary investigation using a different analysis method (LI) was conducted (figure 11) to rule out potential effects of artifact

removal and to suggest that exploring different types of EFRs in a separate study may be worthwhile.

However, regarding the conclusions of this study, the selected sample size is sufficient to support the findings on the utility and applicability of SWEEP stimulation for objective CI fitting.

5. Conclusion

In this work it has been shown that a step-wise sweep over five modulation frequencies in an AM CI pulse train can elicit EFRs. Similar to EASSRs from conventional AM stimulation, these EFRs can potentially be used for objective CI fitting. Using a hybrid EEG dataset, it has been shown that the proposed SWEEP stimulation improves the separation of response and stimulation artifact, which leads to an increase in SNR, that effectively increases detection sensitivity and reduces the required EEG recording time that is needed to detect a response above the noise floor. This greatly benefits the quality of objective fitting in CI users as the accuracy of the *T*-level detection increases while clinical time decreases. Furthermore, the neural response detection can be performed in an online fashion as the response latency, which is here used to differentiate between response and artifact, can be determined within the same recording, allowing for real-time assessment during ongoing acquisition. This makes the SWEEP stimulation resilient to interruptions. Unlike the conventional AM stimulation paradigm, where interruptions can lead to incomplete data collection, the SWEEP stimulation ensures dataset completeness with every stimulation epoch, while interruptions only affect the SNR. This robustness makes it particularly suitable for real-world clinical applications.

In addition to the technical advantages of the SWEEP stimulation, this work has also shown, that CI stimulation with time-varying modulation frequencies can reliably elicit EFRs. Next to the SWEEP stimulation, a stimulation paradigm with 'random' modulating signals (Shuffled-SWEEP) has been proposed and applied in four CI users. Again, neural responses have been elicited and successfully detected in all users. While neural tracking with speech stimuli has been conducted in CI users (Somers *et al* 2019, Nogueira and Dolhopiatenko 2022), there are few results available using a forward modeling approach. Implementing a forward model, as proposed in section 2.3, combined with the proposed stimuli, allows for systematic spatiotemporal assessment of the envelope processing in CI users, as demonstrated by Gransier and Wouters (2021) for normal-hearing individuals. A recent publication (Federici *et al* 2025) underscores the importance of temporal assessment of speech processing in CI users.

It can be concluded that the presented work enables the development of a robust, clinically applicable objective fitting routine in which EFRs can be used to determine the stimulation levels or assess the temporal envelope processing capability of CI users on each individual CI stimulation electrode. Future work should therefore focus on two things. One, investigating the potential difference in neural responses following the different stimulation paradigms, i.e. increasing the sample size and investigating the possible implications of differences in elicited neural responses. Two, the actual development and validation of an objective fitting routine with EFRs.

Data availability statement

The data supporting this study cannot be shared due to internal compliance regulations.

Acknowledgments

This work was funded by Cochlear Technology Centre Belgium and the Flanders Innovation & Entrepreneurship Agency through VLAIO research Grant HBC.2019.2373. The authors would like to thank E Verwaerde for the support with the collection of the data used in this paper.

Conflict of interest

The authors declare no conflict of interest.

ORCID iDs

Julian Schott https://orcid.org/0000-0003-0687-0560
Robin Gransier https://orcid.org/0000-0001-7429-6439
Marc Moonen https://orcid.org/0000-0003-4461-0073
Jan Wouters https://orcid.org/0000-0002-0093-698X

References

Aldag N and Nogueira W 2024 Psychoacoustic and electroencephalographic responses to changes in amplitude modulation depth and frequency in relation to speech recognition in cochlear implantees *Sci. Rep.* **14** 8181

Alemi R, Nozaradan S and Lehmann A 2021 Free-field cortical steady-state evoked potentials in cochlear implant users *Brain Topogr.* **34** 664–80

Bernardi G, Waterschoot T V, Moonen M, Wouters J, Hillbratt M and Verhaert N 2017 Measurement and analysis of feedback and nonlinearities for the codacs direct acoustic cochlear implant *IEEE Access* **5** 8702–13

Brown C J, Hughes M L, Luk B, Abbas P J, Wolaver A and Gervais J 2000 The relationship between eap and eabr thresholds and levels used to program the nucleus 24 speech processor: data from adults *Ear Hear.* **21** 151–63

Caner G, Olgun L, Gültekin G and Balaban M 2007 Optimizing fitting in children using objective measures such as neural response imaging and electrically evoked stapedius reflex threshold *Otol. Neurotol.* **28** 637–40

Carter L, Dillon H, Seymour J, Seeto M and Dun B V 2013 Cortical auditory-evoked potentials (caeps) in adults in

response to filtered speech stimuli *J. Am. Acad. Audiol.* **24** 807–22
David W, Verwaerde E, Gransier R and Wouters J 2023 Effects of analysis window on 40-Hz auditory steady-state responses in cochlear implant users *Hear. Res.* **438** 108882
Deprez H, Gransier R, Hofmann M, van Wieringen A, Wouters J and Moonen M 2017 Characterization of cochlear implant artifacts in electrically evoked auditory steady-state responses *Biomed. Signal Process. Control* **31** 127–38
Deprez H, Gransier R, Hofmann M, Wieringen A V, Wouters J and Moonen M 2017 Template subtraction to remove CI stimulation artifacts in auditory steady-state responses in CI subjects *IEEE Trans. Neural Syst. Rehab. Eng.* **25** 1322–31
Deprez H, Gransier R, Hofmann M, Wieringen A V, Wouters J and Moonen M 2018 Independent component analysis for cochlear implant artifacts attenuation from electrically evoked auditory steady-state response measurements *J. Neural Eng.* **15** 016006
Dijk B V *et al* 2007 Clinical results of autonrt,™ a completely automatic ecap recording system for cochlear implants' *Ear Hear.* **28** 558–70
Dillier N, Lai W K, Almqvist B, Frohne C, Müller-Deile J, Stecker M and Wallenberg E V 2002 Measurement of the electrically evoked compound action potential via a neural response telemetry system *Ann. Otol. Rhinol. Laryngol.* **111**
Dobie R A and Wilson M J 1996 A comparison of t test, f test and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids *J. Acoust. Soc. Am.* **100** 2236–46
Drullman R, Festen J M and Plomp R 1994 Effect of reducing slow temporal modulations on speech reception *J. Acoust. Soc. Am.* **95** 2670–80
Eeckhoutte M V, Wouters J and Francart T 2016 Auditory steady-state responses as neural correlates of loudness growth *Hear. Res.* **342** 58–68
Eeckhoutte M V, Wouters J and Francart T 2018 Electrically-evoked auditory steady-state responses as neural correlates of loudness growth in cochlear implant users *Hear. Res.* **358** 22–29
Farina A 2000 Simultaneous measurement of impulse response and distortion with a swept-sine technique *Proc. AES 108th Conv (Paris, France)*
Federici A *et al* 2025 Resilience and vulnerability of neural speech tracking after hearing restoration *Commun. Biol.* **8** 343
Gordon K A, Papsin B C and Harrison R V 2004 Toward a battery of behavioral and objective measures to achieve optimal cochlear implant stimulation levels in children *Ear Hear.* **25** 447–63
Gordon K, Papsin B C and Harrison R V 2004 Programming cochlear implant stimulation levels in infants and children with a combination of objective measures *Int. J. Audiol.* **43**
Gransier R, Carlyon R P and Wouters J 2020 Electrophysiological assessment of temporal envelope processing in cochlear implant users *Sci. Rep.* **10** 15406
Gransier R, Deprez H, Hofmann M, Moonen M, van Wieringen A and Wouters J 2016 Auditory steady-state responses in cochlear implant users: effect of modulation frequency and stimulation artifacts *Hear. Res.* **335** 149–60
Gransier R, Luke R, Wieringen A V and Wouters J 2020 Neural modulation transmission is a marker for speech perception in noise in cochlear implant users *Ear Hear.* **41** 591–602
Gransier R, van Wieringen A and Wouters J 2017 Binaural interaction effects of 30-50 Hz auditory steady state responses *Ear Hear.* **38** e305–15
Gransier R and Wouters J 2021 Neural auditory processing of parameterized speech envelopes *Hear. Res.* **412** 108374
Herdman A T, Lins O, Roon P V, Stapells D R, Scherg M and Picton T W 2002 Intracerebral sources of human auditory steady-state responses *Brain Topogr.* **15** 69–86
Hofmann M and Wouters J 2010 Electrically evoked auditory steady state responses in cochlear implant users *JARO - J. Associat. Res.Otolaryngol.* **11** 267–82
Hofmann M and Wouters J 2012 Improved electrically evoked auditory steady-state response thresholds in humans *J. Associat. Res. Otolaryngol.* **13** 573–89
Hu H and Ewert S D 2021 Exploring artifact rejection for high-pulse rate electrically evoked auditory steady state responses in cochlear implants users 1311–6
Hughes M L, Brown C J, Abbas P J, Wolaver A A and Gervais J P 2000 Comparison of EAP thresholds with MAP levels in the nucleus 24 cochlear implant: data from children *Ear Hear.* **21** 164–74
Ljung L 1999 *System Identification : Theory for the User* (*Prentice Hall Information and System Sciences Series*) 2nd edn (Prentice Hall/PTR)
Luke R, Deun L V, Hofmann M, Wieringen A V and Wouters J 2015 Assessing temporal modulation sensitivity using electrically evoked auditory steady state responses *Hear. Res.* **324** 37–45
Luke R and Wouters J 2017 Kalman filter based estimation of auditory steady state response parameters *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 196–204
Mao D, Innes-Brown H, Petoe M A, Wong Y T and McKay C M 2019 Fully objective hearing threshold estimation in cochlear implant users using phase-locking value growth functions *Hear. Res.* **377** 24–33
McKay C M, Chandan K, Akhoun I, Siciliano C and Kluk K 2013 Can ecap measures be used for totally objective programming of cochlear implants? *J. Associat. Res. Otolaryngol* **14** 879–90
McKay C M and Smale N 2017 The relation between ecap measurements and the effect of rate on behavioral thresholds in cochlear implant users *Hear. Res.* **346** 62–70
Miller C A, Brown C J, Abbas P J and Chi S-L 2008 The clinical application of potentials evoked from the peripheral auditory system *Hear. Res.* **242** 184–97
Müller S and Massarani P 2001 Transfer-function measurement with sweeps *J. Audio Eng. Soc.* **49** 443–71
Nelles O 2001 Nonlinear system identification
Nogueira W and Dolhopiatenko H 2022 Predicting speech intelligibility from a selective attention decoding paradigm in cochlear implant users *J. Neural Eng.* **19** 26037
Petley L, Blankenship C, Hunter L L, Stewart H J, Lin L and Moorea D R 2024 Amplitude modulation perception and cortical evoked potentials in children with listening difficulties and their typically developing peers *J. Speech, Lang. Hear. Res.* **67** 633–56
Picton T W, John M S, Dimitrijevic A and Purcell D 2003 Human auditory steady-state responses *Int. J. Audiol.* **42** 177–219
Purcell D W and John M S 2010 Evaluating the modulation transfer function of auditory steady state responses in the 65 Hz to 120 Hz range *Ear Hear.* **31** 667–78
Purcell D W, John S M, Schneider B A and Picton T W 2004 Human temporal auditory acuity as assessed by envelope following responses *J. Acoust. Soc. Am.* **116** 3581–93
Schott J, Gransier R, Wouters J and Moonen M 2022 Electrically evoked auditory steady state response detection in cochlear implant recipients using a system identification approach *Euro. Signal Process. Conf.* **2022** 1243–7
Schott J, Gransier R, Wouters J and Moonen M 2024 Detection of electrically evoked auditory steady-state responses in cochlear-implant recipients with a system identification based method *IEEE Trans. Biomed. Eng.* **71** 738–49
Seyle K and Brown C J 2002 Speech perception using maps based on neural response telemetry measures *Ear Hear.* **23** 72S-9S
Shannon R V, Zeng F-G, Kamath V, Wygonski J and Ekelid M 1995 Speech recognition with primarily temporal cues *Science* **270** 303–4
Smoorenburg G F, Willeboer C and Dijk J E V 2002 Speech perception in nucleus ci24m cochlear implant users with processor settings based on electrically evoked compound action potential thresholds *Audiol. Neuro-Otol.* **7** 335–47
Somers B, Verschueren E and Francart T 2019 Neural tracking of the speech envelope in cochlear implant users *J. Neural Eng.* **16** 016003

Távora-Vieira D, Wedekind A, Ffoulkes E, Voola M, Marino R and Delano P H 2022 Cortical auditory evoked potential in cochlear implant users: an objective method to improve speech perception *PLoS One* **17** e0274643

Távora-Vieira D, Wedekind A, Marino R, Purdy S C and Rajan G P 2018 Using aided cortical assessment as an objective tool to evaluate cochlear implant fitting in users with single-sided deafness *PLoS One* **13** e0193081

Thai-Van H *et al* 2004 Modeling the relationship between psychophysical perception and electrically evoked compound action potential threshold in young cochlear implant recipients: clinical implications for implant fitting *Clin. Neurophys.* **115** 2811–24

Undurraga J A, Yper L V, Bance M, McAlpine D and Vickers D 2021 Characterizing cochlear implant artefact removal from EEG recordings using a real human model *MethodsX* **8** 101369

Undurraga J A, Yper L V, Bance M, McAlpine D and Vickers D 2021 Neural encoding of spectro-temporal cues at slow and near speech-rate in cochlear implant users *Hear. Res.* **403** 108160

Visram A S, Innes-Brown H, El-Deredy W and McKay C M 2015 Cortical auditory evoked potentials as an objective measure of behavioral thresholds in cochlear implant users *Hear. Res.* **327** 35–42

Wouters J, McDermott H J and Francart T 2015 Sound coding in cochlear implants: From electric pulses to hearing *IEEE Signal Process. Mag.* **32** 67–80