Detection of electrically evoked auditory steady-state responses in cochlear-implant recipients with a system identification based method

Julian Schott, Robin Gransier, Jan Wouters, and Marc Moonen

***Abstract*— *Objective:* Electrically evoked auditory steady-state responses (EASSRs) can potentially be used as an objective measure to realize the automatic fitting of cochlear implants (CIs). They can be recorded using electroencephalography (EEG) and objectively detected at the modulation frequency of the stimulus. The main roadblock in using EASSRs is the presence of CI stimulation artifacts in the EEG recording. In this paper, we present an improvement of a recently introduced system identification (SI) based artifact removal method. We evaluate its applicability for objective CI fitting on a larger dataset. *Methods:* The parameter estimation problem of the SI is solved using ordinary least squares (OLS), where an additional regularization term is added to the cost function. We compare EASSR latencies as determined by the commonly used linear interpolation artifact removal method and SI, to evaluate the artifact removal and EASSR detection quality on a dataset of 16 CI recipients and four different stimulation levels. *Results:* SI can fully remove stimulation artifacts and detect EASSRs, even for recordings from ipsilateral EEG channels, where all other artifact removal methods fail so far. Using OLS with regularization prevents false positive response detection. *Conclusion:* Using SI, EASSRs can reliably be detected in EEG recordings, even for ipsilateral recording channels and recordings with lower stimulation levels. As the recordings are obtained with clinically relevant settings of the CI, they reveal the potential impact of SI on the objective fitting of CIs. *Significance:* We argue, that SI enables therefore a big step towards automated CI fitting with EASSRs.**

***Index Terms*— Cochlear implant, EEG, Objective fitting, Stimulation artifact, System identification**

This work was funded by Cochlear Technology Centre Belgium and the Flanders Innovation & Entrepreneurship Agency through VLAIO research grant HBC.2019.2373.

Julian Schott is with the Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, and with the Department of Neurosciences, ExpORL, KU Leuven, Belgium.

Robin Gransier and Jan Wouters are with the Department of Neurosciences, ExpORL, KU Leuven, Belgium.

Marc Moonen is with the Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Belgium.

Copyright (c) 2021 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

I. Introduction

COCHLEAR implants (CIs) are used to restore severe to profound hearing loss in people suffering from sensorineural hearing loss, through direct electrical stimulation of the auditory nerve. A CI consists of an external sound processor with a microphone array, which is used to record sounds from the surroundings, preprocess the signals and transmit electrical stimulation patterns via a transmitter coil to the internal part of the CI. The internal part consists of a receiver coil and an electrode array which is inserted into the cochlea. The stimulation patterns, capturing the recorded sound characteristics, are used to electrically stimulate the auditory nerve via the implanted array electrodes. This enables hearing in CI recipients.

To achieve the best possible listening experience and to maximize speech understanding, CIs have to be tuned to the specific needs and characteristics of their recipients. This so-called CI fitting is a tedious process that has to be repeated regularly and contributes immensely to the overall clinical burden. During a fitting, the stimulation parameters of a CI, i.e., the minimum and maximum stimulation current levels, T- and C-level respectively, have to be determined per array electrode. Highly trained clinicians are necessary to perform the task, which inherently leads to high costs per fitting. Additionally, the fitting procedure may vary across clinics and clinicians, as it may be subject to personal preference or experience, and may therefore make a comparison of fitting outcomes across CI recipients impossible. Another challenge during the fitting is the necessity of subjective feedback from the CI recipient. However, this is impossible in certain populations such as infants, people with cognitive disorders, or subjects under anesthesia. Objectively performed fitting could hence lead to a consistently repeatable protocol that guarantees optimal CI performance and would drastically reduce the required effort to perform an initial fitting of new recipients. It could be automated and can therefore highly reduce the costs associated with the fitting.

In recent years, objective measures have gained increased interest in the field of CI fitting. They could be used to perform the fitting procedure objectively without any required interaction between clinician and CI recipient and hence can even

lead to a fully automated fitting. Especially auditory steady-state responses (ASSRs) have shown great potential in normal-hearing subjects to determine hearing thresholds [1] [2] [3] [4] [5], along with good correlation with behavioral loudness growth functions [6], and could therefore be very useful for objective fitting. They are elicited via amplitude-modulated (AM) tones and recorded via electroencephalography (EEG). They can then objectively be detected at the modulation frequency of the stimulus using statistical tests [1] [7]. Electrically evoked auditory steady-state responses (EASSR) are the electrical equivalent of ASSRs in CI recipients, elicited via periodic low-rate or modulated high-rate CI pulse trains [1]. Previous studies have shown that T-levels based on EASSRs correlate well with behaviorally determined T-levels [8] [9], and EASSR magnitudes correlate well with behavioral loudness growth functions in CI recipients [10]. Other advantages of EASSRs include that their magnitude is tightly coupled to the temporal modulation sensitivity of the CI recipient [11], which in turn is an indication for speech perception in noise [12]. A crucial aspect of their clinical relevance is the ability to elicit the steady-state response with carrier rates, that are equal to the stimulation pulse rates as clinically used in CIs. This makes them preferred over other neural responses such as the electrically evoked compound action potential (ECAP) for T-level determination, which is elicited with single stimulus pulses [13]. Objectively determined thresholds with ECAPs show only moderate correlation with behavioral thresholds, especially with an increasing stimulation rate of the behavioral measures [14] [15]. One alternative to EASSRs are cortical auditory evoked potentials (CAEPs). These responses originate from the auditory cortex and are also elicited with clinical stimulation settings of the CI. Studies have shown their feasibility in determining objective threshold levels [16] [13] and even their effectiveness in exploiting them for map adjustments that lead to improvements in speech perception [17]. However, due to EASSRs correlation to modulation encoding ability, the main feature of CI stimulation, and the rather short duration of CAEP stimuli (hence lesser related to continuous speech), we focus our work on EASSRs rather than CAEPs.

The major challenge when measuring EASSRs is the presence of electrical artifacts in the EEG recordings, originating from the CI stimulation itself. These artifacts overlap with the EASSR in the time domain and are, as the EASSR, present at the modulation frequency [18]. They may therefore lead to false-positive response detection when solely checking the spectrum of the recordings. They originate from non-linearities in the stimulation and or recording system and highly depend on the stimulation as well as recording parameters. The most crucial factors of artifact contamination are the CI stimulation mode, stimulation level, and EEG recording channel. Fig. 1 illustrates the fundamental challenge with the stimulation artifact. In Fig. 1(a), the AM stimulus is shown, which does not contain any component at the modulation frequency as seen in Fig. 1(c). Consequently, the stimulus also cannot contain any component at the modulation frequency after passing through a linear system. In this case, any distinct component at this frequency in the EEG recordings can solely be attributed to the EASSR and not to the stimulus. However, even small non-linearities may introduce an artifact component. In Fig. 1(b), a non-linearity is simulated by adding a squared term of the AM stimulus, which leads to an asymmetry in the stimulus that introduces the DC and artifact components, as seen in the spectrum of the signal in Fig. 1(d). This component at the

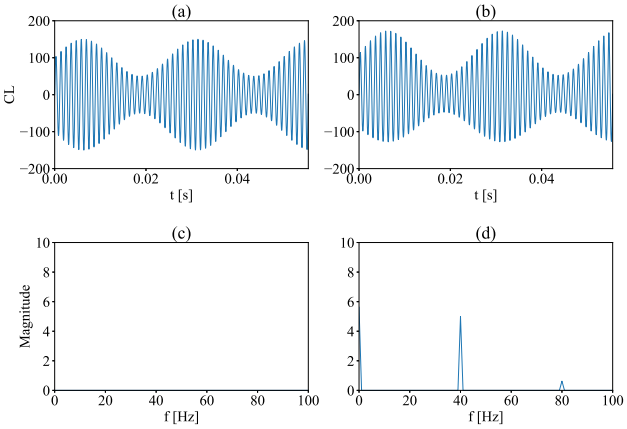


Fig. 1: Illustration of a stimulation artifact. (a) Example AM stimulation $u(t)$ with modulation frequency $f_m = 40Hz$. (b) AM stimulation with small non-linearity: $u(t)_{nl} = u(t) + 0.001 \cdot u(t)^2$. (c) Magnitude spectrum $|U(f)|$ of $u(t)$: No component at f_m. (d) Magnitude spectrum $|U(f)_{nl}|$: Artifact component at 0, f_m and $2 \cdot f_m$ Hz.

modulation frequency may be falsely interpreted as EASSR and can have a significant impact on the threshold estimation. Particularly with clinically used monopolar stimulation (MP), stimulation artifacts are larger in magnitude and endure longer in the time domain. Recordings from ipsilateral EEG channels are more contaminated due to their proximity to the CI. The artifacts are timely bound to the stimulation pulses of the CI and decay exponentially after the pulses [18].

Conventional filtering to remove the artifact is not feasible as it may also remove the EASSR. In the past, different artifact removal methods have been applied with different levels of success. In [8], linear interpolation (LI) of recording samples is used to remove the stimulation artifact. As the artifact decays after the stimulation pulse, the recorded samples that are furthest away from the previous and closest to the next pulse are assumed to be artifact-free. Linearly interpolating between these samples leads to an artifact-free recording sequence, which can then be analyzed in the frequency domain to determine the EASSR [8] [9]. This method has shown great success in stimulation and recording settings, where the artifact sufficiently decays between subsequent stimulation pulses. This is the case for EEG recordings with bipolar (BP) stimulation and lower pulse frequencies ($500pps$) [8] [19], as here the artifacts are lower in amplitude and the inter-pulse interval is longer. Also, results from contralateral EEG recording channels are better, when using this method. LI is relatively simple to apply and is therefore widely used in practice [12] [10] [19] [20]. Recently, good results with LI with MP stimulation and clinically relevant stimulation

rates ($900pps$) were obtained, however only for contralateral recording channels [21].

In [22], the artifact removal method is based on template subtraction. Here, EEG recordings with AM CI pulse trains are acquired, where the stimulation levels are below the subject's threshold levels. The recordings are therefore free of EASSR and any significant component at the modulation frequency is guaranteed to be from the artifact only. These are then used to create an artifact template sequence, which can be scaled and subtracted from subsequent EEG recordings with higher stimulation levels, with the intent to create an artifact-free recording. The underlying assumption is that after the non-linearity, which introduces the artifact component in the recordings, the artifact magnitude scales linearly with the stimulation level. The results presented in [22] indicate effective artifact removal, also for clinically relevant MP stimulation and ipsilateral recording channels. However, the necessity of extra EEG recordings to construct the templates renders this method clinically less interesting, as the ultimate goal is to reduce the time of the fitting procedure.

Spatial filtering methods are widely used in biomedical signal processing. Especially in EEG recordings, where the spatial resolution is poor, techniques such as independent component analysis (ICA), beamforming, and principal component analysis are popular to increase the signal-to-noise ratio by exploiting multichannel EEG recordings. For the detection of transient auditory responses such as ECAPs and electrically evoked auditory brainstem responses (EABRs), these techniques, especially ICA, have been applied with success [23] [24] [25] [26]. However, for EASSR detection, the results, when using ICA to remove the stimulation artifact, are mixed. In [27] and [28], using ICA for EASSR detection is introduced and evaluated. The results indicate, that especially with the clinically relevant $900pps$ stimulation rate, the ICA fails to correctly identify and disentangle the artifact and EASSR and may hence lead to a false removal of EASSR components [28]. In [29], ICA effectively removes the stimulation artifact from EEG recordings of cortical steady-state responses, but not from recordings of subcortical steady-state responses. In [30] and [31], denoising source separation (DSS), another spatial filtering approach akin to ICA, is applied to simulated and real recordings respectively. Results show, that a combination of LI and DSS is able to remove stimulation artifacts from the recordings with high pulse frequencies. However, here, DSS is merely a way to separate and reduce noise and hence increase the signal-to-noise ratio. The actual artifact removal, however, is still mainly performed by the prior LI.

In the aforementioned methods, the EASSR detection is performed in the frequency domain after artifact removal. In [32], a model based approach is used where the EASSR is modeled as a sinusoid at the modulation frequency and the artifact is modeled based on the observed artifact shape in the recordings. A Kalman filter (KF) is then used to estimate the model parameters that result in the best fit between the models and the recording. The estimated EASSR model parameters are then used to determine the phase and amplitude of the EASSR. This method is successful for contralateral recordings from BP stimulation but has not been evaluated for clinically relevant EEG recordings. Additionally, the modeling of the stimulation artifact is not very flexible and not generally applicable to all EEG recordings, especially for MP stimulation.

In [33] we have proposed a new method based on system identification (SI). It is based on the assumption that the EEG recordings can be modeled with two finite impulse response (FIR) systems. These unknown systems are identified based on the recordings, which leads to a separation of the EASSR and artifact in the recordings. The preliminary results show success in EASSR detection and artifact removal. Most significantly, these results are obtained for recordings with CI stimulation rates of $900pps$ in MP mode, in contralateral as well as in ipsilateral recording channels. Comparison to results obtained with LI shows that for recordings where the artifact exceeds the inter-pulse interval, LI fails, but the SI still succeeds to remove the artifact and detect the EASSR. However, the presented results are obtained from recordings of only one subject and can merely serve as proof of concept so far. One disadvantage of the work presented in [33] is the use of a KF to identify the unknown FIRs. This requires careful tuning of the noise covariance matrices and leads to an increase in computation time. A second disadvantage of the KF is that it does not straightforwardly allow the integration of a regularization term, which may be beneficial in the separation of EASSR and artifact. Furthermore, as the recorded data is pre-processed in a way that the time-variance of the systems gets lost, the main advantage of using a KF becomes irrelevant.

In this paper, we use the SI method on a dataset of EEG recordings from 16 CI recipients to evaluate the artifact removal and EASSR detection quality. We apply it to recordings at different stimulation levels to determine the effectiveness of SI regarding objective T-level detection in CI recipients. As in [33], we compare results from the SI with results obtained from LI to substantiate the preliminary results obtained in [33]. Unlike in [33], we determine the unknown model parameters using an ordinary least squares solution (OLS), which also allows for the integration of a regularization term. We show how the proposed model relates to a Hammerstein model, a popular choice for modeling non-linear systems based on linear parameter estimation.

II. Material and Methods

A. Dataset

EEG recordings were collected from 16 adult, unilateral CI recipients, all with Cochlear Ltd. devices, with a mean age of 67 years and at least three years of CI experience. 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.

EASSR stimuli consisted of AM pulse trains with symmetric biphasic cathodic-first pulses, with equal phase width of $25\mu s$ or $37\mu s$, depending on the clinical settings of the recipient, and an interphase gap of $8\mu s$. They were generated using custom software and directly delivered to the implant using the NIC4 research interface platform (Cochlear Ltd.). A clinically relevant pulse rate of $900pps$ was used, and the stimuli were presented between intracochlear electrode 15 and

an extracochlear electrode (MP mode), with some recipients using a combination of ball- and implant plate electrode (MP1+2) and some using the plate electrode only (MP2). A detailed summary can be found in TABLE I.

To reliably detect EASSRs, even in the presence of residual artifact after artifact removal, the apparent latency (further referred to as latency) is calculated from at least two recordings with different modulation frequencies [1] [8]. Here, three distinct recordings with modulation frequencies of $37Hz$, $40Hz$ and $43Hz$ are used and repeated at four different stimulation levels. A fixed modulation depth of 50 current levels (CL) between a defined maximum stimulation level and the level minus 50 CL was used for the AM pulse trains. The maximum stimulation levels were defined as follows:
For each subject, the T- and C-level were behaviorally determined before the EEG recordings. Here, an amplitude-modulated stimulus with a modulation frequency of $40Hz$, a modulation depth of 50CLs and a duration of 1s was used. T-and C-levels were determined based on a 7-point categorical loudness scale (i.e. “inaudible”, “very soft”, “comfortably loud”, “loud”, “very loud”, and “unbearable”). First, the C-level was assessed by gradually increasing CLs from ”inaudible” to the most comfortable loudness level which is defined as the current level prior to the transition between “comfortably loud” and “loud” on the categorical loudness scale. Second, the current was gradually decreased to determine the T-level of the modulated pulse train, which was defined as last the current level at which the participant could perceive the stimulus before it became ”inaudible”. The dynamic range (DR = C - T) was calculated, and the four maximum stimulation levels of the EEG recordings were defined as C, T + 0.5·DR, T and T - 10CL, in the following referred to as $100pDR$ (”100 percent dynamic range”), $50pDR$, $0pDR$ and $Tm10$. The $Tm10$ EEG condition could not be measured for all subjects due to timing constraints in some individual cases.

A Biosemi ActiveTwo Hyper-Rate EEG recording system with a sampling rate of 262,144kHz was used to perform the 8-channel recordings. Six recording electrodes were placed according to the international 10-20 system at $P9$, $P10$, Iz, Cz, Fz, Fpz and two extra electrodes at the left and right mastoid respectively (MaL, MaR). In cases where the EEG recording electrode was located at the CI transmitter/ receiver coil position, the nearest neighbor electrode position was used instead.

A total signal length of 300 epochs of 1.024s duration each was recorded per modulation frequency. The EEG signals were low-pass filtered with the system’s built-in analog third-order antialiasing filter having a -3 dB point at 50 kHz. All off-line processing was performed using Matlab R2020a. Off-line referencing with Cz was performed and detrending with a second-order polynomial was applied, using Matlab’s built-in $detrend$ function. The signals were then, based on the stimulation triggers, split into epochs of 1.024s duration and 5% of the epochs with the highest peak-to-peak amplitudes were rejected. The modulation frequencies and pulse rate of the stimuli were rounded before stimulation so that an integer number of cycles and pulses were contained within each epoch to avoid spectral leakage. The mean epoch for MaR, MaL, $P9$, $P10$, and Iz was used for the further analysis.

B. System identification based EASSR detection

As described in section I, the signal component at the modulation frequency (f_m) contains the EASSR as well as an artifact contribution. To distinguish between these two, the phase information of the component can be used, as the EASSR results from phase-locking of the neurons to the stimulation [1] [34], unlike the artifact. However, due to the ambiguous character of the phase, the latency is used instead [1]. As described in [33], the EASSR latency corresponds to the group delay of the auditory system. To determine the group delay from the recordings, the EASSR and artifact need to be modeled separately. The model definitions in the next subsections are adopted from [33] and shown here for convenience.

1) EASSR model: The EASSR ($r[k]$) can be modeled as the output of an FIR system ($FIR1$)

$$\begin{aligned} r[k] =& \mathbf{m}[k] \cdot \mathbf{b_{resp}} \\ =& \begin{bmatrix} m[k] & m[k-1q] & \dots & m[k-N_1 q] \end{bmatrix} \cdot \mathbf{b_{resp}}, \end{aligned} \tag{1}$$

with the modulation of the stimulus pulse train

$$m[k] = \sin(2\pi f_m k T_s + \frac{\pi}{2}) \tag{2}$$

as its input signal. Here, $T_s = \frac{1}{F_s}$ denotes the sampling interval of the EEG recording system, k is the discrete time index, N_1 is the order of $FIR1$ and q a time-shift parameter that will be chosen later. The system coefficients $\mathbf{b_{resp}} \in \mathbb{R}^{N_1+1}$ are to be estimated from the recordings using the proposed method in subsection II-B.3. The resulting $FIR1$ is then defined as

$$FIR1(z) = \begin{bmatrix} 1 & z^{-q} & z^{-2q} & \dots & z^{-N_1 q} \end{bmatrix} \cdot \mathbf{b_{resp}}. \tag{3}$$

Correct identification of $\mathbf{b_{resp}}$ leads to the determination of the group delay of the auditory system and hence the detection of the EASSR.

This is however only possible in interaction with the simultaneous identification of the stimulation artifact in the recording.

2) Artifact model: The artifact arises from non-linearities in the system (stimulation system, subject head, and/ or recording system). A vast amount of non-linear system identification procedures exist in the literature, and are typically much more involved than linear system identification procedures [35]. A good approach may therefore be to use a linearization of the non-linear system. However, in practice, linearization is not straightforward as the required non-linear models are unknown. Therefore, a linear model, approximating the input and output relation/spectra of the non-linear system, can be formulated and identified using normal system identification procedures [35]. In [33], the non-linear system identification is done with a second FIR system ($FIR2$), of which the output $a[k]$ models the stimulation artifact of the recording.

$$\begin{aligned} a[k] =& \mathbf{u}[k] \cdot \mathbf{b_{arti}} \\ =& \begin{bmatrix} u[k] & u[k-1] & \dots & u[k-N_2] \end{bmatrix} \cdot \mathbf{b_{arti}}. \end{aligned} \tag{4}$$

TABLE I: Subject summary.

Subject	Age	Recorded stimulation levels	Recording electrodes	Phase width [μs]	Return electrode
S1	70	$100pDR$, $50pDR$, $0pDR$	MaR, MaL, $P9$, $P10$, Iz	25	MP1+2
S2	74	$100pDR$, $50pDR$, $0pDR$	MaR, MaL, $P9$, $P10$, Iz	50	MP1+2
S3	65	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	25	MP1+2
S4	69	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	25	MP1+2
S5	65	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	25	MP1+2
S6	48	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $TP7$, $P10$, Iz	25	MP1+2
S7	54	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $TP7$, $P10$, Iz	25	MP1+2
S8	62	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	37	MP2
S9	80	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $TP7$, $P10$, Iz	37	MP2
S10	59	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	25	MP2
S11	70	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	37	MP2
S12	70	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	37	MP2
S13	73	$100pDR$, $50pDR$, $0pDR$	MaR, MaL, $P9$, $P10$, Iz	25	MP2
S14	78	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	37	MP2
S15	66	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	37	MP2
S16	71	$100pDR$, $50pDR$, $0pDR$, $Tm10$	MaR, MaL, $P9$, $P10$, Iz	37	MP2

The crucial aspect of this approximation is that the input signal to this FIR system ($u[k]$), is a monophasic AM pulse train (with only positive values) and consequently contains a component at f_m. Otherwise, in the case of a biphasic input signal, the linear model would not be able to approximate the artifact component at f_m (see Fig. 1). The input signal is defined as follows:

$$u[k] = A_c \cdot (1 + D \cdot m[k]) \cdot s[k], \tag{5}$$

with

$$s[k] = \begin{cases} 1, & \text{for } mod(kT_s, \frac{1}{f_c}) < T_s \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

with pulse rate f_c, and pulse train $s[k]$, which contains 1's at the sample positions k, which are closest to the stimulation pulse positions and otherwise 0. A_c and D are defined by the used stimulation level and modulation depth. For example, for the $100pDR$ condition:

$$A_c = \frac{C + (C - 50)}{2} \tag{7}$$

and

$$D = \frac{C - (C - 50)}{C + (C - 50)}. \tag{8}$$

This definition of $u[k]$ and the resulting artifact model can be related to a Hammerstein model, which is a popular model for non-linear systems. The generalized case consists of a static non-linear element, followed by a linear system [36]. The artifact, modeled as a biphasic AM pulse train converted into a monophasic AM pulse train, followed by a linear FIR filter, indeed then corresponds to a Hammerstein model with *Rectified Linear Unit* (ReLu) as static non-linearity as shown in Fig. 2.

3) System model and parameter estimation: In order to determine the group delay from the recordings, the individual models and recordings of the three distinct modulation frequencies need to be stacked into one combined system model.

$$\underbrace{\begin{bmatrix} \mathbf{z}^{37} \\ \mathbf{z}^{40} \\ \mathbf{z}^{43} \end{bmatrix}}_{\mathbf{z}} = \underbrace{\begin{bmatrix} \mathbf{M}^{37} & \mathbf{U}^{37} \\ \mathbf{M}^{40} & \mathbf{U}^{40} \\ \mathbf{M}^{43} & \mathbf{U}^{43} \end{bmatrix}}_{\mathbf{H}} \cdot \underbrace{\begin{bmatrix} \mathbf{b_{resp}} \\ \mathbf{b_{arti}} \end{bmatrix}}_{\mathbf{x}} + \mathbf{v}, \tag{9}$$

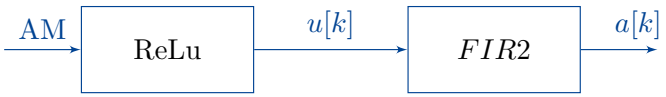


Fig. 2: Hammerstein model for artifact component $a[k]$. *Rectified Linear Unit* (ReLu) non-linearity applied to biphasic AM stimulation leads to monophasic stimulation input sequence $u[k]$. Filtering of $u[k]$ with to be identified $FIR2$ leads to a linear approximation of the non-linear artifact component $a[k]$ in the recording.

with recording $\mathbf{z}^{f_m} = \begin{bmatrix} z^{f_m}[0] & z^{f_m}[1] & \dots & z^{f_m}[K] \end{bmatrix}^\intercal$ being the mean epoch of the recording with modulation frequency f_m of the analyzed channel, with a total number of $K + 1$ samples, and where $\mathbf{v}$ is the similarly defined additive noise. The input matrices $\mathbf{M}^{f_m}$ and $\mathbf{U}^{f_m}$ are similarly defined per modulation frequency f_m, using $\mathbf{m}[k]$ and $\mathbf{u}[k]$ respectively,

$$\mathbf{M}^{f_m} = \begin{bmatrix} \mathbf{m}^{f_m}[0] \\ \mathbf{m}^{f_m}[1] \\ \vdots \\ \mathbf{m}^{f_m}[K] \end{bmatrix}, \tag{10}$$

and

$$\mathbf{U}^{f_m} = \begin{bmatrix} \mathbf{u}^{f_m}[0] \\ \mathbf{u}^{f_m}[1] \\ \vdots \\ \mathbf{u}^{f_m}[K] \end{bmatrix}. \tag{11}$$

The unknown model parameters $\mathbf{x}$ then need to be derived to determine the group delay of the $FIR1$ system. In [33], the parameters are determined recursively with a KF, as the underlying system is assumed to be time-varying. Time variations may arise from adaptation in the auditory pathway, position changes or a change in the wakefulness of the subject during the recordings. However, these dynamics get lost when averaging subsequent recording epochs, as done in the pre-processing here. Therefore, in this paper, we assume time-invariance in the analyzed mean epoch and consequently use a batch processing approach with OLS estimation of the parameters. This way, we also avoid the often difficult

hyperparameter tuning of the KF and at the same time reduce the computation time.

The OLS problem formulation also facilitates the addition of a regularization term. Here, we use Tikhonov regularization in the least squares cost function

$$J = \|\mathbf{Hx} - \mathbf{z}\|_2^2 + \alpha\| \mathbf{\Gamma x}\|_2^2, \tag{12}$$

with the regularization matrix $\mathbf{\Gamma}$ and the regularization constant α. The solution is given as (derivation in the supplementary materials)

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{\mathbf{b}}_{\mathbf{resp}} \\ \hat{\mathbf{b}}_{\mathbf{arti}} \end{bmatrix} = (\mathbf{H}^\intercal\mathbf{H} + \alpha\mathbf{\Gamma}^\intercal\mathbf{\Gamma})^{-1}\mathbf{H}^\intercal\mathbf{z}. \tag{13}$$

The regularization term $\alpha\|\mathbf{\Gamma x}\|_2^2$ penalizes solutions with larger parameter estimates, which is especially beneficial for ill-posed problems, dealing with multicollinearity in the model, such as the one formulated in (9). The multicollinearity is a result of the inherent high correlations among the columns of $\mathbf{M}^{f_m}$, the predictor variables in the EASSR model, as they are time-shifted sinusoids. Furthermore, $\mathbf{U}^{f_m}$ contains the same modulation as $\mathbf{M}^{f_m}$, only sub-sampled at the positions of the stimulation pulses. This further adds to the multicollinearity of $\mathbf{H}$ and hence leads to high condition numbers of the moment matrices of the least-squares solution. Interestingly, for problem formulations with multicollinearity, the overall prediction accuracy of the model may not be impacted as much, only the calculation and hence interpretation of the individual coefficients may be strongly influenced by it [37].

Furthermore, as the EASSR amplitude A with normalized angular modulation frequency $\omega_m = 2\pi f_m/F_s$ is defined as

$$\begin{aligned} A(f_m) &= |FIR1(z = e^{j\omega_m})| \\ &= |\begin{bmatrix}1 & e^{-jq\omega_m} & e^{-j2q\omega_m} & \dots & e^{-jN_1q\omega_m}\end{bmatrix} \cdot \mathbf{b}_{\mathbf{resp}}|, \end{aligned} \tag{14}$$

penalizing large values in $\mathbf{b}_{\mathbf{resp}}$ in $\mathbf{x}$ leads to a regularization of the modeled EASSR amplitude. This is assumed to favor the separation of EASSR and artifact, as the artifact amplitude is expected to be higher than the EASSR amplitude. Unregularized estimated $\hat{\mathbf{b}}_{\mathbf{arti}}$ coefficients, are then larger in magnitude than estimated $\hat{\mathbf{b}}_{\mathbf{resp}}$ coefficients, and thus lead to a higher estimated contribution of the artifact $a[k]$ than the response $r[k]$ in the recording. Moreover, a large first filter coefficient corresponds to a zero-phase filter. Regularizing $\mathbf{b}_{\mathbf{resp}}$ may prevent this and therefore enforce a non-zero phase delay in the estimated $FIR1$. Simply put, regularization of $\mathbf{b}_{\mathbf{resp}}$ prevents overfitting of the estimated EASSR from an artifact-dominated recording.

To regularize $\mathbf{b}_{\mathbf{resp}}$, $\mathbf{\Gamma}$ is defined as a diagonal matrix with $1's$ along the main diagonal elements, corresponding to the $N_1 + 1$ response coefficients and $0's$ elsewhere.

$$\mathbf{\Gamma} = diag(1, 1, ..., 1, 0, ..., 0). \tag{15}$$

4) Model hyperparameters and additional preprocessing: In order to solve the system of equations, the following extra processing steps and hyperparameter definitions are required.

For each EEG recording, a pre-defined number of samples around the stimulation pulses of the mean epoch are skipped and not included in the model creation, as the system is assumed to behave non-linearly mainly in this time window. The skipping is based on the time difference between the sample k at discrete time kT_s and the nearest stimulation pulse at a multiple of period $T_p = 1/f_c$. Equivalently, recording sample k is skipped when

$$T_p - s_{\text{I}} \leq \mod(kT_s, T_p) \leq s_{\text{II}}, \tag{16}$$

with $\mod(kT_s, T_p)$ being the time difference between sample k and the previous stimulation pulse, and s_{I} and s_{II} the skipping duration before the next and after the previous stimulation pulse respectively. The skipping additionally alleviates the multicollinearity problem, as the number of columns in $\mathbf{H}$ and consequently complexity of the model is reduced. After the skipping of the recording samples, the reduced mean epochs are mean-centered again, to obtain a zero-mean signal in $\mathbf{z}$. Consequently, $\mathbf{H}$ is also mean-centered.

The time-shift q is used to further decrease the multicollinearity between the columns in $\mathbf{M}$, as these are time-shifted sinusoids. Especially with the high sampling frequency used in the recordings, sinusoids that are only shifted by one sample are close to linearly dependent and hence lead to identifiability issues in the OLS. The choice of q depends on the number of columns in $\mathbf{M}$, i.e., filter coefficients in $\mathbf{b}_{\mathbf{resp}}$, which is maximally two times the number of used modulation frequencies or stacked recordings respectively, and the expected group delay. As we create $\mathbf{H}$ from three recordings with different modulation frequencies, N_1 is set to its max value 5, resulting in $\mathbf{b}_{\mathbf{resp}} \in \mathbb{R}^6$. For the used f_m range around $40Hz$, an EASSR latency, i.e., a group delay of the auditory system, around $40ms$ is to be expected [38] [39] [12]. The q is chosen so that $FIR1$ can model a group delay of this duration. Furthermore, the phase response of $FIR1$ is also expected to be linear, which leads to the definition of q as follows:

$$q = \lfloor \frac{0.04s \cdot F_s}{N_1/2} \rceil, \tag{17}$$

which allows the estimation of a symmetric FIR filter and hence leads to a linear phase response. This way, the estimated filter taps in $\mathbf{b}_{\mathbf{resp}}$ will correspond to delays around $0ms$, $16ms$, $32ms$, $48ms$, $63ms$ and $80ms$ respectively.

The order of $FIR2$ is chosen, so that it can model stimulation artifacts, that overlap the inter-pulse interval and is therefore set 1.25 times the number of samples in the inter-pulse interval, i.e., $N_2 = \lfloor 1.25 \cdot T_p * F_s \rceil$.

The regularization constant α is empirically determined, using the sub-threshold $Tm10$ recordings. With these recordings, a good artifact estimation is achieved, when after artifact removal only residual noise is left in the remaining signal (the determined EASSR is statistically non-significant), as no response is expected in these recordings. We swept over different values of α to determine the best compromise between the maximum number of non-significant results and the lowest regularization value and then chose $\alpha = 5000$. This value was then used for all supra-threshold recordings.

An overview of the model hyperparameters and their respective values is shown in TABLE II.

C. Linear interpolation based EASSR detection

LI replaces signal samples between a selected pre-stimulus and post-stimulus sample by linear interpolation between these samples to remove the stimulation artifact. This is possible because the artifact is known to be time-bound to the stimulus, decaying with increasing duration after the stimulus [18]. The maximal interpolation length and consequently maximal artifact removal is restricted by the pulse period T_p, which is 1.11 ms for $f_c = 900pps$. In the following, we applied LI between samples from $100\mu s$ before to $1000\mu s$ after each stimulation pulse for each recording, which has proven to be successful for recordings, where the artifact does not exceed the inter-pulse interval [21]. After LI, the spectrum of the interpolated signal of each recording is evaluated at the respective modulation frequency. A simple linear regression is fitted to the unwrapped phases of the f_m components and the latency L is defined as

$$L = \mid \frac{1000 * a}{360^\circ} \mid, \tag{18}$$

with a the determined slope of the line fit.

D. Evaluation

Evaluation of artifact removal quality and consequently EASSR detection is based on the determined group delay values, further also referred to as latency. On the one hand, for recordings where an EASSR is present, a linear phase change is expected with increasing modulation frequencies, due to the frequency specificity of the neural generators of the response [34] [1]. The latency values are expected to be around $40ms$ for the used f_m range around $40Hz$ [38] [39] [12]. On the other hand, the phase stays constant over different modulation frequencies, when the recordings are dominated by artifact, as the artifact is time-locked to the pulses of the stimulus and its propagation is assumed to be instantaneous. The phase is then stable around multiples of 180° and hence the latency remains around $0ms$.

Latency results, where at least one of the three recordings of different modulation frequencies, does not contain a significant component at f_m, are excluded from evaluation and hence not plotted in Fig. 3(a) - 3(d). That means, that for every plotted latency result point in Fig. 3(a) - 3(d), all three components at the respective frequencies $37Hz$, $40Hz$ and $43Hz$ were statistically deemed significant. The statistical significance of the component is based on the one-sample Hotelling T^2 test, where the identified phase and amplitude at f_m of the mean epoch are compared to the variance over all epochs [1] [9] [7]. P values that exceed the significance level of 5% signify that the detected EASSR does not differ from EEG background noise. We therefore do not take these results into account. Nonetheless, non-significant results may be used to compare the quality of artifact removal for lower stimulation levels. In the absence of an EASSR, the stimulation artifact may still lead to a significant component at the modulation frequency, which may be determined as artifact by inspecting the latency value. After successful artifact removal, the component should become non-significant as only noise should remain at the modulation frequency. This way, artifact removal quality can be evaluated in the absence of EASSRs, as seen in Fig. 3(d).

TABLE II: Hyperparameter settings for system model creation.

Parameter	Description	Value
N_1	Order of FIR1	5
N_2	Order of FIR2	364
q	Time shift in samples in FIR1	4194
s_{I}	Beginning skipping interval	$100\mu s$
s_{II}	Ending skipping interval	$450\mu s$
α	Regularization constant	5000

III. Results

A. Latency

The EASSR latencies determined with LI and SI are shown for each stimulation level in Figs. 3(a) - 3(d). Each figure compares both methods based on the latency values, as determined for all recording channels of all subjects. The comparison is separately displayed for contra- and ipsilateral channels in the left and right columns of the figures respectively. An overview of the determined EASSR latencies per individual subject and stimulation level can be found in the supplementary materials. For the $100pDR$ condition (Fig. 3(a)), the latencies are in the expected range (indicated via horizontal dashed lines) for contralateral recording channels for both methods. 39 out of 48 (16 subjects times 3 contralateral recording channels) and 46 out of 48 results for LI and SI respectively, were deemed significant and included in the presented results, as indicated in the textbox in the upper left corner of the figure column. For ipsilateral recording channels, only SI is able to reliably determine latency values in the expected range for the majority of the recordings. LI produces significantly lower latencies, with only 4 of the 28 significant results in the expected range. SI produces 30 significant results out of the 32 possible results (16 subjects times 2 ipsilateral recording channels), most of them in the expected latency range. A Kruskal-Wallis test confirms a significant difference between determined latencies for all four groups, respectively defined as *LI-contra*, *SI-contra*, *LI-ipsi*, *SI-ipsi* ($F(3,139) = 49.3$, $p < 0.0001$). Dunn's test for multiple comparisons finds significant differences in mean latency between *LI-contra* and *LI-ipsi* ($p < 0.0001$), *LI-contra* and *SI-ipsi* ($p < 0.02$), *SI-contra* and *LI-ipsi* ($p < 0.0001$), and *SI-ipsi* and *LI-ipsi* ($p < 0.0001$).

Results for the $50pDR$ condition (Fig. 3(b)) are similar to the results for the $100pDR$ condition. Significant latencies from contralateral recordings are still within or close to the expected range for both methods. With LI, 18 results are significant, whereas with SI 23 results are significant. For ipsilateral recordings, two clear differences can be observed. SI clearly determines higher latencies in the recordings (LI median: 8.0ms, SI median: 47.3ms), indicating the presence of EASSRs, whereas with LI only one result lies within the expected range. With LI 27 results are deemed significant,

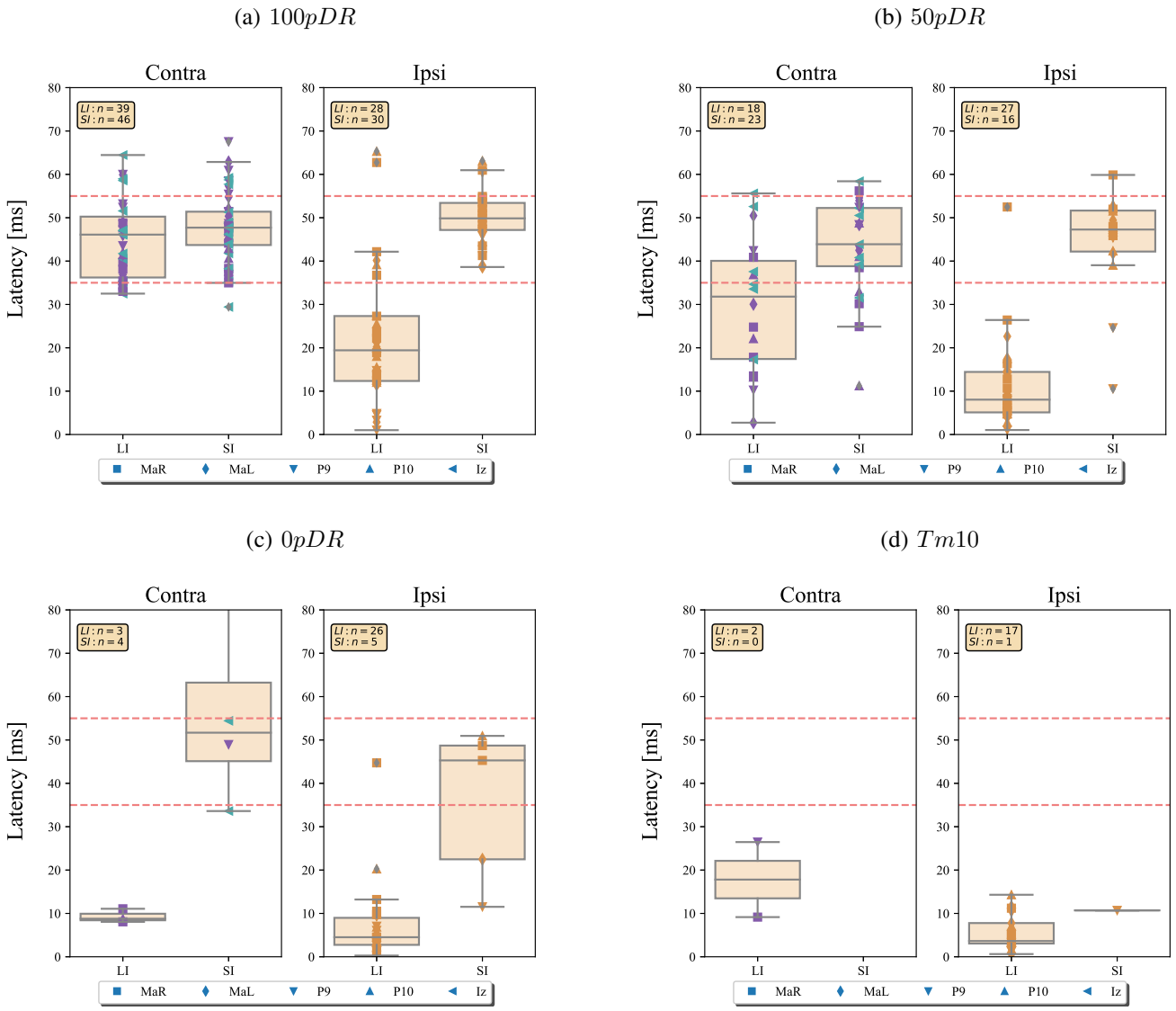


Fig. 3: Determined EASSR latencies for the stimulation levels at $100pDR$ (a), $50pDR$ (b), $0pDR$ (c), and $Tm10$ (d). LI and SI results of all subjects are compared side by side and separated for contralateral (Contra) and ipsilateral (Ipsi) recording channels. The expected latency range between $35ms$ - $55ms$ is indicated with dashed lines. The text box provides information on the number of results, which are statistically significantly different from background noise, and which are used for the box-plot calculation and plotted as individual plot points.

while with SI 16 results are deemed significant, of which 13 clearly contain EASSRs, as indicated by the latencies. A Kruskal-Wallis test confirms a significant difference between the groups (F(3, 80) = 44.76, $p < 0.0001$), and Dunn's test determines pair-wise differences between *LI-contra* and *LI-ipsi* ($p < 0.0019$), *SI-contra* and *LI-contra* ($p < 0.023$), *LI-contra* and *SI-ipsi* ($p < 0.031$), *SI-contra* and *LI-ipsi* ($p < 0.0001$) and *SI-ipsi* and *LI-ipsi* ($p < 0.0001$).

For the $0pDR$ condition (Fig. 3(c)), almost no results indicate the presence of EASSRs in the recordings. With LI 3 results are deemed significant in the contralateral recordings, of which none are within the expected latency range. With SI however, 4 results of the contralateral recordings are significant, of which 3 reach the expected latency range. For ipsilateral recordings, LI determines one latency in the expected range, whereas SI determines 3 latencies within the expected range. Again, with LI more results are deemed significant than with SI, mostly with a low latency value, indicating residual artifact contamination. A Kruskal-Wallis test confirms a significant difference between the groups (F(3, 34) = 20.15, $p < 0.0002$), and Dunn's test determines pair-wise differences between *SI-contra* and *LI-ipsi* ($p < 0.0005$), and *SI-ipsi* and *LI-ipsi* ($p < 0.00012$).

A detailed summary of the determined significant latencies is included in Tab. III.

B. EASSR phase

The phase results, used to calculate the latency, are added for completeness in the supplementary materials.

TABLE III: Determined EASSR latencies in ms for the 25th, 50th, 75th percentile. Values only include significant results as shown in Fig. 3(a) - 3(d).

	LI latencies [ms]			**SI latencies [ms]**		
	25%ile	50%ile	75%ile	25%ile	50%ile	75%ile
100pDR contra	36.2	46.1	50.2	43.7	47.7	51.4
100pDR ipsi	12.4	19.4	27.3	47.2	49.9	53.4
50pDR contra	17.4	31.8	40.1	38.8	43.9	52.3
50pDR ipsi	5.1	8.0	14.4	42.2	47.3	51.6
0pDR contra	8.4	8.8	9.9	45.1	51.7	63.2
0pDR ipsi	2.8	4.5	9.0	22.5	45.3	48.7
Tm10 contra	13.5	17.8	22.1	nan	nan	nan
Tm10 ipsi	3.1	3.6	7.8	10.7	10.7	10.7

C. EASSR amplitude

A comparison of the estimated EASSR amplitudes for both methods and recording sides is shown in Figs. 4(a) - 4(d). An average of the background noise level over recording channels is plotted per subject for each modulation frequency as dotted line. It can be seen that for all stimulation conditions, results from contralateral recordings are similar when using SI or LI. However, the estimated EASSR amplitudes from ipsilateral recordings differ when using SI or LI. With LI, amplitudes are generally higher, also when compared to amplitudes from contralateral recordings, and are higher compared to the amplitudes from ipsilateral recordings, when determined with SI. EASSR amplitudes from ipsilateral channels are in the same range as EASSR amplitudes from contralateral channels when using SI. Additionally, it can be observed, that with decreasing stimulation levels, the determined EASSR amplitudes also decrease. A detailed summary of the results can be seen in Tab. I in the supplementary materials.

IV. Discussion

Auditory steady-state responses recorded via EEG have shown great potential as an objective measure for objective CI fitting. However, in recordings from CI users, the CI stimulation adds a significant signal component at the modulation frequency, which makes objective EASSR detection challenging, due to the ambiguous origin of the signal energy at the frequency of interest. In this paper, we refined the recently introduced system identification based EASSR detection and artifact removal method. We propose to use ordinary least squares with Tikhonov regularization to deal with ill-conditioning and collinearity of the models when estimating the unknown model coefficients. The method is applied to a dataset consisting of 16 subjects with recordings at four different stimulation levels each.

A. Latency

The results show that SI, using OLS with Tikhonov regularization, is able to reliably remove stimulation artifact, and determine EASSRs in the recordings, especially for higher stimulation levels such as the $100pDR$ and $50pDR$ condition. In contralateral recordings, the stimulation artifact is shorter than the inter-pulse interval, and hence LI is also able to remove it and determine similar EASSR latencies. For lower stimulation levels such as $50pDR$, SI performs better, as it determines more significant responses with smaller IQR and higher EASSR latencies for the 25th, 50th and 75th percentile as seen in Fig. 3(b) for the contralateral recordings. P-values from Dunns's test confirm a difference between *SI-contra* and *LI-contra* (p = 2.26e-2). A comparison between *SI-contra-100pDR* and *SI-contra-50pDR* determines that the latencies of these groups are statistically equal (Dunn's test p = 0.44). This is a crucial aspect when the stimulation level decreases further and the expected EASSR amplitude approximates the noise level. As can be seen for the $0pDR$ condition in Fig. 3(c), only a few latency values indicate the presence of EASSRs in the recording. SI is able to objectively determine a response for at least 3 contralateral recordings at T-level. However, with LI, no significant result has a latency that indicates the presence of an EASSR. Most of the results are indistinguishable from the EEG background noise and hence deemed non-significant. This is not unexpected at this stimulation level, as the used CLs are modulated between T-level and T - 50CL, hence a large portion of the stimulus is below the subject's T-level and may therefore merely elicit a small response or no response at all.

In clinical scenarios with an objective fitting routine with a focus on threshold detection, the results obtained with SI indicate superior performance with small responses close to noise level as compared to LI. This might be due to the implementation of LI, as here, most of the actual signal is removed and interpolated between two selected samples in an inter-pulse interval, making the resulting interpolated signal very susceptible to low signal-to-noise ratios (SNRs). The interpolated signal may then follow the noise in the recording, rather than the response, leading to a rejection of the result.

Ultimately, the most significant difference and hence the contribution of SI becomes evident when comparing latencies from ipsilateral recording channels, where the artifact contamination is highest in the recordings. Fig. 3(a) and Fig. 3(b) clearly show artifact domination in these recordings, as the determined latencies with LI are below the expected range. The latencies determined with SI are, however, higher in comparison to those determined with LI and in many cases within the range of the latency of the true neural generators [38]. Dunn's test shows statistical differences for all pairs, where *LI-ipsi* is included, confirming insufficient artifact removal of this group. For groups that use artifact removal with SI, the p-values show no difference between ipsilateral or contralateral recording channel results. This clearly shows that the method can model stimulation artifacts, including

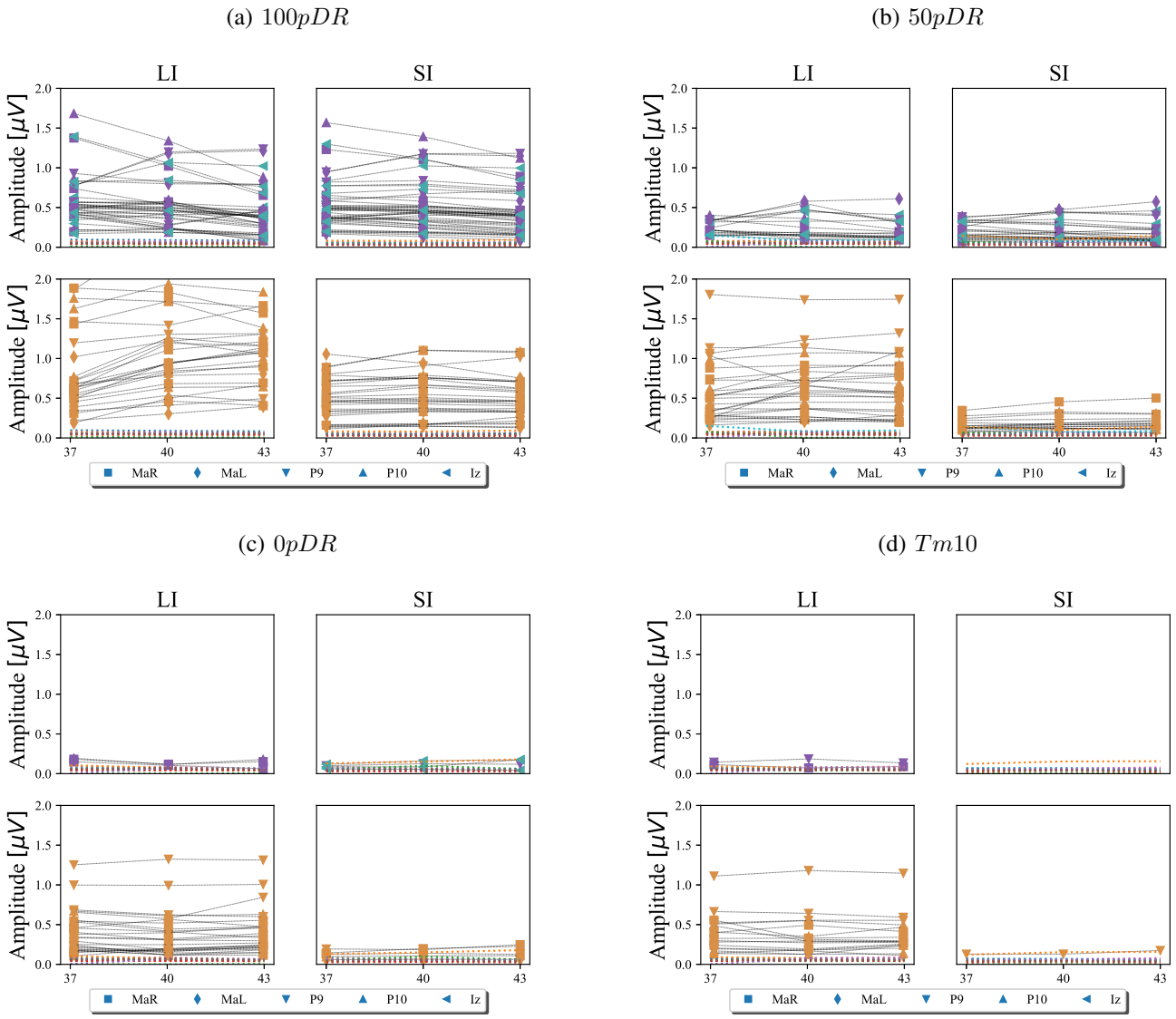


Fig. 4: Determined EASSR amplitudes for the stimulation levels at $100pDR$ (a), $50pDR$ (b), $0pDR$ (c), and $Tm10$ (d). Results for all subjects are plotted against the used modulation frequencies at $37Hz$, $40Hz$, and $43Hz$, per recording site (contralateral channels on top and ipsilateral channels at the bottom) and used artifact removal method (LI and SI). The average of the background noise level over recording channels is plotted per subject for each modulation frequency as dotted line.

artifacts that exceed the inter-pulse interval, and separate them from the EASSR. Most notably, for low stimulation levels such as the $50pDR$ and $0pDR$ condition, SI, applied to ipsilateral recordings, detects EASSRs better than LI, applied to contralateral recordings, as seen in Figs. 3(b) and 3(c), although not confirmed by the p-values of Dunns's HSD.

The obtained results confirm the preliminary findings in [33] at a group level across individuals.

B. Artifact removal

As the amount of skipped recording samples in the model creation as defined in (16), is significantly shorter than the interpolation length as used with LI, i.e., only 50% of the interpolation length, the superior artifact removal with SI is achieved by the modeling and identification of $FIR2$ and not skipping of the samples. This confirms the success of the artifact approximation as depicted in Fig. 2.

Apart from inspecting the determined EASSR latencies, the quality of the artifact removal can also be evaluated as follows.

Firstly, high EASSR amplitudes, especially when they differ between recording channels, may indicate insufficient artifact removal in the channels with higher amplitudes. An example can be seen in Fig. 4, where the EASSR amplitudes from contralateral recordings are similar for SI and LI. For the ipsilateral recordings, amplitudes determined with LI are higher and show higher inter-subject variability, which suggests that even with maximal interpolation length, the artifact cannot be fully removed. However, amplitudes determined with SI achieve similar values as the ones determined from contralateral recordings, indicating an equivalent and successful artifact removal.

Secondly, the number of significant responses after artifact removal in the low stimulation level recordings shows the effectiveness of the SI. For the $Tm10$ condition (Fig. 3(d)), the stimulation remains below the subject's T-level and hence no EASSRs are evoked. The significant component at f_m therefore only originates from the stimulation artifact. After perfect artifact removal, this component should be gone and only noise should remain. However, especially for artifact-dominated ipsilateral recordings, LI still leads to 17 significant results, indicating residual artifact. With SI, only 1 result contains significant residual artifact. For contralateral channels, SI is able to fully remove all stimulation artifacts from the recordings. This also applies to the $0pDR$ ipsilateral recordings, as seen in Fig. 3(c), where SI removes more artifact and hence leads to more non-significant results. In addition, at this stimulation level, EASSRs can still possibly be elicited, as is evident from the few significant results with latencies within the expected range. These results also indicate the success of the regularization. As, undoubtedly, a perfect artifact estimation is not possible, an unregularized OLS solution ($\alpha = 0$) would fit the EASSR model to the residual modeling error and hence lead to more false positive EASSR results. This is prevented through the high chosen value of $\alpha = 5000$, which removes almost all false positive EASSR results in the $Tm10$ condition and still allows accurate EASSR detection in the higher stimulation conditions ($100pDR$ and $50pDR$).

V. Future work

Future work with the SI should focus on the following points. Firstly, the currently used artifact model is only a linear approximation of the non-linear system, which causes the artifact in the recording. In recent years, the use of Neural Networks has drastically increased in non-linear system identification. It should therefore be investigated whether an approach using machine learning and a non-linear model can achieve similar or even better results. Secondly, the SI makes use of three distinct recordings with different modulation frequencies per stimulation level. In the pursuit of reducing expensive measurement time, the goal should be to use as few recordings as possible. However, using for instance only one modulation frequency recording per stimulation level leads to a simplified EASSR model (order $N_1 = 1$), from which the latency cannot be determined anymore. Merely a spectral analysis at the modulation frequency is performed, to determine the respective phase delay. Additionally, the accuracy of the artifact model would suffer from using fewer modulation frequencies. On the one hand, it remains to be investigated whether this is as successful as determining the latency. On the other hand, stimulation paradigms, using several input frequencies per recording at once, could be a promising solution to achieve both, good system identification and a reduction of recording time. Lastly, in an automated fitting procedure, where EASSRs are evoked at T-level, low SNRs become the dominant obstacle in objectively determining responses (assuming successful artifact removal).

VI. Conclusion

In this work, we refined and evaluated the new stimulation artifact removal method based on SI as introduced in [33]. As opposed to [33], we estimated the model parameters using OLS, which allows easy addition of a regularization term in the parameter estimation problem. This prevents false-positive EASSR detection in recordings with low stimulation levels.

We determined the EASSR latencies for 16 subjects and for 4 different stimulation levels. For the higher stimulation levels ($100pDR$ and $50pDR$), the main contribution of SI is seen with recordings from the ipsilateral EEG channels. Here, where the stimulation artifact likely exceeds the inter-pulse interval, LI fails to completely remove the artifact and hence determines lower EASSR latencies. SI, however, results in EASSR latencies within the expected range, which indicates proper modeling and separation of artifact and EASSR. Latencies for contralateral recordings are similar for both approaches.

For the lowest stimulation level of $0pDR$, only SI can reliably remove artifacts and detect responses in the recordings. Although this is only the case for six recordings, it is still very interesting, as for objective T-level determination, the stimulation level of the recordings is decreased, and the expected response amplitude is close to the noise level. The results indicate that SI deals better than LI with these conditions and is therefore an advantage when moving towards the automatic, objective fitting of CIs with EASSRs. The final clinical effectiveness of using EASSRs remains to be investigated but is not hindered by the stimulation artifacts of the CI anymore.

We can conclude that the separate modeling of EASSR and stimulation artifact through FIR systems using SI leads to successful EASSR detection in EEG recordings. SI even works with recordings with clinical stimulation settings of the CI and ipsilateral recording channels, which makes it the first method in the field to achieve this with EASSRs. This is of major significance, not only for objective threshold detection but also for lateralization studies or bilateral CI stimulation studies using this type of neural response.

VII. Acknowledgment

The authors would like to thank E. Verwaerde and W. David for the design of the experiments and collection of the data used in this paper.

References

[1] T. W. Picton, M. S. John, A. Dimitrijevic, and D. Purcell, "Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos," *International Journal of Audiology*, 2003.

[2] G. Rance, F. W. Rickards, L. T. Cohen, S. De Vidi, and G. M. Clark, "The automated prediction of hearing thresholds in sleeping subjects using auditory steady-state evoked potentials," *Ear and Hearing*, 1995.

[3] G. Rance and F. Rickards, "Prediction of hearing threshold in infants using auditory steady-state evoked potentials," *Journal of the American Academy of Audiology*, 2002.

[4] H. Luts, C. Desloovere, and J. Wouters, "Clinical application of dichotic multiple-stimulus auditory steady-state responses in high-risk newborns and young children," *Audiology and Neurotology*, 2006.

[5] J. Alaerts, H. Luts, B. Van Dun, C. Desloovere, and J. Wouters, "Latencies of auditory steady-state responses recorded in early infancy," *Audiology and Neurotology*, 2010.

[6] M. Van Eeckhoutte, J. Wouters, and T. Francart, "Auditory steady-state responses as neural correlates of loudness growth," *Hearing Research*, 2016.

[7] R. A. Dobie and M. J. Wilson, "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," *The Journal of the Acoustical Society of America*, 1996.

[8] M. Hofmann and J. Wouters, "Electrically evoked auditory steady state responses in cochlear implant users," *JARO - Journal of the Association for Research in Otolaryngology*, 2010.

[9] ——, "Improved electrically evoked auditory steady-state response thresholds in humans," *JARO - Journal of the Association for Research in Otolaryngology*, 2012.

[10] M. Van Eeckhoutte, J. Wouters, and T. Francart, "Electrically-evoked auditory steady-state responses as neural correlates of loudness growth in cochlear implant users," *Hearing Research*, vol. 358, pp. 22–29, 2018. [Online]. Available: https://doi.org/10.1016/j.heares.2017.12.002

[11] R. Luke, L. Van Deun, M. Hofmann, A. Van Wieringen, and J. Wouters, "Assessing temporal modulation sensitivity using electrically evoked auditory steady state responses," *Hearing Research*, 2015.

[12] R. Gransier, R. Luke, A. Van Wieringen, and J. Wouters, "Neural Modulation Transmission Is a Marker for Speech Perception in Noise in Cochlear Implant Users," *Ear and Hearing*, 2020.

[13] D. Mao, H. Innes-Brown, M. A. Petoe, Y. T. Wong, and C. M. McKay, "Fully objective hearing threshold estimation in cochlear implant users using phase-locking value growth functions," *Hearing Research*, 2019.

[14] C. M. McKay and N. Smale, "The relation between ECAP measurements and the effect of rate on behavioral thresholds in cochlear implant users," *Hearing Research*, vol. 346, pp. 62–70, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037859551630507X

[15] C. M. McKay, K. Chandan, I. Akhoun, C. Siciliano, and K. Kluk, "Can ECAP measures be used for totally objective programming of cochlear implants?" *JARO - Journal of the Association for Research in Otolaryngology*, 2013.

[16] A. S. Visram, H. Innes-Brown, W. El-Deredy, and C. M. McKay, "Cortical auditory evoked potentials as an objective measure of behavioral thresholds in cochlear implant users," *Hearing Research*, 2015.

[17] D. Távora-Vieira, A. Wedekind, E. Ffoulkes, M. Voola, and R. Marino, "Cortical auditory evoked potential in cochlear implant users: An objective method to improve speech perception," *PLoS ONE*, 2022.

[18] H. Deprez, R. Gransier, M. Hofmann, A. van Wieringen, J. Wouters, and M. Moonen, "Characterization of cochlear implant artifacts in electrically evoked auditory steady-state responses," *Biomedical Signal Processing and Control*, 2017.

[19] A. Bahmer, S. Pieper, and U. Baumann, "Evaluation of an artifact reduction strategy for electrically evoked auditory steady-state responses: Simulations and measurements," *Journal of Neuroscience Methods*, vol. 296, pp. 57–68, 2018. [Online]. Available: http://dx.doi.org/10.1016/j.jneumeth.2017.12.025

[20] B. Intartaglia, A. G. Zeitnouni, and A. Lehmann, "Recording EEG in cochlear implant users: Guidelines for experimental design and data analysis for optimizing signal quality and minimizing artifacts." *Journal of neuroscience methods*, vol. 375, p. 109592, jun 2022.

[21] R. Gransier, R. P. Carlyon, and J. Wouters, "Electrophysiological assessment of temporal envelope processing in cochlear implant users," *Scientific Reports*, vol. 10, no. 1, pp. 1–14, 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-72235-9

[22] H. Deprez, R. Gransier, M. Hofmann, A. Van Wieringen, J. Wouters, and M. Moonen, "Template Subtraction to Remove CI Stimulation Artifacts in Auditory Steady-State Responses in CI Subjects," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2017.

[23] N. Castañeda-Villa and C. J. James, "Objective source selection in Blind Source Separation of AEPs in children with Cochlear Implants." *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2007.

[24] ——, "The selection of optimal ICA algorithm parameters for robust AEP component estimates using 3 popular ICA algorithms," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology"*, 2008.

[25] ——, "Independent component analysis for auditory evoked potentials and cochlear implant artifact estimation," *IEEE Transactions on Biomedical Engineering*, 2011.

[26] P. M. Gilley, A. Sharma, M. Dorman, C. C. Finley, A. S. Panch, and K. Martin, "Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials," *Clinical Neurophysiology*, 2006.

[27] H. Deprez, M. Hofmann, A. Van Wieringen, J. Wouters, and M. Moonen, "Cochlear implant artifact rejection in electrically evoked auditory steady state responses," in *European Signal Processing Conference*. European Signal Processing Conference, EUSIPCO, nov 2014, pp. 1995–1999.

[28] H. Deprez, R. Gransier, M. Hofmann, A. Van Wieringen, J. Wouters, and M. Moonen, "Independent component analysis for cochlear implant artifacts attenuation from electrically evoked auditory steady-state response measurements," *Journal of Neural Engineering*, 2018.

[29] R. Alemi, S. Nozaradan, and A. Lehmann, "Free-Field Cortical Steady-State Evoked Potentials in Cochlear Implant Users," *Brain Topography*, 2021.

[30] J. A. Undurraga, L. Van Yper, M. Bance, D. McAlpine, and D. Vickers, "Characterizing Cochlear implant artefact removal from EEG recordings using a real human model," *MethodsX*, vol. 8, p. 101369, 2021. [Online]. Available: https://doi.org/10.1016/j.mex.2021.101369

[31] ——, "Neural encoding of spectro-temporal cues at slow and near speech-rate in cochlear implant users," *Hearing Research*, vol. 403, p. 108160, 2021. [Online]. Available: https://doi.org/10.1016/j.heares.2020.108160

[32] R. Luke and J. Wouters, "Kalman filter based estimation of auditory steady state response parameters," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2017.

[33] J. Schott, R. Gransier, J. Wouters, and M. Moonen, "Electrically evoked auditory steady state response detection in cochlear implant recipients using a system identification approach," *European Signal Processing Conference*, vol. 2022-Augus, pp. 1243–1247, 2022.

[34] T. W. Picton, A. Dimitrijevic, M. Sasha John, and P. Van Roon, "The use of phase in the detection of auditory steady-state responses," *Clinical Neurophysiology*, 2001.

[35] J. Schoukens, R. Pintelon, and Y. Rolain, "Identification of Nonlinear and Linear Systems, Similarities, Differences, Challenges." *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 122–124, 2006.

[36] A. Novak, L. Simon, and P. Lotton, "Extension of Generalized Hammerstein model to non-polynomial inputs," in *European Signal Processing Conference*, 2016.

[37] G. A. F. Seber and C. J. Wild, "9 - Least Squares," in *Statistical Methods for Physical Science*, ser. Methods in Experimental Physics, J. L. Stanford and S. B. Vardeman, Eds. Academic Press, 1994, vol. 28, pp. 245–281. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0076695X08602598

[38] R. Gransier, H. Deprez, M. Hofmann, M. Moonen, A. van Wieringen, and J. Wouters, "Auditory steady-state responses in cochlear implant users: Effect of modulation frequency and stimulation artifacts," *Hearing Research*, vol. 335, pp. 149–160, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.heares.2016.03.006

[39] R. Gransier, A. van Wieringen, and J. Wouters, "Binaural Interaction Effects of 30-50 Hz Auditory Steady State Responses," *Ear and hearing*, 2017.