bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Estimating Multiple Latencies in the Auditory System from Auditory Steady-State Responses on a Single EEG Channel

Lei Wang, Elisabeth Noordanus, and A. John van Opstal

Abstract

The latency of the auditory steady-state response (ASSR) may provide valuable information regarding the integrity of the auditory system, as it could potentially reveal the presence of multiple intracerebral sources. To estimate multiple latencies from recorded ASSRs, we propose a novel two-stage procedure that consists of a nonparametric estimation method, called apparent latency from phase coherence (ALPC), followed by a heuristic sequential forward selection algorithm (SFS). Compared with existing methods, ALPC requires few prior assumptions, and is straightforward to implement for higher-order nonlinear responses to multi-cosine sound complexes with their initial phases set to zero. It systematically evaluates the nonlinear components of the ASSRs by estimating multiple latencies, automatically identifies involved ASSR components, and reports a latency consistency index (LCI). To verify the proposed method, we performed simulations for several scenarios: two nonlinear subsystems with different or overlapping outputs. We also recorded the EEG from ten normal-hearing adults by bilaterally presenting superimposed tones with four frequencies that evoke a unique set of ASSRs. From these ASSRs, two major latencies were found to be stable across subjects on repeated measurement days. The two latencies are dominated by low-frequency (LF) (near 40 Hz, at around 41-52 ms) and high-frequency (HF) (>80 Hz, at around 21-27 ms) ASSR components. The frontal-central (FC) brain region showed longer latencies on LF components, but shorter latencies on HF components, when compared with temporal-lobe regions. In conclusion, the proposed nonparametric ALPC-SFS method, applied to zero-phase, multi-cosine sound complexes is more suitable for evaluating embedded nonlinear systems underlying ASSRs than existing methods. It may therefore be a promising objective measure for hearing performance and auditory cortex (dys)function.

Index Terms

EEG, ASSR, Latency, Auditory system, Apparent latency from phase coherence (ALPC), Multi-cosine stimuli.

I. Introduction

Auditory steady-state responses (ASSRs) are stable brain oscillations that are locked to the frequencies present in the periodic envelope of acoustic stimuli ([1] [2]). Studies have indicated that ASSRs may be generated at different levels within the auditory pathway, ranging from as early as the cochlear nerve and subcortical sources ([3], [4]), to the neocortex ([5], [6], [7], [8]). Because of their reproducibility, and involuntary nature, ASSRs have been considered a valid objective biomarker for auditory system disorders that feature abnormal sound processing, as well as for evaluating primary and non-primary auditory cortex function ([9], [10]). Its presence could relate to basic bottom-up sound processing ([11]), as well as to more complex cognitive skills, such as auditory learning ([12]), selective attention ([13],) speech and music perception ([7]), mental disorders ([14]), or illnesses such as tinnitus ([15], [9]). Moreover, the latency estimate that can be obtained from the ASSRs, as well as the associated interaural phase-modulation following responses (due to binaural interactions between different carrier frequencies from either ear), can serve as an objective measure of hearing performance, including binaural processing in the human brain ([16]). Finally, since ASSRs may contain signatures of the neural processing from auditory periphery to cortex, the electrically evoked ASSR from a cochlear implant (CI) could provide an objective measure for the responsiveness of different regions of the auditory pathway of hearing-impaired CI-users ([17]).

It should be noted that ASSRs are due to nonlinear mechanisms in the auditory system (see [18], for a review). At the earliest stages in the auditory pathway, the nonlinear character of neural sound encoding has been evaluated from auditory nerve recordings. A stimulus consisting of multiple pure tones (the carrier frequencies) was shown to result in an envelope spectrum that could be characterized by a unique series of 2^{nd}-order nonlinear difference frequencies (or 'beats') between any pair of carrier frequencies, which all showed up in the auditory nerve response. This indicates that these 2^{nd}-order nonlinear distortions already arise at, or before, the auditory nerve ([19]). In the ascending auditory pathway, such harmonic complexes might also evoke higher-order nonlinear components, which in turn could give rise to additional ASSRs. In particular, when two pure tones with frequencies f_1 and f_2, are passed through a nonlinear system with system order R, a series of combination tones is produced at the output, characterized by $nf_1 \pm mf_2$ (> 0), with n and m positive integers, such that $n+m \leq R$ ([15], [20]). So far, few studies have systematically evaluated such higher-order nonlinearities in the EEG responses of the human auditory system.

L. Wang, E. Noordanus, and A.J. van Opstal are with the Department of Biophysics, Donders Centre for Neuroscience, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands. E-mail: L.Wang@donders.ru.nl; E.Noordanus@donders.ru.nl; j.vanopstal@donders.ru.nl

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Symbol & term	Representation
α_{Σ_i}	current phase angle of output frequencies $\in [0, 2\pi]$
φ_{Σ_i}	initial phase angle (at t=0) of output frequencies $\in [0, 2\pi]$
φ_i^{in}	initial phase angle (at t=0) of input frequencies $\in [0, 2\pi]$
ω_{Σ_i}	output angular frequencies
ω_i^{in}	input angular frequencies
τ_e	estimate of real latency
τ_p	pseudo latency
Δ_t	the onset delay of a target EEG epoch
C_τ^i	latency consistency index (LCI) of i-th ASSR
C_τ^{sig}	95% confidence limit of LCI
d_ϕ	the slope of phase lags vs frequencies
ξ	relative amplitude gain ($\xi > 0$)
θ	threshold for PLV
$\phi(f_i)$	compute current phase of a frequency by FT
$\mathrm{mod}(\cdot)$	modulus function: remainder after division
PLV	phase lock value
MPE	mean phase error at the estimated latency $\in [0, 2]$
MSPC	multi-spectral phase coherence
ALPC	apparent latency from phase coherence
SFS	sequential forward selection
MB	monaural beat

Apart from these nonlinear distortion components, also the latency at which they occur in the EEG provides information of the underlying neural mechanisms, and could reveal the contribution from different intracerebral sources ([21], [6], [7]). Although the auditory system consists of both ascending and descending pathways ([5]), the ASSRs recorded at the scalp are assumed to be mainly due to the ascending system. Therefore, the longer the latency, the higher in the auditory system its generators ([1]).

In earlier studies, the latency of ASSRs has been estimated by methods such as the apparent latency ([1], [22]), or by group delay ([23]). Apparent latency is derived from the change of the response phase, $\Delta\phi$, with frequency (i.e., $\tau = \Delta\phi/2\pi f$), and is used as an indirect method to estimate neural response latency for steady-state stimuli, when a direct measurement is not feasible ([23]). This method, however, suffers from the inherent 2π phase ambiguity from the so-called wrap-around effect ([24]), and is only suitable for analyzing low-order ASSR components when the initial phases are known. The more recently-developed multi-spectral phase-coherence (MSPC) method ([20]; or similarly, the cross-spectral coherence (CSC) method, [25]) accounts for the phase-ambiguity problem, and allows evaluation of higher-order nonlinear response components. However, the MSPC method relies on prior assumptions regarding the underlying nonlinear systems: knowledge about the system orders is required for evaluating the expected nonlinear interactions, from which response latencies can be estimated. Therefore, the MSPC method is difficult to implement when the number of underlying systems and their orders are unknown (and thus, the nonlinear interactions are unknown), in combination with the dimensionality explosion of nonlinear interactions when multiple carrier frequencies are used. For given nonlinear systems and multi-tone sinusoidal inputs, a recursive method ([26]) can be used to predict the series of output frequencies. The number of output frequencies is a steeply increasing function of the system order R and the number of input frequencies, I, but its accurate computation is non-trivial because of potentially overlapping distortion products ([27]; see also Appendix C).

Thus, current state-of-the-art methods are limited in estimating multiple latencies for different nonlinear systems from the ASSRs. However, as indicated above, several studies have suggested that ASSR sources can be located at multiple cortical and subcortical regions ([6], [3]), which are likely associated with different latencies. For example, the two-component (brainstem-cortex) model suggests that a cortical generator will dominate low-frequency distortion products (around 45 Hz), whereas a brainstem generator would dominate the higher modulation frequencies (around 90 Hz) ([2], [28], [29]). As the measured EEG will typically contain the components from these multiple generators, a more general hybrid nonlinear model would read:

$$y(t) = \sum_{m=1}^{M} \sum_{r=1}^{R_m} \psi_{mr}(f) x_m^r(t - \tau_m) \tag{1}$$

in which the unknown parameters are the number of subsystems M, the subsystem order R_m, the subsystem latencies τ_m, as well as the potentially frequency-dependent gains, ψ_{mr}. Note that, in general, it will not be possible to uniquely identify each nonlinear system, when the underlying nonlinearities and their inputs are unknown. Yet, for certain cases, system-specific information can be obtained from the ASSRs. Figure 1 illustrates a potential simple scenario of such nonlinear models. The EEG is considered to reflect the (linear) sum of the outputs from multiple nonlinear subsystems (here taken, for simplicity, as homogeneous nonlinearities of different orders, R_m, and each at a different latency, τ_m). The output distortion frequencies of these subsystems will differ when they receive non-overlapping inputs (e.g., systems S_1 and S_2). In this case, the latencies of

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

both subsystems can be accurately estimated. In contrast, the output distortion products could be largely overlapping when the systems receive identical inputs, as in S_2 and S_M. In such a case, it will be much harder to estimate the underlying latencies. We simulated both cases in this paper.

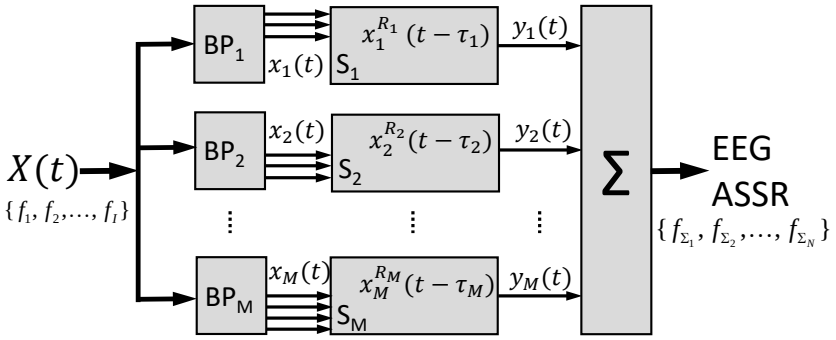

Fig. 1. Simplified scenario of M nonlinear subsystems that underlie the measured EEG response. The input signal $X(t)$ contains I frequencies. BP_m represent different band-pass filters, so that the subsystems (with order R_m and latency τ_m) may be driven by the same or by different input frequencies. The EEG contains the linear sum of all subsystem outputs $y_m(t)$, with N nonlinear distortion components f_{Σ_n} (ASSRs).

There are currently no suitable methods to deal with such models. We here propose and test a nonparametric, semi-heuristic method that requires little prior assumptions regarding the underlying EEG sources, and aims to objectively estimate multiple potential latencies from the ASSRs. We named it 'apparent latency from phase coherence' (ALPC), which is applied to a specifically designed stimulus complex, and implements a heuristic sequential forward selection (SFS) algorithm ([30]) to identify the components of each subsystem. Our method can estimate the contributions from more than one underlying source in the EEG, without the need of prior assumptions, like the number of underlying systems, their input frequencies, and their order. To test our method, we performed simulations for several scenarios, as illustrated in Fig. 1. We also recorded EEG responses from ten normal-hearing participants on two different days to superimposed tones that evoked ASSRs during passive listening.

II. Methods

To estimate multiple latencies from recorded ASSRs, we developed a two-stage procedure that consists of the nonparametric ALPC estimation method, followed by the heuristic SFS algorithm. The ALPC method can apply phase compensation on stimuli with arbitrary initial phases (as used in [20]), or it can use a much simpler time compensation (TC) procedure for a special stimulus complex, in which the initial phases of all stimulus frequencies, and thus their nonlinear distortion products, are set to zero. Note that the former requires prior information regarding the underlying nonlinear orders of systems and the corresponding nonlinear interaction (see Appendix A). In contrast, ALPC with TC requires no such information, and thus can be viewed as a nonparametric method (see the right-hand side of Fig. 3). We combined ALPC with SFS and TC to identify multiple latencies from the ASSRs.

A. Problem description

The EEG is considered as a time-series signal, S, containing a number of distortion frequency components $\{f_{\Sigma_1}, f_{\Sigma_2}, \ldots, f_{\Sigma_N}, \ldots\}$, with corresponding angular frequencies $\omega_{\Sigma_i} = 2\pi f_{\Sigma_i}$, which are the outputs of a nonlinear system. Assuming that N frequency components of them have a common, unknown, delay τ relative to their initial phase φ_{Σ_i} at t=0 (potentially having a common generator), we compute the current phase angle α_{Σ_i} of each component by taking the Fourier transform (FT), where $\alpha_{\Sigma_i}, \varphi_{\Sigma_i} \in [0,\ 2\pi]$. They follow:

$$\begin{aligned} \alpha_{\Sigma_1} + \tau\omega_{\Sigma_1} &= 2\pi n_1 + \varphi_{\Sigma_1} + \varepsilon_1 \\ \alpha_{\Sigma_2} + \tau\omega_{\Sigma_2} &= 2\pi n_2 + \varphi_{\Sigma_2} + \varepsilon_2 \\ &\vdots \\ \alpha_{\Sigma_N} + \tau\omega_{\Sigma_N} &= 2\pi n_N + \varphi_{\Sigma_N} + \varepsilon_N \end{aligned} \tag{2}$$

where n_i are unknown integers, i.e., the number of cycles that a component is delayed, and ε_i is the phase error (PE), caused by random noise.

To estimate the common latency, we minimize the mean-squared phase error across N frequencies ([22]):

$$\tau_e = \min_{\tau}\left(\frac{1}{N}\sum_{i=1}^{N} \operatorname{mod}\left(\frac{\alpha_{\Sigma_i} + \tau\omega_{\Sigma_i} - \varphi_{\Sigma_i}}{2\pi}\right)^2\right) \tag{3}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $\mathrm{mod}(\cdot)$ is the modulus between 0 and 2π. If the nonlinear system order, $R \leq 2$, the initial phases are straightforward to measure from the input stimulus by extracting the phases from the envelope of the input signal. However, when the systems order is unknown, and higher than three, this method will fail, because the unknown φ_{Σ_i} are endowed with the wrap-around problem (i.e., unknown n_i).

B. General framework of ALPC

First, to avoid that a discontinuous 'jump' (from 0 to 2π) would affect the averaging in (3), we represent the components of (2) by Euler's formula according to: $\exp\left(j(\alpha_{\Sigma_i} + \tau\omega_{\Sigma_i})\right) = \exp\left(2\pi n_i + \varphi_{\Sigma_i}\right) + \vec{\varepsilon}_i$ or, equivalently, $\exp\left(j(\alpha_{\Sigma_i} + \tau\omega_{\Sigma_i})\right) = \exp\left(j\varphi_{\Sigma_i}\right) + \vec{\varepsilon}_i$, where $\vec{\varepsilon}_i$ is a vectorized phase error (PE) and its length $|\vec{\varepsilon}_i| \in [0, 2]$. We then minimize the cost, defined by the mean length of PE (MPE) across N frequencies, to estimate τ_e:

$$\tau_e = \min_{\tau}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\exp\left(j(\alpha_{\Sigma_i} + \tau\omega_{\Sigma_i})\right) - \exp\left(j\varphi_{\Sigma_i}\right)\right|\right) \tag{4}$$

where $|\cdot|$ is the length of a complex vector. Using the estimated τ_e, we can compute n_i and the absolute phase lag of each frequency by $2\pi n_i + \varphi_{\Sigma_i} - \alpha_{\Sigma_i}$ (equal to $\tau_e\omega_{\Sigma_i} - \varepsilon_i$). Plotting the absolute phase lags for each frequency component as a function of frequency should yield a line with a slope d_ϕ, where the relationship between τ_e and d_ϕ follows the apparent latency rule ([1]):

$$\tau_e = \frac{d_\phi}{2\pi}. \tag{5}$$

In (4), we estimate τ_e by searching within a period T of signal S, where T is the smallest common multiple of the periods of frequency components $\{f_1, f_2, ..., f_N\}$. Then, the solutions for the latency are given by a set $\{\tau|\tau = \tau_e + mT\}$, with m an arbitrary integer. In a particular application, a *prior* for the possible range of latency values can be imposed to constrain a unique value for τ_e.

1) Time compensation (TC) for stimuli with fixed initial phases: To avoid the need for computing the unknown φ_{Σ_i} when using stimuli with arbitrary initial phases (which constitutes 'phase compensation', see Appendix A), we here propose to use a more practical stimulus complex: a multi-cosine stimulus, in which all initial phases are set to zero at t=0. With such a stimulus complex it is straightforward to apply time compensation (TC) for an arbitrary ongoing interval (starting at t>0) in the following way (see Fig. 2).

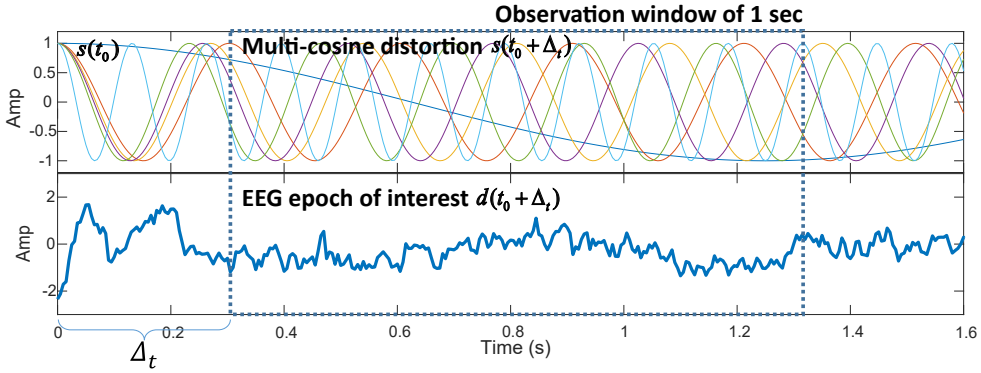

Fig. 2. Illustration of time compensation (TC) on a single-channel EEG. The stimulus complex consisted of a superposition of four cosine pure tones, at 461, 500, 504 and 537 Hz, starting at t=0 with initial phase 0. Here, the six distortion frequencies (i.e., the 2^{nd} order system products) in the top panel at [4, 33, 37, 39, 43, 76] Hz, are shown and time-scaled by a factor 0.1, for illustrative purposes. Note that all six distortions also start with zero phase at t=0 (because of the good property of cosine; see Appendix C). The EEG response to the stimulus complex is shown in the bottom panel (1^{st} trial on EEG electrode P8, subj#1) and has been down-sampled to 200 Hz for easy visualization. The first Δ_t = 0.3 s of the EEG signal is excluded from analysis because of the initial non-stationary ERP components (i.e., non-ASSR).

For a given multi-cosine input stimulus, $s(t) = \sum_{i=1}^{I}\cos(\omega_i^{in}t + \varphi_i^{in})$, with all initial phases $\varphi_i^{in} = 0$ at $t_0 = 0$, the ongoing stimulus interval, starting at $t_0 + \Delta_t$, is denoted as $s(t_0 + \Delta_t)$. The delayed EEG signal (the ASSR) $d(t)$ is considered to be the output of an unknown nonlinear system with an unknown latency, τ_e, i.e., $d(t) = F(s(t - \tau_e))$, where $F(\cdot)$ represents the unknown nonlinear mapping. The EEG epoch of interest will hence follow:

$$\begin{gathered} d(t_0 + \Delta_t) = F(s(t_0 + \Delta_t - \tau_e)) \triangleq F(s(t_0 - \tau_p)) \\ \text{with } \tau_p = \tau_e - \Delta_t \end{gathered} \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Therefore, from the ASSR we can estimate the pseudo-latency τ_p between $d(t_0+\Delta_t)$ and $s(t_0)$. Since $\varphi_i^{in}=0$, the corresponding initial phases φ_{Σ_i} of all high-order responses are also zero (i.e., $\varphi_{\Sigma_i}=\sum_{i=1}^{r} a_i\varphi_i^{in}\equiv 0$, see Appendix A and C). Moreover, it can be shown that for zero-phase sine inputs, all even-order difference interactions (e.g., f_n-f_m, and $2f_n-2f_m$) will yield the same initial phase of $\pi/2$ rad, so that the method can also be applied to these components (after changing the initial phase to $\pi/2$; see Appendix C). Therefore, from (4) and (6), we can estimate the pseudo latency τ_p. Finally, we apply TC to obtain the actual latency:

$$\tau_e=\tau_p+\Delta_t. \tag{7}$$

In our study, Δ_t= 0.3 s, in order to exclude the non-stationary event-related potentials (ERP) in the EEG after stimulus onset (see Fig. 2). The device jitter (resulting to be $<$ 0.5 ms) involved a small trigger jitter of the EEG recording system, and a fixed travel time of the acoustic signal through the silicon tube to the ears. Thus, the estimated latency had around 1-ms system error, which was deemed acceptable for the estimated ASSR latencies.

2) Latency consistency index (LCI): Considering that the estimated latency may vary due to additive noise on the nonstationary signals, we quantified its consistency over K time epochs. We defined the latency consistency index (LCI) for the i^{th} output frequency f_{Σ_i} as C_τ^i.

$$\begin{aligned} C_\tau^i &= |\tfrac{1}{K}\sum_{k=1}^{K}\exp(j\omega_{\Sigma_i}\tau_{e,k})| \\ &= |\tfrac{1}{K}\sum_{k=1}^{K}\exp\left(j(\varphi_{\Sigma_i}-\alpha_{\Sigma_i})_k\right)| \end{aligned} \tag{8}$$

where $\tau_{e,k}$ is the estimated latency for frequency component f_{Σ_i} on the k^{th} epoch, and it follows $\omega_{\Sigma_i}\tau_{e,k}=2\pi n_i+\varphi_{\Sigma_i}-\alpha_{\Sigma_i}$ from (2), assuming Gaussian noise with zero mean. The LCI is hence mathematically equivalent to the phase-coupling strength used in MSPC ([20]), which shows that the higher the phase-coupling strength between inputs and outputs, the more stable the estimated latency. We selected the duration of each time epoch (without overlap) as multiples of the periods of the ASSR frequencies (e.g., integer seconds, given integer ASSR frequencies) so that the initial phase φ_{Σ_i} remains the same across epochs. In this case, $\varphi_{\Sigma_i}\equiv 0$ when the cosine carriers have their initial phases set to zero. Thus, the LCI no longer relies on the stimulus frequencies, and reads:

$$C_\tau^i=|\frac{1}{K}\sum_{k=1}^{K}\exp\left(j\alpha_{\Sigma_i}\right)_k| \tag{9}$$

C_τ^i can vary between 0 and 1. If the latency is perfectly stable over time epochs, $C_\tau^i=1$. Otherwise, if no stable latency exists, C_τ^i is statistically indistinguishable from zero. Therefore, LCI can be used to determine if an ASSR component can be used to estimate latencies. The theoretical threshold for significance is $C_\tau^{sig}=\sqrt{\frac{3}{K}}$ ([31]).

The null hypothesis is that the phase of an ASSR component is randomly distributed in the interval $[0,\ 2\pi]$. If C_τ^i exceeds C_τ^{sig}, the null hypothesis is rejected, and the ASSR component is considered to be significant. C_τ^{sig} approximates the 95% confidence limit, as verified by Monte Carlo simulation ([20]). In our analysis, only significant ASSR components were considered for estimating latencies.

C. ALPC with sequential forward selection (SFS)

A particular set of ASSR components may have a common latency that lead to a low cost of the MPE in (4). In addition, the set of ASSR components may contain several subsets that correspond to different latencies, which would be indicative for different underlying sources. To automatically identify a frequency subset from all significant ASSR components that belong to a common latency, we applied the heuristic SFS algorithm. SFS is commonly used as an alternative to a greedy algorithm, to reduce computational complexity. SFS starts with a frequency (or frequency pair), and then sequentially adds another frequency (from the remaining set) that results in the smallest increase of the cost (i.e., MPE in a range of [0 - 2]). The details of the SFS algorithm are described in ([30]) and ([32]).

The termination rule of SFS should be carefully chosen, depending on the data at hand. Here, we used the following procedure. When starting from a single frequency, SFS continued as long as the following two criteria were met: (i) the increase of the MPE of the current step remained $<$ 0.1, (ii) the current MPE $<$ 0.5. Otherwise, SFS terminates. If SFS started from a pair of frequencies, it continued when the following three criteria were met: (i) the difference of estimated latencies between the current and the previous step $<$5 ms, (ii) the step increase of MPE $<$ 0.1, (iii) the current MPE $<$ 0.5. Otherwise, SFS terminates. These criteria may be further tuned for the particular EEG data set.

Figure 3 summarizes the pipeline for estimating the apparent latency from the EEG by the MSPC and ALPC-SFS methods, respectively. Depending on the initial phases of the carriers (either arbitrary, or all zero), one can use either a parametric (the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

left-hand side, MSPC) or a non-parametric (the right-hand side; ALPC-SFS) approach to estimate the underlying latency. For the parametric approach, phase compensation (see Appendix A) can be used to estimate the unknown initial phases of the ASSRs, which are fed to the MSPC algorithm to estimate the dominant latency. Note that the phase compensation method requires prior assumptions regarding the nonlinear system. For the non-parametric approach as proposed in the present paper, ALPC with SFS can estimate multiple potential latencies, by selecting the involved ASSR frequencies for each subsystem.

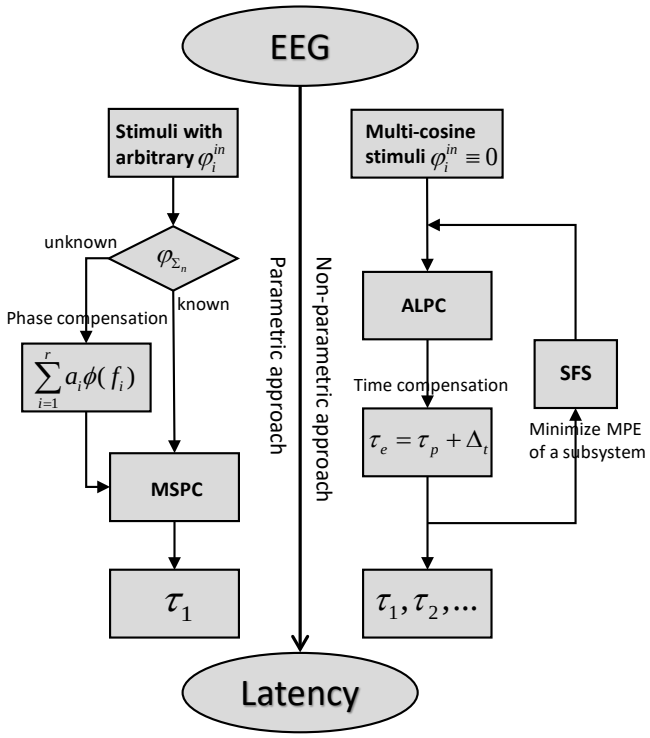

Fig. 3. Pipeline for estimating the apparent latency from EEG, using the parametric (left) or non-parametric (right) approach. φ_i^{in} represents the initial phase of each frequency component in the input stimuli, and φ_{Σ_n} represents the initial phase of each ASSR frequency component (output). By using multi-cosine carriers with their initial phases set to zero, the ALPC-SFS method (right) avoids the need for phase compensation of the MSPC method (left; see Appendix A), which also requires prior assumptions of the underlying nonlinearities. Moreover, the former can identify potentially multiple latencies instead of only one dominant latency.

III. EEG EXPERIMENTS

A. Experimental design and stimuli

We performed EEG recording experiments on 10 normal-hearing adults (1 female, age 29 $\pm$ 13 y) (hearing thresholds $<$20 dB HL). Each subject was measured twice, using the same stimuli on different days (i.e., $>$48 hours interval). The stimulus at the left ear was a superposition of two tones of 461 Hz and 500 Hz at 81 dB SPL, and two tones of 504 Hz and 537 Hz at 56 dB SPL. The stimuli at the right ear consisted of the reverse pattern: two loud tones of 504 Hz and 537 Hz at 81 dB SPL, and two weaker tones of 461 Hz and 500 Hz at 56 dB SPL. The stimuli were fed into both ears simultaneously. The four frequencies were chosen such that the difference frequencies and higher-order distortions were all unique.

In the EEG experiments described here, all input tones were sines with their initial phases at t=0 set to zero. In Appendix C we demonstrate that such carriers will yield irregular phase-shifts for all higher-order distortion products, except for the even-order difference interactions (e.g., $f_1 - f_2$, and $2f_1 - 2f_2$). The initial phase of these even-order distortion products will remain the same. Thus, when only even-order difference interactions in the ASSRs are considered (as in this study), APLC-SFS with TC can also be applied for zero-phase sine stimuli.

In each trial, stimuli with a duration of 12.3 s were played after a silent break with a random duration between 2 and 3 sec. The first 0.3 s of the EEG of each trial was excluded from analysis to avoid the contributions from phasic event-related potentials (i.e., non-ASSRs). The remaining 12 s of the EEG signals allow for an accurate estimate of the neighboring SNR (see section E). We repeated 100 trials while recording the scalp EEG with a 64-channel cap with Ag/AgCl electrodes, with the listener sitting in an anechoic chamber, watching a silent video. The EEG signals were digitized at a sampling rate of 2000 Hz by a Refa amplifier system (TMSi, Twente Medical Systems International B.V., the Netherlands), and stored for offline analysis. At the start of each recording session, the impedances of all EEG electrodes were carefully set to values below 10 $k\Omega$ to ensure good contacts between electrodes and scalp. During the session, these values were regularly checked, and adjusted

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when deemed necessary. Stimuli were generated by TDT (Tucker-Davis Technologies, USA) system 3 hardware, and presented through ER3C insert earphones (Etymotic Research, Elk GroveVillage, IL), which were connected to the participants ears via 30 cm-long plastic tubes and foam earplugs.

B. Preprocessing of EEG

We applied a three-stage preprocessing protocol on the raw EEG recordings, after first removing the three EEG electrodes above the eyes (FP1, FPz, and FP2) because they were contaminated by eye-blinks and eye-movement artifacts. First, the raw EEG signals were re-referenced to a common average reference (CAR) to reduce the noise from the grounding electrode ([33]). Compared with re-reference in relation to a certain EEG electrode that may not be 'neutral' to ASSRs, the CAR montage will keep the phases of ASSRs unchanged on most EEG electrodes. Second, each EEG channel was filtered with a zero-phase shift filter ([34]) that consisted of a 10th-order Butterworth high-pass filter with a cutoff frequency of 1 Hz, and a notch filter at 50 Hz to remove the line noise. Third, to exclude severely contaminated EEG trials from further analysis (mainly by EMG artifacts, which may cause false detections or decreased SNRs of ASSR components), we developed a method to automatically exclude bad EEG trials. To that end, we computed the mean and the variance of the peak-to-peak amplitude range as indicators on each EEG trial. We excluded EEG trials whose indicators were positive outliers on 100 trials, i.e., samples $>$ Q3+1.5*IQR. (Interquartile range (IQR) = upper quartile (Q3) $-$ lower quartile (Q1)). In this way, up to 5% of all trials were automatically excluded from further analysis.

C. Epoch filtering by phase lock value (PLV)

We tested whether excluding poorly phase-locked epochs would improve the accuracy of the phase estimates, and latencies, for the frequencies of interest. To that end, we divided each trial into twelve one-second epochs, and computed their phase-lock value (PLV $\in [0, 1]$) to each frequency of interest ([35]). Epochs with a PLV below a selected threshold, θ (taken as θ = 0, 0.4, 0.6, or 0.8) were excluded. Finally, we computed the phase values on the remaining epochs by using two different averaging methods: AVG EEG, and AVG phase (described below).

D. Phase extraction of EEG

The accuracy of the extracted phases of the frequency components is important for estimating latency. Previous studies often average the EEG signals across a number of trials to reduce the background noise, so that ASSRs have a higher SNR ([36]). Here, we employed two different methods, 'AVG EEG' and 'AVG phase' to extract the phase of a target frequency. The 'AVG EEG' method averages filtered EEG signals of all trials and subsequently computes the phases of the average EEG. The phase values from this method tend to be dominated by epochs with strong ASSRs (i.e., with large amplitudes). Instead, the 'AVG phase' method averages the extracted phases from all EEG epochs with a duration that is an integer multiple of the signal period (1 sec in this study), without considering amplitude information (as used in [20]). For this method, all EEG epochs thus contributed with equal weight to the estimate. For comparison, we used both methods to extract the phases on each EEG channel independently.

E. SNR on target frequencies

It is difficult to separate the signals generated by ASSRs from background noise in the EEG on a single-trial basis. However, after averaging the EEG across a number of trials (100 trials in this study), the power of phase-locked frequencies is enhanced, whereas the power of the background EEG is suppressed due to random phases. To quantify the SNR of each ASSR frequency component, we defined the neighboring SNR as ([3]):

$$SNR(f_t) = 10\log 10(\frac{P(f_t)}{\frac{1}{N}\sum_{i}^{N} P(f_i)}) \tag{10}$$

where $P(f_t)$ is the power of the target frequency bin f_t. The f_i is a neighboring frequency bin within a small range ($\pm$0.5 Hz) of f_t. For a 12 s EEG trial, the width of a frequency bin is 1/12 Hz and N=12. Compared with the frequency spectrum of the EEG, the neighboring SNR is robust to the pink noise present in EEG signals (see Fig. 9).

IV. Results

A. Simulation results

To illustrate the potential of ALPC with SFS in combination with zero-phase cosine stimuli, we simulated different models and scenarios (cf. Fig. 1). In these examples, two parallel nonlinear sub-systems with different delays could have different, or (partially) shared inputs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1) Example 1: Two pure second-order subsystems with non-overlapping inputs: Figure 4 shows the simulation results of our ALPC-SFS algorithm for the simple example of Eq. 11. The total input signal, $x(t)$, consisted of five frequency components: (17, 21, 27, 41, 49) Hz, where x_1=(17, 21, 27) Hz were fed to a pure second-order subsystem y_1, and x_2=(41, 49) Hz to a similar subsystem y_2. In this model, system y_1 generated the distortion products $\{2f_i, f_i \pm f_j\} = \{4, 6, 10, 34, 38, 42, 44, 48, 54\}$ Hz, and system y_2 yielded the non-overlapping components $\{8, 82, 90, 98\}$ Hz. Subsystems y_1 and y_2 had different latencies τ_1 = 51 ms and τ_2 = 21 ms, respectively. Y is the combined output signal, (representing the EEG) with additive Gaussian white noise (WGN), $\varepsilon(t)$. We added noise with different variances, such that the SNRs were 5, 0, -5, -10, -15, and -20 (dB) (using MATLAB's function 'awgn').

$$\begin{aligned} y_1(t) &= {x_1}^2(t-\tau_1) \\ y_2(t) &= {x_2}^2(t-\tau_2) \\ Y(t) &= y_1(t) + y_2(t) + \varepsilon(t) \end{aligned} \tag{11}$$

To estimate the unknown latencies, we assume that only the total input (stimulus) and output (EEG) are available, but that the information regarding the subsystems (e.g., input frequencies of each subsystem, and system orders) is unknown. Figure 4 shows that our ALPC-SFS method separated the mixed outputs into the appropriate contributions from the two underlying subsystems, and correctly estimated their latencies. Here, we took for the additional signal delay, $\Delta_t = 0$, for simplicity, and signal Y had an SNR of 5 dB. Figure 4a shows the input and output signals, and their spectrum. Figure 4b illustrates the SFS procedure, starting from a randomly selected frequency component (here: 38 Hz) from the output frequency set (i.e., those frequencies whose power are higher than the background). The upper panels show the mean phase errors (MPEs) of

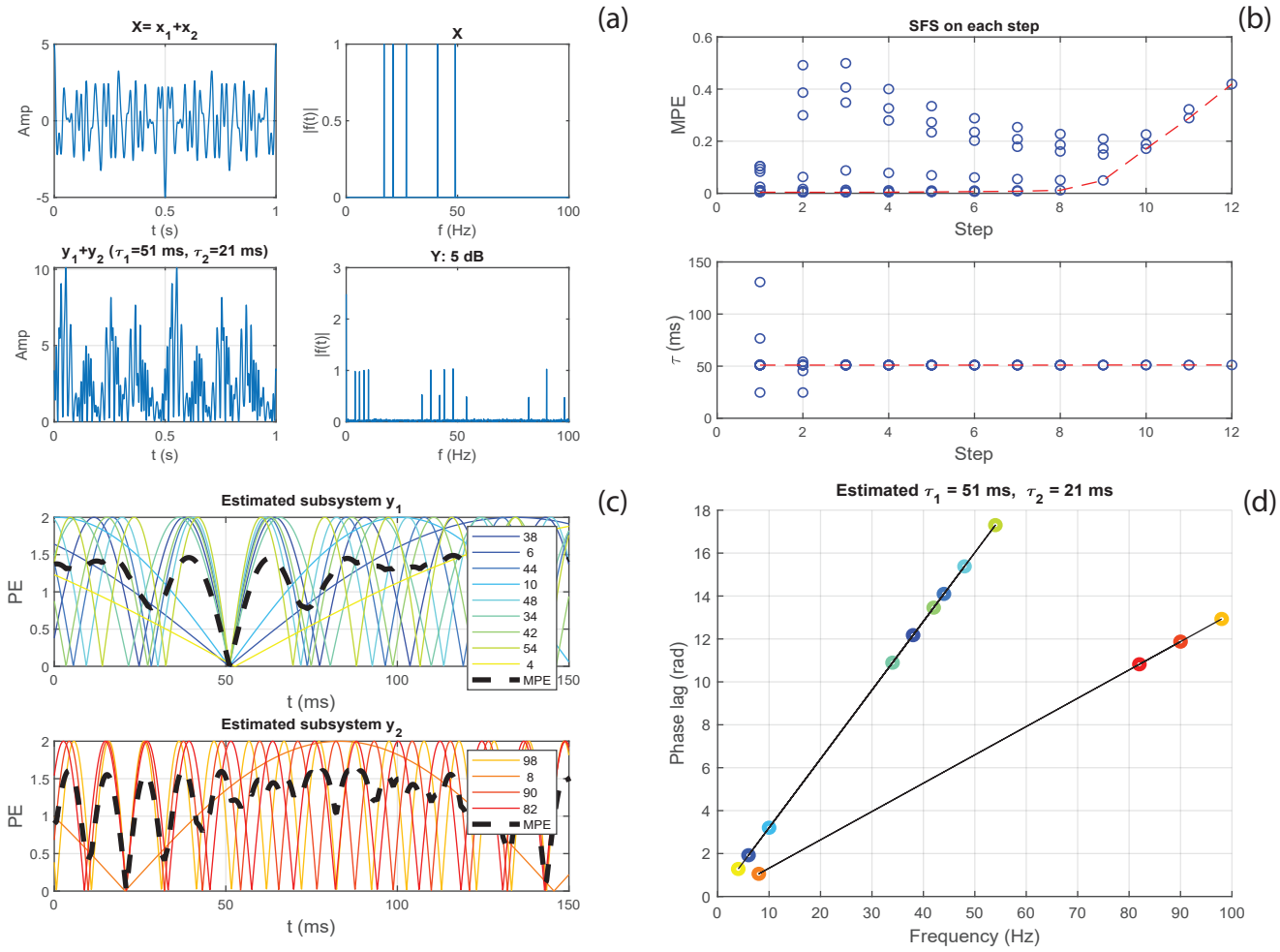

Fig. 4. Simulation of the ALPC-SFS method for Eqn. 11. The two subsystems were successfully separated from the mixed nonlinear response, and their latencies were estimated correctly. (a) Input signal X (sum of five cosine signals with zero phases) and their mixed total response, Y. (b) SFS without applying the termination criterion. Each circle represents a candidate frequency and the dashed line crosses the candidate frequencies with the minimum MPE for each step (top panel) and their corresponding latencies (bottom panel). (c) Phase error (PE, in radians) of each selected frequency over time. The MPE (bold dashed line) reaches its minimum (close to 0) at the estimated latency. (d) Plot of phase lags determined from the estimated latency. The slopes of the best-fit lines correspond to the latencies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

all candidate frequencies on each step in the SFS, the bottom panel shows the corresponding estimated latencies of candidate frequencies at each step. The MPE started to rise significantly after step 8, where a termination criterion of SFS came into action (see Methods). Thus, the first nine (8+1) frequencies were selected, yielding an estimated latency of 51 ms. Subsequently, SFS was performed on the remaining frequency set, which gave an estimated latency for a second subsystem of 21 ms. Without SFS, an estimate based on all frequencies would result in only one dominant latency (51 ms), as shown for step 12 (Fig. 4b).

The relationship between SNR and accuracy of the estimated latencies is provided in Table A1 of Appendix B.

We also considered a scenario for which Δ_t was not zero. We first estimated the pseudo latencies and applied TC to determine the actual latencies. The pseudo latency uses the property that the initial phases of the nonlinear distortions of zero-phase cosine inputs remain the same (see Appendix C for details). In this way, we do not need to apply phase compensation. However, when Δ_t exceeds the actual latency, the corresponding pseudo latency will be negative (Eq. 6), yielding a negative slope for the phase-lag vs. frequency relation (See Fig. 5).

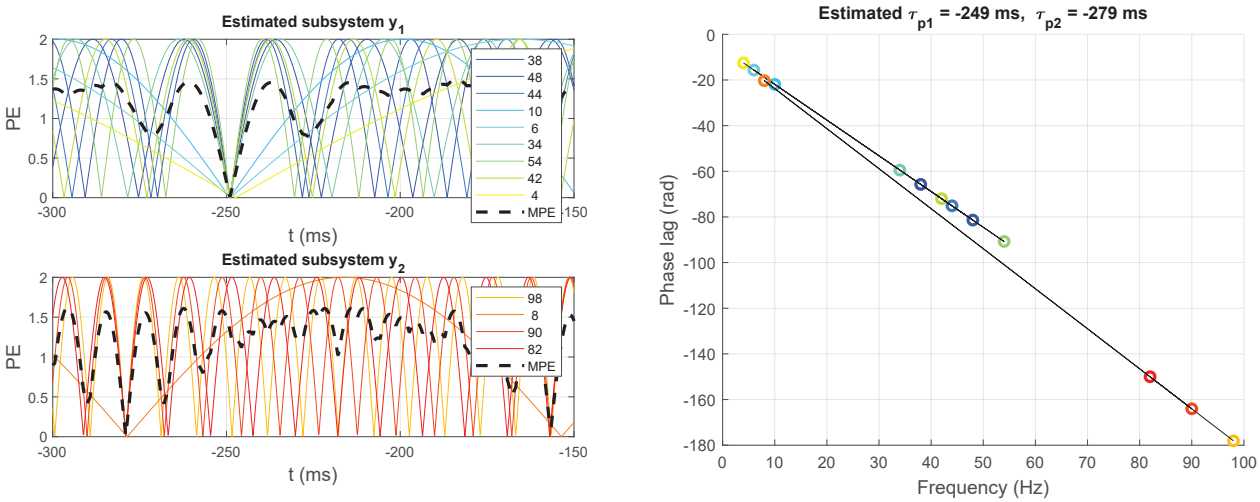

Fig. 5. The process of estimating the pseudo latency when the stimuli are delayed (here: Δ_t = 0.3 s) as shown in Fig. 2. After applying TC (Eq. 7), the estimated latencies are correctly estimated as 51 and 21 ms, respectively. Notice that the negative slopes in the right-hand panel correspond to the negative pseudo latencies of -249 and -279 ms, respectively (see Methods).

For comparison, we also applied the MSPC method ([20]) to this simple model. Also this method estimates a system latency by minimizing the MPE over a time window. It requires the total input (stimuli) and output (EEG), and has to make an assumption about the underlying system orders. In case of a correct selection of inputs and outputs and system-order information, MSPC may perform equally well as ALPC-SFS (Fig. 3). However, when the sub-systems inputs and outputs are not known a priori, MSPC will fail to correctly identify the systems and make a biased estimate. Figure 6 demonstrates this for the same settings as in Fig. 4. The MSPC method estimated only one biased latency (close to the dominant one) for the

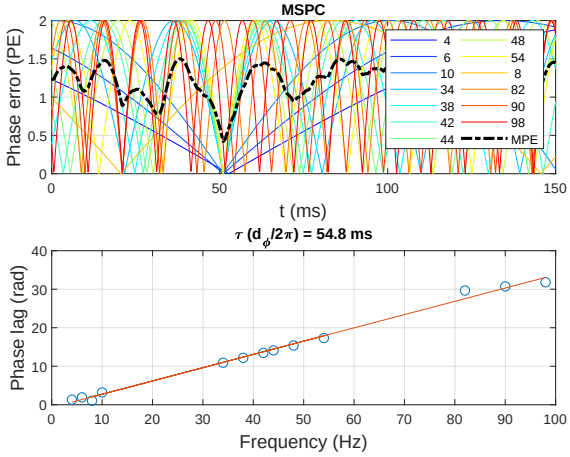

Fig. 6. The MSPC method estimates a system's latency for the same input X and output Y as in Fig. 4. It assumes that the system order equals two. It estimates one biased latency of 54.8 ms with its corresponding MPE of 0.42 (MPE $\in [0,\ 2]$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

assumed 2^{nd}-order system. However, it failed to find the second hidden subsystem, which produced fewer nonlinear distortion products.

2) Example 2: Models containing multiple system orders.: We also applied both methods to a more complex model, in which each subsystem contained a second- and a third-order nonlinearity:

$$\begin{aligned} y_1(t) &= x_1^2(t-\tau_1) + x_1^3(t-\tau_1) \\ y_2(t) &= x_2^2(t-\tau_2) + x_2^3(t-\tau_2) \\ Y(t) &= y_1(t) + y_2(t) + \varepsilon(t) \end{aligned} \tag{12}$$

The total input signal $X(t)$ consisted of four frequency components: {37, 43, 38, 46} Hz, where x_1=(37, 43) Hz were fed to subsystem y_1, and x_2=(38, 46) Hz to subsystem y_2. In this example, subsystem y_1 generated the following distortion product set: {6, 74, 80, 86, 31, 37, 43, 49, 111, 117, 123, 129} Hz, and subsystem y_2 yielded the non-overlapping set: {8, 76, 84, 92, 30, 38, 46, 54, 114, 122, 130, 138} Hz. For both frequency sets, the first four entries are 2^{nd}-order distortion products and the remaining components are 3^{rd}-order outputs. Subsystems y_1 and y_2 had latencies τ_1 = 51 ms and τ_2 = 21 ms, respectively. Y is the combined output signal, representing the EEG.

To estimate the latencies, we assumed that only the total input (X) and output (Y) were available, and that there was no prior information about the subsystems. Figure 7 shows the estimated latencies for the ALPC-SFS and MSPC methods. In this example, signal Y had an SNR of 5 dB. Without knowing the information of both subsystems y_1 and y_2, ALPC-SFS correctly estimated the latencies and separated the two original subsystems. In contrast, the MSPC method relied on an explicit assumption regarding the order of potential subsystems. However, as in this example, within a given system order, not all distortions are associated with a common latency. As a result, the MSPC method makes a biased estimate in this case.

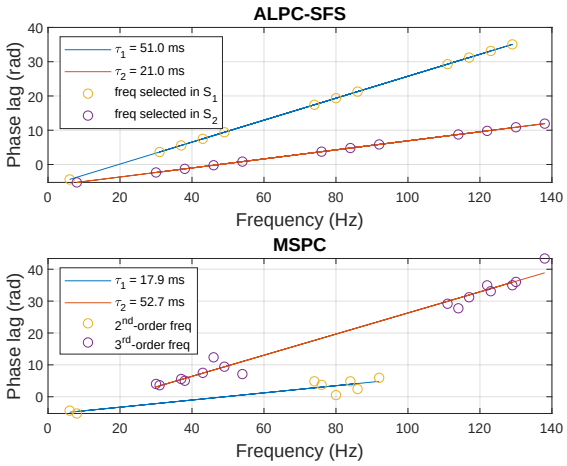

Fig. 7. Estimate latencies when subsystems (y_1 and y_2) are unknown. (Top) by using ALPC and SFS on Y, two latencies (51 and 21 ms) are correctly estimated with corresponding MPEs less than 0.001 and with correctly identified subsystems ($S_1 = y_1$ and $S_2 = y_2$). Here, SFS starts from a randomly selected distortion frequency, 6 Hz, and the same termination rule with scenario 1 is applied here. (Bottom) MSPC estimates latencies for the assumed 2^{nd}-order and 3^{rd}-order subsystems from X and Y. The estimated latencies are 17.9 ms and 52.7 ms with corresponding MPEs of 0.78 and 0.66 for the 2^{nd}-order and 3^{rd}-order subsystems, respectively.

3) Example 3: two mixed subsystems with identical outputs.: We also simulated the more challenging scenario in which the two subsystems had overlapping output components. In this case, only distorted phases can be extracted from the mixed output signals, and the relative amplitude gain of the output components (related to ψ_{mr} in (1)) results to determine the estimated latencies. The stimulated model reads:

$$\begin{aligned} y_1(t) &= {x_1}^2(t-\tau_1) \\ y_2(t) &= {x_2}^2(t-\tau_2) \\ Y(t) &= \xi y_1(t) + y_2(t) + \varepsilon(t) \end{aligned} \tag{13}$$

where the relative amplitude gain $\xi > 0$, and the latencies, $\tau_1 = 15$, $\tau_2 = 20$ ms. Both y_1 and y_2 had identical input frequencies (17, 21, 27) Hz so that they generated nine identical 2^{nd}-order output components: {4, 6, 10, 34, 38, 42, 44, 48, 54} Hz. We varied the relative gain $\xi = [\frac{1}{100}, \frac{1}{10}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{1}{\sqrt{2}}, 1, \sqrt{2}, 2, 3, 4, 6, 10, 100]$, and simulated the mixed signals Y without noise, and under SNRs of (5, 0, -5, -10, -15, -20) dB. The latency estimates as function of the relative gain are shown in Fig. 8.

The simulations show that in case of total overlap of the system outputs, the estimated latency is a weighted average of the underlying true latencies, where the relative gain of the systems serves as the weighting factor. After normalizing the estimated latencies (on the pure, noiseless signal) to a range between 0 and 1, we fitted the estimates with a sigmoid function

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

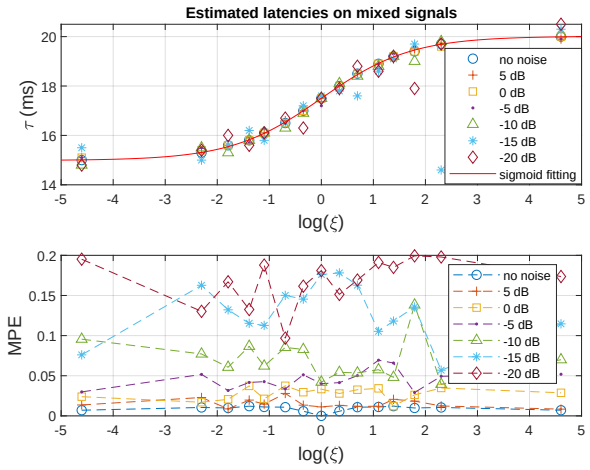

Fig. 8. (Top) Estimated latencies on mixed signals of two subsystems with overlapping output frequencies and varying amplitude scale ratio ξ. The estimated latencies from mixed signals (on the noise-free signals) fit a sigmoid function between two original latencies. (Bottom) The mean phase error (MPE) tends to increase with decreasing SNR. In all cases, SFS started with 10 Hz.

$f(x) = \frac{a}{1+e^{(-bx)}}$, with x the natural logarithm of the relative gain, and fit coefficients a=1.00 and b=1.17, with 95% confidence bounds (0.9944, 1.007) and (1.143, 1.196), respectively. In general, the estimated latency for the superposition of the two system outputs with latencies ($\tau_1 \leq \tau_2$) follows: $\tau_{mix} = \frac{(\tau_2 - \tau_1)}{1+e^{(-1.17\log(\xi))}} + \tau_1$, which is equivalent to,

$$\tau_{mix} = \frac{1}{1+\xi^{(-1.17)}}(\tau_2 - \tau_1) + \tau_1 \tag{14}$$

On the signal without noise, each selected starting frequency of the SFS yielded the same results for latency, MPE, and selected frequency components (i.e., output components). On the signal with noise, selecting a different starting frequency for the SFS yielded similar results for latency and MPE (as shown in Fig. 8), but could slightly differ regarding the selected frequency components, because frequency components with a large phase shift (due to low SNRs) are not selected by SFS.

B. Measured ASSR components in the EEG

Table I shows the nonlinear interactions between the four stimulus input frequencies up to 6^{th} order, and the corresponding frequencies that could potentially be present in the measured ASSR. The odd-order distortion products (around 500 Hz and higher; e.g., $2f_1 - f_2$) fall above the pre-processed EEG bandwidth (<200 Hz). In the measured EEG, the theoretical ASSRs all exceeded the background (i.e., non-ASSR integer frequency components) of 6.3 ($\pm$0.6) dB (see Fig. 9).

TABLE I
ASSR COMPONENTS GENERATED BY THE NONLINEAR INTERACTION

Order	Nonlinear interaction	Output frequencies (Hz)
2^{nd}	$\lvert f_i - f_j \rvert,\ i \neq j$ $i, j \in \{1, 2, 3, 4\}$	33, 39; 4, 37, 43, 76
4^{th}	$2MB_{1/2}$, $\lvert MB_1 \pm MB_2 \rvert$	6, 72, 78, 66
6^{th}	$3MB_{1/2}$, $\lvert 2MB_{1/2} \pm MB_{2/1} \rvert$	99, 117, 45, 27, 111, 105

The four stimulus tones were sinusoids of f_1 =461, f_2= 500, f_3 =504, f_4 =537 Hz, with zero initial phases. The principal monaural beat frequencies MB_1 (33 Hz) and MB_2 (39 Hz) were generated by the two louder tones in each ear (see Methods). We here show a subset (<200 Hz) of the theoretical output frequencies for each system order.

In the ASSR components, two predominant monaural beat (MB) frequencies (i.e., 33, 39 Hz) were generated by the second-order difference of the two louder pure tones in each ear. Their full 2^{nd}-order interactions (including harmonics and their inter-modulations, i.e., a subset of 4^{th}-order interactions of the input frequencies) and 3^{rd}-order interactions (i.e., a subset of 6^{th}-order interactions of the input frequencies) also show high SNRs within the EEG's effective bandwidth of 200 Hz. Note that in addition to the 4^{th}-order output frequencies shown in Table I, more 4^{th}–order output frequencies could potentially be generated among the full set of 2^{nd}-order output frequencies, e.g., 29 and 35 Hz. Such output frequencies, however, were not considered in this study because of much lower SNRs.

The EEG channel with the third-largest SNR value out of all 61 channels represents the median value of the highest five channels, and avoids the potential detrimental effect of EMG artifacts. We thus averaged the top-3^{rd} SNRs on each integer

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

frequency across ten subjects, measured on both recording days. The measurements of day 1 and day 2 showed highly similar SNR values. The grand average (n=20, day 1 and 2) SNR across all 61 EEG channels showed that ASSRs tended to have their maximum activation at both temporal lobes and frontal-central regions (bottom panel of Fig. 9).

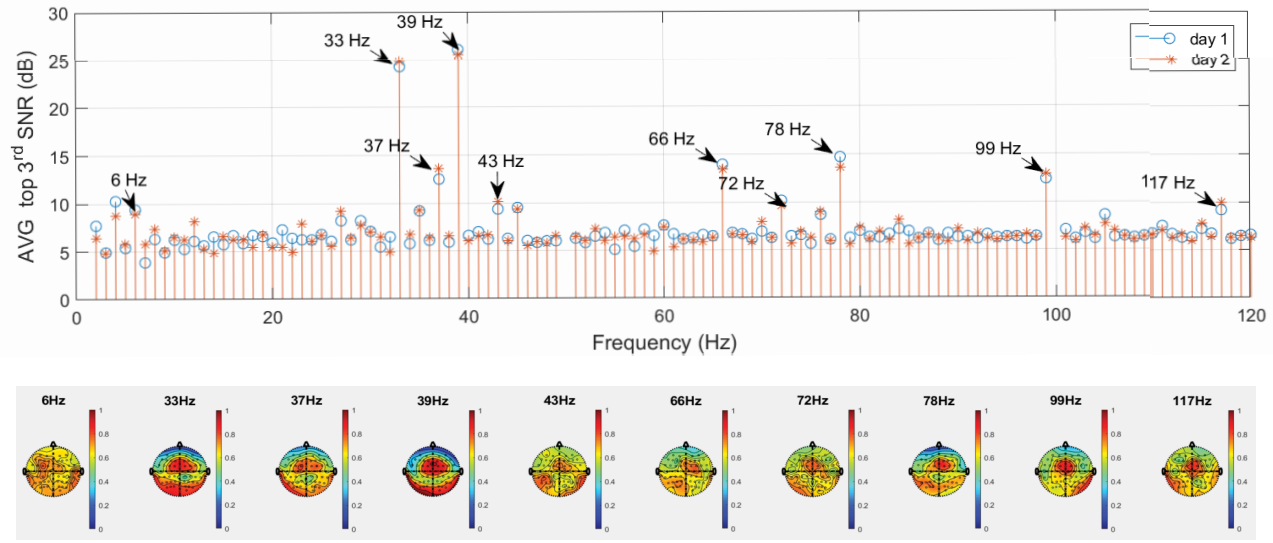

Fig. 9. Grand average SNRs (n=10 subjects on days 1 and 2, top panel) and grand averaging scalp topography (n=20 measurements over two days, bottom panel) on ten major ASSR components indicated by arrows. The third-largest SNR (out of 61 channels) on each integer frequency was averaged across 10 subjects measured on day 1 or day 2. To calculate the grand average of the scalp topography, the SNR of each measurement was normalized between 0 and 1 to enable a direct comparison of each subject's contribution.

C. *Estimated latencies from ASSRs*

The MB frequencies showed a higher SNR (or LCI) than the other ASSR components. Therefore, in applying ALPC-SFS, we estimated latencies by starting with a pair of MB frequencies (i.e., 33 and 39 Hz). The MB frequency pair can determine a search direction for SFS, so that it finds the involved ASSR components that have a common latency. Similarly, we evaluated the underlying latencies dominated by higher frequencies (HF) (>80 Hz) by using SFS starting with the 3MB frequencies (i.e., 99 and 117 Hz). The ten subjects yielded similar results for the measurements on both days. The results of day 2 are shown in Fig. 10, where the phase extraction was performed by taking the 'AVG EEG' procedure (see Methods).

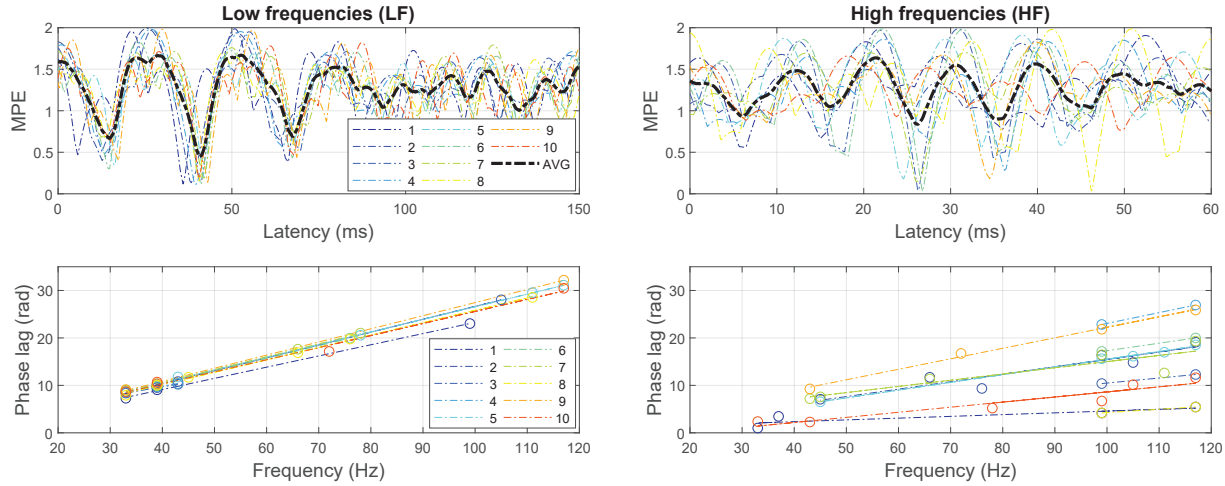

Fig. 10. Estimated latencies of 10 subjects on EEG channel P8. The left-hand panel shows the latencies determined when SFS started with the MB frequencies; the right-hand panel shows the result when SFS started with the 3MB components. The minimum mean phase error (MPE) within the imposed prior range (<100 ms) corresponds to the estimated latency, which is proportional to the slope of the fitted lines in the bottom panels. Given estimated latencies τ_e, we plot absolute phase lags as function of frequency for each subject from Eq. (2).

The estimated latencies resulted to vary with the location of the EEG channels. Table II includes the estimated latencies for three representative EEG channels: P7 (left TL), FCz (FC region) and P8 (right TL). Channels FCz and P8 were chosen because they yielded the largest grand-average LCI. P7 is located contralateral to P8, and was selected for comparison.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE II
ESTIMATED LATENCIES, MPE, AND LCI ON LF- AND HF- DOMINATED SYSTEMS (N=20).

EEG channel	LF (m, ±sd)			HF (m, ±sd)		
	Latency*	MPE	LCI (MBs)	Latency*	MPE	LCI (3MBs)
P8 (right TL)	40.8 (38.8, 41.3)	0.14 (±0.03)	0.29 (±0.12)	26.7 (25.2, 34.6)	0.11 (±0.06)	0.056 (±0.032)
FCz (FC)	51.7 (27.6, 54.4)	0.16 (±0.05)	0.37 (±0.08)	20.6 (16.7, 26.8)	0.14 (±0.04)	0.076 (±0.044)
P7 (left TL)	42.5 (40.6, 68.2)	0.13 (±0.05)	0.26 (±0.14)	-	-	0.047 (±0.038)

TL: temporal lobe; FC: frontal central. MPE: mean phase error at the estimated latency. MPE $\in [0,\ 2]$. Note the low values for the MPE, indicating high confidence in the estimates.
* Median values, and 25% and 75% percentiles, respectively.

The results based on calculating AVG phase, instead of AVG EEG, were similar (see Fig. 11). The estimated latencies for the low-frequency distortion products (LF) on EEG channels FCz and P7 had a non-Gaussian distribution, so that we report their median values and percentiles. The MPEs at the estimated latencies were low (<10%) and similar level for all estimates of both LF and HF dominated systems, which is indicative for a reliable estimation. The most consistent latencies were obtained from EEG channel P8, which yielded a large ASSR response (i.e., high SNRs and LCIs). As the power of high-order responses (e.g., 4^{th} and 6^{th} orders) tended to decrease, the corresponding ASSR components were not significant on certain EEG channels, causing a less reliable estimated latency. We thus selected a subject-specific best channel with the largest LCI value instead of using P7 for estimating the HF-dominated latencies, showing a median latency of 21.1 ms, with 19.9 and 24.3 ms as the 25% and 75% percentiles, respectively. The 'best ch' is mainly selected from the FC region, thus it shows a similar median value as FCz. The HF-dominated latency generally shows a larger variance than the LF-dominated latency, which is due to a smaller LCI of the HF components.

The LCI was computed on K=1200 (100 trials*12 sec) EEG epochs with 1-sec duration for each ASSR component. According to the theoretical threshold of $\sqrt{3/K}$ (see Methods), the 95% confidence threshold for a significant LCI was 0.05. The average LCI of MBs and 3MBs was used for LF and HF components. Table II shows that the LCIs for LF components were significantly higher than for HF components, which is mainly due to the decreased SNRs of the high-order ASSRs (see Fig. 9). The lower LCIs on HF partly explains the larger variance on the estimated latencies for these components.

D. Effect of epoch filtering

We excluded poorly phase-locked epochs by applying epoch filtering (see Methods) and compared the results by using four thresholds (θ = 0, 0.4, 0.6, and 0.8) for the phase-lock values, PLV. The estimated latencies were robust for different values of θ, as shown in Fig. 11. The cause for this stability is that the ALPC-SFS method employs multiple ASSR frequencies to determine a latency, and among them the MB frequencies show generally stable phase values across subjects. As a result, the Average EEG and Average Phase extraction methods both yielded similar results. Around 10% of all EEG epochs (i.e., $\theta >$ 0.8) generated similar results to that using all EEG epochs, which indicates that the latency is mainly determined by EEG

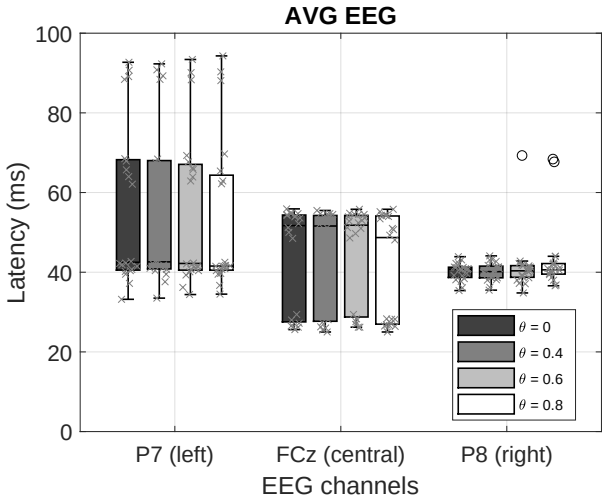

Fig. 11. Estimated LF-dominated latencies (n=20) by applying EEG epoch filtering (1 sec) with four thresholds of the PLV. Only the EEG epochs in which the PLV values exceeded the thresholds were used for phase extraction. The included EEG epochs were about 100%, 60%, 30% and 10% for the thresholds at θ = 0, 0.4, 0.6, and 0.8, respectively. The results were highly similar when phase extraction was performed with AVG phase (not shown here).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

epochs that are strongly phase-locked to the stimuli. On the other hand, using fewer EEG epochs increased the variance of estimated latency across subjects.

E. A comparison with the MSPC method

We compared our ALPC-SFS method with the MSPC method for estimating the ASSR latencies from the recorded EEG. Figure 12 shows the results obtained from one representative subject. The ALPC-SFS method selects a group of ASSR frequencies that share a common latency across different potential system orders. In contrast, the MSPC method computes a latency by assuming a specific system order. All significant ASSRs (i.e., LCI > 0.05) were used. Notice that the LCI is mathematically equivalent to the phase-coupling strength defined in the MSPC method (see Methods). The latter method, however, led to larger MPEs, because ASSRs of the same order could have been generated by subsystems with different latencies, and thus yielded poorer linear regression results for the relation between phase lag vs. frequency (see Fig. 12).

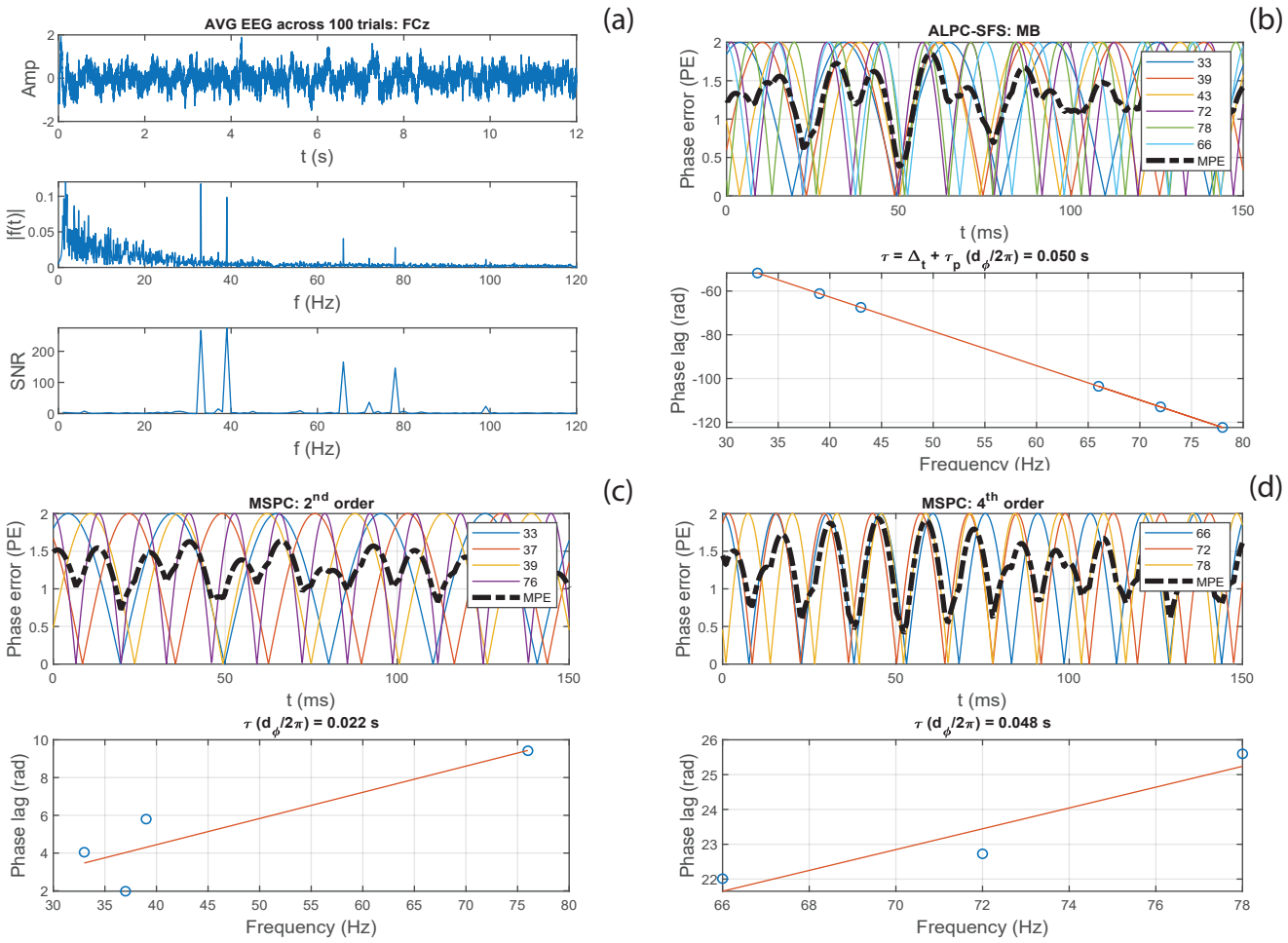

Fig. 12. An example using one channel EEG signals (FCz, 100 trials) from Subj#1 to demonstrate the difference between the ALPC-SFS and MSPC methods. (a) shows the average EEG across 100 trials, its spectrum, and the neighboring SNR (see Methods). (b), ALPC-SFS starts from two MB frequencies (33, 39 Hz), and automatically selects the ASSR frequencies that have a similar latency (error < 5 ms) and lead to a MPE < 0.5. ALPC-SFS applied time compensation, and here shows the negative slope corresponding to the pseudo latency. The MSPC method estimated latencies by assuming a specific system orders (c: order 2) and (d: order 4) (see also Table 1). However, the lumped $2^n d$-order ASSRs did not yield a reliable latency estimate, although all selected ASSRs were significant (i.e., LCI > 0.05). In contrast, the ALPC-SFS procedure identifies latencies across system orders, and generated a subset of ASSRs with the same latency from both 2^{nd} order (33, 39, 43 Hz) and 4^{th} order system outputs (66, 72, 78 Hz).

Table III shows the estimated latencies from the MSPC method for 20 measurements on two EEG channels: FCz and P8, on which subjects showed generally more significant ASSRs. Compared with the ALPC-SFS based MPE values in Table II, the MPE values (> 0.3) in Table III indicate that the estimated latencies from the MSPC method were not reliable. In addition, the low LCI values show that many ASSRs were weak and not significant (i.e., LCI < 0.05), and were excluded for estimating latencies. Note that the median values of the ALPC-SFS vs. MSPC-based latencies were similar (e.g., when based on 2^{nd}-order outputs). The ALPC method becomes mathematically equivalent to the MSPC method when the assumed system information is correct. However, compared with the ALPC-SFS method, the MSPC method lacks the procedure of ASSR frequency selection,

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and instead relies on a prior assumption regarding the system order to group frequencies (see Fig. 12). Such an assumption may be appropriate for the skeletal motor system ([20]), but is not suitable for analyzing the human auditory system. For example, looking at the 4^{th}-order distortion products, the MSPC method would include 2MBs, which are already generated at the cochlea, as well as the difference frequencies (e.g., 6 Hz) of MBs from both ears, which are generated at higher levels in the binaural auditory pathways (e.g., inferior colliculus and cortical levels). The different components may have different latencies, and are therefore not suitable to become lumped into one subsystem for estimating a latency.

TABLE III
ESTIMATED LATENCIES WITH THE MSPC METHOD FOR THREE SYSTEM ORDERS

System order	FCz			P8		
	Latency*	MPE (m, ±sd)	LCI (m, ±sd)	Latency*	MPE (m, ±sd)	LCI (m, ±sd)
2^{nd}	55.3 (52.2, 84.7)	0.68 (±0.24)	0.16 (±0.02)	40.7 (15.8, 41.5)	0.63 (±0.25)	0.12 (±0.04)
4^{th}	40.3 (37.1, 50.4)	0.31 (±0.19)	0.07 (±0.04)	42.3 (32.2, 55.5)	0.48 (±0.26)	0.06 (±0.03)
6^{th}	22.0 (12.1, 31.1)	0.38 (±0.22)	0.05 (±0.02)	45.2 (25.2, 46.6)	0.33 (±0.25)	0.04 (±0.02)

* Median value, and the 25% and 75% percentiles.
MPE: mean phase error, $\in [0,\ 2]$, at the estimated latency.
LCI here is mathematically equivalent to the MSPC value used in [20].

V. DISCUSSION

A. Advantages of the proposed method and stimuli

ALPC combined with SFS and time compensation has three major advantages compared with existing methods: (i) It requires no prior assumptions of the underlying nonlinear systems (e.g., system orders, number of subsystems) to calculate the unknown phase information of the ASSR components through phase compensation, and is therefore easier to implement. (ii) It allows for the estimation of multiple underlying latencies. (iii) It allows to analyze different nonlinear (sub)systems with different nonlinearities in one pass, and automatically selects the contributing ASSR components with the same latency (and potential common generators). Therefore, ALPC with SFS is better suited for analyzing nonlinear, higher-order ASSRs that are thought to underlie complex brain functions.

Moreover, the use of multi-cosine stimuli with all initial phases set to zero provides two additional advantages. (i) Compared to stimuli with arbitrary initial phases, it avoids the complicated procedure of phase compensation, and thus requires no assumptions regarding the underlying system nonlinearities. Note that the stimuli should preferably be cosine (not sine) with initial phase zero, because for multi-cosine stimuli, arbitrary-order (even and odd) nonlinear distortion products all have zero initial phase, so that one can readily apply ALPC-SFS with time compensation. As shown in Appendix C, a sum of sines only maintains a common initial phase for a specific subset of distortion products, because the nonlinear distortions of the sine will generate phase shifts $\in$ [0, $\pi/2$, π, or $3\pi/2$] that depend on system order and nonlinear interactions. (ii) The estimated latency for zero-phase distortions will be more robust against the unavoidable small system errors (e.g., small delays caused by tubes and the measurement devices). It is often difficult to directly measure the exact phase values of the stimulus carriers at the onset of experimental trials, i.e., the exact time at which sound stimuli hit the ear drums. As a result, even a small timing error of one millisecond will already significantly change the phases of high-frequency stimuli (e.g., $>$ 1k Hz). Consequently, small errors in the stimulus phases may lead to an unreliable latency estimate for methods that require exact initial phase values of the stimuli. In contrast, for multi-cosine stimuli the exact onset phase at the ears is not critical, as the estimated latency will be the sum of the true latency and the unknown (small) system delay, which can be readily incorporated as a fixed time compensation (described in Methods).

B. Limitation of the method

The proposed ALPC-SFS method with time compensation can identify multiple latencies, and thus potentially separate multiple sources characterized by different frequency components on a single EEG channel. However, it remains difficult to separate nearby sources that are characterized by the same frequency components. This is due to the nature of EEG signals. The EEG signal represents the superposition of potentials from the local cortex and nearby volume conduction. For each frequency component, the scalp-EEG phase is affected by underlying multiple sources, either between their original phases or dominated by the one with a larger amplitude. Whereas ALPC is a phase-based method, the shifted phases may cause a bias of estimation, which is determined by the relative amplitudes of the two original systems (see Fig. 8). To further separate the nearby sources in deeper brain structures, it is helpful to employ other techniques, such as MEG that also has a better spatial resolution ([7]), or beamforming techniques using multiple EEG channels ([37]). In addition, to take advantage of multiple electrodes (or EEG channels), we will evaluate a subject-specific optimal EEG channel selection ([38]), and a recently-proposed multivariate source separation method, called the rhythmic entrainment source separation (RESS) [39].

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

C. Pitfalls of EEG montage

For the preprocessing of EEG signals, it is detrimental to re-reference all channels to an EEG channel that is not neutral to ASSRs, since a re-reference EEG channel containing strong ASSRs will affect all other EEG channels. For example, Fig. 13 shows phase values of all EEG channels at one MB frequency. After re-referencing to EEG channel Cz (carrying a large SNR of ASSRs), the distribution of SNR and phase values clearly changed. The phase of all EEG channels became similar to that on Cz. Therefore, it is recommended to either use the CAR montage ([33]), or to re-reference to a neutral electrode. On the other hand, also the CAR montage has its drawbacks, as it may cause phase ambiguity on the EEG channels on which ASSRs are relatively low. Certain ASSR components might be reversed, resulting in a phase shift of π. For that reason, the EEG channel P7 shows a larger variance on the estimated latency. To avoid the influence of such phase ambiguity, it is good practice to use the EEG channels that show relatively stronger ASSRs, e.g., at the frontal-central (FC) and the right temporal-lobe regions. Note that completely reversing an EEG channel (i.e., y = -x) will cause the same phase shift of π on all frequency components. It thus only changes the bias term of a fitted line (phase lag vs frequency) without changing the slope. Therefore, it does not affect the results of ALPC, or apparent latency.

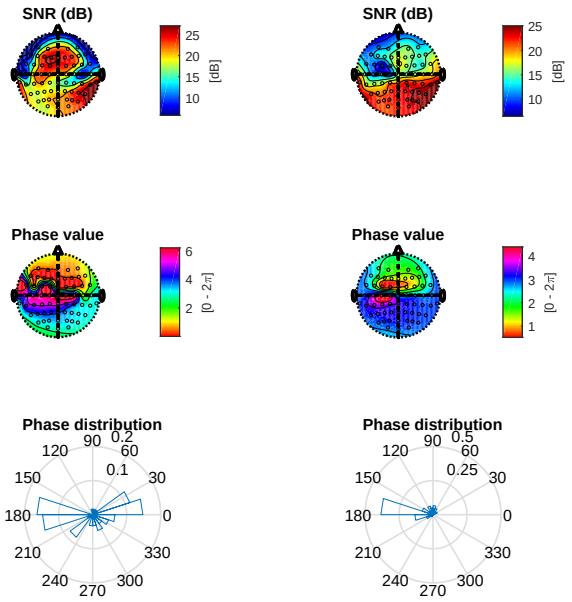

Fig. 13. SNR and phase topography of the average EEG of 100-trials at an MB frequency of 33 Hz (data from subj#1). The left-hand column shows the common average reference (CAR) montage, while the right-hand column shows the result after re-referencing to EEG channel Cz. Note that the phase topography of the CAR shows a typical fronto-occipital phase delay, which is in line with ([40]). In contrast, the phases of many EEG channels (right-hand column) are changed by Cz, thus may leading to biased estimate of latencies.

D. Multiple latencies of ASSRs

To estimate reliable latencies with the ALPC-SFS method, several important issues should be considered: (i) select the starting parameters for the SFS; (ii) use only EEG channels with a relatively high SNR (or LCI), and (iii) use EEG epochs that show high phase-locked values. An important parameter for the SFS is the starting frequency. SFS is a suboptimal method, and the starting frequency (or frequency pair) determines the direction of the heuristic search. It is good practice to start with those ASSR components with the highest SNR (or LCI), like the two major MB frequencies (33 and 39 Hz) in our experiments. Our simulation results demonstrate that the SNR of ASSRs affect the probability that two subsystems can indeed be separated (see Table AI). Starting the SFS with these two MB components also identified the other ASSR components with the same latenency. We also started the SFS from 2MBs or from the (4, 6) Hz pair. However, only when the SFS started from the major MBs or from 3MBs, we obtained consistent results across subjects. ASSRs at 4 and 6 Hz are weak in most subjects, so that no reliable phase values are available for estimating the latency.

ASSR sources locate in multiple cortical as well as subcortical regions: ASSRs ($<$ 20 Hz) originate from the auditory cortex; ASSRs around 40 Hz has multiple locations including the brainstem, thalamus and the auditory cortex, while high frequency ASSRs (80 - 100 Hz) dominantly from the upper region of the brainstem [2], [41]. Our analysis revealed two major latencies from the ASSR components. The LF-dominant system (near 40 Hz) showed longer latencies than the HF-dominant system ($>$80 Hz). This finding is similar with previous studies using amplitude modulation (AM) stimuli on EEG or MEG signals [1], [23]. LF and HF correspond to the frequency ranges where maximum activation of ASSRs arise from the cortex ($\approx$45 Hz) and the brainstem ($\approx$ 90 Hz), respectively [28]. For the LF-dominant system, the two TL regions showed similar latencies

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(≈ 41 ms) whereas the FC region was endowed with a slightly longer latency (≈ 52 ms). This observation is consistent with the results from an EEG-based, binaural hearing study [40]. A longer latency suggests an origin in the upper stages of the ascending pathways of the auditory system. Thus, the longer latency in the FC region could perhaps arise from binaural interactions in the cortex, following monaural interactions. For the HF-dominant system, however, the estimated latency on the FC region (near FCz) was shorter than at both TL regions (P7 and P8), which is similar as reported by [23]). For these HF components, the brainstem has been considered as the major underlying source [28]. In the monaural auditory pathway, the signal may propagate from the lower brainstem, through the upper brainstem (which is anatomically located near the FC region), to the auditory cortex (located near the TLs).

It is further interesting to note that both LF and HF-dominant systems contained some overlapping ASSR components (see Fig. 10). This finding suggests that multiple neural mechanisms, characterized by different frequency ranges, may still enter a phase-locked status and thus shows a common latency. This finding is in line with the observed phase synchronization across multiple brain regions and function stages in the hearing pathways and high-level cortex [3].

ASSRs are right-lateralized, i.e., right-hemisphere response exceeds the left-hemisphere response, which has been reported by many ASSR studies [7], [16]. We also found it in this study for both LF and HF ASSRs (see Table II), thus the EEG channel P8 showed similar latency but smaller variance than on P7 (see Fig. 11). In addition, the frontal-central (FC) brain region (near FCz) shows the strongest ASSRs. However, the latencies estimated from FCz show larger variance than on P8. In particular, Fig. 11 shows two latency clusters, one major around 52 ms and a smaller around 25 ms. The major latency is larger than the TL region latency (around 40 ms), which might represent a binaural cortical interaction, i.e., the 4^{th} and 6^{th} -order distortion frequencies are mainly generated from interaction between the beat frequencies of both ears. The smaller cluster (n=6 out of 20 measurements) should not be simply considered as outliers. It is because that the FC brain region may represents the superposition of responses from multiple underlying sources (yielding an weighted average latency estimate for brainstem and cortex). Some subjects may have stronger ASSRs measured from the upper brainstem than from cortex, and hence affect the estimated latency, i.e., the brainstem related latency is much smaller than cortex related latency [21], [42].

E. Possible applications

The ascending auditory system, from auditory nerve to auditory cortex, contains several parallel pathways of acoustic signal processing, which eventually reach the cortical areas. For example, the localization of a sound source requires the neural processing of three types of acoustic cues, which are extracted by independent neural circuits within the auditory brainstem: interaural time differences (ITDs) and interaural level differences (ILDs) for locations in the horizontal plane, and direction-dependent spectral-shape cues from the pinna for directions in the vertical plane (Fig. 14; [43]).

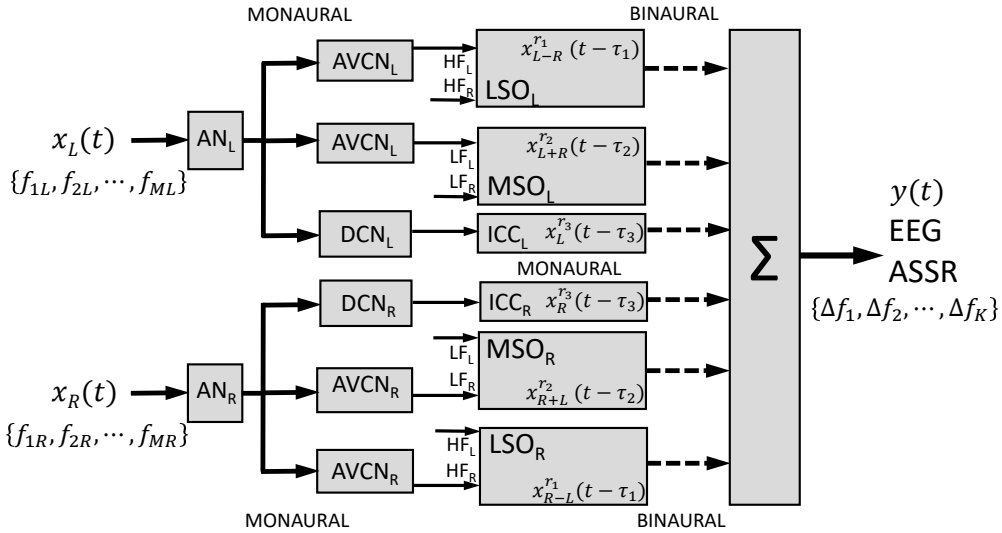

Fig. 14. Sound localization relies on three independent monaural and binaural pathways that use largely non-overlapping frequency bands. All sounds are initially processed by a monaural stage at the auditory nerve (AN) and antero-ventral and dorsal subdivisions of the cochlear nucleus (AVCN and DCN). Low carrier frequencies (LF) are processed by the medial superior olive (MSO), which computes the interaural time difference (ITD); the high frequencies (HF) follow the lateral superior olive (LSO) pathway, which determines the interaural level difference (ILD). The DCN is part of a monaural path to the midbrain Inferior Colliculus (ICC) and cortex. All stages may introduce nonlinearities in the signal paths that are represented as frequency distortion products, Δf_n, in the EEG.

The binaural difference cues operate on complementary frequency bands: as the ITD for humans varies from about -600 μs (sound at the far left) to +600 μs (sound at the far right), this cue only provides unambiguous phase information for

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sound frequencies up to ∼1500 Hz. For these low frequencies, phase-locked neurons in the auditory nerve, the anteroventral cochlear nucleus (AVCN), and the medial superior olive (MSO) carry precise phase information about the carrier frequency. In the MSO, neurons respond to the interaural phase difference at their characteristic frequency. Because the phase difference is frequency-dependent for a given delay ($\Delta\varphi = 2\pi f \Delta\tau$), MSO neurons that are tuned to a specific frequency and phase difference thus encode a particular inter-aural absolute delay, and hence a particular direction in the horizontal plane. Recently, a periodic representation of ITDs has also been demonstrated for human cortex ([44]). Neurons that are tuned to the same direction in azimuth, but are sensitive to other frequencies, will be tuned to different phase differences. In this way, the neuronal processing resembles the ALPC algorithm, which also relies on frequency-dependent phase-differences to estimate the unique underlying (characteristic) delay ([45], [46], [43]).

A second pathway processes the ILDs that arise for frequencies above ∼3 kHz because of the head shadow effect. Cells in the lateral superior olive (LSO) are tuned to the ILD for a particular frequency. The ILD cue depends strongly on the sounds frequency: the higher the frequency, the stronger the head-shadow effect. Combining the responses from a population of neurons, tuned to different frequencies, is needed to estimate the veridical azimuth angle of the sound source. It is conceivable that the final stage in estimating the sound's azimuth angle involves nonlinear integrative mechanisms for both binaural pathways. The dorsal cochlear nucleus (DCN) feeds a monaural signal pathway via the central nucleus of the Inferior Colliculus (ICC) to the cortex, and is possibly involved in the localization of source elevation (spectral cue extraction), and pitch perception.

It is possible that the different auditory processing mechanisms may be endowed with different delays. As each of these mechanisms operates in different frequency bands, and could potentially introduce its own nonlinear distortions, they may in principle be separated by our ALPC-SFS method by applying specific frequency combinations to either ear. Similarly, the method may be used to test quantitative models of the ascending auditory system ([47], [19]).

VI. Conclusions

We proposed a nonparametric method, ALPC-SFS with time compensation, to estimate multiple latencies of the auditory system from ASSRs recorded on a single EEG channel. Compared with existing methods, our method requires no prior assumptions of underlying nonlinearity (e.g., system orders), and is easy to implement to higher-order ASSRs, when combined with a zero-phase cosine stimulus complex. We illustrated our method on recorded EEG signals and it successfully identified two major (LF and HF -dominated) latencies that were stable across subjects and between measurement days. The LF-dominated latencies were longer than the HF-dominated latencies and related to different brain regions (e.g., central brain or temporal-lobe regions). The LF-dominated latency is mainly determined by EEG epochs that are strongly phase-locked to the stimuli. ALPC-SFS is promising as an objective measure for hearing performance and for the underlying nonlinearities of both primary and non-primary auditory cortex (dys)function.

Acknowledgments

We thank Yuan Yang from Northwestern university for his advice on the experiments and for providing the analysis code of the MSPC method. We thank anonymous reviewers for their constructive criticisms on an earlier version of this manuscript, which greatly helped to improve our paper. Also, we thank the students and colleagues in Radboud university and Delft university for the help on collecting the EEG datasets. This research was supported by the Dutch Organisation for Scientific Research, NWO-TTW Perspectief, project 'NeuroCIMT' - Otocontrol, nr. 14689 (LW, EN), and EU Horizon 2020 ERC Advanced Grant 2016 'Orient', nr. 693400 (AJVO).

Appendix A
Phase compensation

For linear systems, the output frequencies are exactly the same as the input frequencies, as well as the initial phases (apart from the input-output delay) of the output frequencies. However, this property does not hold if the system is nonlinear. When a nonlinear system is subject to a mixture of I inputs, the output frequencies generated by the r^{th}-order system nonlinearity consist of all possible combinations of the I input frequencies, $\{-f_I, \ldots, -f_1, f_1, \ldots, f_I\}$ taken r items at a time. This property can be described as a set given by ([26]):

$$\left\{ f_\Sigma \middle| \begin{array}{c} f_\Sigma = |f_{i_1} +, \ldots, + f_{i_r}| \\ f_i \in \{-f_I, \ldots, -f_1, f_1, \ldots, f_I\},\ r = 1, \ldots, R \end{array} \right\} \tag{A1}$$

The corresponding initial phases of f_Σ are $\phi(f_\Sigma) = \mathrm{mod}(\frac{\phi(f_{i_1})+,\ldots,+\phi(f_{i_r})}{2\pi})$, where $\phi(\cdot)$ computes the initial phase of a frequency by FT ([20]). In particular, for a given r^{th}-order system, the initial phases of the output frequency components are a set:

$$\left\{ \phi(f_{\Sigma_r}) \middle| \begin{array}{c} \phi(f_{\Sigma_r}) = \sum_{i=1}^{r} a_i \phi(f_i),\ with\ \sum_{i=1}^{r} |a_i| = r \\ f_i \in \{f_1, f_2, \ldots, f_I\}, \end{array} \right\} \tag{A2}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where a_i is a positive or negative integer. The resulting phase set $\phi(f_{\Sigma_r})$ can be further mapped into the range of [0, 2π] by using the $\mathrm{mod}(\cdot)$ function. We denote the above process as 'phase compensation'. Phase compensation has been used to compute the phase coupling strength between inputs and outputs of nonlinear systems ([20]). However, phase compensation is not needed for stimuli when the initial phases of multi-cosine inputs are set to zero (i.e., $\phi(f_i) \equiv 0$). See Appendix C, for further details.

APPENDIX B
SIMULATION RESULTS OF EXAMPLE 1

The SNR in the first column of Table A1 shows the amount of additive Gaussian noise. The SNR in the second column is computed from (10). The MPE shows the average phase errors (in the 3^{rd} column) of the output frequency components between pure signals and signals with added noise. We simulated 100 trials of 12 sec input signals for each SNR.

TABLE A1
SIMULATION RESULTS WITH DIFFERENT SNRS

SNR (dB)	SNR*(dB) (m, ±sd)	MPE [0 - 2] (m, ±sd)	LCI (m, ±sd)	Latency (m, ±sd)		Acc (%)**	
				y_1	y_2	y_1	y_2
5	31.2 (±0.33)	0.02 (±0.01)	1.00 (±0.00)	51.0 (±0.04)	21.0 (±0.04)	100	100
0	26.3 (±0.40)	0.03 (±0.01)	0.99 (±0.01)	51.0 (±0.09)	24.8 (±18.8)	100	94
-5	21.2 (±0.36)	0.06 (±0.01)	0.96 (±0.03)	50.9 (±0.2)	38.2 (±41.8)	100	82
-10	16.2 (±0.56)	0.10 (±0.03)	0.86 (±0.10)	51.7 (±8.5)	49.9 (±49.3)	99	65
-15	11.4 (±0.74)	0.19 (±0.04)	0.68 (±0.14)	53.0 (±17.9)	67.8 (±54.2)	92	32
-20	6.6 (±0.98)	0.33 (±0.08)	0.44 (±0.13)	58.6 (±28.0)	71.8 (±50.2)	85	15

* Neighboring SNR: computed from neighboring frequency bins (± 0.5 Hz).
** Accuracy: an estimated latency with an error (± 5 ms) was counted as a correct estimation.

APPENDIX C
NONLINEAR DISTORTIONS FOR COSINES AND SINES

Here, we analyze the general properties of the phase shifts of high-order (odd and even-power) nonlinear distortion products for cosine vs. sine inputs. The general power relations for cosine and sine functions are given in (A3) (1^{st} row for odd powers, 2^{nd} row for even powers):

$$\begin{aligned}
\cos^n\alpha &= \tfrac{2}{2^n}\sum_{k=0}^{(n-1)/2}\binom{n}{k}\cos\left((n-2k)\alpha\right) & \sin^n\alpha &= \tfrac{2}{2^n}\sum_{k=0}^{(n-1)/2}(-1)^{\left(\frac{n-1}{2}-k\right)}\binom{n}{k}\sin\left((n-2k)\alpha\right)\\
\cos^n\alpha &= \tfrac{1}{2^n}\binom{n}{\frac{n}{2}} + \tfrac{2}{2^n}\sum_{k=0}^{\frac{n}{2}-1}\binom{n}{k}\cos\left((n-2k)\alpha\right) & \sin^n\alpha &= \tfrac{1}{2^n}\binom{n}{\frac{n}{2}} + \tfrac{2}{2^n}\sum_{k=0}^{\frac{n}{2}-1}(-1)^{\left(\frac{n-1}{2}-k\right)}\binom{n}{k}\cos\left((n-2k)\alpha\right)
\end{aligned} \tag{A3}$$

and together with the multinomial theorem:

$$(x_1+\ldots+x_m)^n = \sum_{k_1+\ldots+k_m=n}\frac{n!}{k_1!k_2!\ldots k_m!}x_1^{k_1}x_2^{k_2}\ldots x_m^{k_m} \tag{A4}$$

a general sum of sines and cosines can be readily computed. As three illustrative examples, we here calculate all phases for the n=2^{nd}, n=3^{rd} and 4^{th}-order powers for a sum of m=3 sines and cosines:

1) Second order: $(x+y+z)^2 = x^2+y^2+z^2+2xy+2xz+2yx$. Applying this to three sine inputs $(\sin\alpha+\sin\beta+\sin\gamma)^2$ yields the following 10 terms:

$$\begin{aligned}
&\frac{3}{2} - \frac{1}{2}\left(\cos(2\alpha)+\cos(2\beta)+\cos(2\gamma)\right) - \frac{1}{2}\left(\cos(\alpha+\beta)+\cos(\alpha+\gamma)+\cos(\beta+\gamma)\right)\\
&+\frac{1}{2}\left(\cos(\alpha-\beta)+\cos(\alpha-\gamma)+\cos(\beta-\gamma)\right)
\end{aligned}$$

If α, α, γ have zero initial phase, the three low-frequency difference distortion components in the output have the same initial phase of $\pi/2$ rad (caused by the transform from sine to cosine). According to (3), this would not affect the estimate from TC, as it will add only a bias to the phase-lag vs. frequency relation, without affecting the slope (latency). However, the six terms following the d.c. all have an initial phase of $3\pi/2$ (due to the negative sign). Thus, if all nine components would be lumped into one set, TC is no longer applicable, because of the different initial phases (and associated biases). Then, phase

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

compensation is required to account for the initial phases of the components. In our EEG example, we employed this property of sines to the second-order difference distortion products in the ASSR.

Whereas three cosine inputs $(\cos\alpha + \cos\beta + \cos\gamma)^2$ yield the following 10 terms:

$$\frac{3}{2} + \frac{1}{2}\left(\cos(2\alpha) + \cos(2\beta) + \cos(2\gamma)\right) + \frac{1}{2}\left(\cos(\alpha+\beta) + \cos(\alpha+\gamma) + \cos(\beta+\gamma)\right) + \frac{1}{2}\left(\cos(\alpha-\beta) + \cos(\alpha-\gamma) + \cos(\beta-\gamma)\right)$$

Each of the nine distortion components now has the same initial phase of zero rad if α, α, γ have phase zero. Thus, TC is readily applicable to the whole set of distortion components.

2) Third order: $(x+y+z)^3 = x^3+y^3+z^3+6xyz+3x^2y+3x^2z+3y^2x+3y^2z+3z^2x+3z^2y$. The ten terms on the right-hand side contain three third-order powers, one tri-linear product, and six quadratic-linear products, for which the following properties hold.

For sines:

$$\begin{aligned}\sin^3\theta &= \frac{3}{4}\sin\theta - \frac{1}{4}\sin 3\theta \\ \sin^2\alpha\sin\beta &= \frac{1}{2}\sin\beta - \frac{1}{4}\left[\sin(\beta+2\alpha) + \sin(\beta-2\alpha)\right] \\ \sin\alpha\sin\beta\sin\gamma &= \frac{1}{4}\sin(\alpha+\beta-\gamma) + \frac{1}{4}\sin(\alpha-\beta+\gamma) \\ &\quad - \frac{1}{4}\sin(\alpha+\beta+\gamma) - \frac{1}{4}\sin(\alpha-\beta-\gamma)\end{aligned}$$

This yields the following 22 terms, with 3 linear components, and 19 distortion products:

$$\begin{aligned}&\tfrac{7}{4}\left(\sin\alpha+\sin\beta+\sin\gamma\right) - \tfrac{1}{4}\left(\sin 3\alpha+\sin 3\beta+\sin 3\gamma\right) \\ &+\tfrac{3}{2}\begin{bmatrix}\sin(\alpha+\beta-\gamma)+\sin(\alpha-\beta+\gamma) \\ -\sin(\alpha+\beta+\gamma)-\sin(\alpha-\beta-\gamma)\end{bmatrix} \\ &-\tfrac{3}{4}\begin{bmatrix}\sin(\alpha+2\beta)+\sin(\alpha-2\beta)+\sin(\alpha+2\gamma) \\ +\sin(\alpha-2\gamma)+\sin(\beta+2\alpha)+\sin(\beta-2\alpha) \\ +\sin(\beta+2\gamma)+\sin(\beta-2\gamma)+\sin(\gamma+2\alpha) \\ +\sin(\gamma-2\alpha)+\sin(\gamma+2\beta)+\sin(\gamma-2\beta)\end{bmatrix}\end{aligned}$$

Only the three positive components will have zero phase if α, α, γ have zero phase, but all 19 negative terms will have a phase of π radians. Therefore, the components cannot all be lumped in the phase analysis.

For cosines:

$$\begin{aligned}\cos^3\theta &= \frac{3}{4}\cos\theta + \frac{1}{4}\cos 3\theta \\ \cos^2\alpha\cos\beta &= \frac{1}{2}\cos\beta + \frac{1}{4}\left[\cos(2\alpha+\beta) + \cos(2\alpha-\beta)\right] \\ \cos\alpha\cos\beta\cos\gamma &= \frac{1}{4}\begin{bmatrix}\cos(\alpha-\beta+\gamma)+\cos(\alpha+\beta-\gamma) \\ \cos(\alpha+\beta+\gamma)+\cos(\alpha-\beta-\gamma)\end{bmatrix}\end{aligned}$$

Each of the resulting 19 distortion components will maintain their zero phase.

3) Fourth-order nonlinearity: $(x+y+z)^4 = x^4+y^4+z^4+4\left(x^3y+x^3z+y^3x+y^3z+z^3x+z^3y\right)+6\left(x^2y^2+x^2z^2+y^2z^2\right)+12(x^2yz+y^2xz+z^2xy)$. The 15 terms on the right-hand side contain three fourth-order powers, six third-order-linear products, three quadratic-quadratic products, and three quadratic-linear-linear products, for which the following properties hold.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For sines:

$$\begin{aligned}
\sin^4\theta =& \frac{3}{8} - \frac{1}{2}\cos 2\theta + \frac{1}{8}\cos 4\theta \\
\sin^3\alpha \sin\beta =& \frac{3}{8}\left[\cos(\alpha-\beta) - \cos(\alpha+\beta)\right] - \\
& \frac{1}{8}\left[\cos(3\alpha-\beta) - \cos(3\alpha+\beta)\right] \\
\sin^2\alpha \sin^2\beta =& \frac{1}{4}\left[\begin{array}{l} 1 - \cos 2\alpha - \cos 2\beta + \\ \frac{1}{2}\cos(2\alpha+2\beta) + \frac{1}{2}\cos(2\alpha-2\beta) \end{array}\right] \\
\cos^2\alpha \cos\beta \cos\gamma =& \frac{1}{4}\left[\cos(\beta-\gamma) + \cos(\beta+\gamma)\right] + \\
& \frac{1}{8}\left[\cos(2\alpha-\beta-\gamma) + \cos(2\alpha+\beta+\gamma)\right] + \\
& \frac{1}{8}\left[\cos(2\alpha-\beta+\gamma) + \cos(2\alpha+\beta-\gamma)\right]
\end{aligned}$$

Now, all terms will yield phase-shifted cosines of either $\pi/2$ (positive terms) or $3\pi/2$ (negative terms) at t=0, which means that they will follow different (shifted) phase-frequency relations and cannot be lumped. Note that the three difference-frequency terms will all have the same $\pi/2$ phase shift as produced by the second-order nonlinearity.

For cosines:

$$\begin{aligned}
\cos^4\theta =& \frac{3}{8} + \frac{1}{2}\cos 2\theta + \frac{1}{8}\cos 4\theta \\
\cos^3\alpha \cos\beta =& \frac{3}{8}\left[\cos(\alpha-\beta) - \cos(\alpha+\beta)\right] + \\
& \frac{1}{8}\left[\cos(3\alpha-\beta) - \cos(3\alpha+\beta)\right] \\
\cos^2\alpha \cos^2\beta =& \frac{1}{4}\left[\begin{array}{l} 1 + \cos 2\alpha + \cos 2\beta + \\ \frac{1}{2}\cos(2\alpha+2\beta) + \frac{1}{2}\cos(2\alpha-2\beta) \end{array}\right] \\
\cos^2\alpha \cos\beta \cos\gamma =& \frac{1}{4}\left[\cos(\beta-\gamma) + \cos(\beta+\gamma)\right] + \\
& \frac{1}{8}\left[\cos(2\alpha-\beta-\gamma) + \cos(2\alpha+\beta+\gamma)\right] + \\
& \frac{1}{8}\left[\cos(2\alpha-\beta+\gamma) + \cos(2\alpha+\beta-\gamma)\right]
\end{aligned}$$

Again, all phases of the 54 different distortion products will remain zero at t=0.

In conclusion: from this analysis and from the general equations (A3) and (A4), it can be readily appreciated that the above properties for sine and cosine inputs will hold, irrespective of the order of the nonlinearity n, and of the number of input frequencies m. We can hence conclude that cosine inputs will always maintain zero phase for all nonlinear distortions, but for sine inputs the situation becomes much more complex because, depending on the contributing nonlinearity, the phases of the distortion components can be either 0, $\pi/2$, π, or $3\pi/2$. However, the phase of the difference frequencies, generated by all even-order nonlinearities, will always be $\pi/2$, and therefore can be unambiguously used in the non-parametric ALPC-SFS method, as has been applied in our EEG data analysis.

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*, vol. 42, no. 4, pp. 177–219, 2003.

[2] T. W. Picton, *Human auditory evoked potentials*. Plural Publishing, 2010, ISBN: 1597566225.

[3] E. D. Farahani, J. Wouters, and A. van Wieringen, "Contributions of non-primary cortical sources to auditory temporal processing," *NeuroImage*, vol. 191, pp. 303–314, 2019.

[4] S. A. Reyes, A. H. Lockwood, R. J. Salvi, M. L. Coad, D. S. Wack, and R. F. Burkard, "Mapping the 40-hz auditory steady-state response using current density reconstructions," *Hearing research*, vol. 204, no. 1-2, pp. 1–15, 2005.

[5] B. Chandrasekaran and N. Kraus, "The scalprecorded brainstem response to speech: Neural origins and plasticity," *Psychophysiology*, vol. 47, no. 2, pp. 236–246, 2010.

[6] G. M. Bidelman, "Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient abr," *Hearing research*, vol. 323, pp. 68–80, 2015.

[7] E. B. Coffey, S. C. Herholz, A. M. Chepesiuk, S. Baillet, and R. J. Zatorre, "Cortical contributions to the auditory frequency-following response revealed by meg," *Nature communications*, vol. 7, p. 11 070, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[8] R. Luke, A. De Vos, and J. Wouters, "Source analysis of auditory steady-state responses in acoustic and electric hearing," *NeuroImage*, vol. 147, pp. 568–576, 2017.

[9] L. E. Roberts, D. J. Bosnyak, I. C. Bruce, P. E. Gander, and B. T. Paul, "Evidence for differential modulation of primary and nonprimary auditory cortex by forward masking in tinnitus," *Hearing research*, vol. 327, pp. 9–27, 2015.

[10] C. S. van der Reijden, L. H. Mens, and A. F. Snik, "Signal-to-noise ratios of the auditory steady-state response from fifty-five eeg derivations in adults," *JOURNAL of the American Academy of Audiology*, vol. 15, no. 10, pp. 692–701, 2004.

[11] P. Korczak, J. Smart, R. Delgado, T. M Strobel, and C. Bradford, "Auditory steady-state responses," *Journal of the American Academy of Audiology*, vol. 23, no. 3, pp. 146–170, 2012.

[12] E. Skoe, J. Krizman, E. Spitzer, and N. Kraus, "The auditory brainstem is a barometer of rapid auditory learning," *Neuroscience*, vol. 243, pp. 104–114, 2013.

[13] A. Lehmann and M. Schnwiesner, "Selective attention modulates human auditory brainstem responses: Relative contributions of frequency and spatial cues," *PloS one*, vol. 9, no. 1, e85442, 2014.

[14] B. F. O'Donnell, J. L. Vohs, G. P. Krishnan, O. Rass, W. P. Hetrick, and S. L. Morzorati, "The auditory steady-state response (assr): A translational biomarker for schizophrenia," in *Supplements to Clinical neurophysiology*. Elsevier, 2013, vol. 62, pp. 101–112, ISBN: 1567-424X.

[15] E Diesch, M Andermann, H Flor, and A Rupp, "Interaction among the components of multiple auditory steady-state responses: Enhancement in tinnitus patients, inhibition in controls," *Neuroscience*, vol. 167, no. 2, pp. 540–553, 2010.

[16] J. A. Undurraga, N. R. Haywood, T. Marquardt, and D. McAlpine, "Neural representation of interaural time differences in humansan objective measure that matches behavioural performance," *Journal of the Association for Research in Otolaryngology*, vol. 17, no. 6, pp. 591–607, 2016.

[17] R. Gransier, M. Hofmann, H. Deprez, M. Moonen, A. van Wieringen, and J. Wouters, "The utility of eassrs for ci fitting: From threshold determination to the assessment of modulation encoding," *Journal of Hearing Science*, vol. 8, no. 2, 2018.

[18] P. Joris, C. Schreiner, and A Rees, "Neural processing of amplitude-modulated sounds," *Physiological reviews*, vol. 84, no. 2, pp. 541–577, 2004.

[19] M. van der Heijden and P. X. Joris, "Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies," *Journal of Neuroscience*, vol. 23, no. 27, pp. 9194–9198, 2003.

[20] Y. Yang, T. Solis-Escalante, J. Yao, A. Daffertshofer, A. C. Schouten, and F. C. Van Der Helm, "A general approach for quantifying nonlinear connectivity in the nervous system based on phase coupling," *International journal of neural systems*, vol. 26, no. 01, p. 1 550 031, 2016.

[21] A. T. Herdman, O. Lins, P. Van Roon, D. R. Stapells, M. Scherg, and T. W. Picton, "Intracerebral sources of human auditory steady-state responses," *Brain topography*, vol. 15, no. 2, pp. 69–86, 2002.

[22] M. John and T. Picton, "Human auditory steady-state responses to amplitude-modulated tones: Phase and latency measurements," *Hearing research*, vol. 141, no. 1-2, pp. 57–79, 2000.

[23] R Schoonhoven, C. Boden, J. Verbunt, and J. De Munck, "A whole head meg study of the amplitude-modulation-following response: Phase coherence, group delay and dipole source analysis," *Clinical neurophysiology*, vol. 114, no. 11, pp. 2096–2106, 2003.

[24] C. L. Witham, C. N. Riddle, M. R. Baker, and S. N. Baker, "Contributions of descending and ascending pathways to corticomuscular coherence in humans," *The Journal of physiology*, vol. 589, no. 15, pp. 3789–3800, 2011.

[25] Y. Yang, T. Solis-Escalante, F. C. van der Helm, and A. C. Schouten, "A generalized coherence framework for detecting and characterizing nonlinear interactions in the nervous system," *IEEE transactions on biomedical engineering*, vol. 63, no. 12, pp. 2629–2637, 2016.

[26] Z.-Q. Lang and S. Billings, "Evaluation of output frequency responses of nonlinear systems under multiple inputs," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 1, pp. 28–38, 2000.

[27] J. Peyton Jones and K Choudhary, "Output frequency response characteristics of nonlinear systems. part ii: Overlapping effects and commensurate multi-tone excitations," *International Journal of Control*, vol. 85, no. 9, pp. 1279–1292, 2012.

[28] D. W. Purcell, S. M. John, B. A. Schneider, and T. W. Picton, "Human temporal auditory acuity as assessed by envelope following responses," *The Journal of the Acoustical Society of America*, vol. 116, no. 6, pp. 3581–3593, 2004.

[29] 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.

[30] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 19, no. 2, pp. 153–158, 1997.

[31] J. Shils, M Litt, B. Skolnick, and M. Stecker, "Bispectral analysis of visual interactions in humans," *Electroencephalography and clinical Neurophysiology*, vol. 98, no. 2, pp. 113–125, 1996.

[32] L. Wang, J. B.A. M. Arends, X. Long, P. J. M. Cluitmans, and J. P. van Dijk, "Seizure pattern-specific epileptic epoch detection in patients with intellectual disability," *Biomedical Signal Processing and Control*, vol. 35, pp. 38 –49, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.27.315614; this version posted October 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[33] K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, and D. R. Kipke, "Using a common average reference to improve cortical neuron recordings from microelectrode arrays," *Journal of neurophysiology*, vol. 101, no. 3, pp. 1679–1689, 2009.

[34] A. Widmann and E. Schrger, "Filter effects and filter artifacts in the analysis of electrophysiological data," *Frontiers in Psychology*, vol. 3, no. 233, 2012.

[35] L. Wang, X. Long, R. M. Aarts, J. P. van Dijk, and J. B. Arends, "A broadband method of quantifying phase synchronization for discriminating seizure eeg signals," *Biomedical Signal Processing and Control*, vol. 52, pp. 371–383, 2019.

[36] P. Prado-Gutierrez, E. Martnez-Montes, A. Weinstein, and M. Zaartu, "Estimation of auditory steady-state responses based on the averaging of independent eeg epochs," *PloS one*, vol. 14, no. 1, e0206018, 2019.

[37] T. Popov, R. Oostenveld, and J. M. Schoffelen, "Fieldtrip made easy: An analysis protocol for group analysis of the auditory steady state brain response in time, frequency, and space," *Frontiers in Neuroscience*, vol. 12, 2018.

[38] B. Van Dun, J. Wouters, and M. Moonen, "Optimal electrode selection for multi-channel electroencephalogram based detection of auditory steady-state responses," *The Journal of the Acoustical Society of America*, vol. 126, no. 1, pp. 254–268, 2009.

[39] M. X. Cohen and R. Gulbinaite, *Rhythmic entrainment source separation: Optimizing analyses of neural responses to rhythmic sensory stimulation. biorxiv*, Generic, 2016.

[40] D. W. Schwarz and P. Taylor, "Human auditory steady state responses to binaural and monaural beats," *Clin Neurophysiol*, vol. 116, no. 3, pp. 658–68, 2005, Schwarz, D W F Taylor, P eng Comparative Study Research Support, Non-U.S. Gov't Netherlands Clin Neurophysiol. 2005 Mar;116(3):658-68. doi: 10.1016/j.clinph.2004.09.014.

[41] E. D. Farahani, T. Goossens, J. Wouters, and A. van Wieringen, "Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway," *Neuroimage*, vol. 148, pp. 240–253, 2017, Farahani, Ehsan Darestani Goossens, Tine Wouters, Jan van Wieringen, Astrid eng Neuroimage. 2017 Mar 1;148:240-253. doi: 10.1016/j.neuroimage.2017.01.032. Epub 2017 Jan 18.

[42] G. Plourde, A. Garcia-Asensi, S. Backman, A. Deschamps, D. Chartrand, P. Fiset, and T. W. Picton, "Attenuation of the 40-hertz auditory steady state response by propofol involves the cortical and subcortical generators," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 108, no. 2, pp. 233–242, 2008.

[43] J. Van Opstal, *The auditory system and human sound-localization behavior*. Academic Press, 2016, ISBN: 0128017252.

[44] N. H. Salminen, S. J. Jones, G. B. Christianson, T. Marquardt, and D. McAlpine, "A common periodic representation of interaural time differences in mammalian cortex," *NeuroImage*, vol. 167, pp. 95–103, 2018.

[45] T. C. Yin, "Neural mechanisms of encoding binaural localization cues in the auditory brainstem," in *Integrative functions in the mammalian auditory pathway*. Springer, 2002, pp. 99–159.

[46] K. Vonderschen and H. Wagner, "Detecting interaural time differences and remodeling their representation," *Trends in neurosciences*, vol. 37, no. 5, pp. 289–300, 2014.

[47] S. Verhulst, A. Altoe, and V. Vasilkov, "Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss," *Hearing research*, vol. 360, pp. 55–75, 2018.