A Procedural Framework for Auditory Steady-State Response Detection

Bram Van Dun*, *Student Member, IEEE*, Geert Rombouts, Jan Wouters, and Marc Moonen, *Fellow, IEEE*

***Abstract*—Auditory steady-state responses (ASSRs) in EEG measurements are currently used for reliable hearing threshold estimation at audiometric frequencies. Especially, newborns with hearing problems benefit from this technique, as with this information, diagnosis can be better specified and hearing aids can be better fitted at an early age. Unfortunately, measurement duration is still very long for clinical widespread use due to the lack of efficient signal detection techniques with sufficient robustness against artifacts. In this paper, a simplified procedural framework for ASSR detection is worked out that allows the development of a multichannel processing strategy, starting from a detection theory approach. It is shown that a sufficient statistic can be calculated that best captures the amount of ASSR in the recorded data. The evaluation is conducted using data from ten normal-hearing adults. It is concluded that most single- and multichannel approaches are similar in performance when applied to uncontaminated EEG. When artifact-rich EEG is used, the proposed detection-theory-based approach significantly improves the number of ASSR detections compared with a noise-weighted common EEG channel derivation (vertex–occiput).**

***Index Terms*—Auditory steady-state response (ASSR), detection theory, electroencephalogram, independent component analysis (ICA), multichannel, multichannel Wiener filtering (MWF), noise-weighted averaging, sufficient statistic.**

I. Introduction

In 1994, the Joint Committee on Infant Hearing (JCIH) stated that universal detection of infants with a hearing loss is recommended before the age of three months [1]. Therefore, the need for objective audiometric techniques has greatly increased over the last decade. As the number of early hearing screened newborns grows worldwide, an appropriate clinical response to the need for rehabilitation of hearing problems is necessary. An effective treatment may consist of the use of a hearing aid. However, the fitting of this device requires a frequency-specific estimation of the newborn's hearing thresholds. These estimates can be provided using auditory steady-state response (ASSR) detection long before first behavioral thresholds can be obtained. Unfortunately, a reliable ASSR-based hearing threshold estimation procedure with adults already takes 45 min and can last several hours when having newborns as subjects [2]. The measured signals are often corrupted by artifacts due to the newborn's uncontrolled activity. The general aim is thus to increase robustness against artifacts and reduce the measurement time.

The ASSR is an electrophysiological response of the brain evoked by acoustic stimuli. These stimuli are generally based on amplitude and/or frequency-modulated sinusoidal carriers. The carriers stimulate specific parts of the cochlea, while the modulations activate certain parts of the auditory system. The appearance of the modulator in the EEG can be used as a label for the carrier. The first significant ASSR study was conducted by Galambos *et al.* in 1981 with modulation frequencies of 40 Hz [3]. Subsequently, extensive research showed that the ASSR can indeed provide an objective and frequency-specific way to determine reliable hearing thresholds, both applicable to adults and infants (refer to [4] for an extensive overview).

Responses to acoustic stimuli in general are being recorded for about 80 years [5]. In these early days, rather straightforward averaging procedures were developed to lower the background EEG and improve the ratio between the observed signal and noise [6]. Nowadays, more advanced detection techniques for certain classes of auditory responses have been considered, like those based on, e.g., wavelets and Bayesian networks for auditory brainstem responses [7], [8]. However, when it comes to ASSR detection, only a relatively small number of studies is available. Improvements have been achieved with techniques like adaptive regularized least squares (RLSs) filtering [9], noise-weighted averaging [10], and independent component analysis (ICA)/multichannel Wiener filtering (MWF) on multichannel EEG data [11], [12]. However, a *general* procedural approach for the detection of ASSRs in EEG noise has never been presented.

In this paper, a simplified procedural framework for ASSR detection is presented. Based on this framework, a multichannel processing strategy can be developed starting from a detection theory approach. It will be shown that a *sufficient statistic* can be calculated that best captures the amount of ASSR in the recorded data. This sufficient statistic can exploit spatiotemporal information present in the EEG measurements. The resulting

Manuscript received June 24, 2008; revised September 19, 2008. First published November 11, 2008; current version published May 6, 2009. This work was supported in part by the Institute for the Promotion of Innovation through Science, and Technology in Flanders (IWT–Vlaanderen), by the Fund for Scientific Research—Flanders (FWO) Project G.0504.04 (Design and analysis of signal processing procedures for objective audiometry in newborns), and by the Concerted Research Action GOA–AMBioRICS. *Asterisk indicates corresponding author.*

*B. Van Dun is with the Experimental Otorhinolaryngology (ExpORL), Department of Neurosciences, Katholieke Universiteit Leuven, 3000 Leuven, Belgium, and also with the SCD–System Identification, Signals, Telecommunication, and Automation (SISTA), Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, 3001 Leuven, Belgium (e-mail: bram.vandun@med.kuleuven.be).

G. Rombouts was with the SCD–System Identification, Signals, Telecommunication, and Automation (SISTA), Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, 3001 Leuven, Belgium (e-mail: geert.rombouts@esat.kuleuven.be).

J. Wouters is with the Experimental Otorhinolaryngology (ExpORL), Department of Neurosciences, Katholieke Universiteit Leuven, 3000 Leuven, Belgium (e-mail: jan.wouters@med.kuleuven.be).

M. Moonen is with the SCD–System Identification, Signals, Telecommunication, and Automation (SISTA), Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, 3001 Leuven, Belgium (e-mail: marc.moonen@esat.kuleuven.be).

Digital Object Identifier 10.1109/TBME.2008.2008395

0018-9294/$25.00 © 2009 IEEE

approach can be linked with the MWF approach and ICA approach of [11] and [12]. The evaluation is conducted using multichannel data from ten normal-hearing adults. Two types of EEG are used: EEG with few artifacts and EEG with a significant number of (controlled) artifacts. The performance of the newly developed sufficient statistic based approach is evaluated and compared with an efficient single-channel processing technique, noise-weighted averaging from [10], and with multichannel approaches as MWF and ICA.

II. Theoretical Background

This section presents a general ASSR signal model and a short overview of detection theory and its application to ASSR detection, including the exploitation of the spatiotemporal structure of the multichannel EEG. Some simplifications with relation to reality are discussed in detail. Links with other multichannel techniques are presented. Finally, a common method used for the evaluation of some of the processing schemes is described.

A. ASSR Signal Model

A simplified ASSR signal model can be given as

$$\mathbf{Z} = \alpha \mathbf{S}\mathbf{D}^T + \mathbf{N} \tag{1}$$

with $\mathbf{Z}^{n\times m}$ being an observation matrix, α a scalar representing the ASSR source amplitude proportional to the applied stimulus intensity, $\mathbf{S}^{n\times 2}$ a desired signal matrix with columns representing the ASSR (sinusoid and cosinusoid, oscillating at the known modulation frequency), $\mathbf{D}^{m\times 2}$ a steering matrix, $\mathbf{N}^{n\times m}$ an additive noise matrix, n the number of data points, and m the number of measured EEG channels.

In (1), given only $\mathbf{Z}$ and $\mathbf{S}$, $\alpha\mathbf{D}$ and $\mathbf{N}$ can be estimated using a QR factorization

$$\overbrace{[\,\mathbf{s}_1^{n\times 1} \quad \mathbf{s}_2^{n\times 1}}^{\mathbf{S}} \quad \mathbf{Z}^{n\times m}] = \mathbf{Q}^{n\times(m+2)}\mathbf{R}^{(m+2)\times(m+2)} \tag{2}$$

with

$$\mathbf{Q} = [\,\mathbf{s}_1 \quad \mathbf{s}_2 \quad \mathbf{Q}_*^{n\times m}\,] \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 & \hat{\mathbf{d}}_1^T \\ 0 & 1 & \hat{\mathbf{d}}_2^T \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_*^{m\times m} \end{bmatrix}. \tag{3}$$

$\mathbf{Z}$ can then be written as

$$\mathbf{Z} = \overbrace{[\,\mathbf{s}_1 \quad \mathbf{s}_2\,]}^{\mathbf{S}} \overbrace{\begin{bmatrix} \hat{\mathbf{d}}_1^T \\ \hat{\mathbf{d}}_2^T \end{bmatrix}}^{\hat{\mathbf{D}}^T} + \overbrace{\mathbf{Q}_*\mathbf{R}_*}^{\hat{\mathbf{N}}}. \tag{4}$$

Note that we have assumed that $\mathbf{S}^T\mathbf{S} = \mathbf{I}$ such that

$$\hat{\mathbf{D}}^T = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{Z} = \mathbf{S}^T\mathbf{Z} \tag{5}$$

which corresponds to a least-squares estimation of $\alpha\mathbf{D}^T$.

The ASSR generator has an unknown source amplitude α that depends on the stimulus level. After propagation through the skull, the distribution of the recorded ASSR on the electrodes, present in the observation matrix $\mathbf{Z}$, can be described by the steering matrix $\mathbf{D}$. The steering matrix $\mathbf{D}$, in contrast with the additive EEG noise, is usually assumed to be stationary as it is merely a representation of the source position, its directivity pattern, the electrode positions, and the propagation attenuation from source to electrode. Physically, no measurable delay occurs between the ASSR source and the electrodes (order of nanoseconds). Therefore, the ASSR delay difference and hence the ASSR phase difference between electrodes is considered zero. On the other hand, the delay between ASSR stimulus and response is measurable (order of milliseconds) but is unknown to the observer due to several physical parameters of the subject. The ASSR phase φ at the electrodes is thus unknown but equal in all channels. The aforementioned assumptions may be simplifications of reality. Any consequences hereof are discussed in Section II-E.

Under these assumptions, the (exact) steering matrix $\mathbf{D}$ is a rank-1 matrix

$$\underbrace{\mathbf{D}^T}_{2\times m} = \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix} \underbrace{\mathbf{d}^T}_{1\times m} \tag{6}$$

so that (1) can be rewritten as

$$\mathbf{Z} = \alpha\mathbf{S}\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}\mathbf{d}^T + \mathbf{N} \tag{7}$$

$$= \alpha\mathbf{s}\mathbf{d}^T + \mathbf{N} \tag{8}$$

where now $\mathbf{s}$ and $\mathbf{d}$ are vectors, and φ corresponds to the ASSR phase.

Based on (4), an estimate of φ can be obtained from

$$\hat{\varphi} = \arg\left\{\max_{\varphi} \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}^T \mathbf{S}^T\mathbf{Z}\mathbf{Z}^T\mathbf{S}\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}\right\} \tag{9}$$

$$= \arg\left\{\max_{\varphi} \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}^T \hat{\mathbf{D}}^T\hat{\mathbf{D}}\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}\right\} \tag{10}$$

or alternatively, including a prewhitening transformation,

$$\hat{\varphi} = \arg\left\{\max_{\varphi} \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}^T \hat{\mathbf{D}}^T(\hat{\mathbf{N}}^T\hat{\mathbf{N}})^{-1}\hat{\mathbf{D}}\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}\right\}. \tag{11}$$

Based on $\hat{\varphi}$, one can then compute

$$\hat{\mathbf{d}} = \hat{\mathbf{D}}\begin{bmatrix} \cos\hat{\varphi} \\ \sin\hat{\varphi} \end{bmatrix}. \tag{12}$$

A spatiotemporal noise covariance matrix $\mathbf{K}^{mn\times mn}$ can be defined as

$$\mathbf{K} = \mathcal{E}\{\mathbf{n}\mathbf{n}^T\} \tag{13}$$

with $\mathbf{n} = \mathrm{vec}(\mathbf{N})$. Here, $\mathcal{E}\{\cdot\}$ is the expected value operator and the $\mathrm{vec}(\cdot)$ operator stacks the columns of a matrix $\mathbf{X}$ into one column vector $\mathbf{x} = \mathrm{vec}(\mathbf{X})$.

If the spatial and temporal correlations are separable, as will be observed here, the spatiotemporal noise covariance matrix $\mathbf{K}$ can be written as [13]

$$\mathbf{K} = \mathbf{K}_{\mathrm{spat}} \otimes \mathbf{K}_{\mathrm{temp}} \tag{14}$$

where $\otimes$ represents the Kronecker product, with a spatial noise covariance matrix

$$\mathbf{K}_{\text{spat}} = \mathcal{E}\{\mathbf{N}^T\mathbf{N}\} \tag{15}$$

and a temporal noise covariance matrix

$$\mathbf{K}_{\text{temp}} = \mathcal{E}\{\mathbf{N}\mathbf{N}^T\}. \tag{16}$$

Based on (4), the noise covariance matrices can be estimated as

$$\hat{\mathbf{K}}_{\text{spat}} = \hat{\mathbf{N}}^T\hat{\mathbf{N}} \qquad \hat{\mathbf{K}}_{\text{temp}} = \hat{\mathbf{N}}\hat{\mathbf{N}}^T. \tag{17}$$

B. Detection Theory: Detecting Signals With Unknown Amplitude in Noise

Detection theory is a means to quantify the ability to discern between the signal and noise [14]. As the ASSR is buried in noise, originating from sources both inside and outside the skull, an approach that finds its origins in the realm of detection theory seems to be a valid one.

Assume that a target signal $\tilde{\mathbf{s}}$ has an unknown amplitude α. The hypotheses H_i can be stated as

$$H_0 : \mathbf{z} = \mathbf{n} \tag{18}$$

$$H_1 : \mathbf{z} = \alpha\tilde{\mathbf{s}} + \mathbf{n}, \quad \alpha = ? \tag{19}$$

with $\mathbf{z}$ being a vector containing a number of observations. As the signal's waveform $\tilde{\mathbf{s}}$ is known exactly, the noise $\mathbf{n}$ is the only stochastic component

$$p_{\mathbf{z}|H_i}(\mathbf{z}|H_i) = p_{\mathbf{n}}(\mathbf{z} - \tilde{\mathbf{s}}_i) \tag{20}$$

where $\tilde{\mathbf{s}}_0 = \mathbf{0}$ and $\tilde{\mathbf{s}}_1 = \alpha\tilde{\mathbf{s}}$.

This way, only the distribution of the noise is of importance. Assume the noise is colored and Gaussian with a distribution

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{\sqrt{\det(2\pi\mathbf{K})}} \exp\left(-\frac{1}{2}\mathbf{n}^T\mathbf{K}^{-1}\mathbf{n}\right). \tag{21}$$

The likelihood ratio $\Lambda(\mathbf{z})$ can be defined as

$$\Lambda(\mathbf{z}) = \frac{p_{\mathbf{n}}(\mathbf{z} - \alpha\tilde{\mathbf{s}})}{p_{\mathbf{n}}(\mathbf{z})} \tag{22}$$

$$= \frac{\exp(-\frac{1}{2}(\mathbf{z} - \alpha\tilde{\mathbf{s}})^T\mathbf{K}^{-1}(\mathbf{z} - \alpha\tilde{\mathbf{s}}))}{\exp(-\frac{1}{2}\mathbf{z}^T\mathbf{K}^{-1}\mathbf{z})}. \tag{23}$$

This likelihood ratio is compared to a certain threshold η for detection

$$\Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{24}$$

If the left-hand side is larger than η, H_1 is chosen, and H_0 otherwise.

The logarithm of both sides is taken, and after simplification, assuming that $\alpha > 0$, one can write [13]

$$\mathbf{z}^T\mathbf{K}^{-1}\tilde{\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{\alpha}\ln\eta + \alpha\tilde{\mathbf{s}}^T\mathbf{K}^{-1}\tilde{\mathbf{s}}. \tag{25}$$

The left-hand side of (25) is defined as the *sufficient statistic* $\Upsilon(\mathbf{z})$

$$\Upsilon(\mathbf{z}) = \mathbf{z}^T\mathbf{K}^{-1}\tilde{\mathbf{s}}. \tag{26}$$

The right-hand side of (25) is the threshold term. Despite its explicit dependence on a variety of factors, it is sufficient to determine the threshold term by specifying a false-alarm probability only, as a *uniformly most powerful test* exists. This special case arises when (as in the current case) both the sufficient statistic and one of the hypotheses (here H_0) does not depend on the unknown parameter α. Otherwise, the detection problem cannot be solved without inserting some value for α [13].

C. Detection Theory Framework for ASSR Processing

The ASSR signal model of (8) can be rewritten as

$$\mathbf{z} = \alpha\overbrace{\mathbf{d}\otimes\mathbf{s}}^{\tilde{\mathbf{s}}} + \mathbf{n} \tag{27}$$

with $\mathbf{z} = \text{vec}(\mathbf{Z})$ and $\mathbf{n} = \text{vec}(\mathbf{N})$, respectively.

The sufficient statistic formula (26) can be applied here [by identifying (27) with (19)]

$$\Upsilon_A(\mathbf{z}) = \mathbf{z}^T\mathbf{K}^{-1}\tilde{\mathbf{s}}. \tag{28}$$

Replacing $\mathbf{K}$ by using (14) and $\tilde{\mathbf{s}}$ by using (27), one can write

$$\Upsilon_A(\mathbf{z}) = \mathbf{z}^T(\mathbf{K}_{\text{spat}}\otimes\mathbf{K}_{\text{temp}})^{-1}(\mathbf{d}\otimes\mathbf{s}) \tag{29}$$

$$= \text{vec}(\mathbf{Z})^T[(\mathbf{K}_{\text{spat}}^{-1}\mathbf{d})\otimes(\mathbf{K}_{\text{temp}}^{-1}\mathbf{s})] \tag{30}$$

$$= \text{vec}[(\mathbf{K}_{\text{temp}}^{-1}\mathbf{s})^T\mathbf{Z}(\mathbf{K}_{\text{spat}}^{-1}\mathbf{d})] \tag{31}$$

$$= \mathbf{s}^T\mathbf{K}_{\text{temp}}^{-T}\mathbf{Z}\mathbf{K}_{\text{spat}}^{-1}\mathbf{d}. \tag{32}$$

Substituting the estimates based on (4), (11), and (17) then leads to a sufficient statistic $\hat{\Upsilon}_A(\mathbf{z})$ suitable for ASSR detection

$$\hat{\Upsilon}_A(\mathbf{z}) = \begin{bmatrix}\cos\hat{\varphi}\\ \sin\hat{\varphi}\end{bmatrix}^T \mathbf{S}^T\hat{\mathbf{K}}_{\text{temp}}^{-T}\mathbf{Z}\hat{\mathbf{K}}_{\text{spat}}^{-1}\hat{\mathbf{D}}\begin{bmatrix}\cos\hat{\varphi}\\ \sin\hat{\varphi}\end{bmatrix}. \tag{33}$$

D. Exploiting ASSR Stationarity and Spatiotemporal EEG Stationarity

In the following paragraphs, both $\mathbf{K}_{\text{spat}}$ and $\mathbf{K}_{\text{temp}}$ are discussed. Different assumptions on the stationarity of the EEG noise lead to different detection procedures. In the first two paragraphs, two approaches to determine $\mathbf{K}_{\text{spat}}$ are considered. In the third and last paragraphs, the structure of $\mathbf{K}_{\text{temp}}$ is studied.

1) $\mathbf{K}_{\text{spat}}$ *Constant Throughout the Experiment:* The steering vector $\mathbf{d}$ may be assumed constant during the same experiment, as described in Section II-A. When the EEG noise is stationary over all channels throughout the experiment, $\mathbf{K}_{\text{spat}}$ is also constant. The term $\mathbf{K}_{\text{spat}}^{-1}\mathbf{d}$ in (32) then represents an optimal weight vector $\mathbf{w}_{\text{opt}}$ that combines the m EEG channels (columns of observation matrix $\mathbf{Z}$) into one channel, such that the SNR of the combined signal $\mathbf{Z}\mathbf{w}_{\text{opt}}$ is maximized

$$\mathbf{w}_{\text{opt}} = \mathbf{K}_{\text{spat}}^{-1}\mathbf{d}. \tag{34}$$

In (33), an estimate of $\mathbf{w}_{\text{opt}}$ is used, namely

$$\hat{\mathbf{w}}_{\text{opt}} = \hat{\mathbf{K}}_{\text{spat}}^{-1}\hat{\mathbf{D}}\begin{bmatrix}\cos\hat{\varphi}\\ \sin\hat{\varphi}\end{bmatrix}. \tag{35}$$

When $\hat{\varphi}$ is computed with (11), this $\hat{\mathbf{w}}_{\text{opt}}$ corresponds to the MWF solution [11], [15].

This optimal weight vector can be expressed alternatively as

$$\mathbf{w}_{\mathrm{opt}} = \mathbf{v}_{\max}\Big\{\mathrm{GEVD}(\mathbf{D}\mathbf{D}^T, \mathbf{K}_{\mathrm{spat}})\Big\} \tag{36}$$

with $\mathrm{GEVD}(\cdot)$ being the generalized eigenvalue decomposition and $\mathbf{v}_{\max}\{\cdot\}$ being the eigenvector associated with the largest eigenvalue (see also Section II-F).

2) $\mathbf{K}_{\mathrm{spat}}$ *Varying During Experiment:* When it is assumed that noise sources emerge and disappear uncorrelated over time on the different recorded channels during an experiment, $\mathbf{K}_{\mathrm{spat}}$ cannot be considered constant anymore, and so, (33) needs to be modified. If the noise in $\mathbf{Z}$ is stationary only in blocks of T samples (rows) and if the noise is assumed uncorrelated between such blocks, then $\Upsilon_A(\mathbf{z})$ can be calculated as the sum of the sufficient statistics $\Upsilon_{A,i}(\mathbf{z}_i)$, with each block $\mathbf{Z}_i$ separately processed

$$\Upsilon_A(\mathbf{z}) = \sum_{i=1}^{\frac{n}{T}} \mathbf{s}_i^T \mathbf{K}_{\mathrm{temp},i}^{-T} \mathbf{Z}_i \mathbf{K}_{\mathrm{spat},i}^{-1} \mathbf{d} \tag{37}$$

$$= \sum_{i=1}^{\frac{n}{T}} \Upsilon_{A,i}(\mathbf{z}_i). \tag{38}$$

For each $\Upsilon_{A,i}(\mathbf{z}_i)$, an approximation as in (33) can then be substituted.

When implemented practically, the steering vector $\mathbf{d}$ is estimated on the entire observation matrix $\mathbf{Z}$ as in (4), (11), and (12) to exploit the ASSR stationarity, encompassing ASSR source position, directivity pattern, electrode positions, and propagation attenuation. Each stationary block $\mathbf{Z}_i$ generates a $\mathbf{K}_{\mathrm{spat},i}$, as it is assumed that the spatial covariance matrix of each block $\mathbf{Z}_i$ is different. An optimal weight vector $\mathbf{w}_{\mathrm{opt},i}$ is calculated for each stationary block $\mathbf{Z}_i$ separately, and (38) sums all $\hat{\Upsilon}_{A,i}(\mathbf{z})$ into one sufficient statistic $\hat{\Upsilon}_A(\mathbf{z})$.

3) Structure of $\mathbf{K}_{\mathrm{temp}}$*:* The temporal noise covariance matrix $\mathbf{K}_{\mathrm{temp}}$ expresses the correlation of noise samples within the same channel. The simplest case is when $\mathbf{K}_{\mathrm{temp}}$ is a diagonal matrix $\sigma\mathbf{I}$, i.e., when it is assumed that the noise is white with power σ^2, uncorrelated, and stationary within the data block. In a more general case, $\mathbf{K}_{\mathrm{temp}}$ is not a diagonal matrix, but assumed to have a Toeplitz structure, such that the correlation between noise samples, and thus the "color" of the noise spectrum can be incorporated. If an experiment is not entirely stationary, it can be divided in stationary blocks. The practical calculation of $\mathbf{K}_{\mathrm{temp}}$ is described in Section III-B.

E. Simplifications of the Framework Compared With Reality

The framework presented in this paper is a simplification of reality. The model assumes only one intracranial source. Healthy adults, however, have at least two intracranial sources [16], [17]. This has some implications. Formula (7) should then be reformulated as

$$\mathbf{Z} = \sum_{i=1}^{q} \alpha_i(t)\mathbf{S}\begin{bmatrix}\cos\,\varphi_i(t)\\ \sin\,\varphi_i(t)\end{bmatrix}\mathbf{d}_i^T + \mathbf{N} \tag{39}$$

with the observation matrix $\mathbf{Z}$ containing q sources from an ASSR with a specific modulation frequency. It is assumed in (7) that an ASSR source is constant in amplitude α and phase φ over time. However, this cannot be guaranteed. For example, ASSR amplitudes mainly originating from the auditory cortex vary significantly with arousal [3]. With more than one ASSR source, the assumption that the resulting phase φ is equal at all electrodes is not valid either, unless these sources are exactly in phase.

The aforementioned concerns are assumed to have only a small effect. According to [17], the general two sources from an ASSR with a specific modulation frequency are a main source in the brainstem and a main source in the auditory cortex. The modulation frequencies used in this study are chosen in a region with the brainstem source being dominant (80–110 Hz). Therefore, the approximation by the model has reasonable validity.

Some aspects can still be addressed by the current model. If present, a varying phase $\varphi(t)$ at the electrodes, resulting from separate $\varphi_i(t)$ from different sources, can be reassessed to each block $\mathbf{Z}_i$ using (38).

The biggest concern is the estimation of the ASSR amplitude $\alpha_i(t)$, being incorporated in the least square estimation of $\alpha_i(t)\mathbf{D}_i$, as $\alpha_i(t)$ is estimated on the entire observation matrix $\mathbf{Z}$ using (1). This way, fluctuating amplitudes cannot be captured. However, using less data compromises the assessment of $\alpha_i(t)$, as the ASSR amplitude is already small compared to the surrounding noise.

F. Alternative Approaches for Calculation of $\mathbf{w}_{\mathrm{opt}}$

The weight vector $\mathbf{w}_{\mathrm{opt}}$, as defined in (34), combines the m-channel EEG signal into one channel. It is interesting to note that there exist alternative approaches to calculate a weight vector $\mathbf{w}'_{\mathrm{opt}}$, while still reflecting a high degree of similarity with (34). For example, an optimal weight vector can also be expressed by

$$\mathbf{w}'_{\mathrm{opt}} = \mathbf{W}\mathbf{A} \tag{40}$$

with $\mathbf{A}$ being a "selection matrix" selecting the weight column $\mathbf{w}'_{\mathrm{opt}}$. Here, $\mathbf{w}'_{\mathrm{opt}}$ is defined as the column of $\mathbf{W}$ that produces a combined signal with the highest SNR. Matrix $\mathbf{W}$ represents the demixing matrix for an ICA approach [18] and can be calculated using [19]

$$\mathbf{W} = \mathrm{GEVD}(\mathbf{Q}, \mathbf{K}_{\mathrm{spat}}) \tag{41}$$

with $\mathbf{Q}^{m\times m}$ being a "cross statistics" matrix containing the sum of the fourth-order cumulants of $\mathbf{Z}$

$$\mathbf{Q} = \mathcal{E}\{\mathbf{Z}^H\mathbf{Z}\mathbf{Z}^H\mathbf{Z}\} - \mathbf{Z}^T\mathbf{Z}\sum \mathrm{diag}(\mathbf{Z}^T\mathbf{Z}) - \mathcal{E}\{\mathbf{Z}^T\mathbf{Z}\}\mathcal{E}\{\mathbf{Z}^H\mathbf{Z}^*\} - \mathbf{Z}^T\mathbf{Z}\mathbf{Z}^T\mathbf{Z}. \tag{42}$$

Here, (42) is valid if the sources are assumed to be non-Gaussian and independent. For other assumptions about the sources, the expression for $\mathbf{Q}$ takes on different forms [19].

The observations in (36) and (41) link the theoretical aspects of ICA [18] and MWF [11], [15]. They are a possible explanation for the similar results when applying ICA and MWF on, e.g., ASSRs [11], [12].

TABLE I
RECORDING ELECTRODE POSITIONS FOR EIGHT-CHANNEL SETUP ACCORDING TO INTERNATIONAL 10–10 SYSTEM [23]

Channel	Position
1	occiput (Oz)
2	P4
3	P3
4	right mastoid (rMa)
5	left mastoid (lMa)
6	F4
7	F3
8	forehead (Fpz)
reference	vertex (Cz)
ground	right clavicle

All channels are referenced to the reference electrode on top of the head (Cz).

G. Detection Using Fast Fourier Transformation (FFT)

In most ASSR studies, an FFT analysis is carried out prior to response detection [2], [20]–[22]. The detection is based on the ratio between the response power P_r at the modulation frequency and the mean noise power σ^2 in $M = 120$ neighboring frequency bins at each side

$$\frac{P_r}{\sigma^2} = \frac{a^2_{\text{response}}}{(1/M)\sum_{p=1}^{M} a^2_{\text{noise},p}} \tag{43}$$

with a_{response} being the amplitude of the modulation frequency and $a_{\text{noise},p}$ being the noise amplitude in the pth adjacent frequency bin. For standard block lengths of 16.384 s ("sweeps"), this is approximately 3.7 Hz on each side [20].

III. METHODS

This section describes a simulation setup that evaluates the framework and the sufficient-statistic-based ASSR detection developed in Section II. The different single- and multichannel schemes that are evaluated in Section IV are presented.

A. Evaluation Design

Ten normal hearing subjects (eight males and two females) with mean age 28.2 years (range 22–32 years) were selected. Their hearing thresholds did not exceed 20 dBHL on the octave audiometric frequencies. All experiments were carried out in a double-walled soundproof room with Faraday cage. Subjects were asked to lie down on a bed and relax or sleep. Lights were switched off. All experiments were identically carried out a second time several days or weeks after the first experiment. This way test–retest comparisons could be done.

Kendall jelly snap electrodes were placed on the positions described in Table I according to the international 10–10 system [23]. This configuration was chosen similar to [24] with some extra channels added to ensure a symmetrical configuration of all electrodes. They were placed on the subject's scalp after the skin was abraded with Nuprep abrasive skin prepping gel. A conductive paste was used to keep the electrodes in place and to avoid that interelectrode impedances exceeded 5 kΩ at 30 Hz. The electrodes were connected to a low-noise eight-channel Jaeger–Toennies amplifier. Each EEG channel was amplified ($\times$50 000) and bandpass filtered between 70 and 170 Hz (6 dB/octave). The sampling rate was set equal to 1000 Hz and downsampled later on to 250 Hz. No artifact rejection was applied initially, but a threshold was determined offline that rejected around 10% of the recorded data blocks ("epochs") that exceeded this threshold. All separate acoustic stimuli were calibrated at 70 dBSPL, using a Brüel & Kjær Sound Level Meter 2260 in combination with a 2-cc coupler DB0138 and an artificial ear 4152. All stimuli were presented to the subject and amplified EEG signals recorded using the Setup ORL for multichannel ASSR (SOMA) program from [25] and an RME multichannel soundcard.

Two combined stimuli, with four 100% amplitude-modulated (AM) and 20% frequency-modulated (FM) carrier frequencies each, were applied to each ear. The carrier frequencies were the same for both ears, namely 0.5, 1, 2, and 4 kHz. The modulation frequencies were taken close to 82, 90, 98, and 106 Hz for the left ear and 86, 94, 102, and 110 Hz for the right ear, respectively. These modulation frequencies were adjusted to ensure that a nonfractional number of modulation cycles fitted into one data block ("epoch") of 1.024 s. For example, 82 Hz is converted into $[\text{round}(82 * 1.024)]/1.024$ Hz [20]. Throughout the rest of the paper, one will refer to the nonadjusted frequencies for reasons of conciseness.

Stimuli were applied at 30 dBSPL (36 sweeps, each sweep lasting 16.384 s). After EEG data collection, each separate channel, or each combination of channels, was reduced to 32 sweeps, using an artifact rejection threshold that removed exactly four sweeps per channel (or per combination of channels). Artifact rejection for multichannel data (i.e., a combination of channels) implies the removal of all simultaneous epochs over different channels. Otherwise, any correlation over simultaneous channels will be removed. This EEG dataset is referred to as *clean EEG* throughout the rest of the paper.

For the performance analysis of the different processing schemes in Section III-B, the number of ASSR detections is counted. It is assumed that all responses to the applied stimuli are present in the EEG as the stimulus application intensity is 30 dBSPL and well above subject's hearing thresholds. There are thus 16 responses per subject (test and retest) available for detection. This way, the maximum number of detections is 160. For single-channel measurements, the detection threshold is fixed at 4.82 dB SNR using (43), which corresponds to 5% allowed false detections. To calculate a detection threshold for multichannel data, all multichannel processing methods are applied on frequencies without a response (e.g., 1 Hz below the modulation frequencies that could contain a response) of the artifact-free multichannel EEG on the current intensity of 30 dBSPL, extended with extra multichannel EEG from the same subjects. The total EEG data length used here was six times longer than the EEG data length for the detection of the responses. Afterwards, the 95 percentile of this noise distribution is defined as the detection threshold.

For the analysis of robustness against artifacts, the EEG data from the previous paragraph are used together with extra measurements that encouraged the generation of artifacts. Subjects were asked to sit on a chair. They carried out a repeating

series of movements in cycles of approximately 6 s while the two combined stimuli described before were applied at an intensity of 30 dBSPL. Measurements were 32 sweeps long. No artifact rejection was applied. The movements were carried out in the following order: look up, down, left, and right. This procedure served as a controlled generator of artifacts on all channels due to muscle activity and electrode cable movement. This EEG dataset is referred to as *dirty EEG* throughout the rest of the paper.

B. Nine Processing Schemes

Nine different processing schemes are evaluated in this section. Five of them are based on the sufficient statistic $\Upsilon_A(\mathbf{z})$ of (37), with varying restrictions on $\mathbf{K}_{\text{spat}}$ and $\mathbf{K}_{\text{temp}}$. The schemes are described later. Optimal channels, channel combinations, and block sizes are determined in Section IV.

1) Scheme 1—Single Channel, Normal Averaging: This is the standard approach to process ASSR data [2], [20], [21]. The provided EEG data stream is divided into 32 sweeps (data blocks of 16.384 s) and averaged. Detection is managed using the method described in Section II-G.

2) Scheme 2—Single Channel, Noise-Weighted Averaging: Each unfiltered epoch (data block of 1.024 s) is transformed to the frequency domain using an FFT. The average power P_i between 77 and 115 Hz is computed after removing the power at the eight frequencies at which responses occurred. The time-domain epoch is then weighted with the average power P_i and concatenated with the preceding epochs to form sweeps. Each epoch of the final summed sweep is then divided by the sum of the weights $P = \sum P_i^{-1}$ of the epochs that has been combined to form that particular epoch (adapted from [10]).

3) Scheme 3—Independent Component Analysis: This approach is presented in [12]. The weight vector $\mathbf{w}'_{\text{opt}}$ is calculated using (40). As a substitute for (41), however, the joint approximate diagonalization of eigenmatrices (JADE) algorithm has been used for better numerical stability [26]. Before application of the algorithm, data are filtered between 77 and 115 Hz and averaged into one sweep to increase the SNR of the ASSRs as ICA does not perform well under conditions with low SNR [18]. After reducing the eight channels to one channel $\mathbf{Z}\mathbf{w}'_{\text{opt}}$, detection is managed according to Section II-G.

4) Scheme 4—Multichannel Wiener Filter: This approach is presented in [11]. The weight vector $\mathbf{w}_{\text{opt}}$ is calculated using (36) on the entire observation matrix $\mathbf{Z}$. After reducing the eight channels to one channel $\mathbf{Z}\mathbf{w}_{\text{opt}}$, detection is carried out following Section II-G.

5) Scheme 5—$\Upsilon_A(\mathbf{z})$, $\mathbf{K}_{\text{spat}}$ Fixed, $\mathbf{K}_{\text{temp}} = \mathbf{I}$ Fixed: In (37), $\mathbf{K}_{\text{spat}}$ is estimated only once on the entire observation matrix $\mathbf{Z}$ using (17), or equivalently, $\mathbf{K}_{\text{spat},i} = \mathbf{K}_{\text{spat}}$ in (37). $\mathbf{K}_{\text{temp}}$ is assumed to be equal to the identity matrix $\mathbf{I}$, implying that the EEG noise is assumed to be white, uncorrelated, and with constant noise power $\sigma^2 = 1$.

6) Scheme 6—$\Upsilon_A(\mathbf{z})$, $\mathbf{K}_{\text{spat}}$ Fixed, $\mathbf{K}_{\text{temp}} = \sigma\mathbf{I}$ Fixed: $\mathbf{K}_{\text{spat}}$ is estimated only once on the entire observation matrix $\mathbf{Z}$ using (17). $\mathbf{K}_{\text{temp}}$ is now scaled by the square root of the noise power σ^2, calculated by applying an FFT to the single-channel result $\mathbf{Z}\mathbf{w}_{\text{opt}}$, using $\mathbf{w}_{\text{opt}}$ from (35), and applying (43) afterwards. This way, $\mathbf{K}_{\text{temp}} = \sigma\mathbf{I}$ accounts for the noise power that is varying from observation to observation. It is important to note that only the noise power of the local spectrum near to the ASSR frequency is relevant, providing an estimate of the noise power at the reference frequency. Estimation of $\mathbf{K}_{\text{temp}}$ using (17) is inferior, as useless information of other frequency bands is taken into account. Note that when $\mathbf{K}_{\text{temp}}$ is chosen to be a Toeplitz matrix (and thus sample correlation is incorporated), performance is found to degrade significantly. While in theory, a corresponding adequate estimation of $\mathbf{K}_{\text{temp}}$ should increase performance; practically, this type of estimation is found to be difficult.

7) Scheme 7—$\Upsilon_A(\mathbf{z})$, $\mathbf{K}_{\text{spat},i}$ Variable, $\mathbf{K}_{\text{temp}} = \mathbf{I}$ Fixed: In (37), $\mathbf{K}_{\text{spat},i}$ is estimated for each block $\mathbf{Z}_i$, based on (17)

$$\hat{\mathbf{K}}_{\text{spat},i} = \mathcal{E}\{\mathbf{N}_i^T\mathbf{N}_i\}. \tag{44}$$

8) Scheme 8—$\Upsilon_A(\mathbf{z})$, $\mathbf{K}_{\text{spat},i}$ Variable, $\mathbf{K}_{\text{temp},i} = \sigma_i\mathbf{I}$ Variable: $\mathbf{K}_{\text{spat},i}$ is estimated for each block $\mathbf{Z}_i$ using (44). $\mathbf{K}_{\text{temp},i}$ is scaled by the square root of the noise power σ_i^2, calculated by applying Section II-G to the single-channel result $\mathbf{Z}_i\hat{\mathbf{w}}_{\text{opt},i}$, using $\hat{\mathbf{w}}_{\text{opt},i} = \mathbf{v}_{\max}\{\text{GEVD}(\mathbf{D}^T\mathbf{D}, \hat{\mathbf{K}}_{\text{spat},i})\}$ from (36). The noise region of 2×3.7 Hz for the calculation of σ^2 is kept constant. The number of noise frequency bins M used for the estimate of σ^2 thus decreases with smaller block lengths of $\mathbf{Z}_i$.

9) Scheme 9—$\Upsilon_A(\mathbf{z})$, $\mathbf{K}_{\text{spat}}$ Fixed, $\mathbf{K}_{\text{temp},i} = P_i P \sigma_n I$ Variable: $\mathbf{K}_{\text{temp},i}$ is calculated based on the rationale of noise-weighted averaging and recalculated each block $\mathbf{Z}_i$. P_i is the average power between 77 and 115 Hz of $\mathbf{Z}_i\mathbf{w}_{\text{opt}}$, with $\mathbf{w}_{\text{opt}}$ calculated using (34). P is the sum of the reciprocals of P_i, i.e., $P = \sum P_i^{-1}$. The mean noise power σ_n^2 is calculated using (43) on a noise-weighted average of $\mathbf{Z}\mathbf{w}_{\text{opt}}$.

IV. Results

Before the different processing schemes of Section III-B are evaluated all together for both "clean" and "dirty" EEG, the channel (combination) with the maximum number of ASSR detections is determined so that comparisons are done correctly. This is done for single-channel EEG and for the combination of multiple EEG channels.

Table II shows the number of detections for each of the eight separate channels of Table I. The test–retest statistic for the "clean EEG" (with few artifacts) over all channels is not significant ($p = 0.129$, no interactions). The data for the "dirty EEG" (with a significant number of artifacts) are not normally distributed. Based on Table II, it is decided to take channel 1 (vertex–occiput) as the reference channel for the "clean EEG" in the global evaluation further on, as this channel returns the most ASSR detections of all channels. For the "dirty EEG," channel 4 (vertex–right mastoid) is withheld as the reference channel. These optimal EEG derivations are confirmed by [24], who quote that a small set of three derivations (Cz-Oz combined with the right mastoid–Cz and left mastoid–Cz) yield the best SNRs in a larger number of participants than would be expected if all derivations were equally efficient. If one discards channel Cz–P4 from the top four list detected in Table II, a channel not being used by [24], the top three from this paper (channels 1,

TABLE II
NUMBER OF ASSR DETECTIONS FOR EACH SEPARATE CHANNEL FROM TABLE I FOR THREE TYPES OF EEG

Channel	'clean EEG' (with artifact rejection)	'dirty EEG' (without artifact rejection)	'dirty EEG' (with artifact rejection)	Sweeps X remaining after artifact rejection
	36 sweeps → 32 sweeps	32 sweeps → 32 sweeps	32 sweeps → X sweeps	$\hat{\mu}_X \pm \hat{\sigma}_X$
1 (Oz)	**116**	17	8	4.9 ± 8.1
2 (P4)	104	36	37	11.0 ± 10.5
3 (P3)	90	19	25	13.2 ± 10.9
4 (rMa)	104	**52**[3]	14	5.2 ± 7.5
5 (lMa)	100	27	16	4.8 ± 5.9
6 (F4)	46[1]	13	4	3.8 ± 6.8
7 (F3)	62[2]	8	3	5.8 ± 8.0
8 (Fpz)	55[2]	12	2	3.0 ± 5.5

First, "clean EEG" generated by relaxed subjects with about 10% of epochs removed exceeding a custom rejection threshold (36 to 32 sweeps). Second, "dirty EEG" generated by controlled head movement (Section III-A) without any artifact rejection (32 to 32 sweeps). Third, "dirty EEG" with the same rejection threshold as for "clean EEG," resulting in a serious length reduction (32 to X sweeps depicted in the rightmost column). The channels in bold are the channels with the highest number of ASSR detections for "clean EEG" and "dirty EEG," respectively.
[1]Significantly different from channel 1 ($p < 0.01$) after pairwise comparison with Bonferroni correction.
[2]Significantly different from channed 1 ($p < 0.05$) after pairwise comparison with Bonferroni correction.
[3]Data not normally distributed. A significant difference ($p < 0.001$) between channels exists.

TABLE III
NUMBER OF ASSR DETECTIONS IN "CLEAN" EEG (WITH ARTIFACT REJECTION) AND "DIRTY" EEG (WITHOUT ARTIFACT REJECTION) FOR INDIVIDUAL CHANNELS 1 AND 4 FROM TABLE I AND SPECIFIC CHANNEL COMBINATIONS CONSTRUCTED USING DECREASING DETECTION ORDER FROM TABLE II

Channel combination	'Clean EEG' (with artifact rejection)	'Dirty EEG' (without artifact rejection)	p–value (%) <SNR (dB)>
1	**116**	**17**[2]	5.0 <4.82>
4	**104**	**52**	5.0 <4.82>
1-4	117	46[3]	1.6 <6.28>
1-2-4	111	57	0.97 <6.75>
1-2-4-5	110	56	0.51 <7.32>
1-2-3-4-5	**114**[1]	**66**	0.37 <7.58>
1-2-3-4-5-7	111	66	0.23 <7.94>
1-2-3-4-5-7-8	105	63	0.10 <8.52>
1-2-3-4-5-6-7-8	101	65	0.063 <8.81>

Channel combinations are processed using (33). The rightmost column shows the detection thresholds used for each channel (combination), The individual channels in bold are the channels with the highest number of ASSR detections for "clean EEG" and "dirty EEG," respectively. The channel combination in bold returns the most ASSR detections in general.
[1]No significant differences with other channel combinations in the same column.
[2]Significantly different from combination 1-2-3-4-5 ($p < 0.01$) after pairwise comparison with Bonferroni correction.
[3]Significantly different from combination 1-2-3-4-5 ($p < 0.05$) after pairwise comparison with Bonferroni correction.

TABLE IV
NUMBER OF ASSR DETECTIONS IN "CLEAN" AND "DIRTY" EEG FOR SCHEMES 7–9 FROM SECTION III-B, WHILE VARYING BLOCK SIZES OF PARTIAL OBSERVATION MATRICES $\mathbf{Z}_i$

Samples	32	64	128	256	512	1024	**2048**	4096
Seconds	0.128	0.256	0.512	1.024	2.048	4.096	**8.192**	16.384
7 (clean)	69	73	70	74	73	74	**74**	75
7 (dirty)	54	58	66	71	75	75	**82**	79
8 (clean)	92	98	96	100	101	104	**104**	105
8 (dirty)	36	48	49	47	46	46	**55**	52
9 (clean)	91	117	116	113	116	114	**114**	115
9 (dirty)	27	61	60	64	69	73	**82**	78

The block size in bold returns the most ASSR detections in general.

4, and 5) correspond to their recommended channels. The table finally shows the effect of applying the same artifact rejection level of the "clean" EEG channels to the "dirty" EEG channels. The number of ASSR detections degrades as only few EEG data sweeps are withheld. This shows that artifact rejection in the presence of many artifacts is not recommended.

Similarly, Table III shows the number of detections for a specific set of channels that are combined using (33). The "clean" and "dirty" EEG test–retest statistics are not significant ($p = 0.413$ and $p = 0.155$, respectively, no interactions). Channels are added incrementally using the order of decreasing number of ASSR detections from the second column of Table II. Due to the statistical multiple testing syndrome, each added channel increases the number of false detections. Therefore, the detection threshold needs to be made more strict, which is depicted in the fourth column of Table III for p-values and SNR. Channel combination 1–3 (121 detections for clean EEG, 23 for dirty EEG) and channel combination 2–3–4–5 (114 detections for clean EEG, 69 for dirty EEG) have the highest detections for clean and dirty EEG, respectively, for all possible combinations (not shown in the table). It is opted for the combination 1–2–3–4–5, however, as this channel combination contains both channels 1 and 4.

For the overall comparison between single-channel and multichannel techniques in the rest of this paper, channels 1 and 4 are used together with channel combination 1–2–3–4–5.

The optimal lengths for partial observation matrices $\mathbf{Z}_i$ for schemes 7–9 are determined in Table IV. Block sizes of $\mathbf{Z}_i$ are varied and the number of ASSR detections is calculated. This way, it is concluded that for schemes 7–9, the optimal block length is 8.192 s (or 2048 samples) to guarantee the highest number of detections for both "clean EEG" (with few artifacts) and "dirty EEG" (with a significant number of artifacts). For "clean EEG," this block length could correspond to the stationarity period of the EEG, but 8 s seems rather long for EEG stationarity in a frequency window of 70–170 Hz (Section III-A). Detection differences between block lengths are not large anyway. Practically the same performance is obtained for data blocks of 256 ms, which could be more plausible. In the case of "dirty EEG" (with the repeated controlled artifact generation), the 8-s period corresponds to the period of repetition. The detection values for this specific block length of 8.192 s are used in the overall comparison.

Table V shows the performance of the schemes, with statistical comparisons, for measurement lengths of 32 sweeps (approximately 9 min, one sweep being 16.384 s). "Clean" EEG (with few artifacts) and "dirty" EEG (with a significant number of artifacts) as described in Section III-A are used.

Noise-weighted averaging improves ASSR detection for both single-channel and multichannel approaches, especially in the

TABLE V
NUMBER OF ASSR DETECTIONS IN "CLEAN" AND "DIRTY" EEG FOR SCHEMES FROM SECTION III-B

Scheme	Description	'Clean EEG'	'Dirty EEG'
1a	normal averaging (ch 1)	116	17[3]
1b	normal averaging (ch 4)	104	52[4]
2a	**weighted averaging (ch 1)**	**120**	**35**[5]
2b	**weighted averaging (ch 4)**	**109**	**65**
3	ICA	124[2]	53
4[1]	MWF	114	66
5	$\mathbf{K}_{\text{spat}}$ fixed, $\mathbf{K}_{\text{temp}} = \mathbf{I}$ fixed	78	98
6[1]	$\mathbf{K}_{\text{spat}}$ fixed, $\mathbf{K}_{\text{temp}} = \sigma\mathbf{I}$ fixed	114	66
7	$\mathbf{K}_{\text{spat},i}$ var, $\mathbf{K}_{\text{temp},i} = \mathbf{I}$ fixed	74	82
8	$\mathbf{K}_{\text{spat},i}$ var, $\mathbf{K}_{\text{temp},i} = \sigma_i\mathbf{I}$ var	104	55[4]
9	**$\mathbf{K}_{\text{spat}}$ fixed, $\mathbf{K}_{\text{temp},i} = P_i P \sigma_n \mathbf{I}$ var**	**114**	**82**

Scheme 7--9 use block sizes of 8.192 s for $\mathbf{Z}_i$, according to Table IV. The single-channel schemes in bold are the schemes with the highest number of ASSR detections for "clean EEG" and "dirty EEG," respectively. The multichannel scheme in bold returns the most ASSR detections in general.
[1]The results from schemes 4 and 6 are identical.
[2]Significantly different from scheme 8 ($p < 0.05$) after pairwise comparison with Bonferroni correction.
[3]Significantly different from schemes 2b, 7, and 8 ($p < 0.05$), schemes 4--6 ($p < 0.01$), and scheme 9 ($p < 0.001$) after pairwise comparison with Bonferroni correction.
[4]Significantly different from scheme 9 ($p < 0.01$) after pairwise comparison with Bonferroni correction.
[5]Significantly different from schemes 5 and 9 ($p < 0.05$) after pairwise comparison with Bonferroni correction.

case of "dirty EEG," where a significant number of artifacts is present. This improvement is obtained for channel 1 (*17* detections for scheme 1a → *35* detections for scheme 2a) and channel 4 (*52* detections for scheme 1b → *65* detections for scheme 2b), and the multichannel combination 1–2–3–4–5 (*66* detections for scheme 6 → *82* detections for scheme 9). The differences are not always significant.

For the dataset used in this study, the number of ASSR detections is highest for channel 1 compared with channel 4 for "clean EEG" with few artifacts (*116* detections for scheme 1a versus *104* detections for scheme 1b). Channel 1 is the Cz–Oz (vertex–occiput) derivation that is used in several studies [24], [27] and is similar to the Cz–neck derivation used in other studies [21], [24], [28], [29]. However, the number of ASSR detections in channel 1 for "dirty EEG" is lower than the number of ASSR detections in channel 4 (*17* detections for scheme 1a versus *52* detections for scheme 1b). Based on this dataset, a difficult choice needs to be made if only three electrodes are available: channel 1 (vertex–occiput) from literature or channel 4 (vertex–right mastoid) that is more robust against artifacts. If more than three electrodes are available, this choice could be avoided using more channels and applying multichannel signal processing to the recorded data. Moreover, the optimal channels with the highest number of ASSR detections cannot be determined beforehand, together with the fact whether the subject will be relaxed (few artifacts) or stressed (lots of artifacts) during the measurement (or both).

The best single-channel results (schemes 2a and 2b) can be compared separately with the best multichannel result (scheme 9), thus discarding the pairwise comparisons and the Bonferroni correction. For the "dirty EEG," scheme 9 produces significantly ($p < 0.001$) more ASSR detections (*82* detections) than scheme 2a (*35* detections), which corresponds to the noise-weighted averaging of channel 1. Compared with the noise-weighted averaging of channel 4 (*65* detections for scheme 2b), the increased number of ASSR detections (*82* detections for scheme 9) is close to significance ($p = 0.063$). Thus, when the reference channel is taken as channel 1 (Cz–Oz) for most single-channel measurements, the application of scheme 9 to channel combination 1–2–3–4–5 outperforms a noise-weighted version of channel 1 (scheme 2a) significantly for EEG corrupted with artifacts. For EEG with few artifacts, the difference between approaches is not significant.

ICA (scheme 3) and MWF (scheme 4) are similar in performance. However, as ICA processing does not take into account any *a priori* information, except the strong assumption of independent sources, it was expected that the MWF approach, incorporating a known reference frequency, would perform better. This could indicate that the ICA approach is sufficient for ASSR detection. Results in previous studies already indicated large similarities between MWF and ICA processing [11], [12]. The biggest drawback of the ICA approach is that the available data need to be filtered and averaged into smaller data blocks first. Otherwise, ASSR amplitudes are too small in the observation matrix and the ICA algorithm cannot separate them from the noise. This shows that ICA does not perform well under conditions with low SNR [18].

The MWF approach (scheme 4) and scheme 6 ($\mathbf{K}_{\text{spat}}$ fixed, $\mathbf{K}_{\text{temp}} = \sigma\mathbf{I}$) are identical in performance. The approaches are characterized entirely by the optimal weight vector $\mathbf{w}_{\text{opt}}$ and the (square root of the) noise power σ^2.

Varying $\mathbf{K}_{\text{spat}}$ reduces detection performance for multichannel EEG recordings (scheme 5 → scheme 7, scheme 6 → scheme 8). This is rather unexpected. Pilot simulations using artificial ASSR data showed that spatially uncorrelated high-intensity noise bursts scattered over different EEG channels containing ASSRs could only be processed efficiently using a varying $\mathbf{K}_{\text{spat},i}$ per data block $\mathbf{Z}_i$. Within this data block, an uncorrupted channel could be selected from the several channels of the multichannel data block. Additional pilot simulations with spatially correlated noise bursts across channels, however, indicated that the effect of varying $\mathbf{K}_{\text{spat}}$ performed similar or even worse than keeping $\mathbf{K}_{\text{spat}}$ fixed. The results on EEG data from real subjects provided in this paper show that the second assumption may be correct. High-intensity noise bursts in the form of muscle or movement artifacts emerge spatially correlated on the different EEG channels and can be reduced by choosing an appropriate $\mathbf{K}_{\text{spat}} \sim \mathbf{w}_{\text{opt}}$ for the entire observation matrix.

$\mathbf{K}_{\text{temp}}$ models the (local) characteristics of the EEG signal. Increasing the precision of the estimation of $\mathbf{K}_{\text{temp}}$ improves ASSR detection performance. When considering schemes 5, 6, and 9, a gradual increase in the total number of detections ("clean EEG" + "dirty EEG") is observed. Scheme 5 does not incorporate any noise information at all in $\mathbf{K}_{\text{temp}}$. Its total number of detections is the lowest (*78* = *98* = *176* detections). The high value (*98* detections for scheme 5) of the "dirty EEG" dataset indicates that EEG with a significant number of artifacts is better without any noise power estimation as the power of the artifacts (noise bursts) is also present in σ^2 (*98* detections for scheme 5 → *66* detections for scheme 6). However, omitting σ^2 has a negative effect on "clean EEG" (*114* detections for scheme 6 →*78* detections for scheme 5). This opposing effect can be solved by a noise-weighted approach, weighting each data block

$\mathbf{Z}_i$ with the noise P_i in that block. The effect of the noise bursts protruding in σ^2 is still present (*98* detections for scheme 5 $\rightarrow$ *82* detections for scheme 9), but better assessed locally by the $P_i P$ terms (*66* detections for scheme 6 $\rightarrow$*82* detections for scheme 9).

A more precise assessment of $\mathbf{K}_{\mathrm{temp}}$ probably can be achieved by making $\mathbf{K}_{\mathrm{temp}}$ nondiagonal, incorporating any frequency-domain information (the "color" of the noise). Perhaps the fine-structured information of the EEG can also be modeled, taking blocks $\mathbf{Z}_i$ with a length close to dimensions where EEG stationarity may be assumed (e.g., less than 100 ms). However, neither approach succeeds on the EEG dataset used in this study.

V. General discussion

Most results have already been discussed in Section IV. Some general remarks are discussed next.

This is the first time, to the authors' best knowledge, that a detection theory approach is investigated for the processing of multichannel EEG data with the main purpose of improving ASSR detection. The presented results support the (simplified) framework and the detection theory approach described in this paper. When appropriate multichannel processing is applied (MWF, ICA, or the newly proposed procedural framework based on detection theory), multichannel measurements demonstrate an improvement compared to (noise-weighted) single-channel ASSR recordings. When its performance is compared to a noise-weighted version of a standard electrode configuration for single-channel measurements (vertex–occiput [2], [24]), this improvement is significant. The aforementioned observations are only valid when the available EEG is corrupted by artifacts originating from realistic head movements. Otherwise, performance is similar.

When focussing on situations where hearing threshold determination is of clinical relevance, the proposed method $\Upsilon_A(\mathbf{z})$ signifies a serious improvement for its robustness against artifacts. Unless the subject is sedated, ASSR measurements are difficult to conduct in a short period of time because of a large number of artifacts due to movement, distress, and agitation, particularly in the patient population where nowadays ASSRs are mostly being applied (in neonates and young children). This type of recording sessions are rarely described in literature, but are unfortunately very current in clinical settings. The proposed procedural framework in this paper could be highly useful in these cases.

VI. Conclusion

In this paper, a simplified procedural framework is proposed that allows the development of a multichannel processing strategy for ASSR detection starting from a detection theory approach. It is shown that a *sufficient statistic* can be calculated that best captures the amount of ASSR in the observation matrix. This sufficient-statistic-based approach can exploit spatiotemporal stationarity present in the EEG measurements and can be linked with the development and application of the MWF approach and ICA based approach to ASSR detection. The evaluation is conducted using data from ten normal-hearing adults. Two types of EEG are used: EEG with few artifacts and EEG with a significant number of (controlled) artifacts. It is concluded that most single- and multichannel approaches are similar in performance when applied to uncontaminated EEG. When artifact-rich EEG is used, the proposed detection-theory-based approach improves the number of ASSR detections compared with the noise-weighted average of the best channel of this dataset (vertex–right mastoid). In general, this best electrode configuration cannot be known beforehand. When compared with a noise-weighted common EEG channel derivation (vertex–occiput), the proposed approach improves ASSR detection significantly.

References

[1] Joint Committee on Infant Hearing, "Joint committee on infant hearing 1994 position statement. American academy of pediatrics joint committee on infant hearing," *Pediatrics*, vol. 95, no. 1, pp. 152–156, 1995.

[2] H. Luts, C. Desloovere, and J. Wouters, "Clinical application of dichotic multiple-stimulus auditory steady-state responses in high-risk newborns and young children," *Audiol. Neurotol.*, vol. 11, pp. 24–37, 2006.

[3] R. Galambos, S. Makeig, and P. J. Talmachoff, "A 40-Hz auditory potential recorded from the human scalp," *Proc. Nat. Acad. Sci. USA*, vol. 78, no. 4, pp. 2643–2647, 1981.

[4] T. W. Picton, M. S. John, A. Dimitrijevic, and D. Purcell, "Human auditory steady-state responses," *Int. J. Audiol.*, vol. 42, no. 4, pp. 177–219, 2003.

[5] P. A. Davis, "Effects of acoustic stimuli on the waking human brain," *J. Neurophsyiol.*, vol. 2, pp. 494–499, 1939.

[6] C. D. Geisler, "Average responses to clicks in man recorded by scalp electrodes," Massachusetts Inst. Technol. (MIT) Research Lab. Electron., Cambridge, Tech. Rep. 380, 1960.

[7] A. P. Bradley and W. J. Wilson, "On wavelet analysis of auditory evoked potentials," *Clin. Neurophysiol.*, vol. 115, no. 5, pp. 1114–1128, 2004.

[8] R. Zhang, G. McAllister, B. Scotney, S. McClean, and G. Houston, "Combining wavelet analysis and Bayesian networks for the classification of auditory brainstem response," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 3, pp. 458–67, Jul. 2006.

[9] Y. Tang and A. M. Norcia, "An adaptive filter for steady-state evoked responses," *Electroencephalogr. Clin. Neurophysiol.*, vol. 96, pp. 268–277, 1995.

[10] M. S. John, A. Dimitrijevic, and T. W. Picton, "Weighted averaging of steady state responses," *Clin. Neurophysiol.*, vol. 122, pp. 555–562, 2001.

[11] B. Van Dun, J. Wouters, and M. Moonen, "Multi-channel Wiener filtering based auditory steady-state response detection," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. II, Honolulu, HI, 2007, pp. 929–932.

[12] B. Van Dun, J. Wouters, and M. Moonen, "Improving auditory steady-state response detection using independent component analysis on multichannel EEG data," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1220–1230, Jul. 2007.

[13] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. London, U.K.: Prentice-Hall, 1993.

[14] D. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

[15] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, ch. 3, pp. 39–60.

[16] 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 Topogr.*, vol. 15, pp. 69–86, 2002.

[17] D. W. Purcell, M. S. John, B. A. Schneider, and T. W. Picton, "Human temporal auditory acuity as assessed by envelope following responses," *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3581–3593, 2004.

[18] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[19] L. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *J. Mach. Learn. Res.*, vol. 4, pp. 1261–1269, 2003.

[20] M. S. John and T. W. Picton, "MASTER: A Windows program for recording multiple auditory steady-state responses," *Comput. Methods Programs Biomed.*, vol. 61, pp. 125–150, 2000.
[21] O. G. Lins and T. W. Picton, "Auditory steady-state responses to multiple simultaneous stimuli," *Electroencephalogr. Clin. Neurophysiol.*, vol. 96, pp. 420–432, 1995.
[22] J. L. Valdes, M. C. Perez-Abalo, V. Martin, G. Savio, C. Sierra, E. Rodriguez, and O. Lins, "Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses," *Ear Hear.*, vol. 18, pp. 420–429, 1997.
[23] F. Sharbrough, G.-E. Chatrian, R. P. Lesser, H. Lüders, M. Nuwer, and T. W. Picton, "American electroencephalographic society guidelines for standard electrode position nomenclature," *J. Clin. Neurophysiol.*, vol. 8, pp. 200–202, 1991.
[24] C. S. van der Reijden, L. H. M. Mens, and A. F. M. Snik, "Signal-to-noise ratios of the auditory steady-state response from fifty-five EEG derivations in adults," *J. Amer. Acad. Audiol.*, vol. 15, no. 10, pp. 692–701, 2004.
[25] B. Van Dun, S. Verstraeten, J. Alaerts, H. Luts, M. Moonen, and J. Wouters, "A flexible research platform for multi-channel auditory steady-state response measurements," *J. Neurosci. Methods*, vol. 169, pp. 239–248, 2008.
[26] J. Cardoso and A. Soloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Electr. Eng. F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.
[27] H. Luts and J. Wouters, "Comparison of MASTER and AUDERA for measurement of auditory steady-state responses," *Int. J. Audiol.*, vol. 44, pp. 244–253, 2005.
[28] A. T. Herdman and D. R. Stapells, "Thresholds determined using the monotic and dichotic multiple auditory steady-state response technique in normal-hearing subjects," *Scand. Audiol.*, vol. 30, pp. 41–49, 2001.
[29] M. S. John and T. W. Picton, "Human auditory steady-state responses to amplitude–modulated tones: Phase and latency measurements," *Hear. Res.*, vol. 141, no. 1–2, pp. 57–79, 2000.



Bram Van Dun (S'07) was born in Zoersel, Belgium, in 1979. He received the Electr. Eng. degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 2003 and 2008, respectively.

He is currently a Doctoral Researcher at both the SCD–SISTA, Department of Electrical Engineering (ESAT) and the Experimental Otorhinolaryngology (ORL) Laboratories, Katholieke Universiteit Leuven. His current research interests include biomedical signal processing in general, and EEG signal processing applied to hearing tests in particular.



Geert Rombouts was born in Turnhout, Belgium, in 1973. He received the Electr. Eng. degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1997 and 2003, respectively.

He was a Postdoctoral Researcher at the SCD–SISTA, Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven. His current research interests include acoustic echo cancelation, acoustic noise reduction, and acoustic feedback suppression.



Jan Wouters was born in 1960. He received the Master's degree in (general) physics and the Ph.D. degree in sciences/physics from the Katholieke Universiteit Leuven (K.U. Leuven), Leuven, Belgium, in 1982 and 1989, respectively.

From 1989 to 1992, he was a Research Fellow with the Belgian National Fund for Scientific Research (NFWO), Institute of Nuclear Physics (UCL Louvain-la-Neuve and K.U. Leuven) and at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (Maryland). Since 1993, he has been a Professor with the Department of Neurosciences, K.U. Leuven. His current research interests include audiology and the auditory system, signal processing for cochlear implants, and hearing aids. He has authored or coauthored about 135 articles published in international peer-review journals.

Prof. Wouters received an Award of the Flemish Ministry in 1989, a Fullbright Award and a North Atlantic Treaty Organization (NATO) Research Fellowship in 1992, the Flemish Vereniging Voor Logopedie (VVL) Speech Therapy-Audiology Award in 1996, and the Corresponding Member Award of the Deutsche Gesellschaft für Audiologie (DGA). He is a member of the Board of the International Collegium for Rehabilitative Audiology and the International Collegium for Otorhinolaryngology (ORL). He is a reviewer for several international journals. He is a member of the Editorial Board of the *Journal of Communication Disorders* and *Belgian—Ear Nose Throat (B-ENT)*.



Marc Moonen (M'94–SM'06–F'07) received the Electr. Eng. degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1986 and 1990, respectively.

Since 2004, he has been a Full Professor in the Department of Electrical Engineering, Katholieke Universiteit Leuven.

Prof. Moonen received the 1994 K.U. Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award, the 2004 Alcatel Bell (Belgium) Award, and was a 1997 "Laureate of the Belgium Royal Academy of Science." He received the Journal Best Paper Award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING and *Elsevier Signal Processing*. He was the Chairman of the IEEE Benelux Signal Processing Chapter (1998–2002). He is currently the President of the EURASIP (2007–2008) and a member of the IEEE Signal Processing Society Technicial Committee on Signal Processing for Communications. He has been a member of the Editorial Board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (2002–2003) and the IEEE SIGNAL PROCESSING MAGAZINE (2003–2005). He was the Editor-in-Chief for the *European Association for Signal Processing (EURASIP) Journal on Applied Signal Processing* (2003–2005). He is currently a member of the Editorial Board of the *EURASIP Journal on Applied Signal Processing*, *EURASIP Journal on Wireless Communications and Networking*, and *Signal Processing*.