An Automatic Sequential Recognition Method for Cortical Auditory Evoked Potentials

Ulrich Hoppe[1,*], *Stephan Weiss*[2], *Robert.W. Stewart*[3], *and Ulrich Eysholdt*[1]

[1] Abteilung für Phoniatrie und Pädaudiologie
Universität Erlangen-Nürnberg, Bohlenplatz 21, 91054 Erlangen, Germany
{ulrich.hoppe,ulrich.eysholdt}@phoni.med.uni-erlangen.de
Tel. ++49(0)9131 8533815, Fax ++49(0)9131 8539272

[2] Department of Electronics and Computer Science
University of Southampton, SO17 1BJ, U.K.
s.weiss@ecs.soton.ac.uk

[3] Department of Electronic and Electrical Engineering
University of Strathclyde, Glasgow G1 1XW, Scotland, UK
r.stewart@eee.strath.ac.uk

Abstract

The detection of cortical auditory evoked potentials (CAEP), which are part of the electroencephalogram (EEG) in reaction to acoustic stimuli, has important applications such as determining objective audiograms. The detection is usually performed by a human operator, with support from often basic signal processing methods. This paper presents a novel mechanism for the detection of CAEPs, which is fully automatic and stops the measurement when a given confidence is reached. This proposed detector comprises of three stages. First, a feature extraction by a wavelet transform parameterizes the time domain EEG signal by only few transform coefficients. This feature vector is then classified by a neural network which yields a binary vote on every EEG segment. Finally, a sequential statistical test is performed on successive classifications; this stops the measurement if a specified decision confidence has been reached. The adjustment of the detector according to a clinical database is discussed. Thus adjusted, the proposed CAEP detection scheme is applied to a study, and compared with a human operator. The results demonstrate that this method can attain similar results, but outperforms the human expert for stimulation levels close to the hearing threshold.

Keywords

cortical auditory evoked potentials, wavelet analysis, neural networks, sequential statistical test, automatic detector.

I. Introduction

Cortical auditory evoked potentials (CAEPs) can be observed as part of the electroencephalogram (EEG) and reflect the neural processing of acoustic stimuli in the cortex. In their clinical application, CAEPs have the advantage of enabling frequency specific testing of the hearing ability by appropriate selection of the stimulus [1]. The recognized presence or absence of a CAEP in response to an acoustic stimulus of specified frequency and level can be used to determine an objective audiogram. Objectivity is achieved in the sense that the active cooperation of an experimentee or patient is replaced by the interpretation of the EEG by an operator [1]. Objective hearing assessment is important to, for example, establish reliable hearing thresholds for persons with low compliance, e.g. small children [2]. The work presented here is concerned with an automatic detector to help and assist in identifying CAEPs.

When an acoustic stimulus reaches the inner ear it elicits neural activity along the neural auditory pathway from the spiral ganglion via nerve fibers over the brainstem up to the auditory cortex, where it results in auditory perception approximately 100 ms after the onset of the stimulus [3]. Triggered by processing in the auditory cortex [4], CAEPs therefore give evidence of correct auditory perception. According to [5] and references therein, for most human adults the signal properties of the CAEP are quite similar when the same measurement paradigm is used and the central auditory processing is normal. The CAEP consists of waveforms with specific latencies with respect to the stimulus onset. The literature names their minima and maxima due to polarity and sequence P1, N1, P2, and N2, whereby N1 and P2 are the most stable waves of a CAEP with peak latencies of about 100 ms and 175 ms, respectively [5]. It is generally the recognition of this N1–P2–complex which is used to establish one of two hypotheses H_0: no CAEP within EEG; and H_1: CAEP within EEG [6].

Variations of amplitudes and latencies of the CAEP waves occur and depend — besides age, sex, ear of stimulation, hearing ability, and vigilance of the experimentee — on the frequency, level and repetition rate of the stimuli [5]. While many of these factors only contribute little or can be controlled during the measurement, the individual sensation level is in general unknown and has an observable influence: As the stimulus level approaches the hearing threshold, the CAEP waves appear weaker and more delayed due to prolonged auditory processing for quieter sounds [3]. An example is given in Fig. 1, where averaged EEGs recorded synchronously to stimuli of different intensities are shown.

Besides this variability, detection of CAEPs is impeded by background EEG due to visual, motoric, and sensoric activities of the experimentee, resulting in a very low signal-

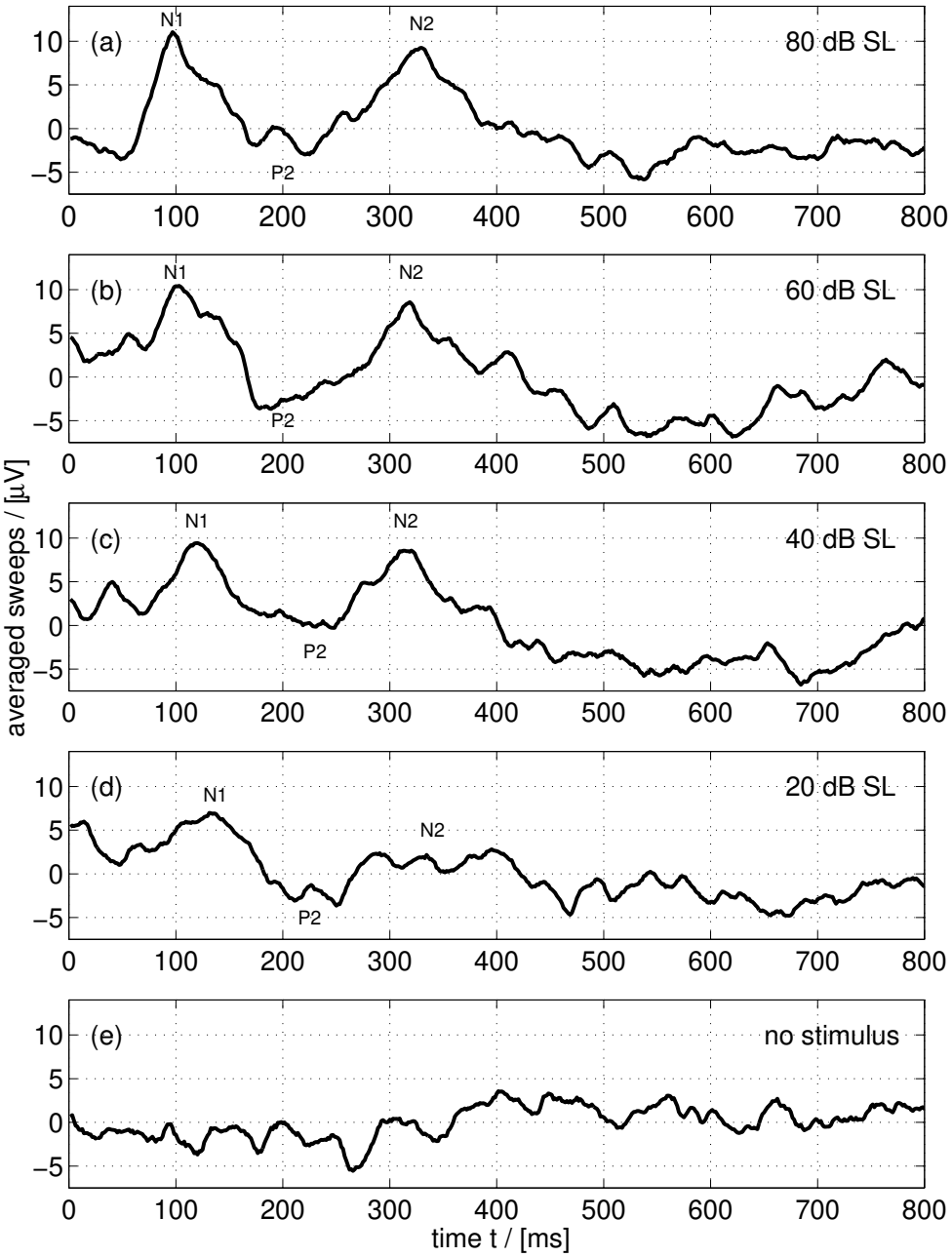


Fig. 1. EEG averages over 100 sweeps obtained from a single individual: (a)-(d) variable stimulation intensities of 80, 60, 40 and 20 dB above hearing threshold (SL) applied at $t = 0$ containing CAEP; (e) no stimulation ($-\infty$ dB) and hence no CAEP.

to-noise ratio (SNR) of less than -10 dB. As CAEP and background EEG cover similar frequency bands, it is impossible to detect a CAEP from EEG measurements with only a single stimulus. Rather, the detection of CAEPs generally has to be based on repeated stimuli and the acquisition of typically about 30 to 80 "sweeps", which here refer to EEG segments recorded under identical conditions synchronously to the given stimulus. Many methods for support in this decision have been proposed, and range from simple averaging to more advanced signal processing.

Simple averaging of successive sweeps is a common, but often sub-optimal CAEP reconstruction method, since stationarity of the background noise, invariance of the CAEP, and the absence of correlation between CAEP and noise have to be assumed. Due to varying vigilance and habituation effects during the measurement, this is only approximately satisfied [7]. Methods have been proposed to take variations of the background EEG into account by estimating the noise power in every sweep [8], [9], [10], [11], and weighting their contribution to the average accordingly. Thus, noisier sweeps and artifacts are disemphasized. Other approaches try to enhance the SNR by considering the variations of the central auditory processing during the measurement [12], [13], [14]. These techniques assume that the latencies of the CAEP vary over time, leading to an approximate affine translation of the signals. The actual translations are estimated by cross correlation of every single sweep with pre-defined patterns. SNR enhancement is then achieved by averaging the latency-corrected sweeps. Clearly limited by the small sample size for variance-type estimations, such methods can only be moderately successful.

A number of parametric approaches aim to replace the subjective interpretation of the averaged CAEP by an automatic classification, which is generally based on the statistics of some sweep parameters. Literature offers a wide range of parameters being used, like CAEP relevant samples of the digitized signals, selected Fourier coefficients, energy of averaged sweeps, or correlation coefficients with 'normal' CAEP [15]. Coarse SNR estimates of the CAEP/EEG based on the comparison of a standard average with a '$\pm$ average', where every second sweep is included with reversed sign, were also used to classify CAEPs [16], [17], [15]. Wavelet coefficients have been used to parameterise evoked potentials successfully [18], and have been shown to perform at least comparable to the above parametric methods for CAEPs [19]. However, such parametric approaches require to a priori set a threshold for classification and cannot provide stopping criteria for the measurement based on a specified error rate, which usually is regarded as major drawback in clinical use.

In the approach proposed here, the good parameterisation property of the wavelet trans-

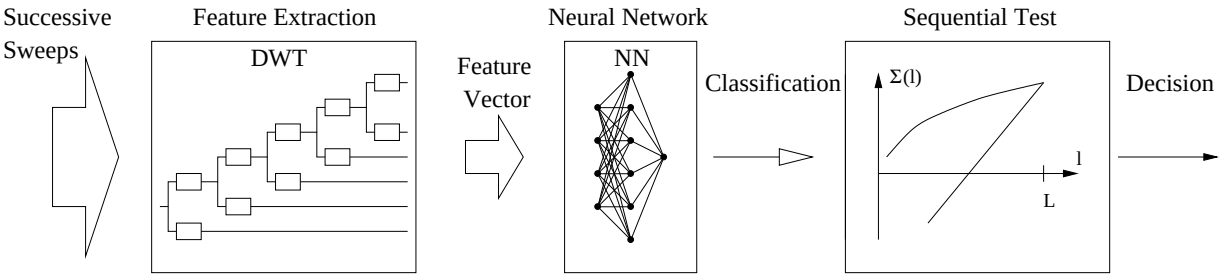


Fig. 2. Block diagram of the CAEP detection scheme.

form is exploited to build a detector with stopping criterion and specified error rate. This detector is outlined schematically in Fig. 2. First, a selected set of discrete wavelet transform (DWT) coefficients, a so called "feature vector" (FV), is calculated for each measured sweep. The dimension of the FV can be much smaller than the original time-domain signal without loss of too much information on a CAEP [18], [19]. The FVs of successive sweeps are each classified by a neural network (NN), which maps the FV onto a single binary $\{0, 1\}$ output. A similar approach is implemented by wavelet networks [20], which combine both a wavelet transform for parameterisation and an NN in a hybrid classifier. As the binary classification of the NN bears a high type-I-error (incorrect acceptance of H_1), a sequential statistical test forms the last stage of the detection procedure, judging the number of positive and negative sweep classifications produced by the NN. This test will provide a decision based on a predefined type-I-error rate, and stops the measurement when this confidence limit is reached.

In Sec. II, the proposed detector is introduced in detail. Sec. III reports on the EEG equipment at hand and the resulting data properties, which are used to adjust the different stages of the detector. This includes the selection of the FV, the dimensioning and training of the NN, and the adjustment of the sequential statistical test. The application of our method to a study using clinical data is reported in Sec. IV, compared to the results achieved by a human operator. Finally, we draw conclusions in Sec. V.

II. Automatic Detector — Methods

This section describes in detail the principles of the proposed detection scheme and its components. The parameterisation by discrete wavelet transform (DWT) coefficients is addressed in Sec. II-A. Sec. II-B introduces a neural network classification of the features thus extracted by the DWT. Finally, a sequential statistical test to assess the classification results of successive sweeps is introduced in Sec. II-C. The adjustment of the detector components with respect to EEG / CAEP data will be performed in the separate Sec. III.

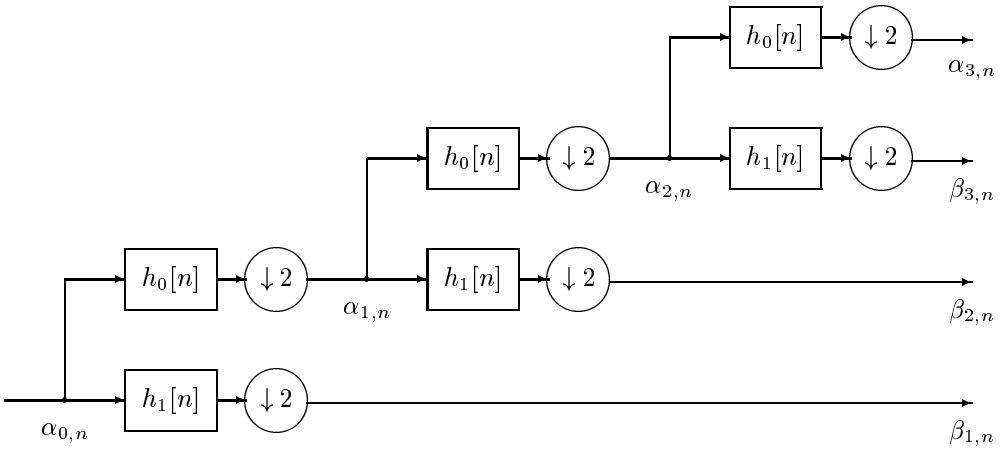


Fig. 3. Octave filter bank to compute an MRA of depth $D = 3$; deeper decompositions are achieved by further splitting $\alpha_{3,n}$.

A. CAEP Feature Extraction by Wavelet Transform

To parameterize a CAEP and condense the information on whether an EEG contains or does not contain a CAEP within only few coefficients, a wavelet transform is invoked. Similar to the Fourier transform, the wavelet transform performs a least-squares fit of an analysis function to the time domain data. However, rather than a sinusoid extending over an infinite interval, here wavelet functions with a limited support are employed as basis functions, which is advantageous when analysing transient signals [21]. For the discrete wavelet transform (DWT) considered here, a set of orthogonal basis functions is obtained by scaling and translation of a mother wavelet.

An efficient calculation of the DWT coefficients in the case of discrete-time data can be achieved with a multiresolution algorithm (MRA, [22]). For a dyadic DWT, this MRA is performed by filtering the signal to be analysed with an octave filter bank as shown in Fig. 3. The highpass filter $h_1[n]$ forms a quadrature mirror filter (QMF) pair [23] with the lowpass $h_0[n]$ of the filter bank. The input sequence to the octave filter bank, $\alpha_{0,n}$, is the signal $x[n]$ to be analysed. Through successive low- and highpass filtering of the samples in the lower frequency band, $\alpha_{m,n}$, and decimation of the resulting signals by a factor of 2 (denoted as $\downarrow 2$), subband samples $\beta_{m,n}$ are obtained, which, except of the lowest frequency band, represent the DWT coefficients. The coefficients $\alpha_{m,n}$ are intermediate values and correspond to a dual basis function of the wavelet, a so called scaling function [22]. The filters $h_0[n]$ and $h_1[n]$ are sampled version of the underlying scaling function and wavelet, and therefore determine which DWT — amongst a large

variety of possible wavelet functions (see e.g. [22], [21], [24]) — is being implemented.

Since the signal to be analysed, $x[n]$, is only defined on a finite interval $n \in [0; N-1]$, suitable signal extensions [25] or boundary filters [26] have to be selected. As both yield identical results [26], for ease of implementation the first possibility will be pursued. Zero padding is not viable for decomposition with a deep octave bank, as information will be blurred over large intervals by the transient behaviour. Periodization of the original data and all intermediate filtering results $\alpha_{m,n}$ however permits to filter in steady-state and thus retain all required information within a preserved period of the signal. With the possibilities of odd or even periodisation [25], the MRA yields N DWT coefficients that completely describe the original time series $x[n]$ comprising N samples. However, restrictions to symmetric wavelets apply, if an odd periodic extension is selected [22].

Based on EEG data, Sec. III will determine, which wavelet function and periodic extension will be used for the proposed automatic detector and which coefficients can be isolated as significant. The task will be to identify a subset of $M \ll N$ coefficients collected in a "feature vector" (FV) $\mathbf{y}_{\mathrm{FV}}$ that can describe the CAEP sufficiently well. Sec. III will also show why these coefficient on their own can only yield a rather unsatisfying indication of a CAEP. This will motivate the requirement of the NN classification and sequential test performed on the selected significant DWT coefficients in the FV described in the next two sections to ensure a reliable and systematic CAEP detection.

B. Neural Network Classification

The aim of this section is to introduce a classification method for single sweeps distinguishing between 'containing a CAEP' or 'not containing a CAEP' based on the set of M DWT coefficients in the feature vector (FV). Due to interindividual variations of CAEP-latencies and -amplitudes these FVs do not exhibit a single specific pattern. Instead, many patterns must be regarded as being specific for CAEPs. So far, there exists no theoretical treatment of this discrimination problem; hence heuristic classifiers have to be used. Artificial neural networks (NN) provide the possibility of learning by groups of sample FVs and require only little computational effort.

A multilayer perceptron was used for classification of the FV calculated for each sweep. This type of NN comprises an input layer with M nodes corresponding to the dimension of the FV, one hidden layer with K neurons and an output layer with a single neuron (M-K-1 network). Since the NN is a non-linear processor, a bias removal and normalisation with respect to the l_∞-norm are applied to the FV $\mathbf{y}_{\mathrm{FV}}$ prior to feeding it into the NN. This operation excludes interindividual amplitude differences that might be incurred due

to slightly different measurement impedances and anatomical differences. The resulting modified FVs $\tilde{\mathbf{y}}_{\mathrm{FV}}$ now lie on the surface of an M dimensional hypercube.

While the number of nodes M in the input layer is determined by the dimension of the FV, the number of neurons in the hidden layer, K, has to be selected appropriately. If K is chosen too small and not enough hidden neurons are used, during training the neural network is unable to classify all FV of the learning set of data. On the other side, a too large number of hidden neurons K will cause the NN to be too specific for the learning data set, and prevent the NN to be able to interpolate (or "generalize") when slightly different CAEP data is presented for classification.

The classification process for a FV $\tilde{\mathbf{y}}_{\mathrm{FV}}$, where l is the sweep index, can be summarized as

$$\gamma(l) = s\big(\mathbf{b}_2 + \mathbf{W}_2 \cdot \tanh(\mathbf{b}_1 + \mathbf{W}_1 \cdot \tilde{\mathbf{y}}_{\mathrm{FV}}(l))\big) \quad , \tag{1}$$

where tanh is a vector valued hyperbolic tangent and $s(x)$ a Heavyside-function ($s(x) = 1$ for $x \geq 0$, $s(x) = 0$ else) to produce a binary classification $\gamma(l) \in \{0; 1\}$. During the training phase of the NN the parameters in weight matrices $\mathbf{W}_1$, $\mathbf{W}_2$, and bias vectors $\mathbf{b}_1$, $\mathbf{b}_2$ are adjusted such that the classification error for certain training FVs will be minimized. This can be done by the backpropagation algorithm[1] [27], [28]. The description of the FV set used for the NN training to appropriately select K and adjust the weight and bias parameters in (1) will be described in detail in Sec. III-C.

It is obvious that this classification based on single sweeps bears large classification errors. This is unavoidable as the time signal in the spontaneous EEG is known to often exhibit patterns similar to a CAEP. In Sec. III-C, we will establish the exact probability p_{NN} for false detection of a CAEP by the appropriately dimensioned and trained NN. This error rate p_{NN} can be interpreted as the fractional area on the surface of the M-dimensional hypercube occupied by those FVs $\tilde{\mathbf{y}}_r(l)$ which will be classified as a CAEP.

The relatively high false positive rate that will be encountered here is due to the low SNR of the AEP within the spontaneous EEG. Enhancement of this SNR could be achieved by sub-averaging over two or more sweeps with subsequent NN-classification of the FV of the averaged data. This however has drawbacks. Firstly, this procedure prolonges measurement time proportionally. Secondly, although the false negative error is reduced, the false positive error classification p_{NN} remains unaffected. A better aproach can be achieved by sequential statistical analysis of the NN classification results. The statistical

[1]To apply the algorithm, for differentiability during training the output non-linearity $s(x)$ is replaced by a hyperbolic tangent; subsequently the desired output of the NN during training was in]-1;1[.

method to perform this final stage of the detector is outlined in the following section.

C. Sequential Statistical Analysis

To assess the classification results yielded by the NN in the last section, usually the statistics of a fixed number L of sweeps is tested. Based on this standard method introduced in Sec. II-C.1, in Sec. II-C.2 we derive a sequential test based on [29], [30], which only defines an upper limit of sweeps, L, but can be stopped earlier if the statistics agree.

C.1 Statistical Test with Fixed Sample Size

Consider a random series of EEG sweeps measured without any acoustic stimulation and therefore without any CAEP. The classification results of the subsequent FVs by the NN gives a sequence of zeros and ones, whereby the probability to misclassify such FVs as to 'containing a CAEP' ($\gamma(l) = 1$) is denoted by $p = p_{\mathrm{NN}}$. After L recorded sweeps the sum $\Sigma(L)$ of all classifications,

$$\Sigma(L) = \sum_{l=1}^{L} \gamma(l) \quad , \qquad \gamma(l) \in \{0, 1\} \tag{2}$$

lies between 0 and L. Since the single classification results are independent, the sum $\Sigma(L)$ is binomially distributed. Thus, the problem is to test the null hypothesis H_0: $p = p_{\mathrm{NN}}$ (no CAEP is inside the EEG) against the alternative H_1: $p > p_{\mathrm{NN}}$ (a CAEP is present). The binomial distribution can be approximated for sufficiently large L by a Gaussian distribution with parameters

$$\mu(L) = L \cdot p_{\mathrm{NN}} \quad , \qquad \sigma(L) = \sqrt{L \cdot (1 - p_{\mathrm{NN}}) \cdot p_{\mathrm{NN}}} \quad . \tag{3}$$

Once the error rate p_{NN} for the NN classification in Sec. II-B is established from test data, the probability for $\Sigma(L) > \mu(L) + z_\alpha \cdot \sigma(L)$ after L classifications can be set to take on a fixed value α by selecting the quantity z_α appropriately. Using the correct values for μ and σ instead of the Gaussian approximations in (3), z_α will become slightly smaller. Hence the approximation is conservative in the sense that the true value for α is slightly smaller when based on the assumption of Gaussianity. With these considerations, a statistical test with a predetermined error of false positive classifications (type-I-error) α can be created.

Using this statistical test after a fixed sample size of L sweep classifications, H_0 can be accepted or rejected based on whether the distribution is inside or outside the expected interval with a pre-specified confidence. However, it is important to note that this test only works with a fixed number of sweep classifications.

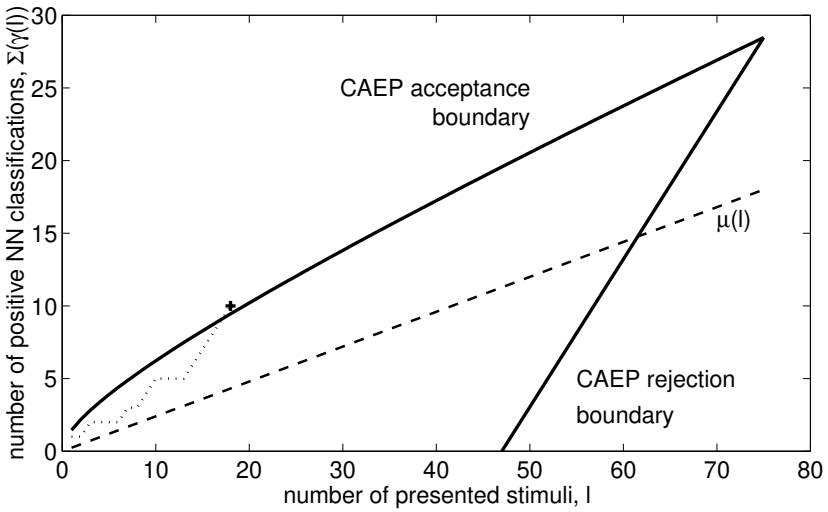


Fig. 4. Sequential test plan with acceptance and rejection boundaries for assessment of NN classification ($L = 75$ and $p_{\mathrm{NN}} = 0.24$); dashed line: mean value for the sum of positive NN classification for random sweeps; dotted line: example of sweep with CAEP detected after $l = 18$ NN classifications.

C.2 Sequential Statistical Test

Sequential testing means that although the sample size has an upper limit, L, the sum $\Sigma(l)$ of NN classifications,

$$\Sigma(l) = \sum_{\lambda=1}^{l} \gamma(\lambda) \quad , \qquad l \in [1; L] \quad , \tag{4}$$

where l is the sweep index, is checked during the entire measurement. The measurement will be stopped immediately if certain statistical limits are exceeded, resulting in an acceptance of H_1.

Due to the complex distribution of $\Sigma(l)$, the statistical limits cannot be directly calculated. Therefore, they were determined using Monte-Carlo simulations based on test data, which are described in detail in Sec. III.

Example. An example expressing the main idea of the sequential test is shown in Fig. 4. There, the abscissa represents the number of single sweep measurements, l, the ordinate the sum $\Sigma(l)$ of binary NN classifications. The upper limit in this example is $L = 75$. If after l sweeps $\Sigma(l)$ exceeds the upper boundary, a CAEP is detected and the measurement will be stopped. If the lower boundary is reached after l sweep recordings, it becomes impossible that the upper boundary is met in the remaining $L - l$ sweeps, a CAEP is rejected and the measurement can be stopped. The dashed line in Fig. 4 shows the mean value $\mu(l) = l \cdot p_{\mathrm{NN}}$ for positive NN-classifications in the absence of a CAEP (type-I-error).

An example for $\Sigma(l)$ is given with the dotted line, which here passes the acceptance boundary ($\Sigma(18) = 10$) after $l = 18$ sweep classifications, and a CAEP is detected. It can be easily calculated that a minimum number of $l = 3$ sweep classifications is required before a CAEP can be detected under the conditions of this example.

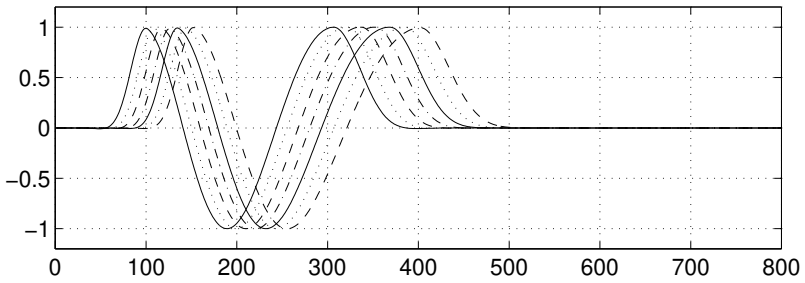
The above sequential test procedure can be regarded as a semi-adaptive decision algorithm [29]. If the SNR of the CAEP inside the ongoing EEG is high enough, the algorithm will considerably shorten the time which is necessary to reach a decision. However, if the SNR is extremely low, a larger number of measurements will be required. In case of the example given above, the test will need an average number of 62 sweeps until a measurement will be rejected as non-AEP, as indicated by the intersection of the mean sum $\mu(l)$ with the rejection boundary in Fig. 4.

III. Adjustment of Detector Using Real Data

Now, step by step, the general automatic detector outlined in the past section will be adjusted with respect to our data acquisition equipment and procedure, which is discussed in Sec. III-A. This has consequences for the tailoring of the DWT and the selection of transform coefficients parameterising CAEPs in Sec. III-B. The subsequent dimensioning and training of the NN classifier is described in Sec. III-C. The false positive classification rate of the NN together with a selected significance finally sets the parameters for tuning the sequential test in Sec. III-D.

A. Equipment and CAEP Data

All EEG measurements employed in this work were obtained under clinically proven conditions in an electrically insulated and acoustically anechoic chamber [31], where patients and experimentees are seated in a comfortable armchair in order to minimize motoric artifacts inside the EEG. The vigilance of persons under test is ensured by monitoring with a video system. The stimulation system (ASTI20, ZLE corp., Munich) used here produces predefined signals, which are presented via a headphone (Beyerdynamics DT48) to the test persons and patients. The EEG is measured as a voltage from the ipsilateral mastoid (M1, M2) and a contralateral forehead position (FP1, FP2) of the stimulated ear against a ground electrode positioned on the centre of the forehead (FZ). After differential EEG amplification by approx. 80 dB, the analogue EEG signal is prefiltered by Bessel filters of 4th order with a passband region between 1.6 Hz and 20 Hz. Finally, the stimulation system triggers a 12-bit analogue-to-digital converter sampling at 640 Hz for synchronous digital recording, yielding $512 = 2^9$ samples for 800 ms segments of EEG represented


1
0.5
0
−0.5
−1
0
100
200
300
400
500
600
700
800