Detection of Change Points in Phase Data: A Bayesian Analysis of Habituation Processes

Z. Mortezapouraghdam[1,2], L. Haab[1], G. Steidl[2], D.J. Strauss[1,3,4]

[1] Systems Neuroscience & Neurotechnology Unit, Faculty of Medicine, Saarland University, Homburg/Saar, Germany

[2] Mathematical Image Processing & Data Analysis Group, Department of Mathematics, Technical University Kaiserslautern, Germany

[3] Leibniz-Institute for New Materials, Saarbruecken, Germany

[4] School of Engineering, Saarland University of Applied Sciences, Saarbruecken, Germany

***Abstract*— Given a time series of data points, as obtained in biosignal monitoring, the change point problem poses the question of identifying times of sudden variations in the parameters of the underlying data distribution. We propose a method for extracting a discrete set of change points from directional data. Our method is based on a combination of the Bayesian change point model (CPM) and the Viterbi algorithm. We apply our method to the instantaneous phase information of single-trial auditory event-related potentials (ERPs) in a long term habituation paradigm. We have seen in previous studies that the phase information enters a phase-locked mode with respect to the repetition of a stimulus in the state of focused attention. With adaptation to an insignificant stimulus, attention tends to trail away (long-term habituation), characterized by changes in the phase signature, becoming more diffuse across trials. We demonstrate that the proposed method is suitable for detecting the effects of long-term habituation on phase information in our experimental setting.**

I. Introduction

Changes in time series data often are the consequence of changes in the parameters of underlying generative processes. In the case of discrete changes, these changes are also known as change points. The detection of change points in time series data is a valuable tool in many different areas such as stock market prices, analysis of DNA sequences, neural activities in the brain. Recently the exploration of synchronicity either within one or between multiple functional parts of the brain when exposed to a particular repetitive stimulus has gained a lot of interest [5], [2]. As an example in [3], it is found that phase information of ERPs enter a precise 'phase-locking' mode under the repetition of a stimulus. When habituation – an endogenous brain mechanism responsible for the drift of attention away from a stimulus – occurs, the synchronicity of phase information diminishes. This can be observed numerically in a decrease in the concentration of phases on the unit circle. We investigate the application of a Bayesian change point algorithm for directional data for estimating long-term habituation in auditory ERPs. We apply it to measurements taken under two stimuli of distinct saliency. The stimuli were presented as pure sinus tones of 1kHz at levels of 50dB (SPL) and 100dB (SPL), representing an insignificant and an aversive stimulus due to their intensity. Contrasting the insignificant stimulus, the aversive tone is likely to bind attention over time, allowing only little or no long-term habituation at all. In this work we detect the most likely time at which the phase information begins to diffuse, such that the parameters of an underlying generative model before and after the detected time are significantly different. In case of phase data for 50dB (SPL) responses, we expect a unique change in the concentration of phase information due to the habituation effect, whereas in 100dB (SPL) such effects are not expected due to the aversiveness of stimulus intensity.

II. Methods

We briefly explain the forward-backward Bayesian change point algorithm from [1] for detecting regime changes in the generative process underlying the given directional time series data. A change point divides the set of data points $\boldsymbol{\Theta} := \{\theta_1, \ldots, \theta_N\}, \theta_i \in [-\pi, \pi)$ into non-overlapping partitions p. The data points in each partition are independently and identically distributed (i.i.d) from a corresponding probability distribution $p(\theta_t|\lambda_p)$ with parameters λ_p. We assume a priori distribution over the interval length between change points, which is called the hazard rate. The number of time steps since the last change point is called the run-length. The hazard rate or the rate at which change points occur, can be a function of the run-length. Our aim is to estimate the posterior distribution over the run-lengths at each time t. We denote the length of the current run at time t as r_t, and $\boldsymbol{\theta}_t^{(r_t=i)} = (\boldsymbol{\theta}_{t-i:t})$ indicates the set of observations associated with the run-length r_t. The run-length r_t can either increase by one at any time step with respect to the run-length at $t-1$, that is $r_t = r_{t-1} + 1$ indicating a non-change point step, or reset to zero in case of a change point $r_t = 0$. We start by formulating the problem in terms of a forward-backward algorithm using the Bayes' rule as follows:

$$\begin{aligned} p\left(r_t|\boldsymbol{\Theta}\right) &= \frac{p\left(\boldsymbol{\Theta}|r_t\right)p\left(r_t\right)}{p\left(\boldsymbol{\Theta}\right)} \\ &= \frac{p\left(\boldsymbol{\Theta}_{t+1:T}|\boldsymbol{\Theta}_{1:t}, r_t\right)p\left(\boldsymbol{\Theta}_{1:t}|r_t\right)p\left(r_t\right)}{p\left(\boldsymbol{\Theta}\right)} \\ &= \frac{p\left(\boldsymbol{\Theta}_{t+1:T}|\boldsymbol{\Theta}_{1:t}, r_t\right)p\left(r_t, \boldsymbol{\Theta}_{1:t}\right)}{p\left(\boldsymbol{\Theta}\right)} \end{aligned}$$

We write $p\left(r_t = i|\boldsymbol{\Theta}\right) = \frac{\alpha_i^t \beta_i^t}{p(\boldsymbol{\Theta})}$, with the forward pass (α-pass) defined as $\alpha_i^t = p\left(r_t = i, \boldsymbol{\Theta}_{1:t}\right)$ and the backward pass (β-pass) as $\beta_i^t = p\left(\boldsymbol{\Theta}_{t+1:T}|r_t = i, \boldsymbol{\Theta}_{1:t}\right)$. In the following, we explain the forward and backward passes as proposed in [1]. For more details see also [6].

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

A. The Forward and Backward Pass

The forward pass is described recursively by

$$\alpha_{r_t}^t = p\left(r_t, \boldsymbol{\Theta}_{1:t}\right) = p\left(r_t, \theta_t, \boldsymbol{\Theta}_{1:t-1}\right) = p\left(\theta_t | r_t, \boldsymbol{\Theta}_{1:t-1}\right) \sum_{r_{t-1}} p\left(r_t | r_{t-1}\right) p\left(r_{t-1}, \boldsymbol{\Theta}_{1:t-1}\right).$$

Similarly, the β-pass (backward-pass) is given by

$$\beta_{r_t}^t = \textstyle\sum_{r_{t+1}} p\left(\boldsymbol{\Theta}_{t+2:T} | \boldsymbol{\Theta}_{1:t+1}, r_{t+1}\right) p\left(\theta_{t+1} | \boldsymbol{\Theta}_{1:t}, r_{t+1}\right) p\left(r_{t+1} | r_t\right).$$

We have chosen a *von Mises*(μ, κ) probability density function (pdf) for the predictive distribution $p\left(\theta_t | r_t, \boldsymbol{\Theta}_{1:t-1}\right)$, where $\mu \in [-\pi, \pi)$ and $\kappa > 0$ are respectively the mean and concentration for the relevant data $\boldsymbol{\theta}_t^{(r_t)} = \left(\boldsymbol{\theta}_{t-r_t:t}\right)$. The von Mises distribution is one of the most popular parametric models for the analysis of *circular data* that resembles to a normal distribution of real-valued data.

A conditional prior on the change points, the *hazard rate* $H(x)$, is incorporated as

$$p\left(r_t | r_{t-1}\right) = \begin{cases} H(r_{t-1}+1) & \text{if } r_t = 0 \\ 1 - H(r_{t-1}+1) & \text{if } r_t = r_{t-1}+1 \\ 0 & \text{otherwise,} \end{cases}$$

with $H(x)$ formulated according to [1] as

$$H(x) = \frac{P_{gap}(g = x)}{\sum_{t=x}^{\infty} P_{gap}(g = t)}.$$

In our problem, we use a Gamma pdf as the prior in order to assign higher likelihoods to run lengths in a certain range: $P_{gap}(g = x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$, where $\alpha > 0$ and $\beta > 0$ represent the shape and rate parameters respectively. The parameters $\alpha = 1$ and $\beta = \frac{300}{\alpha}$ were empirically determined and correspond to an expected run length of 300 samples.

B. Discrete Change Point Extraction

We have used the forward-backward method to estimate the posterior distribution over run-lengths $p\left(r_t | \boldsymbol{\Theta}\right)$ at different times t. In this section, we derive a method to determine a single most likely sequence of run-lengths over the whole time span $t = 1, \ldots, M$ of the observation sequence.

For this we first define a matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ of transition likelihoods. A given entry $a_{i,j}$ of $\boldsymbol{A}$ corresponds to the a-priori likelihood for changing from a run length $r_t = i$ at time t to a run length $r_{t+1} = j$ at $t+1$. As described before, only two run lengths at time $t+1$ are possible given $r_t = i$. Assuming that both cases are equally likely, we define

$$a_{i,j} = \begin{cases} 0.5 & \text{if } j = i+1 \\ 0.5 & \text{if } j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

The most likely sequence of run-lengths is given by

$$\hat{r}_2, \ldots, \hat{r}_N = \underset{r'_2, \ldots, r'_N \in \{0, \cdots, N-1\}}{\arg\max} \left(\prod_{t \in \{2, \cdots, N\}} a_{r'_{t-1}, r'_t} p\left(r_t = r'_t | \Theta\right) \right)$$

and defining the initial run-length $\hat{r}_1 = r'_1 = 0$. Algorithmically, we solve this equation by formulating it as a recurrence relation $W_{t,k} = p\left(r_t = k | \Theta\right) . \max_{x \in \{0, \cdots, N-1\}} (a_{x,k} W_{t-1,x})$ which can be efficiently computed by a dynamic programming algorithm similar to the Viterbi algorithm (see [7]).

Note that while equal likelihoods for both transitions are used in the definition of $\boldsymbol{A}$, the run length likelihoods $p(r_t = i | \boldsymbol{\Theta})$ already incorporate our prior distribution.

The transition matrix ensures that the extracted state sequence is "valid", in the sense that a change point is always a discrete event that resets the run length to 0. It also ensures that the run length increases by exactly 1 in the absence of a change point. In contrast, many simpler techniques, such as taking the maximum likelihood run length $\max_i p(r_t = i)$ locally at each time t, do not lead to a consistent set of run lengths (Fig. 3).

Our approach is equivalent to the Viterbi algorithm (see [7]) applied to a hidden Markov model defined as follows:

The set of states is given by $Q = \{q_0, \ldots, q_{N-1}\}$ where the state q_i corresponds to a run length of i. The state transition likelihoods are given by our matrix $\boldsymbol{A}$. We define the sequence of observations $y_1, \ldots, y_N \in \mathbb{N}$ as $y_i = i$ (an integer t is observed at time t).

It follows

$$p(y_t | q_i) = \frac{p(q_i | y_t) p(y_t)}{p(q_i)} = \frac{p(r_t = i) p(y_t)}{p(q_i)}.$$

By definition, we have $p(y_t) = N^{-1}$ since each time $t \in 1, \ldots, N$ occurs exactly once in the observations $y_1, \ldots, y_N$. If we further assume $p(q_i)$ to be uniform, that is $p(q_i) = N^{-1}$ for every run-length, we get $p(y_t | q_i) = p(r_t = i)$. The state sequence chosen by the recurrence relation $V_{t,k} = p(y_t | q_k) \cdot \max_{\mathrm{x} \in \{0, \cdots, N\text{-}1\}} (a_{x,k} V_{t\text{-}1,\mathrm{x}})$ of the Viterbi algorithm is therefore equivalent to the run-length sequence determined by our algorithm.

III. Experimental Data Acquisition

The experimental data has been acquired from ten healthy subjects at Saarland University with no history of hearing deficits. Each subject received an audiogram test before the experiment and an audiogram check up after the experiment. The experiment was performed in a sound-proof room and subjects had to lie on a bed with their eyes closed. They were instructed to avoid any motion during the experiment.The sound stimuli were presented only to the right ear via a headphone (HDA 200, Sehnheiser) at two different sound levels of 50dB(SPL) and 100dB(SPL). A break of 3 minutes was given between the 50dB(SPL) and 100dB(SPL) sound exposures. The auditory stimuli consisted of pure tones of 1 kHz with a duration of 40ms, and a constant inter-stimulus interval (ISI) of 1s. The recorded EEG data was sampled at 512Hz.

A. EEG Post-Processing

Let $\mathcal{A} = \{\mathbf{s}_n \in \mathbb{R}^M : n = 1, 2, \ldots, N\}$ be a set of N sampled ERP single–trials within the time interval $[0, M/f_s]$ (f_s is the sampling frequency, M is the number of samples representing a trial) of a particular experiment. From $\mathcal{A}$ we

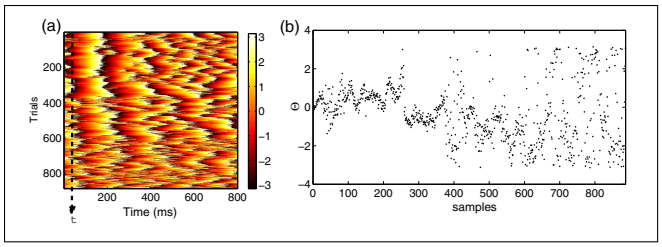

Fig. 1: A phase matrix $\boldsymbol{P}$ is represented in (a). We apply the CP algorithm on the phase information over all single trials at a specified time t. In plot (b) we illustrate the phase modulations for time t as marked in (a).

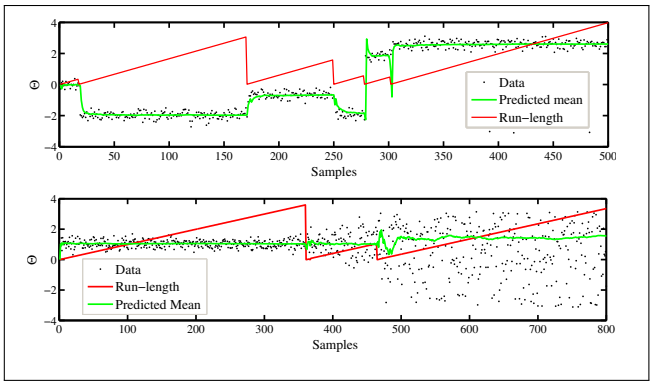

Fig. 2: Detection of change points for simulated circular data. The plot on top (a) shows the results of the change point detection on synthetic circular data with abrupt changes. The plot below (b) shows an example of the simulated ERP with the habituation effect.

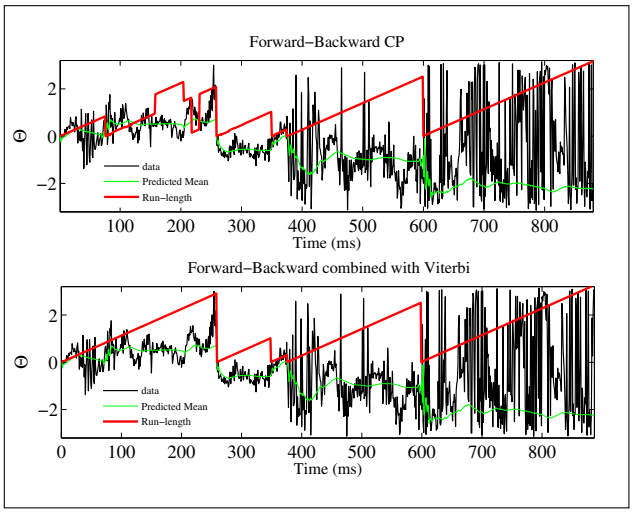

Fig. 3: Maximum likelihood run-length values extracted 1) by taking maximum likelihood values locally at each time (on top), and 2) by using our approach of applying the Viterbi algorithm to extract the globally most likely sequence of run-lengths (below).

can construct the *ERP image* $\mathbf{S} \in \mathbb{R}^{N\times M}$ such that $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N)^T$. For a two dimensional denoising of $\mathbf{S}$, we applied the non–local means scheme as described in [9]. This scheme exploits the event–related activity induced self–similarity in $\mathbf{S}$ for the denoising process. It can be used for an electrophysiologically well founded denoising of the ERP single–trials. We use the very same parameters as in [9] for the numerical experiments in this study.

B. Extraction of Phase Information

For the extraction of the instantaneous phase information, we employed the continuous wavelet transform as described in [8]. Let $\psi_{a,b}(\cdot) = |a|^{-1/2}\psi((\cdot - b)/a))$ where $\psi \in L^2(\mathbb{R})$ is the wavelet with $0 < \int_{\mathbb{R}} |\Psi(\omega)|^2 |\omega|^{-1} \mathrm{d}\omega < \infty$ ($\Psi(\omega)$ is the Fourier transform of the wavelet), and $a, b \in \mathbb{R}$, $a \neq 0$. The wavelet transform $\mathcal{W}_\psi : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R}^2, \frac{\mathrm{d}a\mathrm{d}b}{a^2})$ of a function $f \in L^2(\mathbb{R})$ with respect to the wavelet ψ is given by the inner L^2-product $(\mathcal{W}_\psi f)(a,b) = \langle f, \psi_{a,b}\rangle_{L^2}$. We used the numerical scheme of MITSI for the application of this concept to our discrete-time ERPs $\mathbf{s}_n$ $(n = 1, 2, \ldots, N)$,(See [4]). Note that in the discrete-time version $(\mathcal{W}_\psi \mathbf{s}_n)$, the parameters a and b are discretized on a scale-time grid $a_1, \ldots, a_K \times b_1, \ldots, b_M$. Note also that $b_{m+1} - b_m = 1/f_s$ for the uniformly sampled data with $b_1 := 0.0ms$ (stimulus onset) in the following.

For a fixed scale a and discretized translations b_m $(m = 1, 2, \ldots, M)$, we introduce the mapping $\mathcal{G}_a : \mathbf{S} \mapsto \mathbf{P}$, where $\mathbf{P} \in \mathbb{R}^{N\times M}$ has the entries $\theta_{n,m} = \arg((\mathcal{W}_\psi \mathbf{s}_n)(a, b_m))$. In other words, the $n = 1, 2, \ldots, N$ rows of $\mathbf{P}$ represent the instantaneous phase of the EPRs $\mathbf{s}_n$ for a fixed scale a. In our numerical experiments, we use the 6th-derivative of the complex Gaussian as wavelet ψ. It has been shown that for a physiologically meaningful range of values of a and b, the phase coherence across the trials $\mathbf{p}_n$ with $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n)^T$ constitutes a neural correlate of selective auditory attention ([3]). As in [8], [3] we use $a = 40$ for the numerical experiments in this study. Note that this corresponds to the alpha/theta-border frequency for the chosen wavelet. An example of phase representation is shown in Fig. 1.

IV. Results and Discussion

We first apply the Bayesian CP model combined with the Viterbi algorithm on generated synthetic data with abrupt changes. In Fig. 2, we show two examples of two different synthetically generated data samples with abrupt changes. In the second plot (b) we show the results for a simulated ERP with the habituation effect.

We then apply the method on the measured ERP phase information. We analyze the effect of habituation using the phase information of the different trials at a specific time of $97ms$ (equivalent to sample $t = 50$). The selected time is chosen to capture the N1-P2 wave complex that typically appears between 80 to 100ms after the onset of a stimulus. We illustrate the improvement achieved by applying the Viterbi algorithm compared to an alternative approach of picking maximum likelihood run-lengths locally at each trial in Fig. 3. The local approach suffers from inconsistencies, such as run length increased of more than 1 at a time, or falling back to a value other than 0. The run lengths extracted by our method on the other hand are globally consistent.

We apply the method for all 10 subjects on two different stimulus levels of 50dB and 100dB. In the context of habituation, we are interested predominantly in change points detected due to changes of the von Mises concentration parameter, rather than the mean. We therefore center the

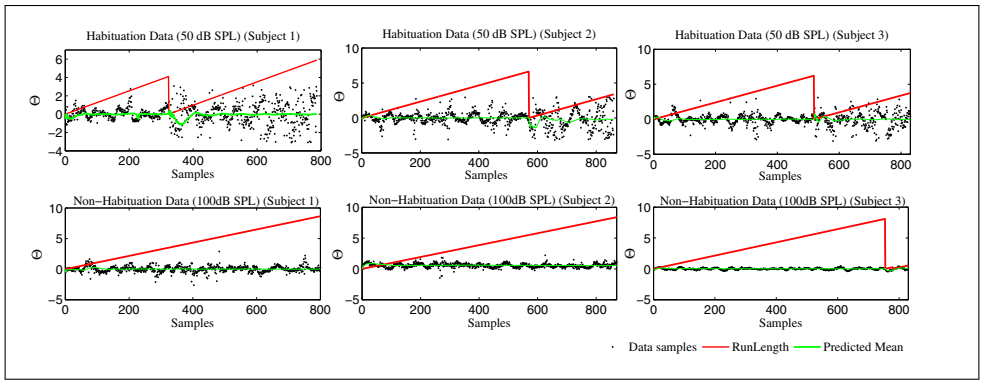

Fig. 4: The top row shows the results of the change point algorithm for 50dB SPL data for three subjects. The corresponding 100dB SPL for individual subject is plotted on the second row.

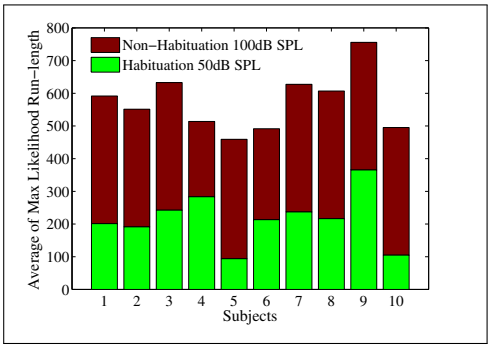

Fig. 5: Average run-length value in the maximum likelihood run-length sequence for all subjects. The 100 dB stimuli consistently lead to higher average run lengths compared to a 50 dB stimulus. The higher average run length corresponds to fewer change points in the time series.

phase data to a mean of zero by applying a Gaussian high pass filter before performing the change point analysis. Fig. 4 shows the detected change points for three subjects at 50dB and 100dB. The change points indicate the time at which the parameters – specifically the concentration – of the underlying distribution undergo a significant change. It is clearly evident that in 100dB stimulus, hardly any changes due to the high synchronization of phase information can be detected. However in 50dB stimulus, the changes in the diffusion of the phase modulations are more tractable and occur at earlier trials in comparison to the 100dB stimulus. There may remain fluctuations at later times due to the sensitization effect with regard to the stimuli, which does not reflect the habituation effect.

We summarize the results by showing the average over run-length values in the maximum likelihood run-length sequence for all subjects in Fig. 5. The 100dB data, which corresponds to a non-habituation process, exhibits a larger average run-length value in comparison to 50dB data (habituation). The difference is a consequence of the absence of change points over long periods (or even entirely) in the case of 100dB, leading to higher run-length values compared to the ones reached with the earlier change points in habituation data.

V. Conclusions

In this paper we combined the forward-backward Bayesian change point algorithm from [1] with the Viterbi algorithm to find the most likely set of discrete change points. Generally, our method can be applied to detect change points in directional data. We specifically applied the method to phase information in the context of a long-term habituation paradigm. We demonstrated that our approach is capable of reliably capturing habituation correlates in auditory ERP signals.

VI. Acknowledgements

This work has partially been supported by DFG-Grant STR 994/1-1, BMBF-Grant 03FH036I3 and DFG-Grant STE 571/11-1.

References

[1] Ryan Prescott Adams and David JC MacKay. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.

[2] Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, Francisco J Varela, et al. Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208, 1999.

[3] Yin Fen Low and Daniel J Strauss. A performance study of the wavelet-phase stability (wps) in auditory selective attention. *Brain research bulletin*, 86(1):110–117, 2011.

[4] Michel Misiti, Yves Misiti, Georges Oppenheim, and Jean-Michel Poggi. Wavelet toolbox. *Matlab User's Guide*, 1997.

[5] Florian Mormann, Klaus Lehnertz, Peter David, and Christian E Elger. Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients. *Physica D: Nonlinear Phenomena*, 144(3):358–369, 2000.

[6] Ulrich Paquet. Empirical bayesian change point detection. *Graphical Models*, 1995:1–20, 2007.

[7] Lawrence Rabiner and Biing-Hwang Juang. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4–16, 1986.

[8] Daniel J Strauss, Wolfgang Delb, Roberto D'Amelio, Yin Fen Low, and Peter Falkai. Objective quantification of the tinnitus decompensation by synchronization measures of auditory evoked single sweeps. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(1):74–81, 2008.

[9] Daniel J Strauss, Tanja Teuber, Gabriele Steidl, and Farah I Corona-Strauss. Exploiting the self-similarity in erp images by nonlocal means for single-trial denoising. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 21(4):576–583, 2013.