Time-frequency and synchrony analysis of responses to steady-state auditory and musical stimuli from multichannel EEG

Tomasz M. Rutkowski
Advanced Brain Signal Processing Lab
RIKEN Brain Science Institute
Wako-shi, Saitama, Japan
tomek@brain.riken.jp

Justin Dauwels
Amari Research Unit
RIKEN Brain Science Institute
Wako-shi, Saitama, Japan
justin@dauwels.com

François Vialatte, Andrzej Cichocki
Advanced Brain Signal Processing Lab
RIKEN Brain Science Institute
Wako-shi, Saitama, Japan
{fvialatte,cia}@brain.riken.jp

Danilo P. Mandic
Department of Electrical and Electronic
Engineering, Imperial College
London, United Kingdom
d.mandic@imperial.ac.uk

Abstract

Brain responses to audio stimuli are analysed using data driven time-frequency analysis. This is achieved based on the electroencephalogram (EEG) recordings and with auditory chirps or music as the audio stimulus. The empirical mode decomposition (EMD) is applied to multichannel EEG recordings, and the insight into the brain responses is provided by the analysis of the dynamics of auditory steady-state responses (ASSR). The proposed approach is further illustrated on the analysis of EEG responses to classical music. A comprehensive synchrony analysis is provided based on the visualization of EMD and spectrogram matching techniques. Simulation results illustrate the potential of the proposed approach in future brain computer/machine interfaces.

1 Introduction

Research on the interpretation of musical experience from neurophysiological recordings, such as electroencephalograms (EEG), has gained considerable attention in neuroscience community. Music is an important part of human daily experience, and it seems natural to explore ways to incorporate it in brain computer/machine interfaces (BCI/BMI). Research on BCI/BMI has been ongoing since three decades with the aim to enable additional or independent communication channel between the brain/mind and a computer/machine. This is achieved without involvement of the peripheral nervous system or muscles activities, as first postulated in [1] and later extensively summarized in [2]. This also allows for computer-aided communication with the outside world [3].

Due to the non-invasive manner there are several important and difficult challenges in the designing of BCI/BMI, however, this technology is envisaged to be at the core of future "intelligent" prosthetics, and is particularly suited to the needs of the handicapped and paralyzed. Other industries which would benefit greatly from the development of BCI include the entertainment, computer games, and automotive industries, where the control and navigation in a computer-aided application is achieved without the use of hands or gestures. The idea to utilize auditory or general musical stimuli for BCI/BMI is very appealing [3] as it allows to: (i) deliver stimuli via headphones or loudspeakers; (ii) integrate the auditory devices in user's environment in a natural way (sound stimuli can be perceived by human spatially - in comparison to visual or tactile stimuli where direct modality channel

must be established - looking at or touching); (iii) minimal user's distraction, since humans can generally concentrate on other tasks while listening to music or simple auditory tones (e.g. applications in cars, since auditory stimulation generally does not distract driver's visual attention); (iv) possible embedding of spell out ASSR stimuli within music, which user can listen to simultaneously; it is also possible to adaptively modify stimuli in order to avoid environmental sounds such as a car engine or other interferences; (v) there is no evidence of health hazard (causing seizures) by auditory steady-state stimuli [4]. This paper presents a study of EEG synchrony as a response to ASSR stimuli; this is used as a reference for further more complex analysis of natural musical stimuli. ASSR is already an established tool in objective hearing levels estimation [4, 5]. The auditory steady-state response (ASSR) is an auditory evoked potential, caused by modulated tones, that can be utilized in BCI/BMI for users of all ages. The ASSR response itself is an evoked neural potential that follows the envelope of a complex auditory stimulus. It is evoked by the periodic modulation, or turning on and off, of a tone or an auditory flutter [6]. The neural response is a brain potential that closely follows the time course of the modulation and thus is perfectly suited for initial study of subject's auditory responses. ASSR is also longer and more user friendly comparing to short clicks or beeps used in classical auditory evoked potential (AEP) studies [7, 8].

The steady–state responses can be recorded over a range of AM and FM modulation rates. Different modulation frequencies rates result in stimulation of different areas of the auditory pathway in the brain [6]. It was summarized in [4] that lower frequency rates ($f_m < 20$Hz) cause activity of the generators responsible for late-latency response, moderate rates ($f_m = 20\text{Hz}, \ldots, 60\text{Hz}$) are responsible for the middle-latency response, and higher rates ($f_m > 60$Hz) reflect activity from the brainstem. For experiments in this paper a frequency set $f_m \in \{7, 10, 13, 17, 21, 27, 31\}$Hz was chosen so we could cover lower frequency ranges which are more vulnerable to subject's state and motivational/attentional control as well as the middle stimuli frequencies which are less affected by subject's mental states.

We use the following novel approach to measure the interdependence of the EEG signals, based on the pairwise alignment ("matching") of their Hilbert-Huang spectra which provide more accurate and sharper time frequency representations comparing to contemporary methods [9]: first, empirical modes are extracted from the signals, which represent oscillatory components with time-varying amplitude and frequency. Second, the empirical modes are Hilbert transformed, resulting in very sharply localized ridges in the time-frequency plane; the obtained time-frequency representations are known as Hilbert-Huang spectra. Finally, the latter are pairwise aligned by means of the stochastic-event synchrony method (SES), a recently proposed procedure to match pairs of multi-dimensional point processes [12]. The level of similarity of two Hilbert-Huang spectra is quantified by three parameters: timing and frequency jitter of coincident ridges, and fraction of non-coincident ridges.

This paper is organized as follows. In the next section, we review the empirical mode decomposition (EMD) method and explain how it can be used to obtain intrinsic mode functions (IMFs) as well as time-frequency maps ("Hilbert-Huang spectra"). In Section 3, the idea behind stochastic event synchrony is outlined, its usability to quantify the similarity of Hilbert-Huang spectra is discussed. In Section 4, several measures to analyze EEG with brain response to steady-state auditory and musical stimuli are reviewed. The paper is concluded with comments on further potential applications of the proposed approach.

2 Empirical model decomposition (EMD)

Empirical Mode Decomposition (EMD) decomposes signals into so called "intrinsic mode functions" (IMF) [9]. They are functions that satisfy the following two conditions: (i) the number of extrema and the number of zero crossings are either equal or differ at most by one; (ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. An IMF represents an oscillatory mode within a given signal: its cycles (defined by its zero crossings) corresponds to *one* (and not more than one) mode of oscillation; both the amplitude and frequency of this oscillation may vary over time, in other words, the oscillation is not necessarily stationary nor narrow-band.

The process of extracting an IMF from a signal $x(t)$ ("sifting process" [9]) consists of the following steps:

1. determine the local maxima and minima of $x(t)$;
2. generate the upper and lower signal envelope by connecting those local maxima and minima respectively by some interpolation method (e.g., linear, spline, piece-wise spline [9] [10]);
3. determine the local mean $m(t)$, by averaging the upper and lower signal envelope;
4. subtract the local mean from the data: $h_1(t) = x(t) - m_1(t)$.

Ideally, $h_1(t)$ can already represent an IMF, however, in practice, $h_1(t)$ still typically contains local asymmetric fluctuations, and the above four steps need to be repeated several times. In order to obtain the second IMF, the sifting process is applied to the residue $\varepsilon_1(t) = x(t) - \text{IMF}_1(t)$, obtained by subtracting the first IMF from $x(t)$; the third IMF is extracted from the residue $\varepsilon_2(t)$ and so on. The sifting process is terminated when two consecutive iterations yield similar results. The empirical mode decomposition of the signal $x(t)$ can be written as:

$$x(t) = \sum_{k=1}^{n} \text{IMF}_k(t) + \varepsilon_n(t), \tag{1}$$

where n is the number of extracted IMFs, and the final residue $\varepsilon_n(t)$ can either be the mean trend or a constant. An EMD example for EEG signal decomposition is given in Fig. 1.

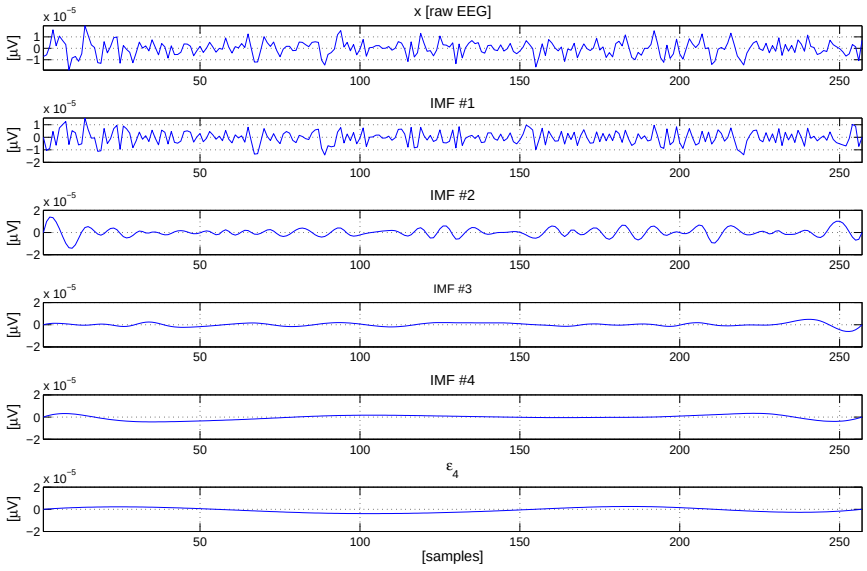

Figure 1: Empirical mode decomposition (EMD): the original signal (top), three IMFs (middle), and residue (bottom).

From (1), the empirical mode decomposition is a complete operator; the original signal can be reconstructed from the IMFs and the final residue. Note that due to their data dependent nature, in practice the IMFs are not guaranteed to be mutually orthogonal, but are often close to orthogonal. Based on IMFs we can construct a time-frequency representation of the signal $x(t)$, i.e., the Hilbert-Huang spectrum (HHS) [9]. The idea is to compute the instantaneous amplitude and frequency for each IMF, for instance derived from the analytic signal:

$$Z(t) = \text{IMF}(t) + i\,Y(t), \qquad i = \sqrt{-1}, \tag{2}$$

with $Y(t)$ the Hilbert transform of the IMF:

$$Y(t) = \frac{1}{\pi}\,\text{P}\int_{-\infty}^{+\infty} \frac{\text{IMF}(t)}{t - t'} dt', \tag{3}$$

where P indicates the Cauchy principal value of the integral [11]. The instantaneous amplitude $a(t)$ and phase $\theta(t)$ of the IMF are defined respectively as the magnitude and angle of $Z(t)$. The instantaneous frequency is then simply defined as:

$$\omega = \frac{d\theta(t)}{dt}. \tag{4}$$

Hilbert-Huang spectra (HHS) are plots of the instantaneous amplitude against instantaneous frequency and time [9]. HHS images are typically sparse and contain sharp ridges, as illustrated in Fig. 2. In the following section, we propose a method to quantify the similarity of two Hilbert-Huang spectra; the key idea will be to match ridges in one HHS to ridges in the other HHS.

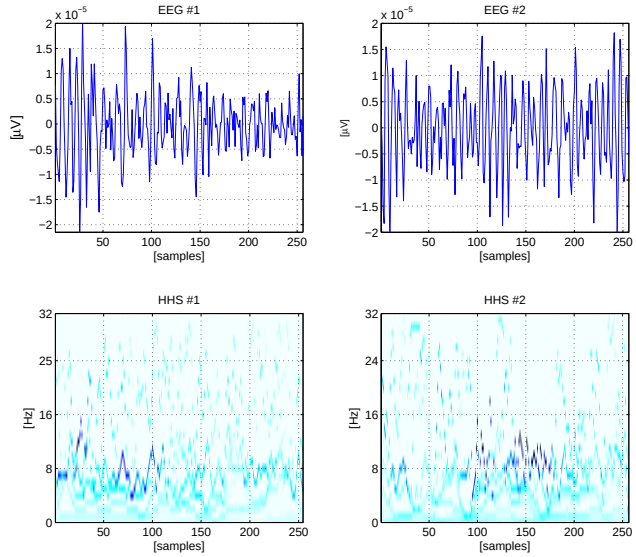

Figure 2: Hilbert-Huang spectra (bottom) of two EEG channels (top)

3 Stochastic Event Synchrony (SES)

In earlier work [12], we developed a measure that quantifies the similarity ("interdependence" or "synchrony") of two (one- or multi-dimensional) point processes, referred to as *stochastic event synchrony* (SES); that measure was applied to point processes on the time-frequency plane, more precisely, to bump models [13]. Here we will use SES to quantify the similarity of two Hilbert-Huang spectra. The first step is to extract a point process ("event string") from an HHS (see Fig. 3 (left)): from the given HHS, we only retain the N largest instantaneous amplitudes (N is typically about 100). Each of the N remaining points ("ridges") $r_j = (t_j, a_j, f_j)$ with $j = 1, \ldots, N$ is viewed as an event, and the sequence $r = (r_1, \ldots, r_N)$ is a three-dimensional point process; in other words, the remaining points r_j in Fig. 3 (left) take the role of the "bumps". Fig. 3 (left) suggests a natural way to define the similarity of two HHS: ridges in one time-frequency map (red) may not be present in the other map (blue) ("non-coincident" ridges); other ridges are present in both maps ("coincident ridges"), but appear at slightly different positions on the maps. Fig. 3 (right) depicts the coincident ridges, obtained after matching the event strings r (red) and r' (blue); the black lines connect the centers of coincident ridges, and hence, they visualize the offset in position between pairs of coincident ridges. Stochastic event synchrony consists of the following parameters: (i) ρ: fraction of non-coincident ridges; (ii) δ_t and δ_f: average time and frequency offset respectively between coincident ridges; (iii) s_t and s_f: variance of the time and frequency offset respectively between coincident ridges. The alignment of the two ridge traces (cf. Fig. 3 (right)) is cast as a statistical inference problem [12]. The associated probabilistic model depends on the SES parameters $\theta = (\delta_t, \delta_f, s_t, s_f)$ besides the following two kinds of latent variables: (i) binary variables $C_{kk'}$, associated to each pair of ridges, where $C_{kk'} = 1$ indicates that event r_k of the first HHS is coincident with event $r'_{k'}$ in

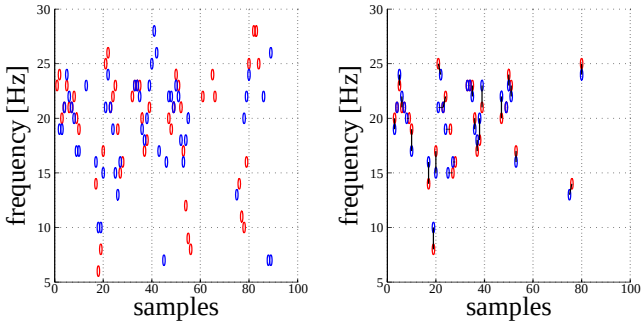

Figure 3: (left) Events ("ridges") extracted from HHS1 (red) and HHS2 (blue) of Fig. 2; (right) coincident ridges.

the other HHS, and where $C_{kk'} = 0$ otherwise; (ii) binary variables B_k and $B'_{k'}$, which indicate whether a ridge is non-coincident. The latent-variable model is of the form:

$$
\begin{aligned}
& p(r, r', b, b', c, \theta) \propto \\
& \prod_{k=1}^{n} (\beta\delta[b_k - 1] + \delta[b_k]) \prod_{k'=1}^{n'} (\beta\delta[b'_k - 1] + \delta[b'_k]) \\
& \cdot \prod_{k=1}^{n} \prod_{k'=1}^{n'} \Big(\mathcal{N}(t'_{k'} - t_k; \delta_t, s_t)\, \mathcal{N}(f'_{k'} - f_k; \delta_f, s_f) \Big)^{c_{kk'}} \\
& \cdot \prod_{k=1}^{n} \big(\delta[b_k + \sum_{k'=1}^{n'} c_{kk'} - 1] \big) \prod_{k'=1}^{n'} \big(\delta[b'_{k'} + \sum_{k=1}^{n} c_{kk'} - 1] \big) p(\delta_t) p(s_t) p(\delta_f) p(s_f), \qquad (5)
\end{aligned}
$$

where β is a constant (which serves as a knob to control the number of non-coincident ridges), n and n' is the total number of ridges in the two HHS, and $\mathcal{N}(x; m, s)$ stands for a univariate Gaussian distribution with mean m and variance s [12]. For convenience, we choose "improper" priors $p(\delta_t) = p(\delta_f) = p(s_t) = p(s_f) = 1$. The SES parameters $\theta = (\delta_t, \delta_f, s_t, s_f)$ and the latent variables C, B and B' are determined jointly by MAP estimation. This may be carried out by cyclic maximization [12]: for fixed θ, one maximizes $\log p$ (cf. (5)) w.r.t. C, B and B' and vice versa. Conditional maximization w.r.t. θ is straightforward, however, the conditional maximization w.r.t. C, B and B' is non-trivial: it involves a mathematically intractable discrete optimization problem. This problem is solved approximately (but successfully) by iterative max-product message passing ("iterative dynamic programming") on a graphical model corresponding to the latent-variable probabilistic model (5) [12]. In Fig. 4, the edges correspond to variables, and the nodes corresponds to factors in (5). The nodes $\mathcal{N}$ corresponds to the Gaussian distributions in (5), whereas the nodes denoted by $\bar{\Sigma}$ represent the factors $\big(\delta[b_k + \sum_{k'=1}^{n'} c_{kk'} - 1]\big)$ (blue) and $\big(\delta[b'_{k'} + \sum_{k=1}^{n} c_{kk'} - 1]\big)$ (red), and the nodes denoted by β correspond to the factors $(\beta\delta[b_k - 1] + \delta[b_k])$ and $(\beta\delta[b'_k - 1] + \delta[b'_k])$. The arrows in Fig. 4 depict "messages" (i.e., probabilities associated with the coincident and non-coincident pairs of ridges) that are iteratively computed at each node according to the max-product computation rules. Intuitively, the nodes may be viewed as computing elements that iteratively update their evaluation about which ridges match and which do not, based on the opinions ("messages") they receive from neighboring nodes. After the algorithm has converged (and the nodes have found a "consensus"), the messages are combined to obtain a decision on C, B and B', and an estimate of ρ and the other SES parameters [12].

4 Experiments and Results

We considered the proposed ridge analysis and SES matching method and applied them to a BCI/BMI problem where subject was asked to concentrate or ignore given auditory stimuli. The EEG signal were recorded from a human subject listening to steady-state auditory and musical examples (introduction of the *Beethoven Symphony No. 7*). The EEG signals were recorded from 12

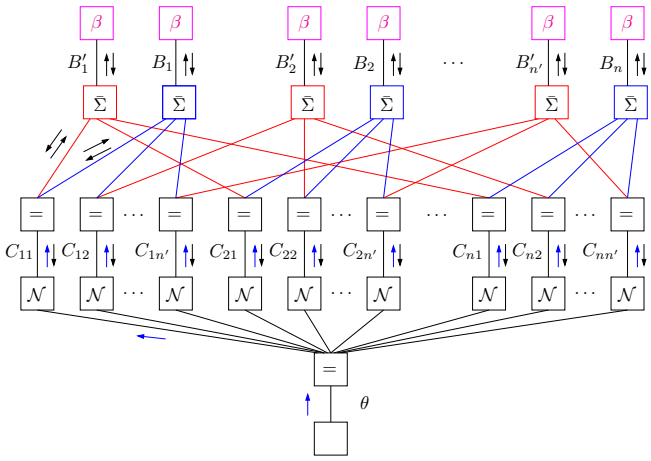

Figure 4: Graphical model of (5).

locations (four frontal [Fp1, Fp2, F3, F4], six auditory cortex areas [C3, C5, T7, C4, C6, T8], and two parietal [P3, P4]) [7]. EMD was used to extract IMF components from those the EEG signals (cf. Section 2). Correlation analysis of IMF components in frequency ranges from 1Hz to 32Hz is presented in Fig. 5, where different patterns of brain responses are presented. Observe, that ASSR stimuli "correlated" across the whole brain, whereas musical stimuli caused synchrony at only a few EEG locations. The lack of auditory stimuli is represented in the from of "non-correlated" patterns. This simple correlation analysis of IMFs did not result in good classification of "musical states". To overcome this problem, as a next step the IMF components were Hilbert transformed (resulting in Hilbert-Huang spectra). Based on the underlying psychophysics, the low-frequency drifts ($<$5Hz) and high-frequency interference parts ($>$32Hz) of the IMFs were removed; one such a segment of two EEG signals (recorded by two auditory channels) together with their Hilbert-Huang spectra is shown in Fig. 2. Finally, the similarity of the so obtained HHS was analyzed by means of stochastic event synchrony (cf. Section 3). Each pair of HHS was matched, and the timing and frequency jitter s_t and s_f of the coincident ridges, and the fraction of non-coincident ridges ρ were computed. Those parameters were then averaged over all pairs of Hilbert-Huang spectra, resulting in three global measures of interdependence. As a benchmark, we also quantified the synchrony of the EEG signals by some classical measures: magnitude squared coherence (COH), partial directed coherence (PDC), and directed transfer function (DTF) [15] (see Table 1).

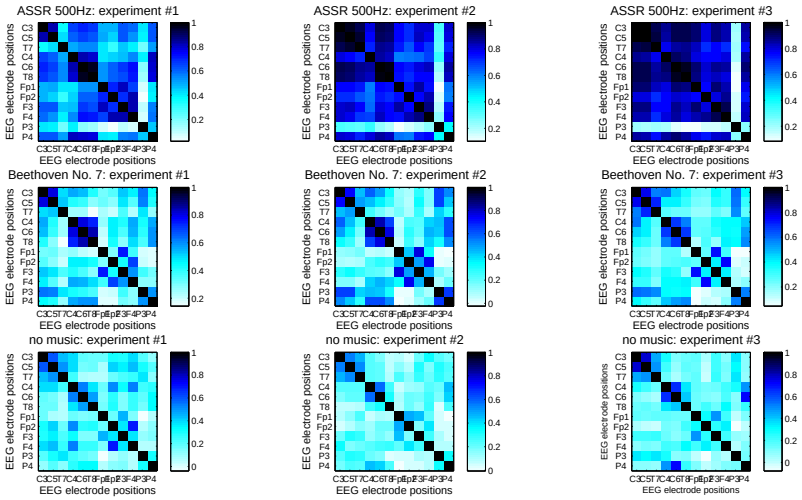

Figure 5: EMD based correlation analysis of the responses related to musical and non-musical events.

Table 1: A comparison of the percentage of incorrectly classified EEG signals using conventional measures for auditory stimuli and the EEG response to steady-state visual evoked potential (SSVEP) experiment for comparison [14].

FEATURES	SYMPHONY	ASSR 500Hz	ASSR 1500Hz	SSVEP
dDTF with PDC	37.5%	38.0%	24.0%	42.2%
PDC with SES ρ	30.0%	40.0%	34.0%	12.5 %
PDC with SES s_t	25.0%	54.0%	30.0%	40.6%
PDC with SES s_f	32.5%	46.0%	30.0%	29.7%
COH with SES ρ	45.0%	50.0%	56.0%	7.8%
DTF	45.0%	42.0%	46.0%	39.1%
PDC	30.0%	46.0%	32.0%	42.2%
COH	42.5%	40.0%	42.0%	25.0%

We next investigated whether fluctuations in EEG synchrony ("interdependence") can be detected in this way. In particular, our aim was to distinguish between EEG signals recorded during stimulation and silent intervals; for each of those two conditions, we considered 32 EEG segments of the same length (2 to 5s). Table 1 shows classification errors obtained by the leave-one-out method; for the sake of simplicity, a linear classifier (hyperplane) was used, whereby synchrony measures were used as single features and as pairs of features. Table 1 only contains the results for the pairs that resulted in the smallest classification errors: for the pairs without SES parameters, the combination of several measures with PDC or even the only use of PDC gave the preliminary results showing minimum 25% of incorrect classification for musical stimuli. Similar results were obtained in the case of ASSR stimuli with carrier frequencies $f_m \in \{500, 1500\}$Hz and modulation frequencies $f_s \in \{7, 10, 13, 17, 21, 27, 31\}$Hz and duration of 1s for each segment showed (see Table 1). The presented preliminary results show significant improvement in classification rates as compared to previously published examples [3]; in our experiment, the stimuli exposure times were shorter, and resulted in clear separation as seen in Figure 6. For comparison, results using the same method applied to the much stronger EEG response of the steady-state visual evoked potential (SSVEP) are also shown.

5 Conclusions

We have analyzed the synchrony in multichannel EEG recordings during both ASSR and musical stimulation. This has been achieved for a number of performance metrics, with most promising results obtained by PDC and the SES parameter s for musical stimuli. The results based on simpler ASSR stimuli have exhibited slightly lower recognition rates, suggesting the high potential of musical stimuli in future BCI/BMI. Previous studies have reported similar findings, emphasizing the level of EEG synchrony is correlated with the subject's attention [13]. This study show that ASSR and BCI/BMI based on musical stimuli may be even used by subjects without prior training. However, state-of-the-art ASSR and musical stimuli based BCI/BMI systems require computationally demanding signal processing and response detection algorithms, yielding relatively low data rates. This is largely due to the fact that those systems typically use the power spectrum as input features only, in particular, the power spectrum at the ASSR and musical stimulation frequencies. Our study has shown (see also [16]), that EEG synchrony has great potential for ASSR detection; despite a relatively short stimulation period (only 2 to 5s), we obtained low classification errors (25% at best). As a consequence, synchrony measures (particularly SES in conjunction with EMD) may prove very useful in the context of future practical BCI/BMI.

References

[1] J. J. Vidal, "Toward direct brain-computer communication," *Annual Review of Biophysics and Bioengineering*, vol. 2, pp. 157–180, June 1973.

[2] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. P. G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, June 2000.

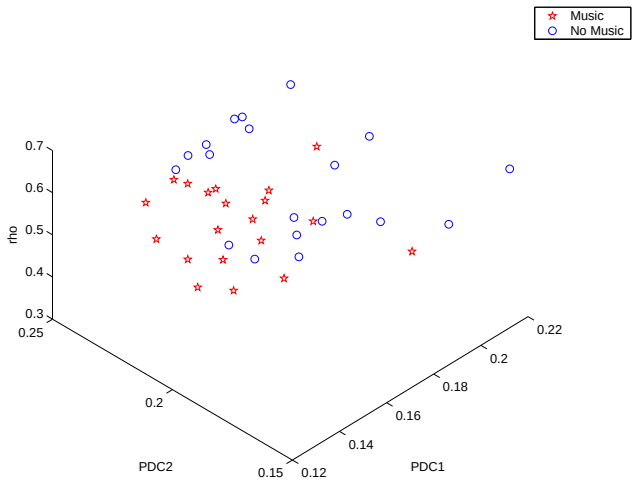

Figure 6: Feature clustering: SES ρ vs. PDC_1 vs. PDC_2 separable features for musical (red stars) and non-musical (blue dots) stimuli periods.

[3] E. R. Miranda, S. Roberts, and M. Stokes, "On generating EEG for controlling musical systems," *Biomedizinische Technik*, vol. 49, no. 1, pp. 75–76, 2004.

[4] B. A. Stach, "The auditory steady-state response: A primer," *The Hearing Journal*, vol. 55, no. 9, pp. 10–18, September 2002.

[5] R. Galambos, S. Makeig, and P. J. Talmachoff, "A 40-Hz auditory potential recorded from the human scalp," *Proceedings of National Academy of Sciences*, vol. 78, no. 4, pp. 2643–92 647, 1981.

[6] D. Bendor and X. Wang, "Differential neural coding of acoustic flutter within primate auditory cortex," *Nature Neuroscience*, p. doi:10.1038/nn1888, November 2007.

[7] E. Niedermeyer and F. L. Da Silva, Eds., *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed. Lippincott Williams & Wilkins, 2004.

[8] G. Plourde, "Auditory evoked potentials," *Best Practice & Research Clinical Anaesthesiology*, vol. 20, no. 1, pp. 129–139, 2006.

[9] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, March 1998.

[10] T. M. Rutkowski, A. Cichocki, and D. P. Mandic, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*. Springer, 2008, ch. Information Fusion for Perceptual Feedback: A Brain Activity Sonification Approach, p. (in press).

[11] S. L. Hahn, *Hilbert Transforms in Signal Processing*. Boston - London: Artech House, 1996.

[12] J. Dauwels, F. Vialatte, T. Rutkowski, and A. Cichocki, "Measuring synchrony by probabilistic message passing," in *Neural Information Processing Systems (NIPS)*, Vancouver, Canada, December 3–6, 2007, p. (in press).

[13] F. Vialatte, C. Martin, R. Dubois, J. Haddad, B. Quenet, R. Gervais, and G. Dreyfus, "A machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics," *Neural Networks*, vol. 20, pp. 194–209, 2007.

[14] J. Dauwels, T. M. Rutkowski, F. Vialatte, and A. Cichocki, "On the synchrony of empirical mode decompositions with application to electroencephalography," in *Proceedings of ICASSP2008*, p. (submitted), 2008.

[15] M. Kamiński and H. Liang, "Causal influence: Advances in neurosignal analysis," *Critical Review in Biomedical Engineering*, vol. 33, no. 4, pp. 347–430, 2005.

[16] E. Gysels, "Phase synchronization for classification of spontaneous EEG signals in brain-computer interfaces," No. 3397, EPFL, Switzerland, 2005.