

Discriminating Multiple Emotional States from EEG Using a Data-Adaptive, Multiscale Information-Theoretic Approach

Yelena Tonoyan

*Research Group Neurophysiology
Laboratory for Neuro- and Psychophysiology
O&N II Herestraat 49 – Box 1021
3000 Leuven, Belgium
yelena.tonoyan@med.kuleuven.be*

David Looney* and Danilo P. Mandic†

*Communication and Signal Processing Research Group
Department of Electrical and Electronic Engineering
Imperial College
Room 813, Level 8, Exhibition Road
London, SW7 2BT, United Kingdom
*david.looney06@imperial.ac.uk
†d.mandic@imperial.ac.uk*

Marc M. Van Hulle‡

*Research Group Neurophysiology
Laboratory for Neuro- and Psychophysiology
O&N II Herestraat 49 – Box 1021
3000 Leuven, Belgium
marc@neuro.kuleuven.be*

Accepted 27 November 2015

Published Online 28 January 2016

A multivariate sample entropy metric of signal complexity is applied to EEG data recorded when subjects were viewing four prior-labeled emotion-inducing video clips from a publically available, validated database. Besides emotion category labels, the video clips also came with arousal scores. Our subjects were also asked to provide their own emotion labels. In total 30 subjects with age range 19–70 years participated in our study. Rather than relying on predefined frequency bands, we estimate multivariate sample entropy over multiple data-driven scales using the multivariate empirical mode decomposition (MEMD) technique and show that in this way we can discriminate between five self-reported emotions ($p < 0.05$). These results could not be obtained by analyzing the relation between arousal scores and video clips, signal complexity and arousal scores, and self-reported emotions and traditional power spectral densities and their hemispheric asymmetries in the theta, alpha, beta, and gamma frequency bands. This shows that multivariate, multiscale sample entropy is a promising technique to discriminate multiple emotional states from EEG recordings.

Keywords: Emotion; complexity; multiscale sample entropy; EMD.

‡Corresponding author.

1. Introduction

Identification of emotional states from EEG in response to associated stimuli has long been sought for diagnosing and treating patients with dysfunctional processing of emotional information.¹ More recently it has been linked to advanced applications such as emotion-sensitive interactive games, affective interfaces, and emotion-sensitive tutoring systems.^{2–4} However, as electromagnetic activity elicited by cortical structures involved in processing emotional information is hard to gauge from EEG electrodes,⁵ the identification and discrimination of emotional states is regarded as notoriously challenging in EEG research. In this paper, we will take up this challenge and develop a new approach for discriminating multiple emotional states from EEG when viewing emotion-inducing video clips.

Emotions versus EEG frequency bands

The traditional approach is to evaluate the overall power within a given frequency band. As already reviewed in the work of Davidson,⁶ alpha band activity (8–13 Hz) is commonly evaluated for changes related to the induction of different emotions. Kostyunina and Kulikov⁷ found that different emotional states correspond to different peak frequencies in the alpha band. Shemyakina and Danko⁸ showed that significant differences in local EEG power and spatial synchronization between electrodes can be observed with different emotions. This effect was largest over the temporal area of the brain. Other researchers did not attempt to distinguish emotions from one another but rather grouped those into positive and negative ones (“valence”). This classification was promoted through the widespread use of the International Affective Pictures System,⁹ a well-validated set of visual stimuli in which valence is further delineated in arousal (and later also dominance). Overall changes in alpha power and lateralization effects related to these changes have fed the Hemispherical Emotional Valence (HEV) hypothesis, the lateralized representation of negative and positive emotions on the human scalp, albeit that the empirical evidence has been called into question.^{10,11} In addition to alpha, changes in the lower theta band have also been noted. Jaušovec and co-workers¹² asked subjects to process emotional content in video clips. They observed that changes in theta occurred

2–3 s into the video clip and differentiated between subjects with low versus high scores on assessments of emotional intelligence. Krause and co-workers¹³ studied differences between EEG bands instead of between brain areas. They showed that the 4–6 Hz band (termed theta 1) elicited a greater synchronization when viewing an aggressive film than when viewing a neutral or sad film. Vecchiato and co-workers¹⁴ observed an increase in theta power in the left frontal brain areas of participants that liked the commercials they viewed compared to those they disliked. Overall, there are two general problems with simply using spectral power changes to measure immediate responses to affective stimuli.¹⁵ First, while studies abound that show differences between negative and positive emotional valences, spectral power patterns that distinguish between emotions of the same valence have not been consistently reported. Spectral power changes seem not to be optimal in this case. Second, the limited temporal resolution could be ineffective to distinguish between different emotional states. To curb the latter, event-related desynchronization/event-related synchronization (ERD/ERS) has been suggested as it can detect rapid amplitude changes within specified frequency bands. ERD/ERS measures have been used for assessing emotional responses to affective stimuli thereby focusing on the lower theta band (3–8 Hz).^{16,17} Patterns of increased theta power using ERD/ERS were also detected in the video clip study of Jaušovec and co-workers.¹² Another approach is to compute density functions of instantaneous amplitudes of wavelet transformed EEG recordings and verify whether for certain wavelet scales these density functions collapse across electrodes and/or subjects (universal scaling behavior), a technique that has been used to distinguish cognitive states including listening to music¹⁸ and mental imagery.¹⁹

Emotions versus EEG-ERPs

Another line of research is based on the event-related potential (ERP), a stereotyped, transient response to a sensory or cognitive stimulus or a motor event.²⁰ Traditionally, ERPs are measured as latencies and amplitudes of positive and negative potentials. However, Paulmann and Kotz²¹ reported that ERP components did not differ based on emotional valence. In contrast, Spreckelmeyer

and co-workers²² showed that ERP components did differentiate between short vocalizations representing happy and sad emotions.

Several researchers reported that ERP measures alone do not provide convincing evidence but some ERP components may have a direct correlation with theta activity. For example, Balconi and Pozzoli¹⁷ observed an increased theta activity in response to affective pictures that correlated directly with the N2 ERP component.

Emotions versus EEG signal complexity

A well-known hypothesis is the decrease in complexity of a physiological or behavioral signal with disease or aging.²³ An observed loss in complexity is attributed to a loss or impairment of functional components and/or their (nonlinear) coupling. This has motivated researchers to look at signal complexity as a diagnostic EEG marker.^{24–34} However, signal complexity has also been related to emotional states. Aftanas and co-workers.³⁵ showed that negative and positive emotions occurred with higher values of EEG dimensional complexity (correlation dimension) estimates compared to the neutral viewing condition. Hosseini *et al.*³⁶ applied two entropy metrics (approximate and wavelet entropy) to discriminate between two emotional states (calm-neutral and negative-excited) in response to viewing sequences of emotion inducing pictures and achieved 73.25% classification accuracy. Jie *et al.*³⁷ applied sample entropy on two binary emotion recognition tasks (positive versus negative emotion both with high arousal and music clips with different arousal levels) and achieved 80.43% and 79.11% classification performance. We will continue with this putative connection between complexity (in particular, sample entropy) and emotion and explore it on a multiscale and data-adaptive level.

Multiscale measure of EEG complexity

Costa *et al.*³⁸ proposed multiscale sample entropy (MSE) which calculates SE, a measure of the degree of randomness of a signal, across multiple scales. This method reveals the interdependence between entropy and scale and led researchers to associate complexity with the ability of the sensed system (i.e. the brain of a healthy subject) to adjust to

a changing environment. In the original (MSE) method, the scales are determined by the so-called coarse-grained procedure where the original signal is averaged over nonoverlapping windows of increasing length.³⁸ The MSE is then computed by applying sample entropy.³⁹ But as the coarse-grained procedure essentially corresponds to a linear smoothing and decimation of the original time series, only low-frequency components are captured as the high-frequency ones at fine scales are lost. A way to overcome this is to apply multivariate empirical mode decomposition (MEMD),⁴⁰ a fully data-driven, time-frequency technique that decomposes a signal into a finite set of amplitude/frequency modulated components, called intrinsic mode functions (IMFs). The successful application of empirical mode decomposition for the detection of epileptic seizures with EEG is shown in the work of Martis *et al.*^{41,42} When applying empirical mode decomposition to EEG recordings, sample entropy can be estimated in each IMF individually. Sharma and co-workers.⁴³ recently used this approach for classifying focal from nonfocal EEG signals of epileptic patients and achieved 87% correctness. A further improvement is *multivariate*, multiscale entropy (MMSE), to account for both within and cross-channel dependencies, and its combination into MEMD-enhanced MMSE.⁴⁴ to exploit multiple data-driven scales in the entropy calculation. In our previous work, we showed the connection between cognitive task performance (based on the participant's behavioral response) and EEG signal complexity using MEMD-enhanced MMSE.⁴⁵ In the present work, we assert that EEG signal complexity measured by MEMD-enhanced MMSE can be used to discriminate between emotional states. We will test our assertion using EEG from 30 participants that viewed four prior-labeled emotion-inducing video clips taken from a publicly available database developed by Schaefer and co-workers.⁴⁶ Our participants were also asked to provide their own emotion labels. We apply mixed models with quadratic functions for the obtained multiscale complexity curves and show that for mid-frontal EEG electrodes we can distinguish between five self-reported emotions (anger, disgust, amusement, tenderness, and sadness). We also compare our complexity results with Schaefer's emotion labels and arousal scores and also with the traditional EEG power spectral densities and their hemispheric asymmetries

in the theta (4–7 Hz), alpha (8–15 Hz), beta (16–31 Hz), and gamma (30+ Hz) frequency bands.

2. Materials and Methods

2.1. Materials

We took our emotion-eliciting video clips from the publicly available database developed by Schaefer and co-workers⁴⁶ (<http://nemo.psp.ucl.ac.be/FilmStim/>). The spoken language of the video clips is French or dubbed into French. Each video clip is labeled in terms of emotional category (fear, anger, sadness, disgust, amusement, tenderness, neutral — further called “standard label”) and scored by their participants on a 7-point scale in terms of emotional arousal: “While I was watching the film...” (1) = “I felt no emotions at all” to (7) = “I felt very intense emotions” (further called “self-reported emotional arousal”). The participants were encouraged to report what they actually felt in reaction to the video clips, not what they believed people should feel. For our study, we selected four video clips with top three mean self-reported arousal levels from four different emotional categories (Table 1). The average duration of the four clips was 3 min.

2.1.1. Participants

The experiment was performed with 30 healthy volunteers (20 female, 10 male, mean age = 32.48, SD = 15.77, age range 19–70) that mastered French language (i.e. French as mother tongue or French-Dutch bilinguals). They were recruited via emails, flyers, and announcements. Some were graduate students, often regular subjects in EEG experiments, and were paid. No participant had any known neurological or psychiatric disorder. The age range was intentionally broad and age was a parameter in our statistical tests (cf., the hypothesis on the decrease in signal complexity with age⁴⁷). Ethical approval for this study was granted by an independent ethical committee (“Commissie voor Medische Ethiek” of UZ Leuven, the university hospital). This study was conducted in accordance with the most recent version of the Declaration of Helsinki.^a

^aOctober 2013, Fortaleza, Brazil, see <http://www.wma.net/en/30publications/10policies/b3/>.

Table 1. Video clips used and their standard labels, their mean self-reported emotional arousal levels, standard deviation (SD) and number of participants (*N*) (data from Schaefer and co-workers, see also Appendix A).

Video clip	Standard label	Mean arousal	SD	<i>N</i>
“Sleepers”	Anger	5.63	1.17	57
“Life is beautiful (4)”	Tenderness	5.59	1.19	50
“City of angels”	Sadness	5.15	1.70	56
“La cité de la peur”	Amusement	4.52	1.75	55

2.1.2. Labels and variables

The experimental paradigm consists of four trials, one for each video clip. The clips were presented in random order. After having watched a video clip, participants were asked to report their emotional category (fear, anger, sadness, disgust, amusement, tenderness, neutral). We further call these the “self-labels”. Note that our participants were not informed about the video clip’s standard labels. To summarize, we have three labels:

- standard label: the emotional category of each video clip as reported in Ref. 46;
- self-label: the emotional category of each video clip as reported by our subjects;
- self-reported emotional arousal scores of each participant in the Schaefer *et al.* study⁴⁶ for the considered video clips (scores provided to us by Alexandre Schaefer).

As our participants’ self-labels could not always be in alignment with Schaefer’s standard labels, we created the variable *standard_self* with “standard” referring to the standard label and “self” to the self-label. Unless noted otherwise, we use self-labels for labeling MMSE curves.

2.1.3. EEG recording and preprocessing

Participants were tested in a sound-attenuated, darkened room with a constant temperature of 20°, sitting in front of an LCD screen. The participant’s task was to watch the four video clips and report their emotional categories. When viewing a clip, EEG was recorded continuously using 32 active electrodes, evenly distributed over the entire scalp (positioning and naming convention following a subset of the

extended 10–20 system) with a BioSemi ActiveTwo system (BioSemi, Amsterdam, the Netherlands) as well as an electro-oculogram (EOG) using the setup of Croft and co-workers.⁴⁸ The EEG signal was re-referenced offline from the original common mode sense reference⁴⁹ (CMS, positioned next to electrode Pz) to the average of two additional electrodes that were placed on the mastoids of the subject. The duration of the experiment excluding electrode setup was 20 min. The EEG signal was filtered using a 4th-order Butterworth filter with range 0.5–30 Hz. Then the initial sampling rate of 2048 Hz was downsampled to 128 Hz (including anti-aliasing), to reduce computational costs. Finally, the EOG signal was used to remove eye artifacts following the method of Croft and co-workers.⁴⁸

2.1.4. Channel selection

As not all EEG electrodes are expected to be relevant for capturing differential emotional responses, we selected electrodes F3 and F4 (mid-frontal areas). Our motivation stems from several studies. For example, in the work of Oschsner and Gross⁵⁰ it was shown that the orbital frontal cortex plays a critical role in cognitive control of emotion (especially in the case of suppressing emotional responses), and activity in this region reflects subsequent appraisal processes related to viewing emotional stimuli. For comparison's sake, besides temporal electrodes T7 and T8, which are thought to gauge memory and imagery processes as well as emotional state modulation,^{51,52} we also consider electrodes O1 and O2 from the occipital pole so as to verify whether observed differences in complexity can be explained by differences in visual processing.

2.2. Methods

2.2.1. Multivariate empirical mode decomposition (MEMD)

Empirical Mode Decomposition (EMD) decomposes a signal into a finite number of narrow-band, amplitude/frequency modulated components known as IMFs⁵³:

$$\text{Signal} = \text{IMF}_1 + \text{IMF}_2 + \text{IMF}_3 + \dots + \text{IMF}_n$$

with IMF_1 corresponding to the highest frequency component and subsequent IMFs to gradually lower, more narrow-banded frequency components. The

last IMF is the trend in the signal and is usually omitted from further analysis. The decomposition operates as follows, starting with IMF_1 : first locate the local maxima and minima in the original signal, then construct an envelope that interpolates between these local minima, respectively the local maxima, and subtract the average of the maximum and minimum envelopes from the original signal yielding the so-called “detail” signal. These steps are then repeated until the “detail” satisfies two IMF criteria.⁵³ When this is the case, the “detail” becomes IMF_1 , and is subtracted from the original signal. This process is repeated until all IMFs are extracted and only a monotonic residue or trend remains. The multivariate extension of EMD (MEMD)⁴⁰ aligns similar frequency bands of multiple channels thus providing an assessment of their possible interdependence (mode alignment property). The MEMD algorithm operates in the same way as the EMD algorithm, but as the average of the maximum and minimum envelopes cannot be defined for multivariate signals directly, MEMD estimates the average from projections along different p -dimensional spaces, with p the number of channels (dimensions).

2.2.2. Sample entropy (SE)

Sample Entropy (SE)³⁹ is the conditional probability that two sequences that are close to each other for m consecutive data points up to a tolerance level r remain so when one more data point to each sequence is added. Formally, SE is expressed as follows:

$$\text{SE}(r, m, N) = -\log \frac{A^m(r)}{B^m(r)}$$

with $B^m(r)$ the probability that two sequences of length m match the given tolerance level r , $A^m(r)$ the probability that two sequences of length $m + 1$ match given r , and N the total length of the data from which the sequences are taken. The tolerance level is usually a percentage of the standard deviation of the normalized data; for our case we took 15%.⁴⁰ In order to estimate sample entropy in the multivariate case (MSampEn), the sequences are formulated as follows. Recalling multivariate embedding theory,⁵⁴ for p -variate time series $\{x_{k,i}\}_{i=1}^N, k = 1, 2, \dots, p$, the multivariate embedded sequence is the composite

delay vector:

$$X_m(i) = [x_{1,i}, x_{1,i+\tau_1}, \dots, x_{1,i+(m_1-1)\tau_1}, x_{2,i}, \\ x_{2,i+\tau_2}, \dots, x_{2,i+(m_2-1)\tau_2}, x_{p,i}, \\ x_{p,i+\tau_p}, \dots, x_{p,i+(m_p-1)\tau_p}]$$

with $X_m(i) \in R^p$, $m = \sum_{k=1}^m m_k$,

$M = [m_1, m_2, \dots, m_p] \in R^p$ the embedding vector, and $\tau = [\tau_1, \tau_2, \dots, \tau_p]$ the time lag vector. The above sample entropy definition is then applied to the composite delay vectors.

2.2.3. MEMD-enhanced MMSE

In order to arrive at a multiscale version of multivariate sample entropy (MMSE), instead of relying on the popular coarse grained procedure,³⁸ which is known to favor low-frequency components, we first apply MEMD to compute our multiple scales. For the purpose of this paper, the MSE software of Ahmed and Mandic was extended towards MEMD-enhanced MMSE and applied to the EEG recordings of each individual participant, for each video clip. The MEMD-enhanced MMSE algorithm can be summarized as follows:

(i) obtain n IMFs over each subject's entire recording length of each clip (approx. 3 min on average) by applying the MEMD method given $N = 2$ EEG channels,

(ii) accumulate the n IMFs (cumulative IMF, CIMF) one by one starting with the first one, which captures the highest frequencies, $\text{CIMF}_1 = \text{IMF}_1$, and then adding the second IMF to the former, $\text{CIMF}_2 = \text{IMF}_1 + \text{IMF}_2$, and the third IMF, $\text{CIMF}_3 = \text{IMF}_1 + \text{IMF}_2 + \text{IMF}_3$, and so on, until all IMFs are added and the original signal is reconstructed,

(iii) for each CIMF, calculate the MMSE for 1000 non-overlapping 100 ms snippets taken from a prior defined EEG signal track and take the average as the final MMSE estimate of that track,

(iv) plot the MMSE estimates as a function of $\text{CIMF}_1, \text{CIMF}_2, \dots, \text{CIMF}_n$ to obtain the MEMD-enhanced MMSE of the EEG recording track under consideration. Note that the algorithm selects by itself the number of IMFs, however, as this is subject-dependent (in our case, the range is 15 to 17 IMFs), we decided to take $n = 15$ IMFs for all subjects.

2.2.4. Statistical analysis

For our statistical modeling, we use linear mixed models.^{55,56} We first need a statistical model for our MMSE curves. We adopt a forward model selection approach. Starting with a linear model and, after an additional use of explorative graphs of the data, the model was extended to a quadratic one. Hence, two regression functions became competitive:

Model_1 (linear model) For example, the MMSE of standard_self=amusement_amusement (am_am) can be modeled as follows:

$$\text{MMSE}_{\text{am_am}} = a_{\text{am_am}} + b_{\text{am_am}} \text{CIMF}_i.$$

Model_2 (quadratic model) For example, the MMSE of "amusement_amusement" (am_am) can be modeled as follows:

$$\text{MMSE}_{\text{am_am}} = a'_{\text{am_am}} \\ + b'_{\text{am_am}} \text{CIMF}_i + b''_{\text{am_am}} \text{CIMF}_i^2.$$

The covariance structure in the model was taken as general unstructured. Using a Likelihood Ratio test (null-hypothesis: there is no difference between the models and the term with the second degree is nonsignificant), we compared the significance of the quadratic versus the linear model. The residuals were satisfactory (normally distributed) for both models (assessed with Q-Q plots, not shown). In addition, a "deleted observation" technique was implemented in order to show that there are no influential observations present in the data. Model 1 had a $-2\log\text{likelihood}$ of -8656.5 , Model 2 $-10,481.4$. The difference is more than 1000 (the threshold value is 4 for the chi-square statistic with 1 degree of freedom) which signifies that Model 2 provides a much better fit than Model 1. Hence, we continue with Model 2. Each statistical test to decide whether pairs of curves are distinguishable is split into two parts: (a) We test if the *shapes* of the modeled curves are the same. For example, for "amusement_amusement" (am_am) and "anger_anger", (an_an) we test the following null-hypothesis: $\mathbf{H}_0: b'_{\text{am_am}} = b'_{\text{an_an}}$ and $b''_{\text{am_am}} = b''_{\text{an_an}}$. (b) We test if the shifts of the curves are the same. Continuing with the same example, we test for the following null-hypothesis $\mathbf{H}_0: a_{\text{am_am}} = a_{\text{an_an}}$. For parts (a) and (b), we apply a chi-square test using the "contrast" statement in "proc mixed" (SAS 9.4). We choose $p = 0.05$ as our

overall significance level and apply Bonferroni correction ($p = 0.05/2$).

3. Results

3.1. Evolution of signal complexity

We computed the MEMD-enhanced MMSE of a typical participant's EEG recordings over the entire length of each video clip and also over the first, middle, and last 100s (Fig. 1). One immediately observes that the complexity curves of the last 100s of the video clips are much better distinguishable than those of the entire length and the first and middle, 100s. This is probably not surprising as the video clips have their climax at the end. Note also that the

overall complexity level seems to settle towards the end of each video clip. All this motivates us to select the EEG recordings of the last 100s for further analysis.

3.2. Signal complexity of video clips

One could argue that the observed differences in EEG data complexity are due to the streamed video clips rather than the emotions they are supposed to elicit. As the video clips are in an entirely different format as the EEG recordings, we computed the MEMD-enhanced MMSE for electrodes O1 and O2, on the occipital pole, which are supposed to be more related to the visual input (Fig. 2, right

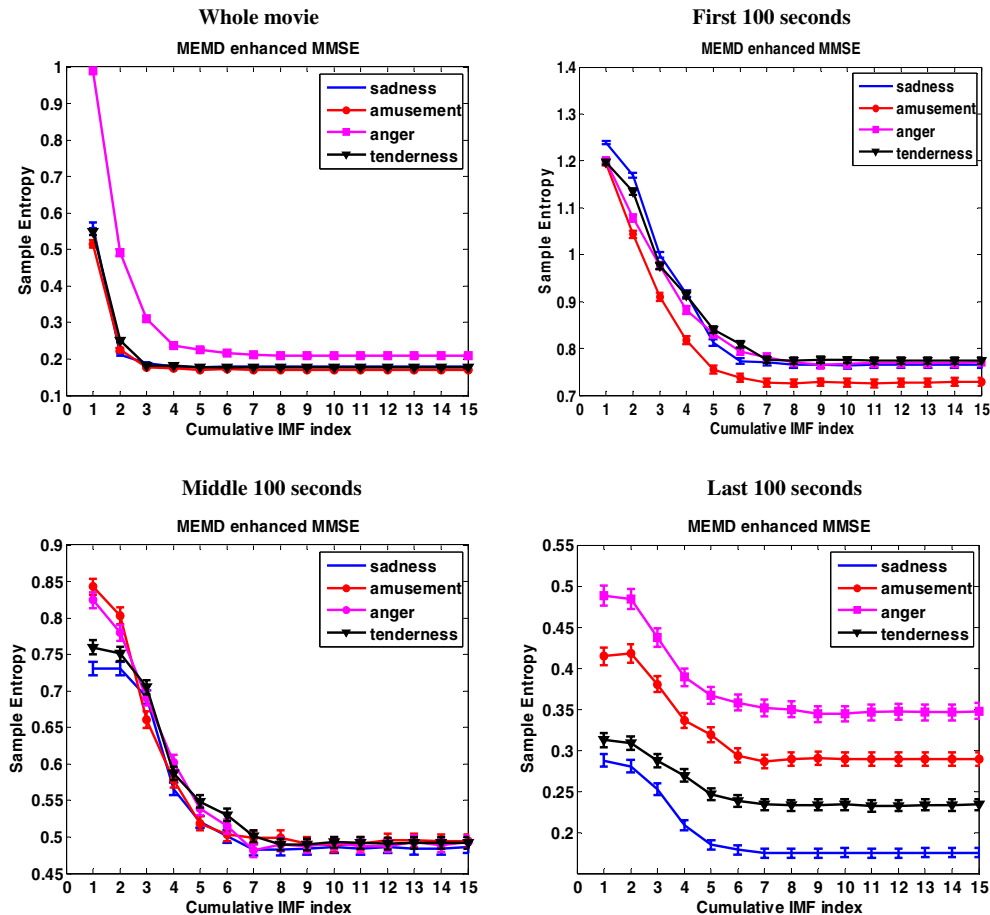


Fig. 1. MEMD-enhanced MMSE curves of participant M. J. when computed over the entire length of each movie, and over the first, middle and last 100s of the video clips. Curves are averages over 1000 segments used for MMSE estimation and labeled according to the participant's self-reported emotional category (self-label, one for each video clip); error bars correspond to standard deviations.

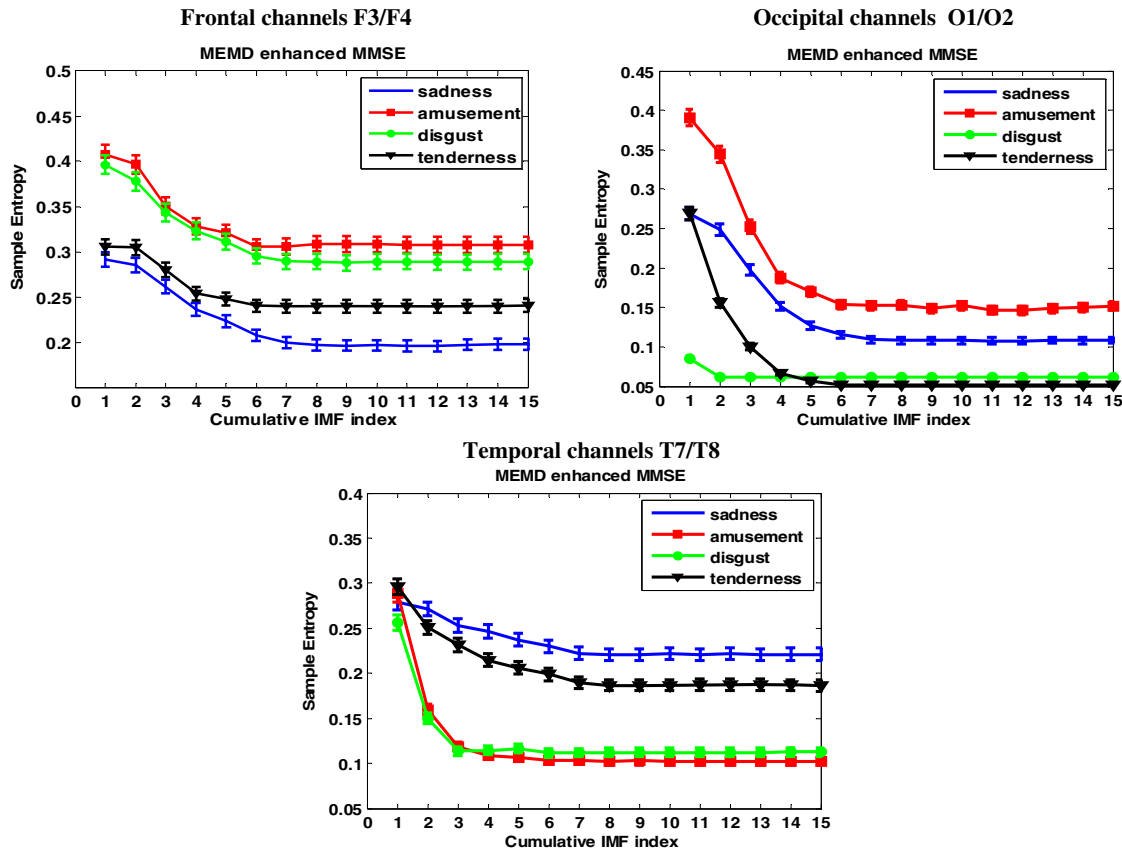


Fig. 2. MEMD-enhanced MMSE curves of participant D. V. calculated for the last 100s of each video clip's EEG recordings of the frontal (left panel), occipital (right panel) and temporal electrode pairs (bottom). Other conventions are as in Fig. 1.

panel). As can be seen from Fig. 2 (left versus right panel), the complexity curves of the occipital channels are ordered differently and less distinguishable. This proves that the frontal channels' complexity curves are not explained by those of the occipital channels. When looking at the curves for T7/T8 (bottom panel) we observe that they are less discernible compared to their F3/F4 counterparts and that the ordering of the curves is reversed but this is a coincidence as, unlike the curves for F3/F4, the ones for T7/T8 (and also O1/O2) are much more subject-dependent (not shown).

The outcome of both comparisons indicates the relevance of mid-frontal electrodes F3/F4 complexity curves in emotional state discrimination.

3.3. Signal complexity versus self-reported emotions

The results of self-labeling are summarized in Table 2. We observe that only in the case of "anger"

(first video) the reports are not univocal as 16 participants reported disgust instead (the video clip was about incest). A similar discrepancy was reported for this particular video in the Supplementary material of Ref. 46: the extent to which their participants felt disgust received a slightly higher rating than anger. The resulting complexity curves for all 30 participants are shown in Fig. 3. Using Model 2, as explained above, with MMSE as dependent variable, *standard_self* and CIMF index as independent variables, and gender and age as covariates, we verified that all complexity curves differ significantly ($p < 0.05$). This shows that with MMSE several emotional states can be discriminated from mid-frontal electrode recordings.

3.4. Arousal scores

Schaefer and co-workers⁴⁶ collected emotional arousal scores from their participants. The question arises whether the video clips we considered can be

Table 2. Emotions reported by our $N = 30$ participants (“Self-label”) with numbers between brackets indicating how many participants reported that emotion. Other conventions as in Table 1.

Video clip	Standard label	Self-label
“Sleepers”	Anger	Anger (14)/ Disgust(16)
“Life is beautiful (4)”	Tenderness	Tenderness (30)
“City of angels”	Sadness	Sadness (30)
“La cité de la peur”	Amusement	Amusement (30)

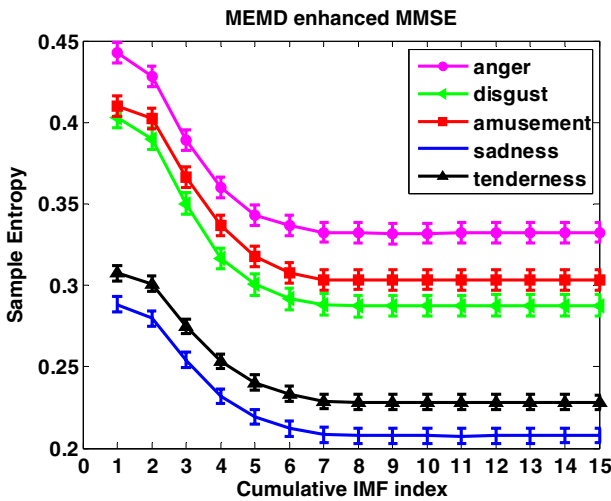


Fig. 3. MEMD-enhanced MMSE curves labeled in terms of our participants’ self-reported emotions.

distinguished from these arousal scores (Appendix A, summarized in Table 1). As this analysis was not reported in Ref. 46, we performed an ANOVA analysis on the arousal scores provided to us by Alexandre Schaefer with “arousal” as dependent variable and “video clip” as independent variable. The analysis was performed with the GLM procedure in SAS. We found that only “La cité de la peur” could be distinguished from “Life is beautiful (4)” ($p = 0.0018$) and also from “Sleepers” ($p = 0.0007$). The other video clips were indistinguishable from each other. Hence, based on arousal scores the four video clips cannot be discerned.

3.5. Complexity versus arousal

Our complexity results can also be compared with Schaefer *et al.*’s arousal scores. As the latter refer to the entire video clips,⁴⁶ we computed the MMSE curves accordingly. We performed a linear mixed

model regression analysis, with MMSE as dependent variable, “arousal” as fixed effect and “subject” as random effect to correct for within subject association. To render the comparison manageable, we considered MMSE for CIMF index 10 only. As our subjects differ from those of Schaefer *et al.*’s (and in their case even between video clips), we adopt a multiple imputations procedure. We simulate arousal scores for each one of our 30 participants and for each video clip by drawing samples from the distribution of Schaefer *et al.*’s arousal scores. We simulate for each video clip 10 arousal scores per participant (in total $4 \times 10 \times 30$ imputations) supposing that the missing mechanism is operating at random. The number of imputations can be lower, but recent publications pointed out that small numbers of imputations lead to results that are affected by Monte Carlo error.^{57,58} After multiple imputation, data sets were analyzed separately using generalized estimating equation (GEE) models, which were applied by using the GENMOD procedure in SAS, and finally the results of the separate analyses combined into a single one using the MIANALYZE procedure. We found no statistically significant association between MMSE and arousal score. For completeness sake, the analysis was redone now using our MMSE results of the last 100s of each video clip (cf., Fig. 3). Also here we found no statistically significant association between MMSE and arousal score. In conclusion, based on our data set, there does not seem to be a clear relation between signal complexity, as assessed with MMSE, and self-reported arousal scores.

3.6. Power spectrum analysis

Finally, given the popularity of the spectral approach to emotion detection with EEG, we also computed the power spectral density (PSD) in the following frequency bands: theta (4–7 Hz), alpha (8–12 Hz), beta (13–24 Hz), gamma (25–40 Hz) (more precisely the area under the PSD curve in each band). We did this for each subject and for the last 100s part of each movie. The question is now: can we separate emotions based on one or more of those bands? And, is there any asymmetry between electrodes F3 and F4? We start with the PSDs of both channels (by concatenating the corresponding F3, F4 recordings) and apply unstructured linear mixed models^{55,56} with the

following variables as a possible response: PSDs of the theta, alpha, beta, gamma bands, and the variable “self-reported emotion” (self-label case) as an independent group variable. As four emotions were reported by each subject, “subject” was taken as random effect. The responses were transformed to a logarithmic scale to reach normality of the residuals. We choose $p = 0.05$ as our overall significance level. For multiple comparisons, Bonferroni correction was used. We found no significant difference between any pair of self-reported emotions. Next, we verified whether there is any asymmetry between the PSDs of the F3 and F4 channels that can be used for separating the emotions. We again apply unstructured linear mixed models with the same possible responses as independent variables, but now with two group variables, “self-reported emotion” and “channel”. In the model, the latter two were taken with an interaction. Since several measurements were done within the same subject, “subject” was taken as random effect. The responses were again transformed to a logarithmic scale to reach normality of the residuals. This meant that for the gamma band, based on the Q-Q plot, five outliers needed to be removed. Again Bonferroni correction was applied. We found no significant differences between the two channels for any of the emotions, except for the gamma band and across emotion labels (albeit marginally: $p = 0.037$). In summary, we found no difference in frequency bands and no asymmetry between the F3 and F4 channels from which emotions (or video clips) could be discerned.

4. Discussion

4.1. Spectral analysis

EEG spectral analysis studies designed to unravel neural correlates of discrete emotions tend to focus on the so-called “frontal asymmetry”, usually in alpha power between left and right frontal regions.⁵⁹ We observed in our study an asymmetry in gamma band activity between electrode F3 and F4, corresponding to the mid-frontal region of the scalp ($p = 0.037$, Bonferroni corrected), but it was not specific for any of the video clips or their self-reported emotions. Contrary to reports on the importance of alpha,^{7,8} gamma,^{60,61} and theta band activity,¹⁴ and differences between these bands,¹³ we could not replicate any of those in our study. We observed a

marginally nonsignificant difference in alpha band activity between the anger and sadness video clips ($p = 0.0833$, Bonferroni corrected) and in beta band activity between electrodes F3 and F4 ($p = 0.0546$, Bonferroni corrected).

4.2. Complexity analysis

Bekkedal and co-workers¹⁵ attribute the failure to clearly demonstrate distinct EEG response patterns for individual emotions to the universal use of visual stimuli (pictures, movies) that tend to carry information irrelevant to their emotionality and whence suggest auditory stimuli instead. However, it still could be that further strides can be made in EEG processing and in the detection of EEG patterns that can be linked to distinct emotional states. In the pertinent paper, we showed that multiscale, multivariate complexity, as assessed with MEMD-enhanced MMSE, is a promising technique for discriminating multiple emotional states from EEG recordings in response to emotion-inducing video clips (Fig. 3). Furthermore, unlike previous studies^{62,63} we used a single-trial design to show our assertion: the emotional state was discriminated even from a single video clip. Jie and co-workers³⁷ performed emotion recognition from univariate, uniscale sample entropy analysis of prefrontal EEG recordings. They considered two emotion recognition tasks and trained classifiers to distinguish positive from negative emotions, both with high arousal, and music clips but with different arousal levels. The EEG recordings of subjects whose emotion self-reports were not in line with the pre-defined ones were excluded from their analysis. Similarly we also had self-reports that were not in line with the standard labels of the video clips used (disgust instead of anger, Table 2). However, contrary to Jie and co-workers, we considered those self-labels and corresponding EEG recordings in our analysis. As a result of this, instead of four emotions, one for each video clip, we could discriminate five emotions (anger, disgust, sadness, tenderness, amusement). This number is also in contrast with previous work where only two or three emotional states could be discriminated.^{2,14,35} Hosseini and Naghibi-Sistani³⁶ stipulate that during an emotional state evoked by pictures, the EEG signal is less complex (assessed with approximate- and wavelet entropy) compared to the normal state. Using video

clips instead of pictures, we observed that complexity decreases and seems to settle towards the end of our recordings (Fig. 1, note that MMSE scales logarithmically). As the amusement inducing video clip “La cité de la peur” has its punchline (i.e. the funny part) only during the very last scene, we split the last 100s into two 50s intervals and repeated the extended MEMD-enhanced MMSE analysis for all video clips (see Appendix B). For all video clips, except “La cité de la peur”, there was no difference. For that clip a significant decrease in complexity was observed for the last 50s (t -test on squared distance between the complexity curves, $p = 0.014$). One might therefore be tempted to perform complexity analysis in a scene-based manner but bearing mind that the overall complexity level also decreases as the subject is viewing the video clip.

4.3. Arousal versus complexity

We also examined the relation between EEG complexity and emotional arousal. The connection was not significant (using a linear mixed model regression analysis) neither for the whole video clip (as the self-reports were collected in this way.⁴⁶) nor for the last 100s (targeted in our EEG complexity analysis). Visual inspection of Fig. 3 reveals that the ordering of the complexity curves labeled anger, tenderness, sadness, and disgust corresponds to that of the mean arousal levels of the video clips (Table 1). The only exception is for the amusement curve. The self-reported arousal level of the amusement inducing video clip “La cité de la peur” is the lowest of all video clips whereas its complexity level is one of the highest. This discrepancy could be due to the fact that the punchline is in the very last scene. As suggested in the previous paragraph, one could restrict the complexity analysis of this video clip to the last 50s. This time the ordering of the complexity curves is completely in alignment with the ordering of Schaefer’s mean arousal levels. On the other hand, it should be noted that our statistical analysis of Schaefer’s arousal levels indicated that only “La cité de la peur” could be distinguished from two other video clips (“Life is beautiful (4)”, “Sleepers”); the other video clips could not be distinguished. Perhaps the collection of the arousal scores should be improved, e.g. by letting participants press a button when an individual scene raises their emotional

arousal.⁶⁴ Another argument in favor of a connection between complexity and emotional arousal is the work of Hepach and co-workers.⁶⁵ They asked 100 healthy individuals to indicate the valence and arousal levels of 62 emotion words in a questionnaire. They showed that the arousal level of anger was higher than that of disgust (8.06 versus 6.87). It can be observed from our results in Fig. 3 that the complexity level of anger is *significantly* higher than that of disgust. In summary, there seems to be a relation between complexity and emotional arousal, but more research is needed to be conclusive. If successful, it would confirm Cannon-Bard’s theory⁶⁶ that physiological arousal and feelings occur together.

5. Conclusion

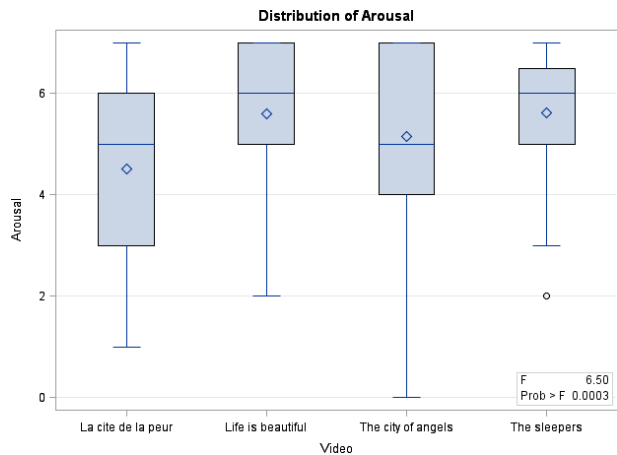
We have shown that the MEMD-enhanced MMSE approach to EEG signal complexity was successful in discriminating five emotional states from mid-frontal recordings in response to four emotion-inducing video clips. When applied to temporal electrode recordings the results turned out to be subject dependent. We also observed that, when viewing the video clip, the complexity level of the mid-frontal recordings decreases and then levels off.

Acknowledgments

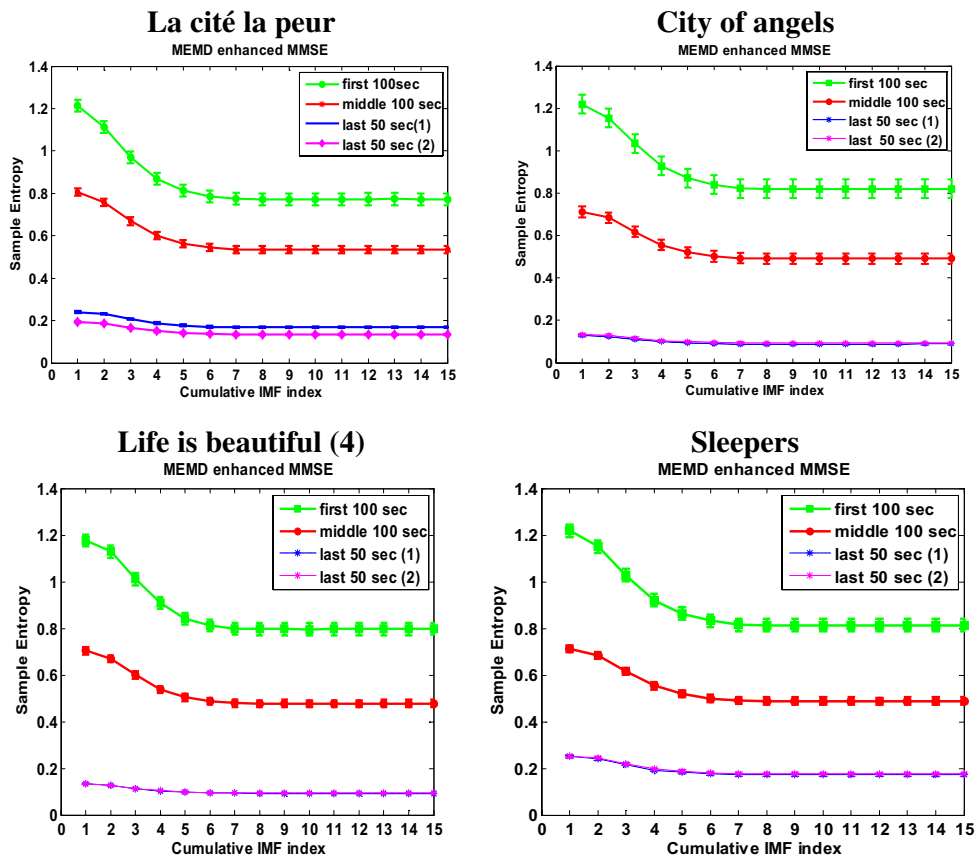
The authors are grateful to Anna Ivanova (L-BioStat, KU Leuven) for her expert advice in the statistical analysis and Prof. Alexandre Schaefer (School of Business, Monash University, Victoria, Malaysia) for his valuable input and for sharing his arousal data with us. YT is supported by the Interuniversity Attraction Poles Programme — Belgian Science Policy (IUAP P7/11). MMVH is supported by research grants received from the program Financing program (PFV/10/008) of the KU Leuven, the Belgian Fund for Scientific Research — Flanders (G088314N, G0A0914N), the Interuniversity Attraction Poles Programme — Belgian Science Policy (IUAP P7/11), the Flemish Regional Ministry of Education (Belgium) (GOA 10/019), the Hercules Foundation (AKUL 043), the Flemish Agency for Innovation by Science and Technology (TETRA project Spellbinder), and by the SWIFT prize of the King Baudouin Foundation of Belgium. Part of the reported work was done while MVH was a Visiting Academic at Imperial College London.

Appendix A

Distribution of arousal scores of 4 video clips used



Appendix B



References

1. E. R. John, L. S. Prichep, J. Fridman and P. Easton, Neurometrics: Computer-assisted differential diagnosis of brain dysfunctions, *Science* **239**(4836) (1988) 162–169, doi:10.1126/science.3336779.
2. A. Choppin, EEG-based human interface for disabled individuals: Emotion expression with neural networks, *Unpubl. Master's Thesis* (2000).
3. J. D. R. Millán, R. Rupp, G. R. Müller-Putz *et al.*, Combining brain–computer interfaces and assistive technologies: State-of-the-art and challenges, *Front. Neurosci.* **4** (2010) 161.
4. A. Nijholt, BCI for games: A “state of the art” survey, *Entertainment Computing-ICEC 2008* (Springer, 2009), pp. 225–228.
5. P. L. Nunez and B. A. Cuttillo, *Neocortical Dynamics and Human EEG Rhythms* (Oxford University Press, USA, 1995).
6. R. J. Davidson, Anterior cerebral asymmetry and the nature of emotion, *Brain Cogn.* **20**(1) (1992) 125–151, doi:10.1016/0278-2626(92)90065-T.
7. M. B. Kostyunina and M. A. Kulikov, Frequency characteristics of EEG spectra in the emotions, *Neurosci. Behav. Physiol.* **26**(4) (1996) 340–343, doi:10.1007/bf02359037.
8. N. V. Shemiakina and S. G. Dan'ko, Influence of the emotional perception of a signal on the electroencephalographic correlates of the creative activity, *Fiziol. Cheloveka* **30**(2) (2004) 22–29.
9. M. M. Bradley and P. J. Lang, The International Affective Picture System (IAPS) in the study of emotion and attention, *Handb. Emot. Elicitation Assess.* (Cambridge University Press, 2007), pp. 29–46.
10. F. Murphy, I. Nimmo-Smith and A. Lawrence, Functional neuroanatomy of emotions: A meta-analysis, *Cogn. Affect. Behav. Neurosci.* **3**(3) (2003) 207–233, doi:10.3758/CABN.3.3.207.
11. F. Schiffer, M. H. Teicher, C. Anderson *et al.*, Determination of hemispheric emotional valence in individual subjects: A new approach with research and therapeutic implications, *Behav. Brain Funct.* **3**(1) (2007) 13.
12. N. Jaušovec, K. Jaušovec and I. Gerlič, Differences in event-related and induced EEG patterns in the theta and alpha frequency bands related to human emotional intelligence, *Neurosci. Lett.* **311**(2) (2001) 93–96.
13. C. M. Krause, V. Viemerö, A. Rosenqvist, L. Silanmäki and T. Åström, Relative electroencephalographic desynchronization and synchronization in humans to emotional film content: An analysis of the 4–6, 6–8, 8–10 and 10–12 Hz frequency bands, *Neurosci. Lett.* **286**(1) (2000) 9–12.
14. G. Vecchiato, L. Astolfi, F. D. V. Fallani *et al.*, Changes in brain activity during the observation of TV commercials by using EEG, GSR and HR measurements, *Brain Topogr.* **23**(2) (2010) 165–179, doi:10.1007/s10548-009-0127-0.
15. V. Bekkedal MY, J. Rossi and J. Panksepp, Human brain EEG indices of emotions: Delineating responses to affective vocalizations by measuring frontal theta event-related synchronization, *Neurosci. Biobehav. Rev.* **35**(9) (2011) 1959–1970.
16. L. I. Aftanas, A. A. Varlamov, S. V. Pavlov, V. P. Makhnev and N. V. Reva, Affective picture processing: Event-related synchronization within individually defined human theta band is modulated by valence dimension, *Neurosci. Lett.* **303**(2) (2001) 115–118.
17. M. Balconi and U. Pozzoli, Arousal effect on emotional face comprehension: Frequency band changes in different time intervals, *Physiol. Behav.* **97**(3) (2009) 455–462.
18. J. Bhattacharya and H. Petsche, Universality in the brain while listening to music, *Proc. R. Soc. London B Biol. Sci.* **268**(1484) (2001) 2423–2433.
19. J. Bhattacharya, Increase of universality in human brain during mental imagery from visual perception, *PLoS One* **4**(1) (2009) e4121, doi:10.1371/journal.pone.0004121.
20. S. J. Luck, *An Introduction to the Event-Related Potential Technique* (MIT Press, 2014).
21. S. Paulmann and S. A. Kotz, An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context, *Brain Lang.* **105**(1) (2008) 59–69.
22. K. N. Spreckelmeyer, M. Kutas, T. Urbach, E. Altenmüller and T. F. Münte, Neural processing of vocal emotion and identity, *Brain Cogn.* **69**(1) (2009) 121–126.
23. A. L. Goldberger, C.-K. Peng and L. A. Lipsitz, What is physiologic complexity and how does it change with aging and disease? *Neurobiol. Aging* **23**(1) (2002) 23–26, doi:10.1016/S0197-4580(01)00266-4.
24. M. Bandarabadi, J. Rasekhi, C. A. Teixeira, T. I. Netoff, K. K. Parhi and A. Douardo, Early seizure detection using neuronal potential similarity: A generalized low-complexity and robust measure, *Int. J. Neural Syst.* **25**(5) (2015) 1550019.
25. Y. Zhang, W. Zhou and S. Yuan, Multifractal analysis and relevance vector machine based automatic seizure detection in intracranial EEG, *Int. J. Neural Syst.* **25**(6) (2015) 1550020.
26. S. Aydin, N. Arica, E. Ergul and O. Tan, Classification of obsessive compulsive disorder by EEG complexity and hemispheric dependency measurements, *Int. J. Neural Syst.* **25**(3) (2015) 1550010.
27. M. Ahmadi, H. Adeli and A. Adeli, Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder, *J. Clin. Neurophysiol.* **27**(5) (2010) 328–333.

28. M. Ahmadlou, H. Adeli and A. Adeli, Fractality analysis of frontal brain in major depressive disorder, *Int. J. Psychophysiol.* **85**(2) (2012) 206–211.
29. M. Ahmadlou, H. Adeli and A. Adeli, New diagnostic EEG markers of the Alzheimer's disease using visibility graph, *J. Neural Transm.* **117**(9) (2010) 1099–1109.
30. M. Ahmadlou and H. Adeli, Visibility graph similarity: A new measure of generalized synchronization in coupled dynamic systems, *Phys. D Nonlinear Phenom.* **241**(4) (2012) 326–332.
31. M. Ahmadlou, A. Adeli, R. Bajo and H. Adeli, Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task, *Clin. Neurophysiol.* **125**(4) (2014) 694–702.
32. M. Ahmadlou and H. Adeli, Fuzzy synchronization likelihood with application to attention-deficit/hyperactivity disorder, *Clin. EEG Neurosci.* **42**(1) (2011) 6–13.
33. M. Ahmadlou, H. Adeli and A. Adeli, Fuzzy synchronization likelihood-wavelet methodology for diagnosis of autism spectrum disorder, *J. Neurosci. Methods* **211**(2) (2012) 203–209.
34. M. Ahmadlou, H. Adeli and A. Adeli, Graph theoretical analysis of organization of functional brain networks in ADHD, *Clin. EEG Neurosci.* **43**(1) (2012) 5–13.
35. L. I. Aftanas, N. V. Lotova, V. I. Koshkarov, V. P. Makhnev, Y. N. Mordvintsev and S. A. Popov, Non-linear dynamic complexity of the human EEG during evoked emotions, *Int. J. Psychophysiol.* **28**(1) (1998) 63–76, doi:10.1016/S01678760(97)00067-6.
36. S. A. Hosseini and M. B. Naghibi-Sistani, Emotion recognition method using entropy analysis of EEG signals, *Int. J. Image, Graph. Signal Process.* **3**(5) (2011) 30, doi:10.5815/ijgsp.2011.05.05.
37. X. Jie, R. Cao and L. Li, Emotion recognition based on the sample entropy of EEG, *Biomed. Mater. Eng.* **24**(1) (2013) 1185–1192, doi:10.3233/BME-130919.
38. M. Costa, A. L. Goldberger and C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, *Phys. Rev. Lett.* **89**(6) (2002) 068102, doi:10.1103/PhysRevLett.92.089803.
39. J. S. Richman and J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am. J. Physiol. Heart Circ. Physiol.* **278**(6) (2000) H2039–49.
40. N. Rehman and D. P. Mandic, Multivariate empirical mode decomposition, *Proc. R. Soc. A Math. Phys. Eng. Sci.* **466**(2117) (2009) 1291–1302, doi:10.1098/rspa.2009.0502.
41. R. J. Martis, U. R. Acharya, J. H. Tan et al., Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction, *Int. J. Neural Syst.* **23**(5) (2013) 1350023.
42. R. J. Martis, U. R. Acharya, J. H. Tan et al., Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals, *Int. J. Neural Syst.* **22**(6) (2012) 1250027.
43. R. Sharma, R. B. Pachori and U. R. Acharya, Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals, *Entropy* **17**(2) (2015) 669–691.
44. M. U. Ahmed and D. P. Mandic, Multivariate multi-scale entropy: A tool for complexity analysis of multichannel data, *Phys. Rev. E* **84**(6) (2011) 061918, doi:10.1103/PhysRevE.84.061918.
45. Y. Tonoyan, D. Looney, D. P. Mandic and M. M. Van Hulle, Predicting subject performance level from EEG signal complexity when engaged in BCI paradigm, *2014 IEEE Int. Work. Mach. Learn. Signal Process.* (2014), pp. 1–5, doi:10.1109/MLSP.2014.6958897.
46. A. Schaefer, F. Nils, X. Sanchez and P. Philippot, Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers, *Cogn. Emot.* **24**(7) (2010) 1153–1172, doi:10.1080/02699930903274322.
47. A. L. Goldberger, C.-K. Peng and L. A. Lipsitz, What is physiologic complexity and how does it change with aging and disease? *Neurobiol. Aging* **23**(1) (2002) 23–26.
48. R. J. Croft, J. S. Chandler, R. J. Barry, N. R. Cooper and A. R. Clarke, EOG correction: A comparison of four methods, *Psychophysiology* **42**(1) (2005) 16–24, doi:10.1111/j.1468-8986.2005.00264.x.
49. A. C. Metting van Rijn, A. Peper and C. A. Grimbergen, High-quality recording of bioelectric events, *Med. Biol. Eng. Comput.* **29**(4) (1991) 433–440.
50. K. N. Ochsner and J. J. Gross, The cognitive control of emotion, *Trends Cogn. Sci.* **9**(5) (2005) 242–249.
51. W. Heller, Neuropsychological mechanisms of individual differences in emotion, personality, and arousal, *Neuropsychology* **7**(4) (1993) 476.
52. H. J. Crawford, S. W. Clarke and M. Kitner-Triolo, Self-generated happy and sad emotions in low and highly hypnotizable persons during waking and hypnosis: Laterality and regional EEG activity differences, *Int. J. Psychophysiol.* **24**(3) (1996) 239–266.
53. N. E. Huang, Z. Shen, S. R. Long et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. A Math. Phys. Eng. Sci.* **454**(1971) (1998) 903–995, doi:10.1098/rspa.1998.0193.
54. T. Gautama, D. P. Mandic and M. M. Van Hulle, A differential entropy based method for determining the optimal embedding parameters of a signal, in *Acoustics, Speech, and Signal Processing, 2003. Proc. (ICASSP'03). 2003 IEEE Int. Conf. Vol. 6* (IEEE, 2003), p. I-29.

55. N. M. Laird and J. H. Ware, Random-effects models for longitudinal data, *Biometrics* **38** (1982) 963–974.
56. G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data* (Springer Science & Business Media, 2009).
57. I. R. White, simsum: Analyses of simulation studies including Monte Carlo error, *Stata J.* **10**(3) (2010) 369.
58. S. Van Buuren, *Flexible Imputation of Missing Data* (CRC Press, 2012).
59. R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis and W. V. Friesen, Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I, *J. Pers. Soc. Psychol.* **58**(2) (1990) 330.
60. M. Li and B.-L. Lu, Emotion classification based on gamma-band EEG, *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual Int. Conf. IEEE* (IEEE, 2009), pp. 1223–1226.
61. M. Balconi and C. Lucchiari, Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis, *Int. J. Psychophysiol.* **67**(1) (2008) 41–46.
62. M. Soleymani, S. Koelstra, I. Patras and T. Pun, Continuous emotion detection in response to music videos, *Face Gesture 2011* (2011), pp. 803–808, doi:10.1109/FG.2011.5771352.
63. M. Soleymani, M. Pantic and T. Pun, Multimodal emotion recognition in response to videos, Available at: http://ibug.doc.ic.ac.uk/media/uploads/documents/tac_multimodal_emotion_recognition.pdf. [Accessed on 18 March, 2015].
64. A. Angrilli, P. Cherubini, A. Pavese, S. Manfredini and S. Mantredini, The influence of affective factors on time perception, *Percept. Psychophys.* **59**(6) (1997) 972–982, doi:10.3758/BF03205512.
65. R. Hepach, D. Kliemann, S. Grüneisen, H. R. Heekeren and I. Dziobek, Conceptualizing emotions along the dimensions of valence, arousal, and communicative frequency—implications for social-cognitive tests and training tools, *Front. Psychol.* **2** (2011), doi:10.3389/fpsyg.2011.00266.
66. W. B. Cannon, The James-Lange theory of emotions: A critical examination and an alternative theory, *Am. J. Psychol.* **39**(1/4) (1927) 106–124.