identifying one or more features from each feature set for differentiating attention/non-attention signals;

determining respective classification scores for each feature set based on the identified one or more features; and

combining the classification scores to obtain a combined attention score for said detecting attention.

12. A method of detecting attention of a subject, the method comprising,

determining a plurality of feature sets from Electroencephalographic (EEG) signals of the subject, wherein the plurality of feature sets comprise pool data feature sets based on EEG data from a pool of subjects;

identifying one or more features from each feature set for differentiating attention/non-attention signals;

determining respective classification scores for each feature set based on the identified one or more features;

combining the classification scores to obtain a combined attention score for said detecting attention; and

wherein a first one of the feature sets is based on each feature of the feature set being an average band-power over a time-window of a differential potential between two EEG channels coupled to the subject and a second one of the feature sets is based on each feature of the feature set being an average band-power over a time-window of a differential potential between two EEG channels coupled to the subject after normalization to uni-variance.

13. The method as claimed in claim 12, wherein the feature sets comprise user specific feature sets associated to the subject.

~~14. The method as claimed in claims 12 or 13, wherein the feature sets comprise pool data feature sets based on EEG data from a pool of subjects.~~

1415. The method as claimed in any one of claims 12 to 1314, wherein the step of identifying one or more features from each feature set comprises maximising mutual

information I between brain conditions C (either attention or in-attentiveness) and EEG features A from the features sets using:

$$I(A,C) = H(C) - H(C|A) = \sum_{\omega} \int_{a} p_{a,\omega}(a,\omega) \log \left[\frac{\rho_{a,\omega}(a,\omega)}{\rho_{a}(a) P_{\omega}(\omega)} \right] da$$

where $H(C)$ is an attention condition variable ω, and $H(C|A)$ is a conditional entropy of an attention condition if the features are known, and *a* is a particular feature vector, and *P* (or *p*) denotes a probability (or probability density) function.

15~~16~~. The method as claimed in any one of claims 12 to 14~~15~~, wherein the step of combining the classification scores comprises using a Fishier linear discriminant (FLD) to combine the classification scores.

16~~17~~. A system for attention detection of a subject, the system comprising

a processing module, the processing module capable of

determination of a plurality of feature sets from Electroencephalographic (EEG) signals of the subject;

identification of one or more features from each feature set for differentiating attention/non-attention signals;

determination of respective classification scores for each feature set based on identified one or more features; and

combination of the classification scores to obtain a combined attention score for said attention detection; and

wherein a first one of the feature sets is based on each feature of the feature set being an average band-power over a time-window of a differential potential between two EEG channels coupled to the subject and a second one of the feature sets is based on each feature of the feature set being an average band-power over a time-window of a differential potential between two EEG channels coupled to the subject after normalization to uni-variance.

17~~18~~. The system as claimed in claim 16~~17~~, wherein the feature sets comprise user specific feature sets associated to the subject.

1819. The system as claimed in claims 1617 or 1718, wherein the feature sets comprise pool data feature sets based on EEG data from a pool of subjects.

1920. The system as claimed in any one of claims 1617 to 1819, wherein the processing module identifies one or more features from each feature set by maximisation of mutual information I between brain conditions C (either attention or in-attentiveness) and EEG features A from the features sets by utilisation of:

$$I(A,C) = H(C) - H(C|A) = \sum_{\omega} \int_{a} p_{a,\omega}(a,\omega) \log \left[\frac{\rho_{a,\omega}(a,\omega)}{\rho_{a}(a) P_{\omega}(\omega)} \right] da$$

where $H(C)$ is an attention condition variable ω, and $H(C|A)$ is a conditional entropy of an attention condition if the features are known, and *a* is a particular feature vector, and *P* (or *p*) denotes a probability (or probability density) function.

2021. The system as claimed in any one of claims 17 16 to 2019, wherein the processing module combines the classification scores by utilisation of a Fishier linear discriminant (FLD).

2122. A computer readable data storage medium having stored thereon computer code means for instructing a computer processor to execute a method of detecting attention of a subject, the method comprising,

determining a plurality of feature sets from Electroencephalographic (EEG) signals of the subject;

identifying one or more features from each feature set for differentiating attention/non-attention signals;

determining respective classification scores for each feature set based on the identified one or more features;

combining the classification scores to obtain a combined attention score for said detecting attention; and

wherein a first one of the feature sets is based on each feature of the feature set being an average band-power over a time-window of a differential potential between two EEG channels coupled to the subject and a second one of the feature sets is based on each feature of the feature set being an average band-power over a time-window of a

differential potential between two EEG channels coupled to the subject after normalization to uni-variance.

5