

RESEARCH ARTICLE

Attention emotion recognition via ECG signals

Aihua Mao^{1,*}, Zihui Du¹, Dayu Lu², Jie Luo^{3,*}

¹ School of Computer Science & Engineering, South China University of Technology, Guangzhou 510006, China

² School of Medicine, South China University of Technology, Guangzhou 510006, China

³ School of Fine Art and Artistic Design, Guangzhou University, Guangzhou 510006, China

* Correspondence: ahmao@scut.edu.cn; jieluo@gzhu.edu.cn

Received February 5, 2021; Revised April 3, 2021; Accepted April 24, 2021

Background: Physiological signal-based research has been a hot topic in affective computing. Previous works mainly focus on some strong, short-lived emotions (*e.g.*, joy, anger), while the attention, which is a weak and long-lasting emotion, receives less attraction. In this paper, we present a study of attention recognition based on electrocardiogram (ECG) signals, which contain a wealth of information related to emotions.

Methods: The ECG dataset is derived from 10 subjects and specialized for attention detection. To relieve the impact of noise of baseline wandering and power-line interference, we apply wavelet threshold denoising as preprocessing and extract rich features by pan-tompkins and wavelet decomposition algorithms. To improve the generalized ability, we tested the performance of a variety of combinations of different feature selection algorithms and classifiers.

Results: Experiments show that the combination of generic algorithm and random forest achieve the highest correct classification rate (CCR) of 86.3%.

Conclusion: This study indicates the feasibility and bright future of ECG-based attention research.

Keywords: affective computing; attention recognition; ECG signals

Author summary: Our work aims to discover the connection between ECG signals and attentive emotion, and proves the feasibility of applying ECG signal in attention recognition.

INTRODUCTION

Affective computing is a comprehensive research area, which involves knowledge from multi-disciplines such as medicine, biology and computer science. Since the concept of affective computing was firstly introduced by Picard [1], a large number of efforts have been made to construct emotion model to access the human's psychological perception. In 1971, Ekman and Friesen [2] found that there are six basic emotions (happiness, surprise, anger, fear, sadness, disgust) that can be generally recognized among people with different ages and cultural backgrounds. Ekman's model has been widely used to recognize emotions from facial expressions [3,4], and even from text material [5]. In 1980, Plutchik *et al.* [6] proposed a wheel model

including a total of 8 basic emotions: surprise, joy, trust, anticipation, sadness, fear, anger and disgust. More complex emotions, such as contempt, optimism, love, can be mixed from these basic emotions. In 1995, Lang [7] investigated that emotions can be categorized in a 2D space by valence and arousal. In this theory, valence ranges from unpleasant (negative) to pleasant (positive), and arousal ranges from passive (low) to active (high), which indicates how strongly human feels.

Currently, most of the affective computing studies target at clearly distinguishable emotions, such as anger, sadness, happiness and neutral. Many efforts have made in emotion analysis [8,9]. Especially, Liu *et al.* achieved a series of advancements in using multichannel EEG signals for emotion recognition [10–13]. However, some uncommon emotion, like approval, prohibition,

attention, antipathy, receives few attentions in research. As a micro-emotion, attention is substantial to human's observation, memory and imagination. In literature, attention is closely related to human's learning process [14] and would affect how people process visual information [15]. Some works have even found that attention has priority in controlling affective stimuli [16].

In order to analyze the attention process, the EEG signal has been mostly used in the previous works. Susan *et al.* [17] detected user's attention level by EEG in an auditory oddball task. Liu *et al.* [18] accessed whether students are attentively learning or not by observing their EEG signals. Hamadicharef *et al.* [19] proposed to improve accuracy of attention detection by learning spectral-spatial pattern from EEG. Alchalabi [20] used EEG signals to recognize ADHD patients' attention level to improve the effect of treatment. These experiments are conducted on specialized application (e.g., online learning, gameplay) and the best classification accuracy of their work ranges from 76% to 89%. Since the channel of electrodes chosen for classification does have great impact on accuracy, heavy experiment steps must be taken to find best placement of electrodes. Though EEG signal can classify emotions, it is limited to clinical environment because of its noise sensibility and time-consuming setting (require wearing head-mounted equipment).

Recently, the use of physiological signals such as ECG, EMG (electromyography), HF (heart rate) [21], has become a promising research topic in emotion recognition. The physiological signals are different from voice or posture which can be controlled subjectively. Their changes are controlled by nervous and endocrine system of human body and reflect by the emotional state more accurately. Though some of physiological signals like EMG, EDA (electrodermal activity) only measure in one dimension (valence or arousal), ECG allows measure independently in both valence and arousal dimensions and usually results in high accuracy in emotion classification [22]. ECG has been proved to be effective for a wide range of emotion detection. As one of the most common physiological signals, ECG signal reflects the potential difference produced by heart beating. There are also mature acquisition facilities in the industry for ECG signal like smart watch/clothes, which make ECG to be one of the most convenient physiological signals to access and to be a good emotion indicator.

In this paper, we apply ECG signals for attention recognition, which is still a novel area awaiting to explore. Based on the basic procedure of affective research, we propose an ECG-based attention recognition method, including data acquisition, denoising, feature extraction, and classification. Our contributions

can be summarized as follows:

(1) We propose a method of attention recognition based on the ECG signals, and shows that there is strong correlation between ECG signals and attentional emotion.

(2) We demonstrate signal processing workflow for the raw ECG signals, which removes a variety of noise interference by wavelet denoising and extracts rich features for attention detection.

(3) We evaluate the performance of different feature selection methods and classifiers, and find that the combination of GA and random forest obtains the best accuracy in attention recognition task.

EXPERIMENTAL RESULTS

Comparison between different feature dimensions

Different feature selection algorithms (FSA) may probably generate optimal features with different dimensions. In order to evaluate the impact of feature selection algorithms on the accuracy of attention recognition, we need to obtain a group of features which perform well between classifiers for each FSA.

The generic algorithm (GA) is a global optimum searching algorithm. To find the optimal dimensions for classification, we can repeatedly perform GA to search the last surviving individual (*i.e.*, the optimal feature sequence) and records the features dimension generated by GA. In our experiment, we set the amount of population to 50, the crossover probability to be 50% and the mutation probability to be 20%. The roulette mechanism is used to choose surviving individuals causing uncertainty to the generated features. Figure 1 shows the histogram of number of features generated by GA after repeating sufficient amount of times (7000 times in this case). It can be clearly seen that the number of generated features range in [23] share the highest proportion. We randomly choose one of the features group in this range to evaluate correct classification rate (CCR).

For the PCA and relief algorithms, we compare the performance of KNN classifier between different dimensions as shown in Fig. 2. For the PCA algorithm, with the number of target features increasing, the classifier also performs better. When this number reach to about 30, CCR tends not to increase any more. For the relief algorithm, the number of generated features and CCR present a valley-like relationship. The highest CCR is reached in the range of [24,25]. If the number exceeds this range, the performance of classifier would not improve anymore and even start to decline. The redundant features may increase the difficulty of attention recognition.

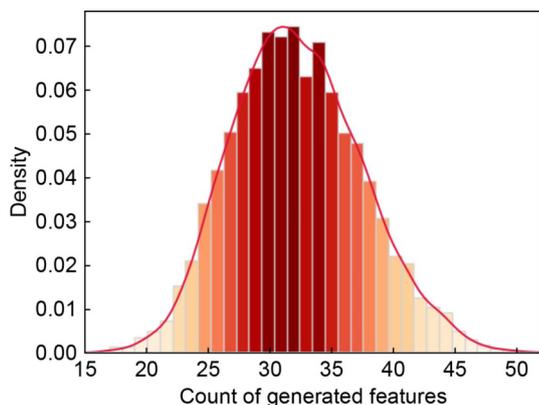


Figure 1. Histogram of number of features generated by GA.

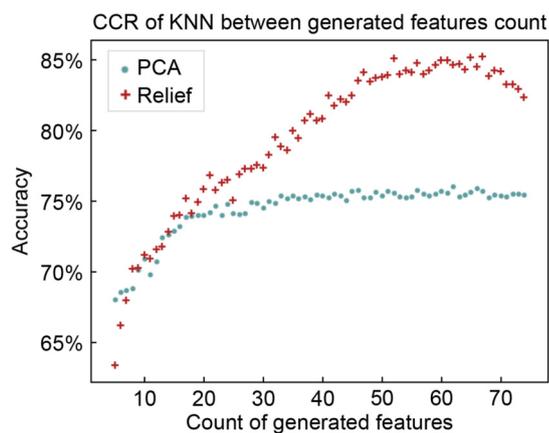


Figure 2. Performance of KNN classifier between different features dimensions.

The best number of generated features differ for different FSA which may indicate that it is hard to find a group of features fitting all classifiers. In the following experiment, we randomly choose the features group in the above range to compare the performance of classifiers.

Comparison between different FAS

Table 1 gives the average CCR between different feature

Table 1 Average CCR between different FSA and classifiers

Method	Raw (%)	PCA (%)	GA (%)	Relief (%)
SVM	76.33±4.75	76.80±4.28	77.48±5.40	77.74±5.15
MLP	67.68±4.39	64.19±1.58	70.05±2.93	62.73±6.64
KNN	81.14±4.45	75.68±4.50	82.30±5.09	84.58±4.61
CART	68.28±5.49	57.19±5.87	68.91±6.77	69.72±5.96
RandomForest	81.90±4.59	67.64±5.34	82.27±5.11	82.25±5.14
Naivebayes	61.92±3.85	58.45±4.64	60.28±5.48	61.24±3.63

selection algorithms (FSA) and classifiers with 5-fold validation. And Fig. 3 compares the performance of FSA based on Table 1.

In Fig. 3, we find that some classifiers like SVM, KNN can obtain high average CCR with the raw features. PCA has little improvement in the average CCR among classifiers compared to using raw features. For the random forest classifier, the average CCR even witness a serious decline. This indicates that PCA may be not the best for attention recognition. The relief algorithm run fast and does improve the performance among some classifiers like SVM, KNN, while the amount of increment is relatively low. Finally, with the help of GA, most classifiers obtain performance boost, and the magnitude of improvement is obviously. This phenomenon shows the power of GA in searching global optimal solutions. However, the convergence speed of GA is slow and unstable, which results in long searching time.

Comparison between different classifiers

Figure 4 compares the performance of classifiers based on Table 1. In Fig. 4, we found that KNN, random forest classifier can achieve an average CCR over 80%. In particular, the combination of GA+ random forest obtain the highest accuracy of 86.28%±3.81%, which may be due to the strong ability of random forest to fit high-dimensional and nonlinear dataset. Although MLP is a powerful tool, it suffers from the dilemma of overfitting. The Naïve Bayes classifier performs worst and only obtains an accuracy of about 60%. The reason may be the fact that Naïve Bayes assumes all features are independent of each other, which is not in line with reality.

The receiver operating characteristic (ROC) curves are plotted in Fig. 5 to compare the classification ability between binary classifiers. The ROC curve closer to upper left corner means that the classifier performs better. From Fig. 5, we can observe the same trending as CCR. The random forest and KNN classifiers both work great, but the random forest mostly receives higher true positive (TP) rate than KNN, which makes it outperform KNN under the AUC (area under the ROC curve) metric.

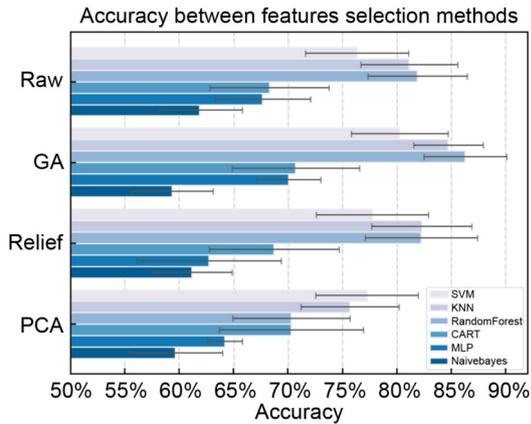


Figure 3. Accuracy comparison between FSA.

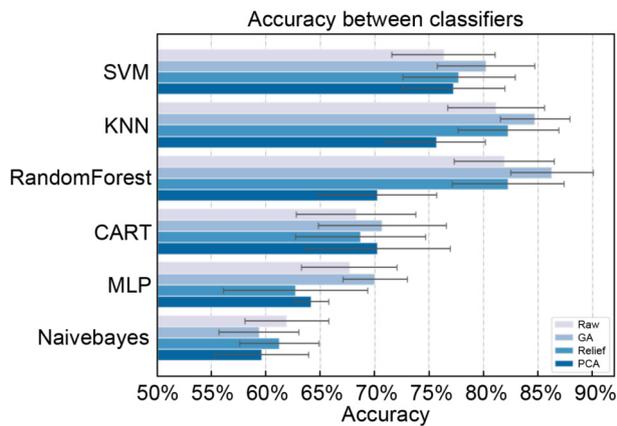


Figure 4. Accuracy comparison between different classifiers.

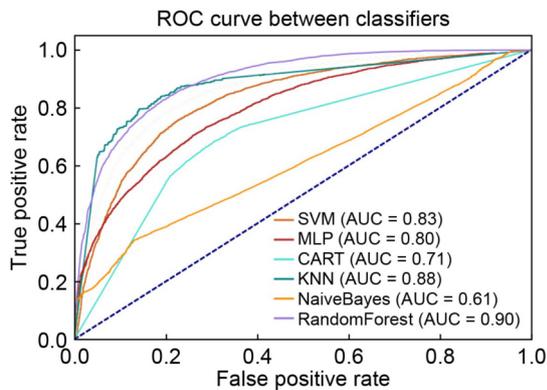


Figure 5. ROC curve between different classifiers.

CNN-based evaluation

Besides the traditional classifiers, we also evaluate the performance of CNN-based classification. Since each ECG sample after denoising contains a length of 61,440 points (a minute long of heartbeat), this length is too

long as input of general neural network and it is hard to improve accuracy with this input length. Therefore, we further split each sample into 10 segments and each segment contain 6144 points (about 3 heartbeats). And we also group all ECG samples into training, validation and test sets with the ratio of 0.5, 0.2, 0.3, respectively. The training process use 5-fold cross validation for fine tuning. Figure 6 shows the changes of accuracy and loss with the increase of training epochs. We can observe that with the training epochs increase up to about 100, the network approximately reaches the highest validation accuracy. After 100 epochs, the model starts to overfit the training set and does not bring improvement to validation accuracy. So, in our case we use the model at 100 epochs as the best trained network and evaluate the its performance at test set with an accuracy of 80.2% and loss of 0.497.

DISCUSSION

In all the experiments of ECG signals for attention recognition above, we observe that the better classification performance can be achieved by using traditional classification algorithms. It is feasible to apply feature selection methods to optimize features, however, their effect may have a large difference. In our experiment, using GA to optimize features can improve the performance for most of classifiers, while the projected features generated by PCA increase the difficulty of recognition task. The final average CCR is greatly affected by the type of classifier, indicating the non-linearity and high complexity of this task. When applying random forest combined with GA for features optimization to build the recognition model, we can achieve the best average CCR of 86.28%+3.81% (as shown in Table 2). That indicates the strong correlation between ECG signals and subject’s attention.

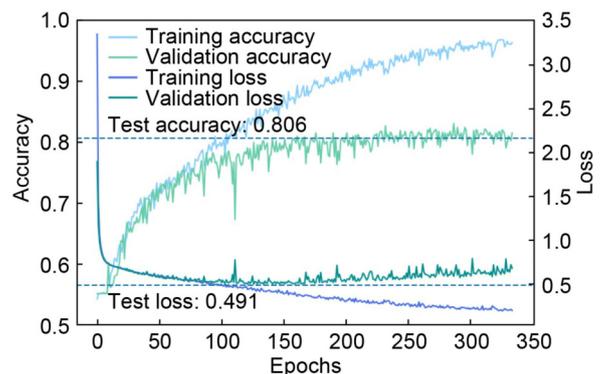


Figure 6. Accuracy trend of CNN-based classification with different training epochs.

Table 2 Average CCR of different classifiers combined with GA

Classifier	SVM	KNN	RandomForest	CART	MLP	Naivebayes	CNN
CCR	80.24%	84.72%	86.28%	70.70%	70.05%	61.92%	80.2%

MATERIALS AND METHODS

Our work aims to identify the attention status of subjects by processing on the ECG data. Given the ECG signals collected from subject as input, our system can recognize the corresponding attentive state. This pipeline is shown in Fig. 7, which includes ECG signal acquisition, preprocessing, attention recognition and performance estimation.

Compared to other physical signals (facial expression, gesture, speech, etc.) collected from camera or perceptual sensors, ECG data acquisition requires stricter experimental environment. Currently, there is still no standard internationally approved ECG database available for attention recognition. Thus, in our study we design an ECG signals acquisition experiment, which is designed especially for attention detection and collects data samples from 10 subjects to obtain raw ECG signals. Since the ECG signals are prone to be suffering from different kinds of noise, we use discrete wavelet transform algorithm to remove noise, including baseline wandering and power-line interference. Then we split the signals into multiple segments as samples and obtain feature points by pan-tompkins algorithm. Morphological and statistics features are then extracted from the feature points of each sample. After that, we adapt several feature selection methods including PCA, relief and generic algorithm, to optimize the extracted features. These features are finally sent to a group of classifiers including SVM, KNN, random forest, for attention recognition. For comprehensive comparison, we also introduce a CNN model to classify the serialized

ECG signals. The classification performances between different features selection methods and classifiers are further evaluated.

ECG signals denoising

ECG whose amplitude is typically between 10 μ V and 5 mV is a relative weak signal and is featured as a continuous waveform including P-wave, QRS complex and T-wave, as shown in Fig. 8. The QRS complex is the dominant component of ECG signal. It can further be decomposed into Q, R and S-wave, with a tendency of falling, rising and falling respectively. In general, it is more convenient to recognize R-wave, since the R-wave contains the largest amplitude. During the detection of ECG, the interference mainly comes from three kinds of noise sources: baseline wandering, myoelectric interference and power-line interference [26,27].

Since the frequency range is extremely high, and far beyond that of real ECG signals, it is difficult to remove myoelectric interference. In practice, the interference it makes is relatively low due to its low amplitude. Therefore, in our experiment, we neglect the effect of myoelectric interference and focus on the denoising of baseline wandering and power-line interference.

Discrete wavelet transform

Concerning the discrete wavelet transform (DWT) can analyze multiresolution signal with a good representation for shape and feature of local region, it is adopted to decompose the non-stationary ECG signal. The

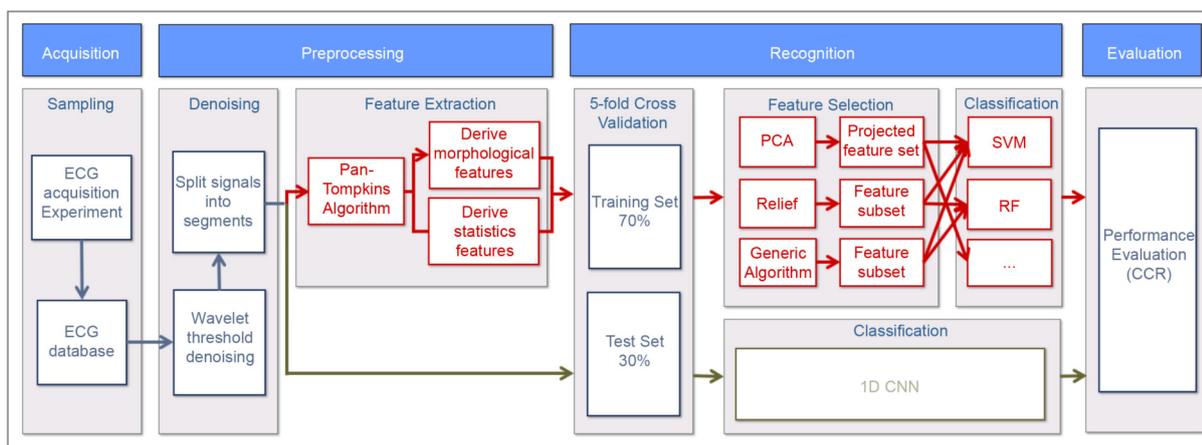


Figure 7. Pipeline of our method.

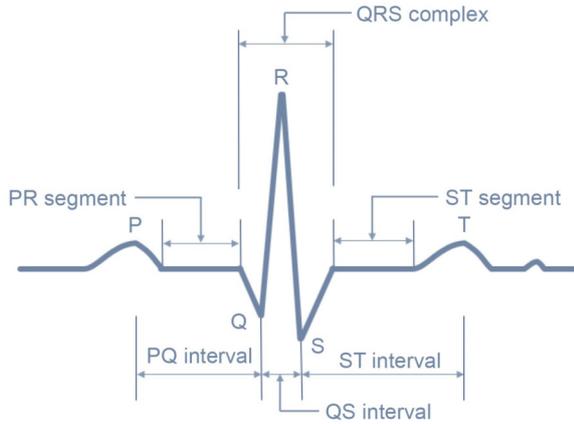


Figure 8. A typical ECG signal waveform [6].

wavelet decomposition can be regard as an iterative process to extract information from both time and frequency domains. Two symmetric filters: high-pass filter (HPF) and low-pass filter (LPF) are built with the mother wavelet and scaling function. These filters provide a set of orthonormal basis to segment the original spectrum and generate signals with high and low frequency respectively in each iteration. The main procedure of DWT is described as follows:

Suppose the measured ECG signals $f(k)$ are combined

with ground-truth signals $s(k)$ and noise:

$$f(k) = s(k) + n(k), \quad k = 0, 1, 2, \dots, N - 1 \quad (1)$$

where $n(k)$ is the Gaussian white noise and it typically subjects to normal distribution with zero mean value and unknown variance σ^2 , N is the signal length.

We can apply discrete wavelet transform to $f(k)$:

$$W_f(l, k) = 2^{-\frac{l}{2}} \sum_{k=0}^{N-1} f(k) \Psi(2^{-l} - k), \quad l = 0, 1, 2, \dots, L - 1 \quad (2)$$

where $W_f(l, k)$ is the wavelet coefficient, j is the scaling factor, $\Psi(t)$ is a chosen wavelet basis function. These wavelet coefficients are composed of real ECG signal and noise.

The denoising process is operated on the wavelet coefficients. The detail of them will be discussed in the next section. After that, we perform reconstruction (climbing up the decomposition tree) to obtain the denoised signals:

$$S_f(l - 1, k) = S_f(l, k) * \tilde{h}(l, k) + W_f(l, k) * \tilde{g}(l, k) \quad (3)$$

where $S_f(0, k)$ is the initial ECG signal, \tilde{h}, \tilde{g} are the conjugate function of h and g .

Figure 9 shows the DWT decomposition results in 7 level using db3 wavelet family.

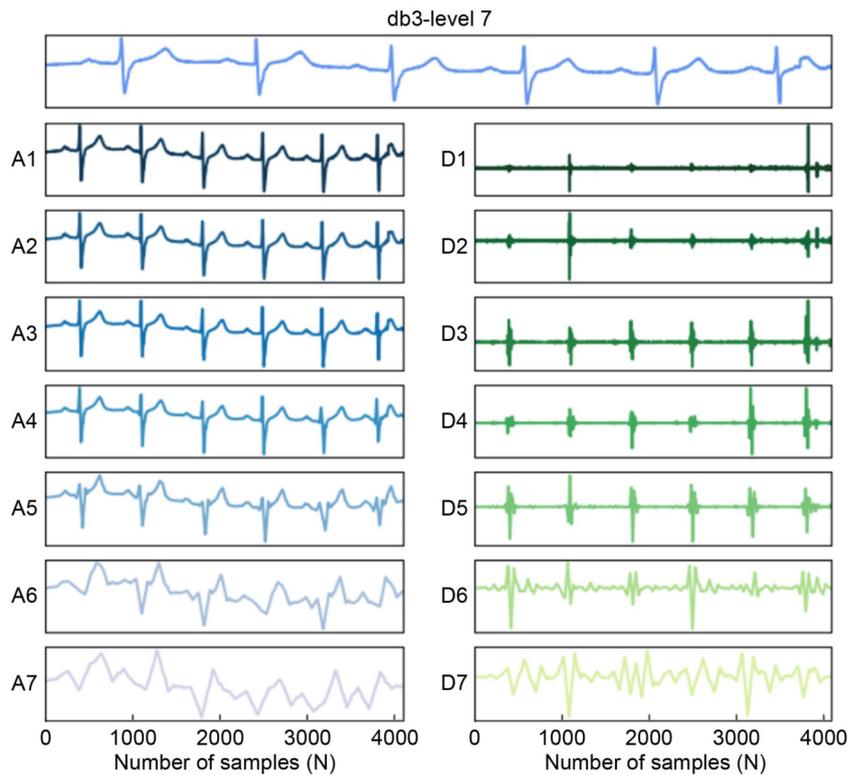


Figure 9. Decomposition results in db3.

Wavelet threshold denoising

To denoise power-line interference, the wavelet threshold denoising algorithm is used in our method. The basic assumption of this algorithm is: considering the continuity of ECG signals in time domain, its amplitude of wavelet coefficients after wavelet decomposition comes to be larger, while the power-line interference shows relatively small amplitude and strong randomness due to its discontinuity in time domain as Gaussian white noise. Based on this assumption, we perform wavelet decomposition on the measured ECG signals in various scales to obtain the detail and approximate coefficients of each layer. When the detail coefficients at a certain scale is smaller than a selected threshold λ , we can consider these coefficients are mainly produced by noise. By discarding them and using the remaining wavelet coefficients (which are produced by real ECG signals), we can finally reconstruct the signal without noise interference [28].

The threshold function, which determines how to resolve the wavelet coefficients and has a great impact on ECG signal denoising, mainly include the hard threshold method and the soft threshold method. When using the hard threshold method, the continuity of wavelet coefficients may probably be disrupted and thus result in local jitter and oscillation in the reconstructed

signal. To obtain smoother wavelet coefficients and reconstructed signal, we apply the soft threshold method in our experiment.

We compare the effect of power-line interference denoising between the raw ECG signal and reconstructed signal (decompose with haar, sym5 and db5 wavelet in 4 level) as shown in Fig. 10. It can be observed that the power-line interference is removed by all three wavelets. The haar wavelet appear to be stepwise, while sym5 and db5 both preserves the smooth clinical detail. We also compare the effect of power-line interference denoising between in different decompose levels. With the increase of decompose level, the reconstructed signals have stronger ability to remove power-line interference, but also increase the probability of distortion. Thus in our method, we use 4 level decomposition and db5 as wavelet basis in denoising since this setting preserves the most clinical detail.

Baseline cancellation

We use wavelet baseline cancellation to remove baseline wandering noise, since it is one of the best performing approaches [29] among them and can be easily integrated into our processing workflow. According this method, we simply set the approximation coefficients at lowest frequency band to zero and the amplitude of all

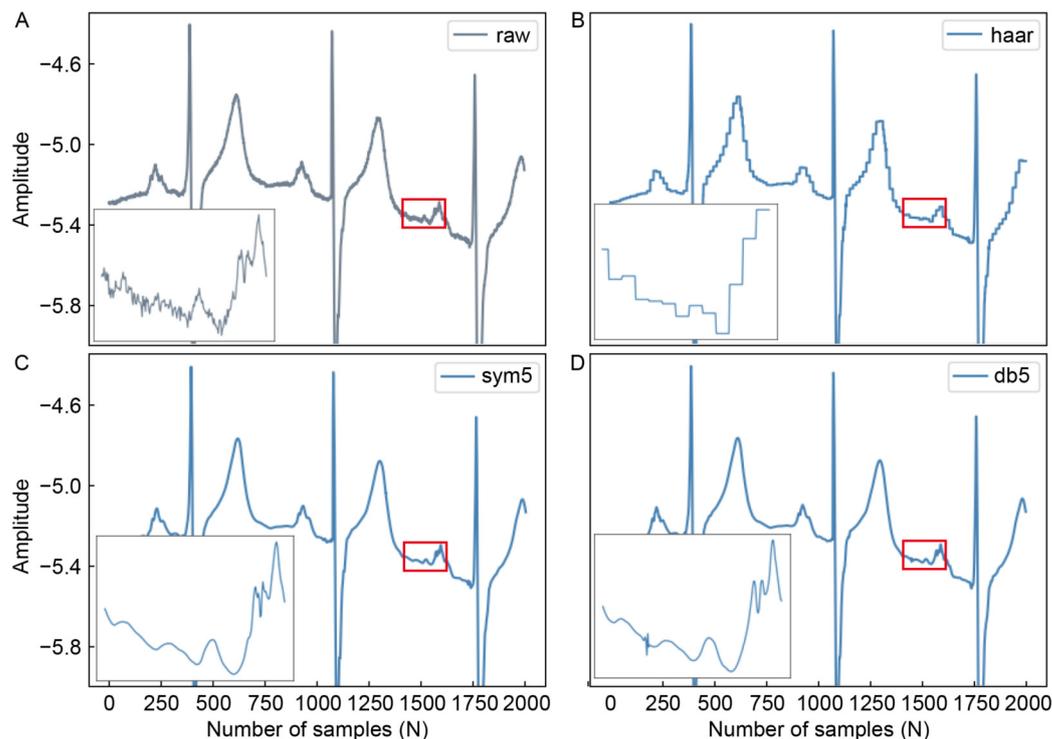


Figure 10. Comparison of power-line interference denoised result between different wavelet basis functions. (A–D) Original, haar, sym5, db5.

ECG signal cycles will be normalized into range $[-1, 1]$ after this algorithm. In our experiment, we tested the reconstruction effect with different L_d . As L_d increases, we can separate the low and high frequency components of signals more easily. Figure 11 shows the effect of baseline cancellation between the raw ECG signal and reconstructed signal. With the decompose level increases, the reconstructed signals are more capable to fit the trend of original signals. When the number of level increases to 11, the reconstructed signals fails to keep its baseline. Thus we set L_d to 9 achieves the best trade-off. After the denoising procedure is completed, we split the signals into segments with around 1 minute, and regard each segment with corresponding attention label as an ECG samples.

Features extraction

The QRS complex detection and feature extraction are the key processes to the success of ECG signal analysis. Since the QRS complex is the dominant component in ECG, the detection of heartbeat depends on the localization of the R-wave. The accuracy of QRS recognition will affect the subsequent detection of P-wave and T-wave.

In our method, we use Pan-Tompkins algorithm [30] for QRS detection, since it can achieve both real-time performance and high accuracy (99.3% accuracy for detecting MIT/BIH arrhythmia database [31]). The Pan-Tompkins algorithm consists of a series of filtering processes, including a band-pass filter cascaded by low-pass and high-pass filter to remove low frequency interference and high frequency artifacts. Then, the moving-window integration is used to acquire more

information about waveform feature. Finally, the locations of QRS complex is marked by applying an adaptive threshold method to the output stream. Whenever a R-wave is recognized, the algorithm recalculates the mean value of RR-interval, which is a dynamic estimation of ECG cycle. The location of other peaks (*i.e.*, Q, R, S, T-peak) can be derived by searching the local maxima or minima from the both side of R-peak in an RR-interval. Figure 12 shows the result of feature point detection of an ECG sample.

After all the waveforms are successfully recognized, we can extract morphological and statistics features from each ECG sample. The morphological features contain the average, median, standard deviation, minimum, maximum and fluctuation of all detected peaks in each ECG sample, and could be extracted from P, Q, R, S, T wave, and PQ-, QS- and ST-intervals. The statistics features could be extracted by wavelet decomposition, including the average, median, standard deviation, minimum and maximum features. Table 3 summarizes the extracted features.

Wavelet decomposition can effectively separate the low and high frequency components from signals. From the view of energy, the low frequency component (*i.e.*, the approximate coefficient) is almost the same with original signal, meaning that it preserves abundant information of original signal. In contrast, the high-frequency component is more vulnerable to various interferences (*e.g.*, muscle noise) and has little contribution to subsequent affect recognition. Therefore, we only consider features extracted from low frequency component and discard high frequency counterpart.

Based on this idea above, we apply the wavelet transform to decompose the signal into 5 levels. The

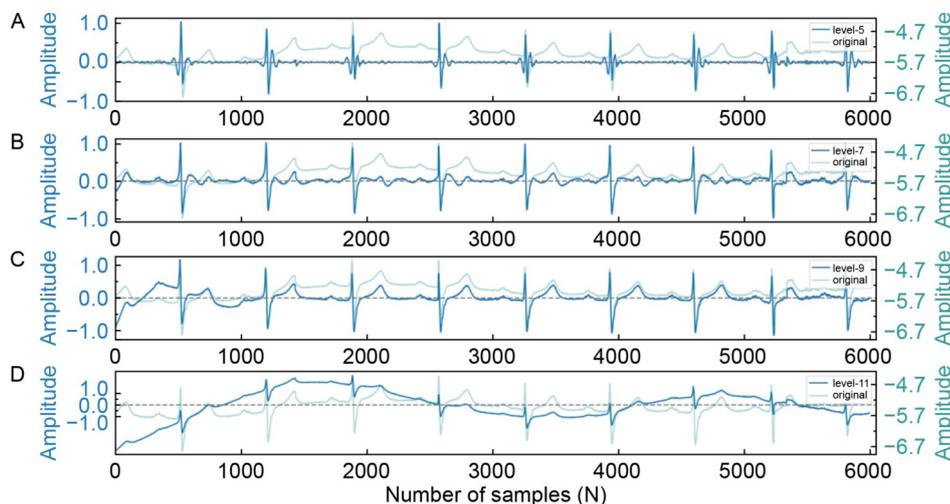


Figure 11. Comparison of baseline wandering denoised result between original and reconstructed ECG signal in multiple decomposed level. (A–D) Level 5, level 7, level 9, level 11.

approximate coefficients of i -th level is denoted by \mathcal{S}_i , while \mathcal{S}_0 is the original signal sequence. Then the first and second order differences are performed on \mathcal{S}_i , indicating the trending of ECG sample:

$$\Delta^2\Psi = (\Delta^2s_i | i = 0, \dots, 4), \quad (4)$$

where Δ^2s_i is the first order difference of \mathcal{S}_i . Finally, we extract features from \mathcal{S}_i by a vector:

$$F_{si} = (S_{i,avg}, S_{i,med}, S_{i,sd}, S_{i,min}, S_{i,max}), \quad (5)$$

where $S_{i,avg}, S_{i,med}, S_{i,sd}, S_{i,min}, S_{i,max}$ are the average, median, standard deviation, minimum and maximum of \mathcal{S}_i respectively. We flatten features of all levels and result in a feature vector with 90 dimensions.

Attention classification

In the task of attention recognition, given the samples of ECG signal as input, our goal is to construct a model which can output the corresponding label (attentive or non-attentive) for each sample. In Section ‘‘Wavelet threshold denoising’’, we extract features with 138-dimensions for each ECG sample by combining morphological and statistics features. These features are normalized before classification. In the task of attention recognition, a certain degree of redundancy may probably exist between these features which may limit the ability of generalization. Thus, several feature selection algorithms are used to the optimize the features. We further compare their performances on different classification algorithms.

In our experiment, three typical feature selection algorithms are selected: PCA, relief and generic

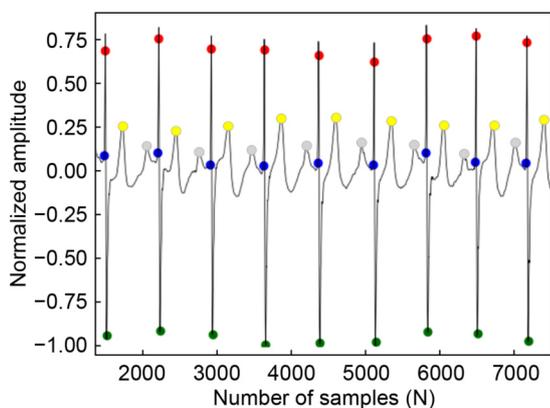


Figure 12. Result of feature point detection of an ECG sample.

algorithm (GA). We choose 7 classifiers to build the attention recognition model: support vector machine (SVM), Multi-Layer perception (MLP), K-Nearest neighbor (KNN) [32], classification and regression tree (CART), random forest (RF), Naïve Bayes. Besides, as neural network become one of the most powerful tools in physiological processing, we also propose an end-to-end CNN for attention classification. The CNN network takes the time-series denoised ECG samples as input and outputs the label of the corresponding signals.

As shown in Fig.13, the proposed CNN consists of two stages. The first stage encodes the signals into low dimensional features by convolution operations, while the second stage tries to incorporate the extracted features to classify heartbeats into target labels. To be specific, the first stage of CNN takes the original signals with size of 6144×1 as input and follows a convolutional process with a convolution and average pooling layer (with pool size of 1×2). This process continues for 3 times each with the kernel sizes of 1×201 , 1×141 and 1×141 respectively. Then the feature maps of all filters are cascaded into a full connected layer (FC) of 980 neurons as input of the second stage. Finally, another FC layer follows containing 128 neurons and the network output one neuron indicating the attention category. The dropout layers added between FC layers help to avoid overfitting.

EXPERIMENT PROCEDURE

To collect enough data for attention recognition, 10 male subjects are participated in the experiment. All subjects were healthy, aged from 20 to 25 and had no history of mental and heart disease. Their mean age was 22 years (std = 1.81 years). Compared with other emotion-stimulated materials such as pictures, audio, video clips can attract one’s attention from both vision and hearing. In our sampling experiment, we use video clips to elicit attention state. Considering that attention is more sensitive than other emotion (*e.g.*, happy, angry), we carefully design the sampling procedure to capture the ECG signal in attentive or non-attentive state during video watching. The subjects were free to choose the video clips that they were most interested and non-interested.

The ECG collecting kit produced by Shimmer Company [33] was used for ECG sampling. The sampling procedure was conduct in a quiet and closed room. Before the experiment begun, the subjects were

Table 3 Morphological and statistics features extracted from ECG signals

Morphological features	P, Q, R, S, T wave, and PQ-, QS- and ST-intervals	Average, median, standard deviation, minimum, maximum and fluctuation
Statistics features	Low frequency of signals	Average, median, standard deviation, minimum and maximum

asked to rest for 10 minutes to restore normal emotional state. Then each subject started to watch a video clip in 10 minutes with the data collecting launched simultaneously. The next session would start in the next 10 minutes giving the subject enough time to relax himself. The experimental procedure is shown in Fig. 14. Inspired by self-assessment manikin (SAM) [34], when each session was done, the subject was asked to fill out a self-assessment questionnaire to evaluate their emotion during experiment with labels of low, media and high level of concentration. It helps to avoid collecting signals of false label. Combining the results of questionnaire and the integrity of corresponding waveforms, a total of 200-minutes samples was collected for attentive and non-attentive state respectively. Finally, an ECG dataset with 360 samples were finally collected. Half of them are attentive states, and the remaining are non-attentive states. We send all the samples to the preprocessing pipeline and obtain feature vectors in 138-dimensions for each ECG sample.

We adopt stratified strategy to split the training set and test set, with 70% of samples for training, 30% for testing. To obtain more accurate and stable result, we use the average classification rate (CCR) as metric to evaluate the performance for classifiers in 5-fold cross validation.

ACKNOWLEDGEMENTS

The work of this paper is financially supported by NSF of Guangdong Province (No. 2019A1515010833) and the Fundamental Research Funds for the Central Universities (No. 2020ZYGXZR089), and the Social Science Research Base of Guangdong Province-Research Center of Network Civilization in New Era of SCUT.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Aihua Mao, Zihui Du, Dayu Lu and Jie Luo declare that they have no conflict of interest or financial conflicts to disclose. All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

OPEN ACCESS

This article is licensed by the CC By under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted

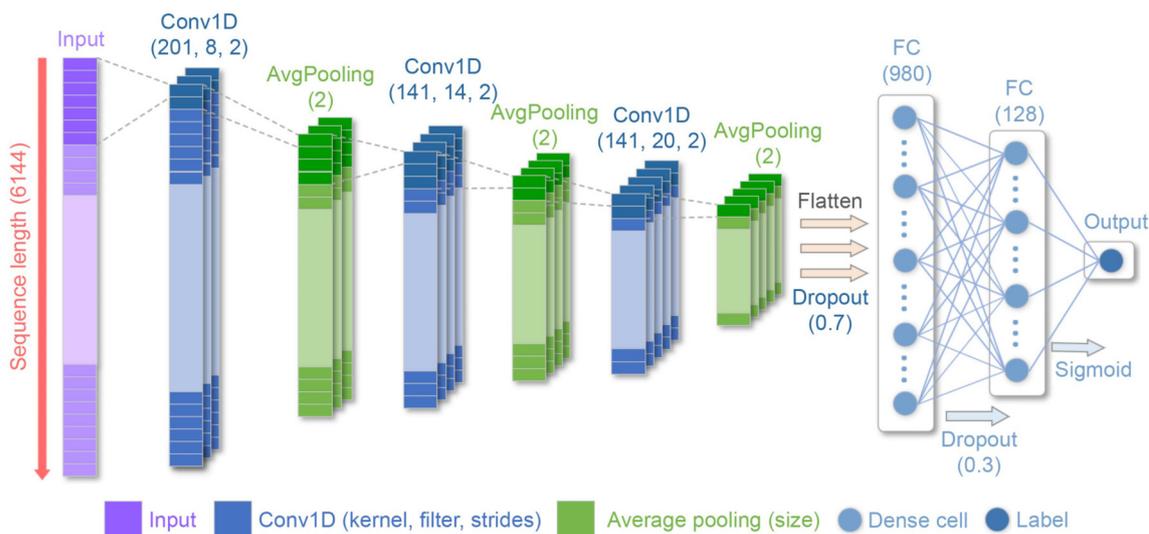


Figure 13. Architecture of proposed CNN.

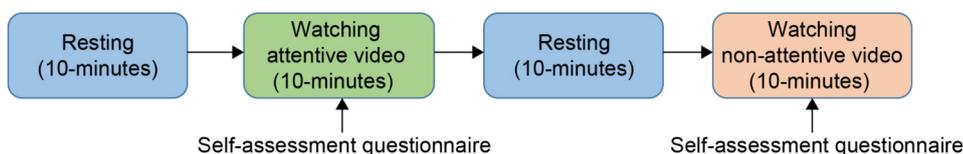


Figure 14. Diagram of the experimental procedure.

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Picard, R. W. (2000) *Affective Computing*. Cambridge: MIT press
- Ekman, P. and Friesen, W. V. (1971) Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.*, 17, 124–129
- Fox, C. J. and Barton, J. J. (2007) What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Res.*, 1127, 80–89
- Batty, M. and Taylor, M. J. (2003) Early processing of the six basic facial emotional expressions. *Brain Res. Cogn. Brain Res.*, 17, 613–620
- Perikos, I. and Hatzilygeroudis, I. (2016) Recognizing emotions in text using ensemble of classifiers. *Eng. Appl. Artif. Intell.*, 51, 191–201
- Plutchik, R. (2001) The nature of emotions. *Am. Sci.*, 89, 344–350
- Lang, P. J. (1995) The emotion probe. *Studies of motivation and attention. Am. Psychol.*, 50, 372–385
- Kuo, Y.-C., Chu, H.-C. and Tsai, M.-C. (2017) Effects of an integrated physiological signal-based attention-promoting and English listening system on students' learning performance and behavioral patterns. *Comput. Human Behav.*, 75, 218–227
- Song, T., Zheng, W., Song, P. and Z. Cui. (2020) EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.*, 11, 532–541
- Hsu, Y., Wang, J., Chiang, W. and Hung, C. (2020) Automatic ECG-based emotion recognition in music listening. *IEEE Trans. Affect. Comput.*, 11, 85–99
- Liu, Y., Yu, M., Zhao, G., Song, J., Ge, Y. and Shi, Y. (2018) Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Trans. Affect. Comput.*, 9, 550–562
- Ding, Y., Hu, X., Xia, Z., Liu, Y.-J. and Zhang, D. (2021) Inter-brain EEG feature extraction and analysis for continuous implicit emotion tagging during video watching. *IEEE Trans. Affect. Comput.*, 12, 92–102
- Du, X., Ma, C., Zhang, G., Li, J., Lai, Y.-K., Zhao, G., Deng, X., Liu, Y.-J. and Wang, H. (2020) An efficient LSTM network for emotion recognition from multichannel EEG signals. *IEEE Trans. Affect. Comput.*, 3013711
- Zhang, G., Yu, M., Liu, Y.-J., Zhao, G., Zhang, D. and Zheng, W. (2021) SparseDGCNN: Recognizing emotion from multichannel EEG signals. *IEEE Trans. Affect. Comput.*, 3051332
- Pourtois, G., Schettino, A. and Vuilleumier, P. (2013) Brain mechanisms for emotional influences on perception and attention: what is magic and what is not. *Biol. Psychol.*, 92, 492–512
- Taylor, J. G. and Fragopanagos, N. F. (2005) The interaction of attention and emotion. *Neural Netw.*, 18, 353–369
- Aliakbarhosseinabadi, S., Kamavuako, E. N., Jiang, N., Farina, D. and Mrachacz-Kersting, N. (2017) Classification of EEG signals to identify variations in attention during motor task execution. *J. Neurosci. Methods*, 284, 27–34
- Liu, N. H., Chiang, C. Y. and Chu, H. C. (2013) Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors (Basel)*, 13, 10273–10286
- Hamadicharef, B., Zhang, H., Guan, C., Wang, C., Phua, K. S., Tee, K. P. and Ang, K. K. (2009) Learning EEG-based spectral-spatial patterns for attention level measurement. In: *IEEE Inter. Symp. Circ. Syst.*, pp. 1465–1468
- Eddin Alchalabi, A., Elsharnouby, M., Shirmohammadi, S. and Nour Eddin, A. (2017) Feasibility of detecting ADHD patients' attention levels by classifying their EEG signals. In: *2017 IEEE Inter. Symp. Medic. Measur. Applic. (MeMeA)*, pp. 314–319
- Ghanadian, H., Ghodratioghar, M. and Al Osman, H. (2018) A machine learning method to improve non-contact heart rate monitoring using an RGB camera. *IEEE Access*, 6, 57085–57094
- Egger, M., Ley, M. and Hanke, S. (2019) Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.*, 343, 35–55
- Emanet, N. (2009) ECG beat classification by using discrete wavelet transform and Random Forest algorithm. In: *2009 Fifth Inter. Confer. Soft Comput., Comput. Words Percept. Syst. Anal., Decis. Contr., Famag.*, 5379457
- Zhang, Y. D., Yang, Z. J., Lu, H. M., Zhou, X. X., Phillips, P., Liu, Q. M. and Wang, S. H. (2016) Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4, 8375–8385
- Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, 18, 193–222
- Agante, P. M. and Marques de Sa, J. P. (1999) ECG noise filtering using wavelets with soft-thresholding methods. In: *Proc. Comput. Cardiology 1999*, pp. 535–538
- Lu, G., Brittain, J. S., Holland, P., Yianni, J., Green, A. L., Stein, J. F., Aziz, T. Z. and Wang, S. (2009) Removing ECG noise from surface EMG signals using adaptive filtering. *Neurosci. Lett.*, 462, 14–19
- Donoho, D. L. and Johnstone, I. M. (1995) Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.*, 90, 1200–1224.
- Lenis, G., Pilia, N., Loewe, A., Schulze, W. H. and Dössel, O. (2017) Comparison of baseline wander removal techniques considering the preservation of ST changes in the ischemic ECG: A simulation Study. *Comput. Math. Methods. Med.*, 2017, 9295029
- Pan, J. and Tompkins, W. J. (1985) A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.*, 32, 230–236
- Hamilton, P. S. and Tompkins, W. J. (1986) Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Trans. Biomed. Eng.*, 33, 1157–1165
- Liu, C., Rani, P. and Sarkar, N. (2005) An empirical study of machine learning techniques for affect recognition in human-robot interaction. In: *2005 IEEE/RSJ Inter. Confer. Intellig. Robots Syst.*, pp. 2662–2667
- Website: <https://www.shimmersensing.com/products/>. Accessed: January 5, 2021
- Bradley, M. M. and Lang, P. J. (1994) Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry*, 25, 49–59