

# Bimodal Recognition of Cognitive Load Based on Speech and Physiological Changes

Dennis Held, Sascha Meudt<sup>(✉)</sup>, and Friedhelm Schwenker

Institute for Neural Information Processing, Ulm University, 89069 Ulm, Germany  
{dennis.held,sascha.meudt,friedhelm.schwenker}@uni-ulm.de

**Abstract.** An essential component of the interaction between humans is the reaction through their emotional intelligence to emotional states of the counterpart and respond appropriately. This kind of action results in a successful interpersonal communication. The first step to achieve this goal within HCI is the identification of these emotional states.

This paper deals with the development of procedures and an automated classification system for recognition of mental overload and mental underload utilizing speech and physiological signals. Mental load states are induced through easy and tedious tasks for mental underload and complex and hard tasks for mental overload. It will be shown, how to select suitable features, build uni modal classifiers which then are combined to a bimodal mental load estimation by the use of early and late fusion. Additionally the impact of speech artifacts on physiological data is investigated.

## 1 Introduction

The interaction between humans and machines is ubiquitous in today's world. Whether at the station, while solving a web card or when using a mobile phone, it has become the aim of making this interaction between man and machine to the user so intuitive, effective and pleasant as possible. While this is desirable in most cases, but not always reach satisfactory levels. A key factor that affects the human interaction significantly, has not yet been taken into account - the emotional intelligence [6]. By the term emotional intelligence one understands the ability of the regulation, the use of knowledge and the expression of emotions in interactions with others or with themselves. A system that is centered on the satisfaction of persons needs should offer these expertise to react appropriately to the emotional state [10].

Recognition of affective states on multiple modalities, such as facial expressions, speech, gestures, and to a lesser extent on physiological changes. The recognition of these states is possible because by state changes the expressions adjust accordingly such as the pitch of the voice when is excited.

The consideration in this work are speech and physiological signals. The speech is in terms of research a dominant modality, as for the people it is representing the most natural way of communicating and a very quick indicator for an emotional state of the counterpart. A closer look at speech turns out that not

only the content of what is said is relevant, but also the way on how its said and which emotions are placed in an utterance. The recognition of the content by machines, with a high detection rate of about 90%, is already realized, however, the rate of recognition of speech emotions is only about 60% [11]. Physiological signals like electrocardiography (ECG), electromyography (EMG) and electrodermal activity (EDA) provide a relatively new and growing area of research in terms of affective state recognition compared to audiovisual emotion recognition. The advantages of physiological signals are that the regulation of the values such as heart rate or the activity of the sweat glands cannot be simply consciously influenced, such as speech and gestures. This leads to an almost undistorted image on the emotional state of the person.

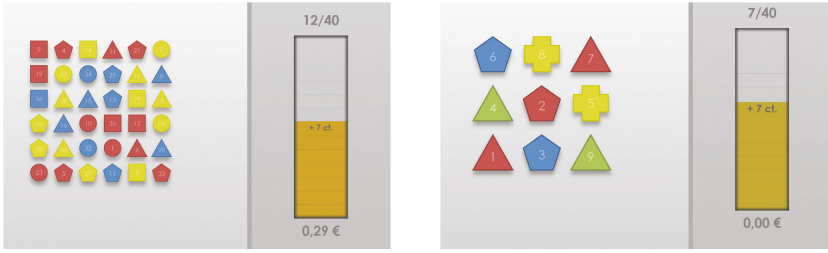
In the past there had been a lot of research on detecting emotional states utilizing single modalities, like detecting affect from speech with detection rates from 50% up to 90% [7,8]. For physiological signals there are many studies on the impact and the detection of mental load [2,3]. For bimodal approach with speech and physiological signals J. Kim shows an improvement in recognition rate compared to unimodal classification [4]. Most of the studies are based on acted or even strong expressive emotional datasets on basic emotions for example defined by Eckman [1]. This methods can not be easily transferred to typical weak expressive behaviours in HCI which in addition typically does not refer to basic emotions.

The utilized dataset contains data from natural behaving, none acting, users interacting with a HCI system which is able to induce different mental load levels. The investigated affective states are mental overload and underload that both have a negative impact on the performance of a person. Firstly, even simple tasks work in a period of excessive demand as relatively difficult. On the other tasks are perceived as boring during a mental underload, which leads to a lack of concentration and thus to a decrease in performance. Based on the dataset this paper develops an classification system for these affective states with the use of ensemble classifiers. The interests are the unimodal classification based on speech and physiological signals, the fusion of these to a bimodal classification result and the investigation of the influence of speech artifacts on the classification of physiological signals.

## 2 Experimental Setting

The dataset is based on an experiment done within the Transregional Collaborative Research Centre SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems”.

Participants were asked to play a series of games based on the Interaction paradigm of Schüssel et al. [12]. The task of each game sequence was to identify the singleton element, i.e. the one item that is unique in shape and color (number 36 and 2 in Fig. 1). The difficulty was set by adjusting the number of shapes shown and the maximum time to answer. If the given answer was incorrect, they received no reward for that particular round. After an introduction each participant completed five game sequences of decreasing difficulty. The first sequence



**Fig. 1.** Screen shot of the difficult level (left) with target element 36 and the easiest one (right) with target element 2. (Color figure online)



**Fig. 2.** Overview of the setting with sensors: (1) MS Kinect 2, (2) frontal webcam, (3) wireless headset, (4) GTec g.MOBILab+ biophysical sensor with sensors attached to the users body.

was designed to induce overload ( $6 \times 6$  board, 6 s to answer, see Fig. 1 left), the second was  $5 \times 5$  with 10 s, the third was set to  $3 \times 3$  with 100 s, sequence four again was  $3 \times 3$  mode with 100 s time (underload). As the sequences 1 and 4 are explicitly designed to cause mental overload and underload, we focused on those two.

After each sequence played, the participants answered a Self Assessment Scale questionnaire (SAM). The aim of those questions was to determine valence, arousal and dominance experienced in the particular sequence. A total of 60 participants were recorded. Of those were 30 male and 30 female. Their age spanned from 17 to 27 (mean 21.97,  $\sigma^2 \approx 2.6$ ).

During the experiment, participants were monitored by several sensors providing multimodal synchronous data. See Fig. 2. The sensory system contains two webcams (Logitech C9100), one in front looking towards the users face and one from the rear providing an overview of the scene, a wireless headset, a Microsoft Kinect v2 camera in front recording rgb, infrared, depth, skeleton/postural and audio information and finally an GTec g.MOBILab+ biophysical sensor recording ECG, EDA, EMG (trapezius muscle), Respiration and Temperature. In this work we focus on the audio and physiological data.

### 3 Unimodal Recognition of Cognitive Load

At first, two unimodal classifiers for speech and two for every channel of the physiological signals were trained. This results in a total of twelve ensemble classifiers, where six are specialized on mental overload (OL) and the other six are trained for recognising the amount of mental underload (UL). These unimodal classifiers are evaluated using individual classification (SELF) method with a  $10 \times 10$ -fold cross validation and leave one subject out (LOSO) method with 10-fold cross validation.

In Chap. 4 fusion approaches for the bimodal classification are shown and the influence of speech on the classification of physiological signals.

#### 3.1 Speech

For cognitive load recognition from speech, random forest ensembles were used to classify. Only the utterances contains useful parts without speaking breaks (silence) are dictated by the experimental settings, this reduces the average length of recorded overload sequence from 362 to 22 s. and for the underload sequence from 324 to 18 s. The average amount of utterances within the overload sequence is 45 and for the underload sequence it is 43.

The extracted features of the audio data is subdivided into the feature extraction methods:

- Linear predictive filter coefficients (LPC)
- Mel-frequency cepstral coefficients (MFCC)
- Relative spectral perceptual linear prediction (Rasta-PLP)
- Modulation spectrum (ModSpec)

The window size for each frame ranges between the different feature extraction methods from 40 to 200 ms. The window shift is 20 ms for every feature instance. Overall for every frame 57 features are extracted consisting of 8 LPC, 20 MFCC, 21 Rasta-PLP and 8 ModSpec.

The evaluation with the SELF and LOSO method results in Table 1 show the average classification accuracy based on different features and the early fusion of these features. Classification based on utterances, containing an aggregation of severe frame level decisions, for SELF and LOSO method shows better average classification accuracies than based on single frame level. The highest result with respect to accuracy are achieved through early fusion of the features and the classification of utterances for the SELF and LOSO method.

#### 3.2 Physiology

In order to recognize mental overload on the basis of physiology, a random forest classifier for each of the following physiological signals was trained: Electrocardiography, trapezius trapezius Electromyography, Electrodermal activity, respiration and body temperature. For the physiological signals 58 statistical features

**Table 1.** Evaluation results of all speech classifiers using SELF and LOSO method for every used extraction method. The results are subdivided into classification results based on frame and utterances level.

| Analysis     | SELF   |           | LOSO   |           |
|--------------|--------|-----------|--------|-----------|
|              | Frame  | Utterance | Frame  | Utterance |
| LPC          | 66.52% | 77.75%    | 61.22% | 68.93%    |
| MFCC         | 76.27% | 83.80%    | 68.54% | 75.53%    |
| Rasta-PLP    | 74.49% | 82.75%    | 68.70% | 74.40%    |
| ModSpec      | 70.20% | 78.82%    | 63.90% | 71.29%    |
| Audio-Fusion | 81.67% | 86.58%    | 72.72% | 77.50%    |

were extracted, which results in 290 features over all. The features are extracted on 5s data chunks with an overlap of 4.9s. The preprocessing for the physiological signals is subdivided into two groups, one for ECG and one for the rest of the physiological signals.

For the ECG signal the preprocessing consists of linear detrending and the normalization of the signals with the mean R-peaks, because the sensors are not exactly at the same position at each participant. This creates a different mean peak value for every participant which has no information about the mental load of a participant. For EMG, EDA, respiration and temperature the preprocessing consists of the use of a butterworth filter with low and high cutoff frequencies of 10 Hz and 125 Hz and the order of four.

The features extracted from the ECG signal are based on 25 features from wavelets [15] and 33 statistical features from the PQRST complex [13]. Features extracted from EMG, EDA, respiration and temperature are based on statistical and mathematical features in time and fequency domain [5,9,14].

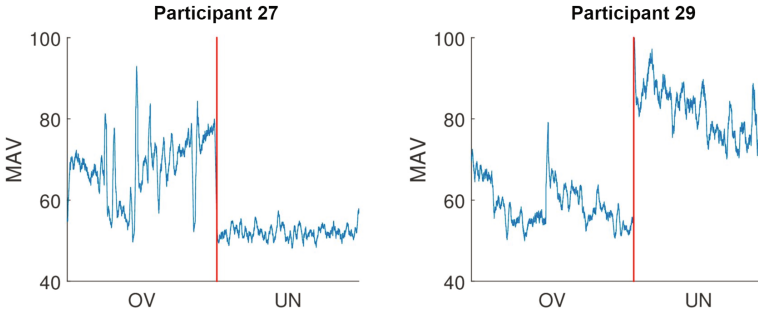
All four classifiers and an additional fusion of these classifiers were evaluated using the SELF and LOSO method. Table 2 shows the average classification accuracies over all 60 participants. The weights for the fusion which are used to fuse the outputs together are calculated through a Moore-Penrose pseudoinverse. The SELF classification results for ECG are promising and achieve an average classification accuracy of 91.52%. The EMG channels shows that there are three different activity patterns for the trapezius muscle:

- More activity in the overload sequence (e.g. in Fig. 3 left)
- less activity in the overload sequence (e.g. in Fig. 3 right)
- almost the same activity in the overload and underload sequence

The results show that the difference between mental overload and underload and therefore the classification can be highly accomplished for SELF, such that the person has a characteristic behaviour for these affective states. Figure 3 shows that for both participants on their own the development of the MAV feature is easily distinguishable between OL and UL sequence. But for the LOSO

**Table 2.** Evaluation results of all physiological classifiers using SELF and LOSO method for every used extraction method.

| Analysis    | SELF   | LOSO   |
|-------------|--------|--------|
| ECG         | 91.52% | 66.53% |
| EMG         | 74.53% | 51.03% |
| EDA         | 64.68% | 51.51% |
| Respiration | 64.68% | 51.50% |
| Temperature | 60.89% | 54.23% |
| Fusion      | 91.91% | 65.61% |



**Fig. 3.** Mean absolute value (MAV) feature from the EMG signal of two participants. Shows the course of MAV feature in the overload (OL) and underload (UL) sequence.

classification Fig. 3 and the results in Table 2 show that there is not an general feature expression in respect to the physiological signals.

## 4 Bimodal Recognition of Mental Overload and Underload

In order to achieve bimodal classification results, two different fusion approaches were used to aggregate the unimodal results. The input consists of two classification results from the speech classifiers for (OL and UL) and two classification results from the physiological classifiers. The output of these classifiers is continuous between 0 and 1, e.g. where 1 for an OL classifier corresponds to that every weak learner classified it as mental overload.

For the bimodal evaluation the data was taken from both modalities, if a participant speaks. If there is no classification from speech the classification result is just based on the physiological classification. The classification structure which is used produces for every utterance for both, the physiological and the audio signal, one classification result.

#### 4.1 Moore-Penrose Pseudoinverse

The pseudoinverse is used to create weights for the OL and UL classifier. Two calculated weights are needed. One for the OL where the extend of mental overload is classified and one for the UL where the extend of mental underload is classified. For the OL classifier a mental overload frame should be classified as 1 and an underload frame should be classified as 0 but for the UL classifier it should work the other way around such that for an underload frame it should be 1.

For the calculation of the weights e.g. for the OL classifier follows these steps:

1. Train the ensembles
2. Take a train set for the fusion with  $N$  frames which are not used to train the ensemble system and which have no intersection with the testset
3. Classify the train set for the fusion thought the trained ensemble system, results in two unimodal classification results
4.  $X$  ( $N \times 2$ ) for OL (two stands for both modalities)
5. Calculate the pseudoinverse  $X^+ = (X^t X)^{-1} X^t$
6. Calculate the weights  $w = X^+ Y$

The same steps should be done for the UL part of the classification system. After theses steps the weights for the bimodal fusion layer are set. The calculated weights strongly depend on the data presented to train the fusion layer. For the SELF classification there are 284 frames and for the LOSO there about 16450 frames to train this layer. This corresponds to about 11% of the data used to train the ensembles.

**Numerical Evaluation.** Table 3 shows the classification accuracy, for SELF classification it is 94.07%. The weights for the input channels through the pseudoinverse for the SELF classification have a  $\sigma^2 = 0.38$  and a mean weight for speech of 0.23 and for physiological of 0.77. The difference of the weights for the modalities is based on the greater variance of the physiological classification between participants.

For the LOSO classification the weight for speech is 0.76 and for physiological it is 0.24. The variance of the weights for LOSO classification is far smaller because only about 1.7% of the data within the training set changes for the classification for a new participant. The overall average accuracy for the LOSO classification is 76.31%. The difference for the weights between the OL and the UL classifier are  $-0.01$  for the speech and  $+0.01$  for the physiological inputs.

#### 4.2 Modified Max-Voting

The second fusion method used is a modified max-voting (MMV). This method does not have any learning phase and can be used immediately to fuse the unimodal classification results. The input consists, as described, of two values for OL and two for UL. For the OL and the UL classification results from the different modalities are averaged. This results in two confidences, one for OL and one

**Table 3.** Bimodal classification accuracy with the Moore-Penrose pseudoinverse method for SELF and LOSO classification.

| Analysis      | SELF   | LOSO   |
|---------------|--------|--------|
| Pseudoinverse | 94.07% | 76.31% |
| OL            | 95.59% | 75.80% |
| UL            | 92.55% | 76.82% |

for UL for each modality. The last step calculates a MMV on these OL and UL classifiers and the bimodal classification result is based on the higher confidence of OL or UL. The bimodal classification confidence is calculated through the difference of confidence of the OL and UL within the step before.

For this fusion method, weights for audio and physiology classification results are the same. It just differentiates how confident the bimodal OL and UL classifiers are. It takes advantage of the classification result for one modality if it has an higher confidence than the confidence of the classifier for the other modality. The more secure a classifier is that it is a specific affective state the less important the other modality is.

**Numerical Evaluation.** Table 4 shows that SELF classification has an average accuracy of 96.41%. The difference between the recognition of mental overload to underload is 3.04%. For 83% of participants the accuracy improves from the unimodal classification results to the bimodal results from 1% up to 29%. This effect is particularly pronounced for a participant which has the lowest classification rate of 67.7% for the physiological signals. His rate increases through the bimodal classification to 95.4%. Another effect is the decrease of variance for the classification results for the different participants: audio ( $\sigma^2 = 2.9$ ), physiological ( $\sigma^2 = 5.6$ ) and bimodal fusion through MMV ( $\sigma^2 = 1.6$ ).

LOSO classification shows an average accuracy of 72.55% but a greater difference between the recognition of mental overload to underload. With a difference of 21.41% between OL and UL it is far more unbalanced in respect to the fusion approach with pseudoinverse.

**Table 4.** Bimodal classification accuracy with the modified max-voting method for SELF and LOSO classification. The accuracies for the mental overload and underload are shown.

| Analysis | SELF   | LOSO   |
|----------|--------|--------|
| MMV      | 96.41% | 72.55% |
| OL       | 97.93% | 82.56% |
| UL       | 94.89% | 61.15% |



**Table 5.** Evaluation results of all physiological classifiers and the comparison between the classification of physiological signals while the participants speak and does not speak using SELF and LOSO method.

| Analysis    | SELF     |             | LOSO     |             |
|-------------|----------|-------------|----------|-------------|
|             | Speaking | No speaking | Speaking | No speaking |
| ECG         | 91.08%   | 92.72%      | 65.70%   | 67.85%      |
| EMG         | 74.54%   | 75.19%      | 49.55%   | 53.62%      |
| EDA         | 64.18%   | 66.08%      | 49.95%   | 54.15%      |
| Respiration | 57.78%   | 60.27%      | 49.72%   | 54.73%      |
| Temperature | 60.47%   | 62.27%      | 53.47%   | 55.65%      |
| Fusion      | 91.73%   | 93.00%      | 65.82%   | 68.42%      |

### 4.3 Influence of Speech to Physiological Classification

The investigation deals with the influence of speech in the classification of physiological signals. The aim is to discover the impact on different physiological channels like EMG and conclude how serious these impacts are.

The change in physiological signals through speech is based on two factors. The first is the direct influence for example through the vibration of the vocal cords or the change of air volume within the lungs and therefore influence on the physiological signal values like for the respiration signal a higher measured respiration rate while speaking. The second is the real influence if there is an direct influence through speaking on the physiological signals e.g. a person speaks if there is significant influence to change the physiological signals and therefore are not based on the emotion to recognize rather than on speaking itself. For example if speaking has a negative influence on the recognition of emotion through physiological signals the reason for this could be that if somebody speaks there is an physiological signature for this within the physiological signals and therefore in all sequence similar. The reason for an influence within the signal is hard to assign based on these two influences therefore it is investigated to what extend the values are changing while speaking and if there is an positive, negative or no influence on the classification.

The training and evaluation is done with extracted frames from the physiological signals where the participant did not speak in any of the 5s length of these frames against the others where the physiological signals are influenced by speaking. This is done for all physiological channels as well as for the unimodal fusion of these channels.

**Numerical Evaluation.** Table 5 shows the classification result for the physiological signals with and without any speaking involved. For the SELF classification the average classification accuracy improves for ECG by 1.64%, EMG by 0.65%, EDA by 1.90%, respiration by 2.49%, temperature by 1.80% and for fusion by 1.27%. 42 participants improve their SELF classification through the

**Table 6.** T-Test between speaking and not speaking for SELF and LOSO classification of the physiological signals.

| T-Test      | SELF                 | LOSO                  |
|-------------|----------------------|-----------------------|
|             | Speaking/no speaking | Speaking/no speaking  |
| ECG         | 0.004                | 0.002                 |
| EMG         | 0.465                | $4.4 \text{ e}^{-5}$  |
| EDA         | 0.041                | $2.2 \text{ e}^{-4}$  |
| Respiration | 0.007                | $2.3 \text{ e}^{-10}$ |
| Temperature | 0.009                | $4.0 \text{ e}^{-4}$  |
| Fusion      | 0.003                | $4.9 \text{ e}^{-9}$  |

fusion up to 8.18%. The improvements are based primarily on the change of classification accuracy of mental underload (+1.52%).

For the LOSO classification the average classification accuracy improves for all channels: ECG by 2.15%, EMG by 2.59%, EDA by 4.07%, respiration by 4.20%, temperature by 2.18% and for the fusion by 2.60%. 47 participants improve their LOSO classification through fusion up to 12.84%. The increase of accuracy is based on the improvement of the recognition for both mental overload and underload. For both SELF and LOSO classification the improvements are the same for male and female.

To investigate if there is a statistical significant difference between the classification of all physiological signals if the participant speaks and did not speak a T-Test is used. The results are shown in Table 6. The results support that there is an difference between speaking and not speaking. For example at an  $\alpha = 0.05$ , all null hypothesis are rejected except the EMG for SELF classification. The highest support for an difference show the respiration channel with LOSO classification.

#### 4.4 Discussion

The investigation of the influence of speech artifacts while the classification of physiological signals shows that the classification accuracy improves both for SELF and LOSO classification if parts containing speech are removed from the physiological data pool. This could be based on the two influences described in Subsect. 4.3. The first could be due to movement of the subject which influences the physiological signal like noise. The second could have more impact on the change in classification results based on my observation. The reason for this is that the participants behave during and shortly after speaking with respect to their posture and activity differently between speaking and not speaking. This change in behaviour while speaking, however is noted in both sequences almost equally, which means that a similar behaviour in the OL and UL sequence is shown and therefore the meaningfulness of the physiological signals while speaking are not so depend on the emotional state as the participant does not

speak. The reason might be clear especially towards the end of the utterance that the participants are waiting for the response of the system, so if their answer was accepted and whether the answer was correct or incorrect.

For SELF classification shows that the MMV is 2.34% better than the fusion with pseudoinverse. A reason for this could be that the weights for pseudoinverse method are dependent on the trainings set for this fusion layer and the training set could show another kind of behaviour than the rest of the data. This could lead into questionable weights for both modalities. For the LOSO classification the pseudoinverse is 3.76% better than the fusion with MMV. The classification of the underload sequence is  $\approx 15\%$  better with the pseudoinverse. The reason for this could be the amount of data which could be used to train this method. Because about 8.3% of the data from 59 out of 60 participants could be used to train the bimodal fusion layer. The probability to use data to train the bimodal fusion layer that does not represent the remaining data to train the ensembles is shrinking compared to the SELF classification.

The LOSO accuracies for the speech data achieve a higher accuracy than the bimodal classification and the reason for this could be the high deviation of classification results from the physiological data. The LOSO classification accuracies for the physiological data reaches from 35.72% up to 89.62%. The reason for this could be the different behaviour patterns for the participants. The classification of participants which have the behaviour pattern of no different movement within the overload and underload sequence improves by 1.91%. The amount of participants which have more activity within the overload sequence are approximately double as frequent as participants which have more activation within the underload sequence. This results in the lower accuracy for the classification of participants with lower activity within the overload sequence for the physiological signals.

The bimodal classification results show that for the SELF classification average classification accuracy of 96.41% could be achieved and for the LOSO classification 76.31%.

**Acknowledgments.** The authors of this paper are partially funded by the Transregional Collaborative Research Centre SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

## References

1. Ekman, P.: An argument for basic emotions. *Cogn. Emot.* **6**(3–4), 169–200 (1992)
2. Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G.: Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Hum. Factors J. Hum. Factors Ergon. Soc.* **40**(1), 79–91 (1998)
3. Haapalainen, E., Kim, S., Forlizzi, J.F., Dey, A.K.: Psycho-physiological measures for assessing cognitive load. In: *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pp. 301–310. ACM (2010)
4. Kim, J.: Bimodal emotion recognition using speech and physiological changes. *Cite-seer* (2007)

5. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(12), 2067–2083 (2008)
6. Mayer, J.D., Geher, G.: Emotional intelligence and the identification of emotion. *Intelligence* **22**(2), 89–113 (1996)
7. Meudt, S., Zharkov, D., Kächele, M., Schwenker, F.: Multi classifier systems and forward backward feature selection algorithms to classify emotional coloured speech. In: *Proceedings of the 15th ACM on International conference on multi-modal interaction*, pp. 551–556. ACM (2013)
8. Pantic, M., Rothkrantz, L.J.: Toward an affect-sensitive multimodal human-computer interaction. *Proc. IEEE* **91**(9), 1370–1390 (2003)
9. Phinyomark, A., Limsakul, C., Phukpattaranont, P.: A novel feature extraction for robust emg pattern recognition. *arXiv preprint [arXiv:0912.3973](https://arxiv.org/abs/0912.3973)* (2009)
10. Picard, R.W.: *Affective Computing*, vol. 252. MIT Press, Cambridge (1997)
11. Scherer, K.R.: Speech and emotional states. In: *Speech Evaluation in Psychiatry*, pp. 189–220 (1981)
12. Schüssel, F., Honold, F., Bubalo, N., Huckauf, A., Traue, H., Hazer-Rau, D.: In-depth analysis of multimodal interaction: an explorative paradigm. In: Kurosu, M. (ed.) *HCI 2016. LNCS*, vol. 9732, pp. 233–240. Springer, Cham (2016). doi:[10.1007/978-3-319-39516-6\\_22](https://doi.org/10.1007/978-3-319-39516-6_22)
13. Tang, X., Shu, L.: Classification of electrocardiogram signals with rs and quantum neural networks. *Int. J. Multimedia Ubiquit. Eng.* **9**(2), 363–372 (2014)
14. Tkach, D., Huang, H., Kuiken, T.A.: Study of stability of time-domain features for electromyographic pattern recognition. *J. Neuroengineering Rehabil.* **7**(1), 1 (2010)
15. Zhao, Q., Zhang, L.: ECG feature extraction and classification using wavelet transform and support vector machines. In: *International Conference on Neural Networks and Brain, ICNN&B 2005*, vol. 2, pp. 1089–1092. IEEE (2005)

Multimodal Pattern Recognition of Social Signals in  
Human-Computer-Interaction  
4th IAPR TC 9 Workshop, MPRSS 2016, Cancun, Mexico,  
December 4, 2016, Revised Selected Papers  
Schwenker, F.; Scherer, S. (Eds.)  
2017, VIII, 161 p. 61 illus., Softcover  
ISBN: 978-3-319-59258-9