Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review

Mouhannad Ali, Ahmad Haj Mosa, Fadi Al Machot and Kyandoghere Kyamakya

Abstract Emotions play an extremely important role in how we make decisions, in planning, in reasoning, and in other human mental states. The recognition of a driver's emotions is becoming a vital task for advanced driver assistance systems (ADAS). Monitoring drivers' emotions while driving offers drivers important feedback that can be useful in preventing accidents. The importance comes from the fact that driving in aggressive moods on road leads to traffic accidents. Emotion recognition can be achieved by analyzing facial expression, speech, and various other biosignals such as electroencephalograph (EEG), blood volume pulse (BVP), electrodermal skin resistance (EDA), electrocardiogram (ECG), etc. In this chapter, a comprehensive review of the state of-the-art methodologies for emotion recognition based on physiological changes and speech is presented. In particular, we investigate the potential of physiological signals and driver's speech for emotion recognition and their requirements for ADAS. All steps of an automatic recognition system are explained: emotion elicitation, data preprocessing such as noise and artifacts removal, features extraction and selection, and finally classification.

1 Introduction

Human emotion is a state involving various physical structures; it is either gross or fine-grained behavior, and it occurs in particular situations [21]. The ability to understand and discern a driver's emotions while driving and to perform the

M. Ali (✉) · A.H. Mosa · F.A. Machot · K. Kyamakya
Institute of Smart System Technologies, Alpen-Adria University, Klagenfurt, Austria
e-mail: Mouhannad.Ali@aau.at

A.H. Mosa
e-mail: AhmadHaj.Mosa@aau.at

F.A. Machot
e-mail: FadiAl.Machot@aau.at

K. Kyamakya
e-mail: Kyandoghere.Kyamakya@aau.at

appropriate actions has been identified as one of the key focus areas listed by international research groups for improving intelligent transportation systems [4]. However, recognition of the emotional state and response during driving is an extremely difficult task and is still a scientific challenge. One of the main difficulties is that the emotion-relevant signal patterns may differ widely from person to person or from one specific situation to another. Moreover, it is hard to find an exact correlation between classes (patterns) due to the problem of precise definition of emotions and their meanings [22].

Nevertheless, the emotions and reaction(s) of a driver can be captured and measured using appropriate biosensors. Most researchers in the field of emotion recognition have focused on the analysis of data originating from a single sensor, such as audio (speech) or video (facial expression) data [9]. Lately, many studies in the emotion recognition field have begun to combine multiple-sensor data in order to build a robust emotion recognition system. The main target of using a combination of multiple sensors is that we as humans use a combination of different modalities in our body to express emotional states during human interaction. The human modalities are divided into audiovisual (facial expression, voice, gesture, posture, etc.) and physiological (respiration, skin temperature, etc.) [21].

Computers can be made to understand human emotions by capturing these modalities, extracting a set of useful features from them, and fusing those features in order to infer an accurate emotional state. There is a growing number of sensors that can capture various physical manifestations of emotion: video recordings of facial expressions [13], vocal inflection changes [2], EEG, skin-surface sensing of muscle tension, electrocardiogram (ECG), electrodermal activity (EDA), body temperature, etc.

In this chapter, an overview of the recent state of emotion recognition approaches involving speech signals and different physiological signals is presented. In particular, it is focused on emotion elicitation scenarios, features extraction, and selection and classification methodologies. The main goal of this review is to get an idea about the current situation (state of the art) of emotion recognition approaches and the current advancement in this filed.

This chapter is organized as follows: Sect. 2 describes the theories of emotion and the different categories of emotion types. Section 3 presents the physiological measures of human emotion recognition. Section 4 discusses the implementation steps of an emotion recognition system using physiological and speech signals. The overview of previous research work on emotion recognition using speech and physiological signals is presented in Sect. 5. Finally, a set of concluding remarks is given in Sect. 6.

2 Theories of Emotion

What is an emotion? "Everyone knows what an emotion is, until asked to give a definition" [14]. Emotion by definition is awareness of situations as relevant, urgent, and meaningful with respect to ways of dealing with it. According to cognitive

theory, people's experience of emotion depends on the way they appraise or evaluate the events around them [28]. For example, a person sees a snake. His brain starts to process the situation as a dangerous one, then his heart rate increases, and he then feels afraid. In general, emotion is a complex concept involving three components [40]:

- Subjective experience: There is a number of basic universal emotions [12] experienced by all humans regardless of culture and race. However, the way of experiencing these emotions is highly subjective [22].
- Emotion expressions: These are observable and nonverbal behaviors that illustrate an affective or internal emotion state. For example, a smile indicating happiness or pleasure and a frown indicating sadness or displeasure. In general, expressions include audiovisual such as face, gesture, posture, voice intonation, breathing noise.
- Physiological response: This is a biological arousal or a physical reaction the body experience during an emotion. For example, when we are frightened, our heart races, our breathing becomes rapid, our mouth becomes dry, our muscles tense, our palms become sweaty, and we may want to run [28].

Emotions can be categorized into various types. The two most frequently applied models for emotion classification are the "discrete emotion model" proposed by Ekman [12] and the "two-dimensional valance arousal model" proposed by Lang [27]. The discrete emotional model categorizes emotions into six basic emotions—happiness, sadness, surprise, anger, disgust, and fear [12]. These emotions are biologically fixed and universal to all humans. They are widely accepted. The dimensional model assumes that emotions are a combination of several psychological dimensions. The best-known dimensional model is the "valance arousal dimensional model." Valance represents the pleasure level and ranges from negative to positive. Arousal indicates the physiological and psychological level of being awake and ranges from low to high [23].

3 Physiological and Speech Signals

The general way to recognize the emotional state of a subject is through his speech, facial expression, or gesture. The speech signal can carry the emotional state of the speaker [31]. Williams and Stevens [43] found that when the sympathetic nervous system is aroused with the emotions of anger, fear, or joy, speech becomes loud, fast, and enunciated with strong high-frequency energy. Moreover, when a subject is sad, his parasympathetic nervous system is aroused and his speech becomes slow with high-frequency energy. According to the cited authors, emotion affects overall energy, energy distribution across the frequency spectrum, and the frequency and duration of pauses of speech signals.

However, detecting the physiological patterns of a subject can also give information about his emotional state, because when a subject is positively or negatively excited, the sympathetic nerves of the autonomic nervous system are activated [43].

This sympathetic activation increases respiration rate, raises heart rate, decreases heart rate variability, and raises blood pressure [41]. The most common physiological signals used for emotion recognition include:

- **Electromyography (EMG)**: This refers to the muscle activity or frequency of muscle tension of a certain muscle. EMG detects the electrical potential generated by muscle cells when these cells are electrically or neurologically activated [36]. High muscle tension often occurs under stress. It can also be measured on the face to distinguish between negative and positive emotions. Figure 1 shows an example of an EMG signal recorded at the smiling muscle (zygomaticus major muscle) during three consecutive muscle contractions.

- **Electrodermal activity (EDA)**: This refers to skin conductivity (SC); it basically measures the conductivity of the skin, which increases if the skin is sweaty. This signal was found to be a good and sensitive indicator of stress as well as other stimuli and also helps to differentiate between conflict and no-conflict situations or between anger and fear. The problem with this signal, however, is that it is also influenced by external factors such as outside temperature. It therefore needs reference measurements and calibration [16]. Figure 2 illustrates the skin conductance response (SCR) in an EDA signal that is occurring in reaction to a stimulus [20].
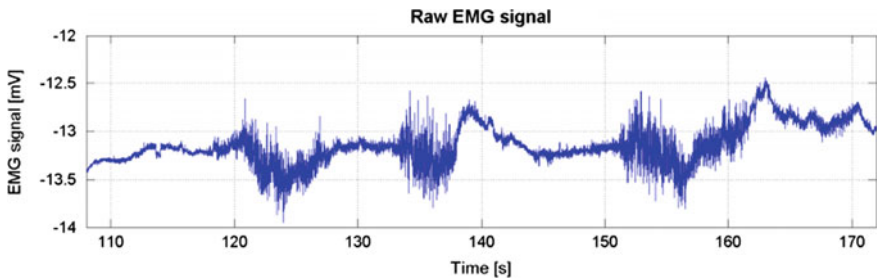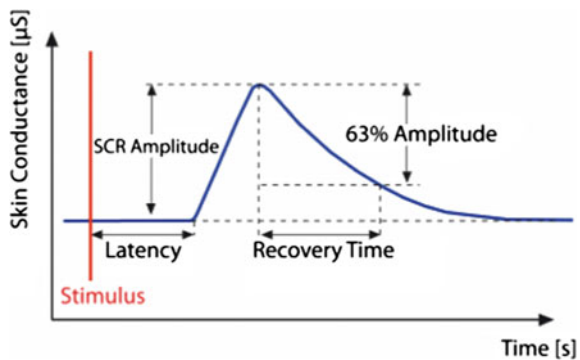

Fig. 1 Example of an EMG signal recorded at the zygomaticus major (smiling muscle), showing three muscle contractions [19]

Fig. 2 Ideal skin conductance response (SCR) in the EDA signal [20]

- **Skin temperature**: This is a measure of the peripheral skin temperature. The skin temperature depends on the blood flow in the underlying blood vessels. Since muscles are tense under strain, the blood vessels will be contracted, and therefore the temperature will decrease. In general, it is a relatively slow indicator of changes in emotional state [16]. Nevertheless, during happiness or anger, the temperature increases, and it decreases during sadness or fear.

- **Blood volume pulse (BVP)**: This is a measure to determine the amount of blood currently running though the vessels using a photoplethysmogram (PPG). A PPG consists of a light source and a photo sensor, which are attached to the skin. The source bounces infrared light against a skin surface and measures the amount of reflected light. BVP is used for emotion recognition, whereby the BV increases during anger or stress and decreases during sadness and relaxation. Moreover, BVP can be used to measure vasoconstriction and heart rate [16].

- **Electrocardiogram (ECG)**: Each healthy heartbeat has an orderly progression of depolarization that begins in the sinoatrial node, which generates an electrical impulse. This impulse spreads through the heart muscle and causes the contraction of the heart. The accumulation of action potentials traveling along the heart muscle generates electrical potential fluctuations. The electrical impulses generated by the heart can be measured on the surface of the skin over a period of time with electrodes [15]. This process of recording is called ECG. The ECG signal is a recurring pattern, as schematically depicted in Fig. 3. The ECG signal consists
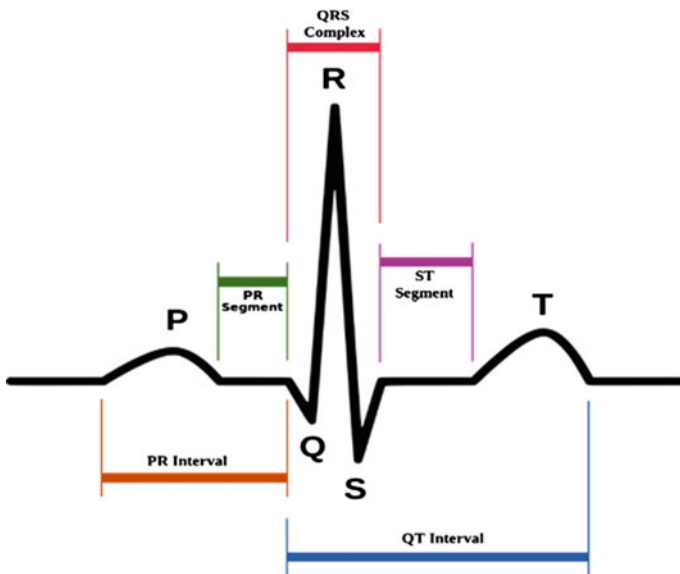

Fig. 3 Typical ECG signal with P, QRS, and T waves. [3]

of three main waves. The first is known as the P wave, which corresponds to the depolarization of the atrium. The is the QRS wave, which indicates the start of ventricular contraction. Finally, after the ventricles have contracted for a few milliseconds, the third wave, known as the T wave, occurs when the ventricular muscle repolarizes [15].

The R-peak represents the most prominent attribute of the ECG, and its time stamp can be precisely determined. It can be used to measure heart rate (HR) and interbeat intervals (IBI) to determine heart rate variability (HRV). A low HRV can indicate a state of relaxation, whereas an increased HRV can indicate a potential state of mental stress or frustration [19].

- **Electroencephalogram (EEG)**: An electroencephalography signal is the measurement of brain waves. It can be used to evaluate brain disorders. The brain waves are generated by current flow during synaptic excitation of the dendrites of many pyramidal neurons in the cerebral cortex [38]. The EEG signals are measured using small, flat metal disks (electrodes) attached to the scalp. There are five major brain waves distinguished by their different frequency ranges. These frequency bands from low to high frequencies are called respectively [38]:

1. Delta (δ) waves, which lie within the range of 0.5–4 Hz. They are usually present during deep sleep and may appear in the waking state.
2. Theta (θ) waves, which lie within the range of 4–7.5 Hz. They are usually present during drowsiness and are associated with increased learning, creativity, and deep
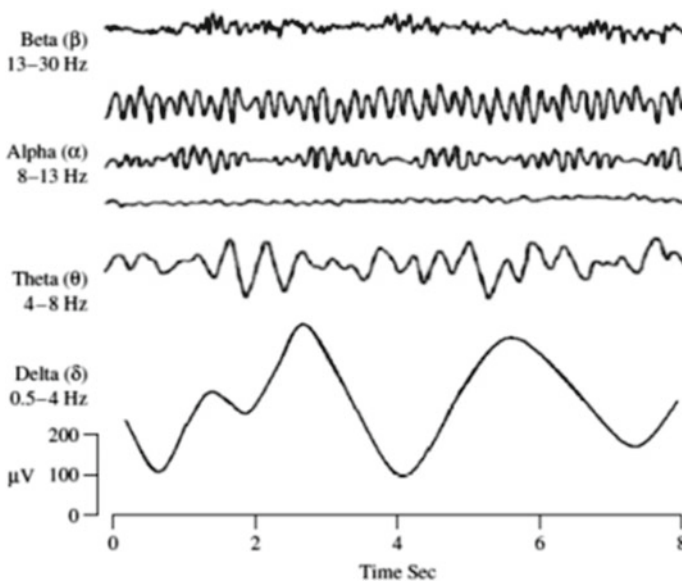

Fig. 4 Four typical brain waves, from high to low frequencies. [38]

meditation. They allow access to the unconscious. Moreover, a theta wave seems to be related to the level of arousal.

3. Alpha (α) waves lie within the range of 8–13 Hz. In general, an alpha wave appears as a round or sinusoidal signal; they are usually associated with relaxation and super-learning.
4. Beta (β) waves lie within the range of 14–26 Hz. They are usually associated with active thinking, active attention, or solving concrete problems.
5. Gamma (γ) waves correspond to the frequencies above 30 Hz. The detection of these waves can be used for confirmation of certain brain diseases.

Figure 4 illustrates the first four brain waveforms with their usual amplitude levels.

- **Respiration**: this measurement indicates the breathing rhythm of a person. It is captured be applying a rubber band around the chest. Usually, fast and deep breathing can point to anger or fear but sometimes also joy. Rapid shallow breathing can indicate fear, panic, or concentration. Moreover, slow and deep breathing refers to relaxation, and slow and shallow breathing can indicate depression or calm happiness [16].

The combination of these signals can be used to derive a set of features that can be used to build a robust classifier. This is then used to automatically detect the emotional state of different subjects.

4 The Implementation Steps of an Emotion Recognition System Based on Physiological and Speech Signals

4.1 Emotion Elicitation

The creation of a high-quality dataset (i.e., a reference database) of speech and biological signals still represents a necessary task for researchers in the field of emotion recognition. Different scenarios have been used to elicit emotion. Current rsearch highlights seven criteria that help with the selection and use of these emotion elicitation scenarios [7]:

1. Intensity: Do the elicitation scenarios lead to intense negative and positive emotion?
2. Complexity: Is the scenario a simple one such as showing a silent image such as a fixation cross, or is it more complex, such as a dynamic visual and auditory sequence?
3. Attention capture: Do the scenarios require much attention?
4. Demand characteristics: Does the scenario include a specific instruction such as "please watch this video carefully." In general, it is content-dependent?

5. Standardization: Is the scenario going to affect all participants in the same way? For example, is it possible to guarantee that showing a specific video will be equally effective for all participants?
6. Temporal consideration: Emotions can be considered relatively rapid phenomena with onsets and offsets over seconds. Generally, film clips are much lower in temporal resolution and range from about 1 to 10 min.
7. Ecological validity: To what extent will emotion elicitation procedures elicit emotions in the way that many stimuli encountered in real life do?

Moreover, the most-used scenarios for emotion elicitation are the following [7]:

Film clips: This scenario works by showing many short film clips to participants. The advantage of this method is that it is a rich source of discrete emotions (love, anger, fear, joy, etc.), which can be self-reported by participants. On the other hand, the disadvantages of this method are that it is necessary to extract particular periods of interest from the film. Furthermore, due the fact that emotions are considered evanescent phenomena, any delay between the activation of emotion and the assessment of it by an experimenter can introduce an error in the measurement.

Pictures: This scenario works by showing a set of pictures to participants. The advantages of this method are that it is easy and fast to apply and can also be self-reported by participants. However, the disadvantages of the method are that it is not a rich source of discrete emotions compared to film clips, and the time for stimulating emotion is too short.

Music: This scenario applies by playing music to participants. The advantages are that it is simple and highly standardized, and emotions develop over time (15–20 min). The drawbacks are that the music tastes of the participant might influence experienced emotions. In addition, this scenario gives only the moods (positive or negative), not the discrete emotions.

Emotional behaviors as emotional stimuli: This scenario involves the manipulation of the target person's behavior or the person's understanding of that behavior in order to change his/her feelings. The advantage of this method consists in the large sources to produce emotions (posture, eye gaze, tone of voice, breathing, and emotional actions). On the other hand, the disadvantage is that in some cases, it would be easy to manipulate the subject, but others might be more difficult (for example, making the participant angry).

Dyadic interaction tasks: Emotion is here elicited through interaction with different types of dyads (friends, romantic partner, family member, etc.). The advantages of this method are that it elicits a range of emotional responses and it studies emotion in social contexts. The disadvantages of this method are that (1) it requires significant resources, for example, dyadic interaction procedures can take 2–4 h; (2) some procedures my not be completed (the participant changes the topic to avoid a high level of emotional intensity; finally, (3) it provides just a snapshot sampling of emotions.

In general, choosing the appropriate elicitation scenario or stimuli depends on the target emotions and available sensors. For example, if we need to extract a speech signal from a subject, then music and picture scenarios will not be useful. The useful scenarios in this case are "emotional behaviors as emotional stimuli" and "dyadic interaction" tasks. Moreover, if we want to extract discrete emotions from a subject, we cannot choose the music scenario, because it just induces moods (positive or negative).

4.2 Preprocessing of Involved Signals

Both speech and physiological signals always contain unwanted modifications during capture, processing, or transmission. These unwanted modifications are noises and other external interferences such as artifacts that appear because of the electrostatic devices and muscular movements [24]. These noises and artifacts should be removed from the signal by the use of different types of appropriate filtering techniques. The appropriate filter is defined according to the signal's type and noise patterns. Low-pass filters such as adaptive filters, elliptic filters, and Butterworth filters generally are used to preprocess the raw ECG and facial EMG signals. And smoothing filters are used to preprocess the raw GSR signals [5, 21]. Also, low- and high-pass filters are used to preprocess EEG signals. Moreover, different methods have been used to remove artifacts from physiological signals. Principal component analysis (PCA) [37] and independent component analysis (ICA) [11] are the best-known methods for removing artifacts from EEG, ECG, and EMG.

Moving to speech signals, preprocessing including preemphasis, framing, and windowing processes have to be integrated [25].

4.3 Feature Extraction

Once the signals have been preprocessed, it is necessary to extract useful information or features from these signals in order to use them in pattern classification to detect the emotional state. The features to be extracted are chosen according to the signal type. Some features (such as mean, standard deviation, minimum, maximum, and range) can be extracted from most of the recorded sensor signals, whereas some special features are extracted only from a specific signal type.

From the speech signal: the best-known speech features usually extracted for emotion recognition include prosodic and spectral features. Prosodic features include pitch, pitch histogram, intensity, formant frequency, and voice quality [25]. Spectral features include Mel-frequency cepstral coefficients (MFCC), Daubechies wavelets, coefficient histogram [44], linear prediction cepstral coefficients (LPC), log fre-

Table 1 ECG features

Feature	Domain
Standard deviation of all RR interval	Time
Root mean Square of differences between RR intervals	Time
Power in low frequency	Frequency
Power in high frequency	Frequency

quency power coefficients (LFPC), and perceptual linear prediction (PLP) coefficients [44].

From the electrocardiogram signal: The most-used ECG features are heart rate (HR) and heart rate variability (HRV). Moreover, [1] used the Hilbert instantaneous frequency and a measure of local oscillation as feature extraction. Table 1 lists different ECG features in the frequency and time domains.

From the electrodermal activity signal: The extracted features from EDA are the average, skin resistance, zero crossing rate of skin conductance, average of absolute derivative, skin conductance response (SCR), and the nonspecific skin conductance response [21].

From the respiration signal: Respiration rate, average, and breathing rhythm are the best-known respiration features. Moreover, the average breath depth and spectral power are also commonly used [19, 21].

From the electromyogram signal: The most frequently extracted features from the EMG are mean value, root mean square, and the power [19, 21].

From the electroencephalogram EEG signal: EEG power spectra at distinct frequency bands, such as delta, theta, alpha, beta, and gamma, are commonly used as indices for assessing the correlates of specific ongoing cognitive processes in EEG research [25, 38], fast Fourier transform analysis, wavelet analysis, and high-order crossing [33].

4.4 Feature Reduction

After the features are extracted from the signals, it is useful to determine which features are most relevant to differentiate well between emotional states. Reducing the dimension of the feature space has two advantages:

- The computational costs are lowered, which leads also to shorter training times [18].
- Overfitting is reduced and prediction performance is improved by excluding irrelevant or noisy features in the learning process [46].

The most used methods for selection of features are principal component analysis (PCA) [37], independent component analysis (ICA) [11], random subset feature selection (RSFS) [35], and sequential floating forward selection (SFFS) [34].

4.5 Classification

After extracting and selecting the features that are appropriate for a best possible differentiation of emotional states, the next step is to use these features to train a classifier and test whether it can classify different emotional states. This section describes the most popular classification algorithms from the literature. We have selected the following classifiers:

4.5.1 Quadratic Discriminant Analysis (QDA)

Quadratic discriminant analysis (QDA) is the most commonly used method. It assumes that the likelihood of each class is normally distributed and uses the posterior distributions to estimate the class for a given test point [17]. The normal (Gaussian) parameters of each class are usually estimated from training points with maximum likelihood (ML) estimation [39].

4.5.2 k-Nearest Neighbor (KNN)

KNN classifies unlabeled samples (testing data) by their similarity with the training data. In general, given an unlabeled sample X, the KNN classifier finds the K closest neighborhood samples in the training data, and it labels the sample X with the class label that appears most frequently in the K closet neighborhood samples in the training data [32].

4.5.3 Support Vector Machines (SVMs)

The SVM is a classifier that separates a set of objects into classes so that the distance between the class borders is as large as possible. The idea of SVM is to separate two classes with a hyperplane so that the minimal distance between elements of both classes and the hyperplane is maximal [8].

4.5.4 Artificial Neural Network

Artificial neural networks are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown [45]. Artificial neural

networks are a computational model that uses the idea of natural neurons that receive signals and exchange messages between each other. Each connection between neurons (nodes) has a numeric weight that can be tuned based on experience, thus making the network capable of learning. In general, ANNs start out with randomized weights for all their neurons. For that, they must be trained to solve a particular problem. The training phase can be applied using two methods, depending on the problem the ANNs must solve. The first method is self-organizing ANN. It uses a large amount of data and tries to discover patterns and relationships in those data. The second method is back-propagation ANN, which is trained by humans to perform specific tasks [6, 47].

5   Previous Works

Emotion recognition has become an important research topic mainly in the field of human–machine interaction. Several studies have been done on emotion recognition using speech and physiological signals. Initially, researchers tried with subject-dependent approaches, where the emotion recognition system is performed only on one user, and it needs to be retrained or recalibrated in order to perform well on another user/subject. Nowadays, the focus has shifted in the direction of subject-independent approaches, where the emotion recognition system is tested on fully different subjects and not those on which it had been trained; i.e., it is tested with unknown speech and physiological signals. Table 2 illustrates a short review of previous works in emotion recognition using speech and physiological signals. The table shows which signals were analyzed, which stimuli were used for emotion elicitation, which emotion were recognized, the number of subjects involved in the experiments, and which features and classification method were applied. Also, the accuracy of the recognition approaches is included in the table.

We can observe that regarding the subject-dependent approaches, the maximum accuracy values reached are 96.58% for recognizing three arousal levels (high, medium, and low), 95% for four emotions (joy, anger, sorrow, and pleasure) and 91.7% for six emotions (amusement, frustration, anger, fear, sadness, and surprise). On the other hand, for subject-independent approaches, the maximum accuracies were 99.5% for recognizing one emotion (stress), 86% for two emotions (joy and sadness), and 70% for detecting four emotional states (joy, anger, sadness, and pleasure). We can also notice, for physiological signals, that besides the feature extraction and classification approaches, the emotion stimulus type involved also has an effect on the classification accuracy. In general, the sensors used, number of subjects, emotional states, stimuli used, feature extraction and classification methods are the required building blocks and parameters to build a robust and reliable emotion recognition system.

Table 2 Literature review on emotion recognition using physiological and speech signals

Ref. No	Signals	Features	Classifiers	Emotions	Stimuli	No of subjects	Accuracy in %
[23]	EMG ECG EDA RSP	Statistical, energy, subband spectrum, entropy	Linear discriminant analysis	Joy, anger, sadness, pleasure	Music	3, MIT-database	95 (sub dependent) 70 (sub independent)
[29]	GSR HR ST	No specific features stated	KNN, discriminant function analysis, Marquardt backpropagation	Sadness, anger, fear, surprise, frustration, amusement	Movies	14	91.7 (sub dependent)
[16]	EMG EDA BVP ECG RSP	Running mean running standard deviation slope	NN	Arousal, valance	IAPS (visual affective picture system)	1	96.58 Arousal 89.93 valence (sub dependent)
[42]	ECG	Fast fourier	Tabu search	Joy, sadness	Movies	154	86 (sub independent)
[10]	EDA HR	No specific features stated	fuzzy logic	Stress	Hyperventilation Talk preparation	80	99.5 (sub independent)
[30]	BVP EMG ST EDA RSP	Statistical features	SVM, Fisher LDA	Amusement, contentment, disgust, fear, sad, neutral	IAPS	10	90 (sub dependent) 92 (sub dependent)
[22]	EMG EDA ECG BVP ST RSP SPEECH	Statistical features, BRV, zero-crossing, MFCCs	KNN	Arousal, valance	Quiz dataset	3	92 (sub dependent) 55 (sub independent)
[26]	EDA HR EMG	No specific features stated	HMM	Arousal, valance	Robot actions	36	81 (sub dependent) 66 (sub independent)

6 Conclusion

This chapter has reviewed and presented the main steps toward human emotion recognition systems using physiological and speech signals. It should be emphasized that building a generalized system (i.e., subject-independent) for classifying different emotional states is still a big challenge, particularly because emotions are highly subjective. Most of the state-of-the art methodologies are based on the subject-dependent approach. Upgrading to a subject-independent approach needs more sophisticated features, more robust classifiers, and eventually more sensory data for training and testing. For good referencing, the collection of sensor data is done during a specific emotion elicitation scenario. Hence, the emotion elicitation scenario plays an important role in defining the target emotional states and how strongly those emotions should be elicited.

Moreover, the respective techniques for feature extraction, feature selection and classification are also very important steps and building blocks towards a robust and reliable subject-independent emotion recognition system.

Some future research avenues that are worth mentioning are related to:

- Emotion state forecasting for short-, middle-, and long-term horizons. Here the human emotional system is considered a dynamical system that is externally excited by emotion elicitation-related elements of the contextual environment. The short-term time horizon covers some seconds to several minutes. The middle-term horizon should cover hours. And the long-term horizon should cover several days. It is evident that a reliable forecasting of the emotional states is a core enabling unit for some form of emotion-related early warning system.
- To improve robustness, reliability, and accuracy, our hypothesis is that neurocomputing-based classifier concepts offer the greatest potential for best-possible performance. A series of our won ongoing works is developing, optimizing, and benchmarking cellular neural network-based neurocomputing classifier concepts, which integrate related recent paradigms such as deep learning, reservoir computing, and echo state.

References

1. Agrafioti, F., Hatzinakos, D., Anderson, A.K.: ECG pattern analysis for emotion detection. IEEE Trans. Affect. Comput. **3**(1), 102–115 (2012)
2. Al Machot, F., Mosa, A.H., Fasih, A., Schwarzlmüller, C., Ali, M., Kyamakya K.: A novel real-time emotion detection system for advanced driver assistance systems. In: Autonomous Systems: Developments and Trends, pp. 267–276. Springer, Berlin (2012)
3. Atkielski, A.: Schematic diagram of normal sinus rhythm for a human heart as seen on ECG. http://commons.wikimedia.org/wiki/File:SinusRhythmLabels.png/ (2006). Accessed 17-March 2016
4. Burns, P.C., Lansdown, T.C.: E-distraction: the challenges for safe and usable internet services in vehicles. In: Internet Forum on the Safety Impact of Driver Distraction When Using In-Vehicle Technologies (2000)

5. Chang, C.-Y., Zheng, J.-Y., Wang, C.-J.: Based on support vector regression for emotion recognition using physiological signals. In: The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2010)

6. Cigizoglu, H.K., Kişi, Ö.: Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. Hydro. Res. **36**(1), 49–64 (2005)

7. Coan, J.A., Allen, J.J.B.: Handbook of Emotion Elicitation and Assessment. Oxford University Press, Oxford (2007)

8. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)

9. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human-computer interaction. IEEE Signal Process. Mag. **18**(1), 32–80 (2001)

10. de Santos Sierra, A., Sánchez Ávila, C., Casanova, J.G., Del Pozo, G.B.: A stress-detection system based on physiological signals and fuzzy logic. IEEE Trans. Ind. Electron. **58**(10), 4857–4865 (2011)

11. Ekenel, H.K., Sankur, B.: Pattern Recognit. Lett. Feature selection in the independent component subspace for face recognition. **25**(12), 1377–1388 (2004)

12. Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., et al.: Universals and cultural differences in the judgments of facial expressions of emotion. J. Personal. Soc. Psychol. **53**(4), 712 (1987)

13. Essa, I.A., Pentland, A.: A vision system for observing and extracting facial action parameters. In: Proceedings CVPR'94. 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994, pp. 76–83. IEEE (1994)

14. Fehr, B., Russell, J.A.: Concept of emotion viewed from a prototype perspective. J. Exp. Psychol.: General **113**(3), 464 (1984)

15. Fox, S.I.: Human Physiology 9th edn. McGraw Hill, New York (1996)

16. Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion recognition using bio-sensors: first steps towards an automatic system. In: ADS, pp. 36–48. Springer, Berlin (2004)

17. Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., Tibshirani, R.: The Elements of Statistical Learning, vol. 2. Springer, Berlin (2009)

18. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning, vol. 112. Springer, Berlin (2013)

19. Kappeler-Setz, C.: Multimodal emotion and stress recognition. Dissertation, Eidgenössische Technische Hochschule ETH Zürich, Nr. 20086 (2012)

20. Kappeler-Setz, C., Gravenhorst, F., Schumm, J., Arnrich, B., Tröster, G.: Towards long term monitoring of electrodermal activity in daily life. Pers. Ubiquitous Comput. **17**(2), 261–271 (2013)

21. Katsis, C.D., Katertsidis, N., Ganiatsas, G., Fotiadis, D.I.: Toward emotion recognition in car-racing drivers: a biosignal processing approach. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **38**(3), 502–512 (2008)

22. Kim, J.: Bimodal emotion recognition using speech and physiological changes. Citeseer (2007)

23. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. **30**(12), 2067–2083 (2008)

24. Kim, K.H., Bang, S.W., Kim, S.R.: Emotion recognition system using short-term monitoring of physiological signals. Med. Biol. Eng. Comput. **42**(3), 419–427 (2004)

25. Konar, A., Chakraborty, A.: Emotion Recognition: A Pattern Analysis Approach. Wiley, New York (2014)

26. Kulic, D., Croft, E.A.: Affective state estimation for human–robot interaction. IEEE Trans. Robot. **23**(5), 991–1000 (2007)

27. Lang, P.J.: The emotion probe: studies of motivation and attention. Am. Psychol. **50**(5), 372 (1995)

28. Lazarus, R.S.: Progress on a cognitive-motivational-relational theory of emotion. Am. Psychol. **46**(8), 819 (1991)

29. Lisetti, C.L., Nasoz, F.: EURASIP J. Adv. Signal Process. Using noninvasive wearable computers to recognize human emotions from physiological signals. **2004**(11), 1–16 (2004)
30. Maaoui, C., Pruski, A.: Emotion recognition through physiological signals for human-machine communication. INTECH Open Access Publisher (2010)
31. Nwe, T.L., Foo, S.W., De Silva, L.C.: Speech emotion recognition using hidden markov models. Speech Commun. **41**(4), 603–623 (2003)
32. Peterson, L.E: K-nearest neighbor. Scholarpedia **4**(2), 1883 (2009)
33. Petrantonakis, P.C., Hadjileontiadis, L.J.: Emotion recognition from EEG using higher order crossings. IEEE Trans. Inf. Technol. Biomed. **14**(2), 186–197 (2010)
34. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recognit. Lett. **15**(11), 1119–1125 (1994)
35. Räsänen, O., Pohjalainen, J.: Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech. In: INTERSPEECH, pp. 210–214 (2013)
36. Robertson, G., Caldwell, G., Hamill, J., Kamen, G., Whittlesey, S.: Research Methods in Biomechanics, vol. 2E. Human Kinetics, Champaign (2013)
37. Rocchi, L., Chiari, L., Cappello, A.: Feature selection of stabilometric parameters based on principal component analysis. Med. Biol. Eng. Comput. **42**(1), 71–79 (2004)
38. Sanei, S., Chambers, J.A.: EEG Signal Processing. Wiley, New York (2013)
39. Scholz, F.W.: Maximum likelihood estimation. Encyclopedia of Statistical Sciences (1985)
40. Thoits, P.A.: The sociology of emotions. Annu. Rev. Soc. **15**, 317–342 (1989)
41. Van Der Vloed, G., Berentsen, J.: Measuring emotional wellbeing with a non-intrusive bed sensor. In: Human-Computer Interaction–INTERACT 2009, pp. 908–911. Springer, Berlin (2009)
42. Wan-Hui, W., Yu-Hui, Q., Guang-Yuan, L.: Electrocardiography recording, feature extraction and classification for emotion recognition. In: 2009 WRI World Congress on Computer Science and Information Engineering, vol. 4, pp. 168–172. IEEE (2009)
43. Williams, C.E., Stevens, K.N.: Vocal correlates of emotional states. Speech Evaluation in Psychiatry, pp. 221–240 (1981)
44. Yang, Y.-H., Chen, H.H.: Music Emotion Recognition. CRC Press, Boca Raton (2011)
45. Yegnanarayana, B.: Artificial Neural Networks. PHI Learning Pvt. Ltd., New Delhi (2009)
46. Yoon, H.J., Chung, S.Y.: EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. Comput. Biol. Med. **43**(12), 2230–2237 (2013)
47. Zhang, J.-R., Zhang, J., Lok, T.-M., Lyu, M.R.: A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. Appl. Math. Comput. **185**(2), 1026–1037 (2007)