1

INTRODUCTION TO EMOTION RECOGNITION

AMIT KONAR AND ANISHA HALDER

Artificial Intelligence Laboratory, Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India

ARUNA CHAKRABORTY

Department of Computer Science & Engineering, St. Thomas' College of Engineering & Technology, Kolkata, India

A pattern represents a characteristic set of attributes of an object by which it can be distinguished from other objects. Pattern recognition aims at recognizing an object by its characteristic attributes. This chapter examines emotion recognition in the settings of pattern recognition problems. It begins with an overview of the well-known pattern recognition techniques, and gradually demonstrates the scope of their applications in emotion recognition with special emphasis on feature extraction, feature reduction, and classification. Main emphasis is given to feature selection by single and multiple modalities and classification by neural, fuzzy, and statistical pattern recognition techniques. The chapter also provides an overview of stimulus generation for arousal of emotion. Lastly, the chapter outlines the methods of performance analysis and validation issues in the context of emotion recognition.

1.1 BASICS OF PATTERN RECOGNITION

A pattern is a representative signature of an object by which we can recognize it easily. Pattern recognition refers to mapping of a set of patterns into one of several object classes. Occasionally, a pattern is represented by a vector containing the features of
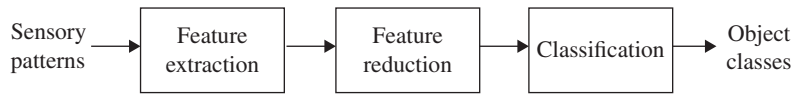
FIGURE 1.1 Basic steps of pattern recognition.

an object. Thus, in general, the pattern recognition process can be described by three fundamental steps, namely, feature extraction, feature selection, and classification. Figure 1.1 provides a general scheme for pattern recognition. The feature extraction process involves using one or more sensors to measure the representative features of an object. The feature selection module selects more fundamental features from a list of features. The classification module classifies the selected features into one of several object classes.

The pattern recognition problem can be broadly divided into two main heads: (i) supervised classification (or discrimination), and (ii) unsupervised clustering. In supervised classification, usually a set of training instances (or data points) comprising a set of measurements about each object along with its class is given. These data points with their class labels are used as exemplars in the classifier design. Given a data point with unknown class, the classifier once trained with the exemplary instances is able to determine the class label of the given data point. The classifier thus automatically maps an unknown data point to one of several classes using the background knowledge about the exemplary instances.

Beginners to the subject often are confronted with the question: how does the classifier automatically determine the class label of an unknown data point, which is not present in the exemplary instances. This is due to the inherent generalization characteristics of the supervised classifier.

In unsupervised classification, the class labels of the data points are not known. The learning system partitions the whole set of data points into (preferably) nonoverlapping subsets based on some measure of similarity of the data points under each subset. Each subset is called a class/cluster. Because of its inherent characteristics of grouping data points into clusters, unsupervised classification is also called clustering.

Both statistical decision theory and machine learning have been employed in the literature to design pattern recognition algorithms [1, 2]. Bayes' theorem is the building stone of statistical classification algorithms. On the other hand there exists a vast literature on supervised and unsupervised learning [3], algorithms, which capture the inherent structural similarity [4] of the data points for application in pattern recognition problems.

1.2 EMOTION DETECTION AS A PATTERN RECOGNITION PROBLEM

Emotion represents the psychological state of the human mind and thought processes. Apparently, the process of arousal of emotion has a good resemblance with its manifestation as facial, vocal, and bodily gestures. This phenomenon has attracted researchers to determine the emotion of a subject from its manifestation. Although the

one-to-one correspondence from manifestation of emotion to a particular emotional state is yet to be proved, researchers presume the existence of such mapping to recognize the emotion of a subject from its manifestation.

Given the manifestation of an emotion, the task of recognizing the emotion, thus, is a pattern recognition problem. For example, facial expression–based emotion recognition requires extraction of a set of facial features from the facial expression of a given subject. Recognition of emotion here refers to classification of facial features into one of several emotion classes. Usually, a supervised classifier pretrained with emotional features as input and emotion class as output is used to determine the class of an unknown emotional manifestation.

Apparently, the emotional state of the human mind is expressed in different modes including facial, voice, gesture, posture, and biopotential signals. When a single mode of manifestation is used to recognize emotion, we call it a unimodal approach. Sometimes all modes are not sufficiently expressed. Naturally, recognition from a less expressed mode invites the scope of misclassification. This problem can be avoided by attempting to recognize an emotion from several modalities. Such a process is often referred to as multimodal emotion recognition.

1.3   FEATURE EXTRACTION

Feature extraction is one of the fundamental steps in emotion recognition. Features are obtained in different ways. On occasion features are preprocessed sensory readings. Preprocessing is required to filter noise from measurements. Sensory readings during the period of emotion arousal sometimes have a wide variance. Statistical estimates of the temporal readings, such as mean, variance, skewness, kurtosis, and the like, are usually taken to reduce the effect of temporal variations on measurements. Further, instead of directly using time/spatial domain measurements, frequency domain transforms are also used to extract frequency domain features. For example, frequency domain information is generally used for EEG (electroencephalogram) and voice signals. Frequency domain parameters are time invariant and less susceptible to noise. This attracted researchers to use frequency domain features instead of time domain.

Frequency domain features have one fundamental limitation in that they are unable to tag time with frequency components. Tagging the time with frequency contents of a signal is important, particularly for a certain class of signals, often labeled as nonstationary signals. EEG, for instance, is a nonstationary signal, the frequency contents of which change over time because of asynchronous firing of the neurons. Wavelet transform coefficients of an EEG signal represent time–frequency correlations and thus deserve to be one of the fundamental features for nonstationary EEG signals. We now briefly outline the features used in different modalities of emotion recognition.

1.3.1   Facial Expression–Based Features

The most common modality of emotion recognition is by facial expression analysis. Traditionally there exist two major classes of techniques for face/facial expression

representation and relevant feature extraction. The first one is called geometrical features. They rely on parameters of distinctive facial features such as eyes, mouth, and nose. On the other hand appearance-based approach considers a face as an array of intensity values suitably preprocessed. This array is then compared with a face template using a suitable matrix.

1.3.1.1   Geometric Model–Based Feature Extraction

Deformable templates have been used for locating facial features. For example, Kass *et al.* [5] suggested the use of active contour models—snakes—for tracking lips in image sequence. They initialized snake on the lips in the facial image and showed that it is able to track lip movement accurately. It however fails, if there exists occlusion or other structure in the image.

Yuille *et al.* [6] employed deformable templates based on simple geometrical shapes for locating the eye and mouth. Yuille's model incorporated shape constraints, but there is no proof that the form of a given model is sufficiently general to capture the deformable geometric shapes.

Researchers are taking a keen interest to represent geometric relations between facial information to extract facial features. In Reference 7, Craw, Tock, and Bennett, considered positional constraints in facial expressions to extract necessary features for emotion recognition.

Brunelli and Poggio considered a number of high dimensional [8] measurements or location of a number of key points in a single image or an image sequence for facial image interpretation.

Kirby and Sirovich [9] took attempts to decompose facial image into a weighted sum of basis images or eigen faces using Karhunen and Loeve expansion. They considered 50 expansion coefficients and were able to reconstruct an approximation of the facial expansion using these parameters.

1.3.1.2   Appearance-Based Approach to Feature Extraction

Appearance-based approach involves preprocessing followed by a compact coding through statistical redundancy reduction. The preprocessing in most cases is required to align the geometry in face image, for instance, by having the two eyes and nose tip at fixed positions through affine texture warping [10]. Optical flow or Gabor wavelets are used to capture facial appearance motion and robust registration, respectively, for successful recognition.

Pixel-based appearance is often represented by a compact coding. Usually statistical reduction principle is used to represent this coding. The unsupervised learning techniques used for compact coding include Principal Component Analysis (PCA), Independent Component Analysis (ICA), Kernel-PCA (KPCA), local feature analysis, and probability density estimation. Supervised learning techniques including Linear Discriminant Analysis (LDA) and Kernel Discriminant Analysis (KDA) are also used for compact coding representation.

The main drawback of PCA-based compact coding is that it retains some unwanted variations. It is also incapable of extracting local features that offer robustness against changes in local region or occlusions. ICA produces basis vectors that are more spatially local than those of PCA. Thus, ICA is sensitive to occlusion and pose

variations. ICA retains higher order statistics and maximizes the degree of statistical independence of features [11, 12].

Recently Scholkopf *et al.* [13] extended the conventional PCA to KPCA, which is able to extract nonlinear features [14]. However, like PCA, KPCA captures the overall variance of 11 patterns and is not necessarily optimal for discrimination.

Statistical supervised learning such as LDA attempts to find the basis vectors maximizing the interclass distance and minimizing the intraclass distance. Similarly, KDA determines the most significant nonlinear basis vector to maximize the interclass distance while minimizing the intraclass distance. Among the other interesting works, the following need special mentioning.

Cohn *et al.* [15] proposed a facial action recognition technique that employs discriminant function analysis of individual facial regions, including eyebrows, eyes, and mouth. They used two discriminant functions for three facial actions of the eyebrow region, two discriminant functions for three facial actions of the eye region, and five discriminant functions for nine facial actions of the nose and mouth region. The classification accuracy for the eyebrow, eye and nose, mouth regions are 92, 88, and 88%, respectively.

In Reference 16 Essa and Pentland proposed a novel control-theoretic method to extract the spatiotemporal motion–energy representation of facial motion in an observed expression. They generated the spatiotemporal templates for six different facial expressions, considering two facial actions, including smile and raised eyebrows for two subjects. Templates are formed by averaging the patterns of motion generated by two subjects exhibiting a certain expression. The Euclidean norm of the difference between the motion–energy template and the observed motion energy is defined as a metric of similarity of the motion energies. A recognition accuracy of 98% is achieved while experimenting with 52 frontal-view image sequences of eight people having distinct expressions.

Kimura and Yachida in Reference 17 modeled facial images by a potential net and attempted to fit the net to each frame of a facial image sequence. The deformed version of the potential net is used to match the expressionless face, typically the first frame of the sequence. The variation in the nodes of the deformed net is used for subsequent processing. In their own experiments, the authors in Reference 17 considered a six image sequence of emotional expressions experienced by a subject with gradual variation in the strength of expressions from expressionless (relaxed) to a maximum expression. The experiment was repeated for three emotions: anger, happiness, and surprise. PCA has been employed here to classify three emotions using standard eigen space analysis.

In Reference 18 Lucey *et al.* detected pain from the movement of facial muscles into a series of action units (AUs), based on the Facial Action Coding System (FACS) [19]. For this novel task, they considered three types of Active Appearance Model (AAM) features: (i)similarity-normalized shape features (SPTS), (ii) similarity-normalized appearance features (SAPP), and (iii) canonical-normalized appearance features (CAPP). AAM features are used here to track the face and to extract visual features, based on facial expressions using the FACS. They obtained classification accuracy of 75.1, 76.9, and 80.9% using SAPP, SPTS, and CAPP,

respectively, using Support Vector Machine (SVM) classifier first and then improving the performance (Fusion of Scores) by linear logistical regression (LLR).

Tian *et al.* in Reference 20 proposed a new method for recognizing AUs for facial expression analysis. They used both permanent and transient features for their work. Movement of eyebrow, cheek, eyes, and mouth are considered permanent features. On the other hand, deeping of facial furrows is considered a transient features. They used different feature extraction algorithms for different features. For the lips, they used lip-tracking algorithm. For eyes, eyebrows, and cheeks, they considered Lucas–Kanade algorithm, and for the transient features they employed Canny edge detector algorithm. Two neural network (NN) based classifiers are considered to recognize the changes in AUs: one for six upper-face AUs and the other for ten lower-face AUs of the FACS. A percentage accuracy of 95.4% is obtained for upper-face AUs and 95.6% for lower-face AUs.

In Reference 21 Kim and Bien designed a personalized classifier from facial expressions using soft computing techniques. They used degree of mouth openness (f_1), degree of eye openness (f_2), the vertical distance between the eyebrows and the eyes (f_3), degree of nasolabial root wrinkles (NLR) (f_4), and degree of nasolabial furrows (NLF) (f_5) in their classifier design. These features are extracted from facial expressions by different techniques. For example, f_1 and f_2 are extracted by a human visual system based approach. f_1 is measured by combining global features (the height ratio and the area ratio between the whole face and the mouth region) and a local feature (Gabor–Gaussian feature). For f_2 they used the "dip" feature in log-polar mapped image and the Gabor-filter coefficients. For f_3, f_4, and f_5, which are transient components, they used the Euclidean distance (f_3) and the Gabor-filtered coefficients (f_4 and f_5). Image features are extracted from four sets of facial expression data to show effectiveness of the proposed method, which confirms considerable enhancement of the whole performance by using Fuzzy Neural Nets (FNN) based classifier.

Huang *et al.* in Reference 22 proposed a novel approach to recognize facial expression using skin wrinkles. They considered many features like eyes, mouth, eyebrows, nostrils, nasolabial folds, eye pouches, dimples, forehead, and chin furrows for their research and used Deformable Template Model (DTM) and Active Wavelet Network (AWN) for extracting those features. Classification accuracy obtained by using Principle Component Analysis and Neural Network is around 70%.

In Reference 23, Kobayashi and Hara recognized basic facial expressions by using 60 facial characteristic points (FCP) from three components of the face (eyebrows, eyes, and mouth). These features are extracted by manual calculation and emotions are classified by neural network.

In Reference 24 a real-time automated system was modeled by Anderson and McOwan, for recognition of human facial expressions. Here, the muscle movements of the human face are considered as features after tracking the face. A modification of spatial ratio template tracker algorithm is used here for tracking the face first and later to determine the motion of the face by optical flow algorithm. A percentage accuracy of 81.82% has been obtained by using SVM as classifier.

Otsuka and Ohya [25] considered a matching of temporal sequence of the 15D feature vector to the models of the six basic facial expressions by using a specialized

Hidden Markov Model (HMM). The proposed HMM comprises five states, namely, relaxed (S1, S5), contracted (S2), apex (S3), and relaxing (S4). The recognition of a single image sequence here is realized by considering transition from the final state to the initial state. Further, the recognition of multiple sequences is accomplished by considering transitions from a given final state to initial states of all feasible categories. The state-transition probability and output probability of each state are obtained from sampled data by employing Baum–Welch algorithm. The k-means clustering algorithm here has been used to estimate the initial probabilities. The method was tested on the same subjects for whom data was captured. Consequently, the feasibility of the proposed technique for an unknown subject is questionable. Although the proposed method was labeled as good, no justification was given in favor of its goodness. Besides the above, the works of Ekman [27–38], Pantic [39–46], [48–51], Cohn [52–57], Konar [58–69] and some others [70–73] deserve special mention.

1.3.2   Voice Features

Voice features used for emotion recognition include prosodic and spectral features. Prosodic features are derived from pitch, intensity, and first formant frequency profiles as well as voice quality measures. Spectral features include Mel-frequency cepstral coefficients (MFCC), linear prediction cepstral coefficients (LPC), log frequency power coefficients (LFPC), perceptual linear prediction (PLP) coefficients. We now briefly provide an overview of the voice features below.

1. Pitch represents the perceived fundamental frequency of a sound. Fundamental frequency is defined as the frequency at which the vocal cords vibrate during speech.

2. A formant is a peak in a frequency spectrum that results from the resonant frequencies of any acoustical system. For human voice, formants are recognized as the resonance frequencies of the vocal tracts. Formant regions are not directly related to the fundamental frequency and may remain more or less constant as the fundamental changes. If the fundamental is low in the formant range, the quality of the sound is rich, but if the fundamental is above the formant regions, the sound is thin. The first three formants: F_1, F_2, and F_3 are more often used to disambiguate the speech.

3. Power spectral density describes the distribution of power of a speech signal with frequency and also shows the strength (signal energy is strong or weak at different frequency) of the signal as a function of frequency. The energy or power (average energy per frame) in a formant comes from the sound source (vibration of the vocal folds, frequency of the vocal tract, movement of lips and jaw). The energy in the speech signal $x(n)$ is computed as

$$E_x = \sum_{n=0}^{N-1} x^2(n) \tag{1.1}$$

The power of the signal $x(n)$ is the average energy per frame:

$$P_x = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n) \tag{1.2}$$

where N is total no. of samples in a frame.

4. "Jitter is defined as perturbations of the glottal source signal that occurs during vowel phonation and affect the glottal pitch period" [75]. Let $u[n]$ be a pitch period sequence. Then we define absolute jitter by

$$\frac{1}{N+1} \sum_{n=0}^{N-1} |u(n+1) - u(n)|$$

5. "Shimmer is defined as perturbations of the glottal source signal that occur during vowel phonation and affect the glottal energy" [75]. Let $u[n]$ be a peak amplitude sequence of N samples. Then absolute shimmer is given by

$$\frac{1}{N+1} \sum_{n=0}^{N-1} |u(n+1) - u(n)|$$

6. MFCC [76] is a widely used term in speech and speaker recognition. However, the definition of MFCC requires defining two important parameters: "Mel scale" and Mel-frequency spectrum. The Mel scale is defined as

$$f_{mel} = 1125 ln(1 + f/700)$$

where f is the actual frequency in Hz. Mel-frequency cepstrum (MFC) is one form of representation of the short-term power spectrum of sound, based on a linear cosine transformation of a log-power spectrum on a nonlinear "mel" scale of frequency.

Mel-frequency cepstral coefficients are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly spaced frequency bands used in the normal cepstrum.

7. A speech sample can be modeled as a linear combination of its past samples. A unique set of predictor coefficients is determined by minimizing the sum of the squared differences between the actual speech samples and the linearly predicted ones. These predictor coefficients are referred to as linear prediction-based cepstral coefficient (LPCC) [77].

Busso *et al.* in Reference 78, presented a novel approach for emotion detection from emotionally salient aspects of the fundamental frequency in the speech signal. They selected pitch contour (mean, standard deviation, maximum, and minimum range of sentence- and voice-level features of pitch) as features for their experiment. Pitches obtained from emotional and neutral speech are compared first by symmetric KLD (Kullback–Leibler Distance). Then pitch features are quantified by comparing nested Logistic Regression Models. They used GMM (Gaussian Mixture Model) and LDC (Linear Discriminant Classifier) for classification process and obtained accuracy over 77%.

In Reference 79, Lee *et al.* detected emotions in spoken dialogues. Features, they used in the paper are pitch, formant frequencies, energy, and timing features like speech duration rate, ratio of duration of voiced and unvoiced region, and duration of the longest voiced speech. In this paper, irrelevant features are eliminated from the base feature set by forward selection (FS) method, and then a feature set is calculated by PCA. This novel approach improved emotion classification by 40.7% for males and 36.4% for females using LDC and k-NN (k-Nearest Neighborhood classifier) for emotion classification.

Wu *et al.* [80] proposed a new method for emotion recognition of affective speech based on multiple classifiers using acoustic–prosodic information and semantic labels. Among the acoustic–prosodic features, they selected pitch, intensity, formants and formant bandwidth, jitter-related features, shimmer-related features, harmonicity-related features, and MFCC. They derived Semantic Labels from HowNet (Chinese Knowledge Base) to extract EAR (Emotion Association Rules) from the recognized word sequence of the affective speech. They used multiple classifiers like GMM, SVM, MLP, (Multilayer Perceptron), MDT (Meta Decision Tree), and Maximum Entropy Model (MaxEnt) and got an overall accuracy of 85.79%.

Kim *et al.* [81] have developed an improved emotion recognition scheme with a novel speaker-independent feature. They employed orthogonal–linear discriminant analysis (OLDA) for extracting speech features, that is, ratio of a spectral flatness measure (SFM) to a spectral center (RSS), pitch, energy, and MFCC. They used GMM as a classifier for emotion recognition. An average recognition rate of 57.2% ($\pm$5.7%) at a 90% confidence interval can be obtained by their experiment. Among the other research works on speech, the work of Mower [82–86], Narayanan [78–94], Wu [95–98], Schuller [99–116], and some others [117–133] deserve special mention.

1.3.3   EEG Features Used for Emotion Recognition

Electroencephalogram is an interesting modality for emotion recognition. Under a hostile environment, people sometimes attempt to conceal the manifestation of their emotional states in facial expression and voice. EEG, on the other hand, gives a more realistic modality of emotion recognition, particularly, due to its temporal changes during arousal of emotion, and thus, concealment of the emotion in EEG is not feasible.

Usually the frontal lobe of the human brain is responsible for cognitive and emotion processing. There exists an internationally accepted 10-20 system for electrode placement on the scalp. Such placement of electrodes ensures that most of the
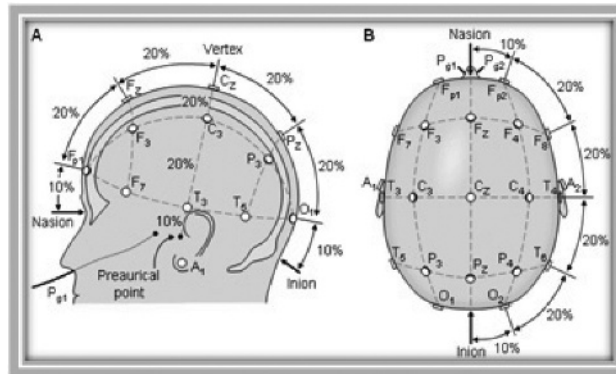
FIGURE 1.2 The international 10-20 electrode placement system.

brain functions, such as motor activation, emotion processing, reasoning, etc., can be retrieved correctly from the EEG signals obtained from these channels. In the 10-20 system (shown in Figure 1.2) of electrode placement, the channels F_3, F_4, F_{p_1}, and F_{p_2} are commonly used for emotion recognition.

Both time- and frequency-domain parameters of EEG are used as features for the emotion classification problems. Among the time-domain features, adaptive auto-regressive (AAR) and Hzorth parameters, and among the frequency-domain features power-spectral density is most popular. EEG being a nonstationary signal, its frequency contents change widely over time. Time–frequency correlated features thus carry essential information of the EEG signal. Wavelet transform coefficients are important examples of time–frequency correlated features. In our study [134, 135], we considered wavelet coefficients, power spectral density, and also AAR parameters [136] for feature extraction. Typically the length of such feature vectors is excessively high, and thus, a feature reduction technique is employed to reduce the length of vectors without losing essential features.

Petrantonakis *et al.* [137] proposed a novel approach to recognize emotion from brain signals using a novel filtering procedure, namely, hybrid adaptive filtering (HAF) and higher order crossings (HOC) analysis. HAF was introduced for an efficient extraction of the emotion-related EEG characteristics, developed by applying Genetic Algorithms (GA) to the Empirical Mode Decomposition (EMD) based representation of EEG signals. HOC analysis was employed for feature extraction from the HAF-filtered signals. They introduced a user-independent EEG-based emotion recognition system for the classification of six typical emotions, including happiness, surprise, anger, fear, disgust, and sadness. The EEG signals were acquired from F_{p_1}, F_{p_2}, F_3, and F_4 positions, according to the 10-20 system from 16 healthy subjects using three EEG channels through a series of facial-expression image projections, as a Mirror Neuron System based emotion elicitation process. For an extensive evaluation of the classification performance of the HAF–HOC scheme, Quadratic Discriminant Analysis (QDA), k-Nearest Neighbor (k-NN), Mahalanobis Distance (MD), and Support Vector Machines (SVMs) were adopted. For the individual-channel case, the best

results were obtained by the QDA (77.66% mean classification rate), whereas for the combined-channel case, the best results were obtained using SVM (85.17% mean classification rate).

In Reference 138, Petrantonakis *et al.* proposed a novel method for evaluating the emotion elicitation procedures in an EEG-based emotion recognition setup. By employing the frontal brain asymmetry theory, an index, namely Asymmetry index (AsI), is introduced, in order to evaluate this asymmetry. This is accomplished by a multidimensional directed information analysis between different EEG sites from the two opposite brain hemispheres. The proposed approach was applied to three-channel (F_{p_1}, F_{p_2}, and F_3/F_4 10/20 sites) EEG recordings drawn from 16 healthy right-handed subjects. For the evaluation of the efficiency of the AsI, an extensive classification process was conducted using two feature-vector extraction techniques and an SVM classifier for six different classification scenarios in the valence/arousal space. This resulted in classification results up to 62.58% for the user-independent case and 94.40% for the user-dependent one, confirming the efficacy of AsI as an index for the emotion elicitation evaluation.

Yuan-Pin Lin *et al.* [139] developed a new idea to recognize emotion from EEG signals while listening to music. In this study, EEG data were collected through a 32-channel EEG module, arranged according to the international 10-20 system. Sixteen excerpts from Oscar's film soundtracks were selected as stimuli, according to the consensus reported from hundreds of subjects. EEG signals were acquired from 30 channels. The features selected include power spectrum density (PSD) of all the 30 channels. SVM was successfully employed to classify four emotional states (joy, anger, sadness, and pleasure) using the measured PSDs. The best result for classification accuracy obtained is found to have mean 82.29% with a variance of 3.06% using a 10 times of 10-fold cross-validation scheme across 26 subjects. A few more interesting works on EEG-based emotion recognition that need special mention include References 140–153.

1.3.4 Gesture- and Posture-Based Emotional Features

Gestures are expressive and meaningful motions, involving hands, face, head, shoulders, and/or the complete human body. Gesture recognition has a wide range of applications, such as sign language for communication among the disabled, lie detection, monitoring emotional states or stress levels of subjects, and navigating and/or manipulating in virtual environments.

Recognition of emotion from gestures is challenging as there is no generic notion to represent a subject's emotional states by her gestures. Further, the gestural pattern has a wider variation depending on the subject's geographical origin, culture, and the power and intensity of his or her expressions.

Gestures can be static, considering a single pose or dynamic with a prestroke, stroke, and poststroke phases [154]. Automatic recognition of continuous gestures requires temporal segmentation. The start and end points of a continuous gesture are often useful to segregate it from the rest. Segmentation of a gesture sometimes is difficult as the preceding and the following gestures often are similar.

The most common gestural pattern, often used in emotion recognition, is the hand movements. Glowinski *et al.* [155] proposed an interesting technique for hand (and head) gesture analysis for emotion recognition. They considered a "bounding" triangle formed by the centroids of the head and hands, and determined several parameters of the 3D triangle to extract the features of individual hand gesture representative of emotions. A set of triangles obtained from a motion cue is analyzed to extract a large feature vector, the dimension of which is reduced later by PCA. A classification technique is used to classify emotion from a reduced 4D data space. Methodologies of feature reduction and classification will be discussed in a subsequent section later.

Camurri *et al.* [156] classified expressive gestures from the human full body movement during the performance of the subject in a dance. They identified motion cues and measured overall duration, contraction index, quantity of motion, and motion fluency. On the basis of these motion cues, they designed an automated classifier to classify four emotions (anger, fear, grief, and joy).

Castellano *et al.* [157] employed hand gestures for emotion recognition. They considered five different expressive motion cues, such as quantity of motion and contraction index of the body(degree of contraction and expansion of the body), velocity, acceleration, and fluidity (uniformity of motion) using Expressive Gesture Processing Library [158], and determined emotion from the cues by direct classification of time series.

In Reference 154, the authors considered both static and dynamic gesture recognition. While static gestures can be recognized by template matching, dynamic gesture is represented as a collection of time-staggered states, and thus can be modeled with HMM, sequential state machines (SSM), and discrete time differential neural nets (DTDNN). Preprocessing of the dynamic gestures includes tracking of important points or regions of interest in the image frames describing temporal states of the gesture. The tracking is often performed using particle filtering and level sets. Although particle filters and level sets have commonality in the usage of tracking in images, the principles used therein have significant difference. Usually a particle filter tracks a geometric-shaped region, circle or rectangle, on the image. So, in a nonrigid video, if the points enclosed in a region (of an image frame) have different directions of velocity, all the points of the reference frame cannot be tracked in subsequent frames. However, in a level set, the points enclosed in a region (of a frame) can be tracked by a nonlinear boundary, which may change its shape over the subsequent frames to keep track of all the points in the region of interest. Both the methods referred to above can track the hand and head gestures. For other important works on emotion recognition using gesture features, readers may consult the references aside 159–169.

1.3.5 Multimodal Features

Multimodality refers to analysis of different manifestations of emotion, including facial expression, voice, brain signals, body gesture, and physiological reactions. A few well-known multimodal schemes for emotion recognition are outlined below.

1.3.5.1   AudioVisual   Zheng *et al.* [170] proposed an interesting approach to audio-visual emotion recognition. They considered 12 predefined motions of facial features, called Motion Units (MU) and 20 prosodic features, including pitch, RMS energy, formants F1–F4 and their bandwidths, and all of their corresponding derivatives. They achieved a person-independent classification accuracy of 72.42% using multi-stream HMM.

Mower *et al.* [171] designed a scheme for audiovisual emotion recognition. They considered both distance features between selective facial points and prosodic/spectral features of speech to recognize the emotion of an unknown subject. The distance features are selected within and between two of the following four regions: cheek, mouth, forehead and eyebrow. For example, the metrics considered refer to the distance from top of the check to the eyebrow, lower part of the cheek to mouth/nose/chin, relative distance features between pairs of selected points on cheek, and average positional features. Similarly they considered mouth features, such as mouth opening/closing, lips puckering, and distance of lip corner and top from the nose. The prosodic features extracted from speech include pitch and energy, and the spectral features employed include MFCC. They employed Emotion-Profile Support Vector Machines (EP-SVM) to obtain classification accuracy of 68.2%.

Busso *et al.* in Reference 172, proposed a new method to recognize emotion using facial expressions, speech, and multimodal information. They considered two approaches indicating fusion of two modalities at decision and feature levels. They employed 102 markers on the given facial image to determine the motion and alignment of the marked data points during utterance of 258 sentences expressing the emotions. They also considered prosodic features, such as pitch and intensity and reduced the feature dimensions by PCA. They classified the emotions by considering both individual visual and voice features, and obtained a classification accuracy of 85% for face data, 70.9% for voice data, and 89.1% for bimodal (face plus voice together) information.

1.3.5.2   Facial Expression–Body Gesture   In a recent paper, Gunes and Picardi 173 consider automatic temporal segment detection and affect recognition from facial and bodily manifestation of emotional arousal. The main emphasis of the paper lies in the following thematic study. First, they demonstrate through experiments that affective faces and bodily gestures need not be strictly synchronous, although apparently they seem to occur jointly. Second, they observed that explicit detection of the temporal phases improves the accuracy of affect recognition. Third, experimental results obtained by them reveal that multimodal information including facial expression and body gesture together perform a better recognition of affect than only facial or body gestures. Last, they noticed that synchronous feature-level fusion achieves better performance than decision-level fusion.

1.3.5.3   Facial Expression–Voice–Body Gesture   Nicolaou, Gunes, and Pantic, in Reference 174, have taken 20 facial feature points as facial features, MFCC, energy, RMS energy, pitch as speech features, and 5 shoulder points as body gesture feature for Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities

in Valence-Arousal Space. They introduced Bidirectional Long Short-Term Memory Neural Networks (BLSTM-NN) and Support Vector Regression (SVR) classifier for emotion classification and concluded that BLSTM-NN gives better performance than SVR.

Castellano *et al.* in Reference 175 introduced a novel approach to emotion recognition using multiple modalities, including face, body gesture, and speech. They selected 19 facial feature points as facial features, MFCC, pitch values, and lengths of voiced segments as speech features, and 80 motion features for each gesture as body gesture features. They trained and tested a model with a Bayesian classifier, using a multimodal corpus with 8 emotions and 10 subjects. To fuse facial expressions, gestures, and speech information, two different approaches were implemented: feature-level fusion, where a single classifier with features of the three modalities is used; and decision-level fusion, where a separate classifier is used for each modality and the outputs are combined a posteriori. Lastly, they concluded that the fusion performed at the feature level provided better results than the one performed at the decision level.

1.3.5.4   *EEG–Facial Expression*

Chakraborty *et al.* [176] correlated stimulated emotion extracted from EEG and facial expression using facial features, including eye-opening, mouth-opening, and eyebrow constriction and EEG features, including frequency domain, time domain, and spatiotemporal features. They considered frequency domain features, such as peak power and average powers of $\alpha, \beta, \gamma, \theta$, and δ bands, time domain features, including 16 Kalman filters coefficients, and spatiotemporal features including 132 wavelet coefficients. They employed a feed-forward neural network to train it with a set of experimental instances using the well-known Back-propagation algorithm. The resulting network on convergence is capable of classifying instances to a level of 95.2% classification accuracy.

1.3.5.5   *Physiology*

In Reference 177, Picard *et al.* proposed that for developing a machine's ability to recognize human affective state, machines are expected to possess emotional intelligence. They performed an experiment considering four physiological signals: electromyogram (EMG), blood pressure volume, Hall effect, and respiration rate, taken from four sensors. One additional physiological signal, heart rate (H) here has been calculated as a function of the inter-beat intervals of the blood volume pressure. In their analysis, they used a combination of sequential floating forward search (SFFS) and Fisher projection (FP), called SFFS-FP, for selecting and transforming the features. A classification accuracy of 81% was obtained by using maximum a posteriori (MAP) classifier for SFFS-FP.

1.3.5.6   *Facial Expression–Voice–Physiology*

Soleymani *et al.* [178] introduced a multimodal database for affect recognition and implicit tagging. They chose 27 subjects and recorded their videos of facial and bodily responses while watching 20 emotional videos. Features used for their experiment include distance metrics of the eye, eyebrow, and mouth as facial features, audio and vocal expressions, eye

gaze, pupil size, electrocardiograph (ECG), galvanic skin response (GSR), respiration amplitude, and skin temperature as physiological features. The main contribution of this research lies in the development of a large database of recorded modalities with high qualitative synchronization between them making it valuable to the ongoing development and benchmarking of emotion-related algorithms. The resulting database would provide support to a wide range of research on emotional intelligence, including data fusion, synchronization studies of modality, and many others.

1.3.5.7   EEG–Physiological Signals   Takahashi, in Reference 179, undertook an interesting research on emotion recognition from multimodal features including EEG, Pulse, and Skin Conductance. They collected psychological data of 12 subjects. The experimental setup contains a set of three sensors and two personal computers; one PC being used to present stimulus to a subject, while the other is used to acquire biopotential signals stated above. They used SVM classifier for emotion recognition from biopotential signals and acquired a classification accuracy of 41.7% for five emotions including joy, anger, sadness, fear, and relaxation. Among the other works on emotion recognition References 180–188 deserve special mention.

1.4   FEATURE REDUCTION TECHNIQUES

EEG and voice features usually have a high dimensionality and many of the experimentally obtained features are not independent. The speed of classifiers is often detrimental to the dimension of features. Feature reduction algorithms are required to reduce the dimensionality of features. Both linear and nonlinear feature reduction techniques are employed for emotion recognition.

Linear reduction techniques employ the characteristics of real symmetric matrices to extract independent features. In other words, the eigen vectors of real symmetric matrices are orthogonal (independent) to each other. Further, the larger eigen values of a system carry more information than the others. So, the eigen vectors corresponding to the large eigen values are used to reduce data dimensionality of a given linear system.

Linear reduction principles have gained popularity for their simplicity in use. However, on occasion researchers prefer nonlinear reduction techniques to their linear counterparts to improve precision and reliability of the classifier. Among the nonlinear feature reduction techniques, the most popular is rough set–based feature reduction. In this section, we briefly outline a few well-known linear and nonlinear feature reduction techniques.

1.4.1   Principal Component Analysis

Principal Component Analysis is one of the most popular linear feature reduction techniques. PCA represents N measurements for M subjects as a $M \times N$ matrix A, and computes $A^{T}A$ to obtain a real symmetric matrix B of dimension $N \times N$. Now the N eigen values of B are evaluated, and the results are sorted in descending order.

It is known that the larger eigen values have higher contribution in representing system characteristics, and thus to reduce features, we take k no. of eigen vectors corresponding to the first k eigen values of the list. The eigen vectors are arranged in columns, and the matrix is called Eigen Vector (EV) matrix of $(N \times k)$ dimension. Now, for each measurement vector a_i, taken from the ith row of the matrix A, we take a projection of a_i on the eigen space by multiplying a_i by EV matrix, and thus obtain $a_i^/$, where $a_i^/$ has dimension $(1 \times k)$. This is repeated for all $i = 1$ to N, and thus the feature vectors a_i are mapped to k-dimensional vectors, where $k \ll N$.

PCA is good for feature reduction of linear systems. If PCA is used for feature reduction of nonlinear systems, where the functional relationship between any two features is nonlinear, PCA sometimes loses important information. Researchers use PCA for its high efficiency and accuracy. Some important work using PCA include References 17, 79, 87, and 189–193.

1.4.2   Independent Component Analysis

Independent Component Analysis is a good choice to separate the sources from mixed signals. Particularly, in EEG and ECG the time series data x_t at time t is a nonlinear function $f(.)$ of previous time samples, that is, $x_t = f(x_0, x_1, \ldots, x_{t-1})$. EEG signals taken from the forehead of a subject are often contaminated with eye-blinking signals, called electrooculogram (EOG). Further, the signal obtained at an electrode located on the scalp/forehead is due to the contribution of a number of signal sources at the neighborhood of the electrode. Elimination of EOG from EEG data and identification of the source signals can be performed together by ICA. One fundamental (although logical) restriction of using ICA lies in the inequality: $C >= S$, where C represents the number of EEG channels, and S stands for the number of independent signal sources. ICA has been widely used in the literature [194–198] to recognize emotion from facial expressions.

1.4.3   Evolutionary Approach to Nonlinear Feature Reduction

Evolutionary algorithms are population-based meta-heuristic optimization algorithms, which rest on the Darwinian principle of the survival of the fittest. The primary aim of this class of algorithms is to determine near optimal solutions, if not global, from a set of trial solutions through an evolutionary process determined by a set of operators like crossover, mutation, and selection. The most popular member of this class is Genetic Algorithm [199–201], devised by Prof. Holland approximately a half century ago. Among other members Differential Evolution (DE) [202, 203] is most popular for its structural and coding simplicity and exceptional performance in optimization problems.

Given a set of feature vectors (also called data points) and class labels for each vector, to implement the evolutionary feature reduction, we first use a supervised learning–based classifier to classify the data points into a fixed number of classes c. Next we reduce the dimension of data points by dropping one feature randomly at a time, and again classify the data points into c number of classes. If the resulting

classes do not differ significantly with the previous classes, then the dropped feature has no major significance. The GA or DE is used to randomly select k number of features at a time, and classify the data points into c classes, and test whether the classes generated have significant difference with the classes obtained from the original dataset after classification. Since k is randomly selected in $[1, n]$, where n is the total number of features in the original dataset, we at the end of the search process expect to find a suitable value of k, for which the classes would be similar with the classes of the original dataset. Thus high dimensional features are reduced to k-dimension for $k \ll n$.

1.5   EMOTION CLASSIFICATION

This section provides principles of several approaches to emotion classification by pattern recognition techniques.

1.5.1   Neural Classifier

Neural networks have widely been used in emotion classification by facial expressions and voice. Both supervised and unsupervised neural architectures are employed in emotion classifiers. The supervised neural networks require a set of training instances. During the training process, the network encodes the connection weights in a manner, such that for all the input components of the training instances, the network can reproduce the output components of the corresponding training instances correctly as listed in the training instances. After the encoding is completed, the trained network can be used for testing. In the testing phase, an unknown input instance is submitted to the network, and the network generates the output instance using the encoded weights. In case of emotion recognition, the output of the neural net usually represents emotion classes, whereas input of the neural net represents a set of features extracted from facial expression/voice/gesture of the subjects. Naturally, a neural network pretrained with emotional features as the input and emotion classes as the output would be able to classify a specific emotional expression into one of several emotion classes.

1.5.1.1   Back-Propagation Algorithm   Among the well-known neural topologies, the Back-propagation is most common. Weight adaptation in the Back-propagation neural net is performed by Newtonian Gradient/Steepest descent learning principle. Let W_{i_j} be the connection weight between neuron N_i and neuron N_j, and E be the error function representing the root mean square error between desired output and computed output for each input training instance. Then the weight adaptation policy is formally given by

$$\delta W_{i_j} = -\eta \delta E / \delta W_{i_j}$$

where ΔW_{i_j} denotes change in weight and η is the learning rate. In Back-propagation algorithm, the weight adaptation for each layer is derived using the above equation.

18   INTRODUCTION TO EMOTION RECOGNITION

Computation of ΔW_{i_j} in the output layer of a multilayered feed-forward neural network is straightforward as the error function E involves the weight W_{i_j}. However, computation of ΔW_{i_j} in the intermediate and input layers is not easy as the error function E does not involve W_{i_j}, and a chain formula of known partial derivatives is used to compute ΔW_{i_j}.

Once ΔW_{i_j} computation is over, we add it to W_{i_j} to obtain its new value. The process of layerwise computation of weights always starts at the output layer, and continues up to the input layer, and this is usually referred to as one pass. Several passes are required for convergence of weights toward steady-state values. After convergence of the weights, the trained network can be used for the application phase. During this phase, the network is excited with a new instance.

One fundamental limitation of the back-propagation algorithm is trapping at local optima on the error (energy) surface. Several methods have been proposed to address the issue. The most common is adding momentum to the weight adaptation dynamics. This helps the dynamics to continue movement even after coming in close vicinity of any local optima. Once the dynamics pass the local optima, their speeds are increased so that the motion is continued until the global optimum is identified. Among the enormous work on emotion recognition using Back-propagation neural network, References 204–208 need special mention.

1.5.1.2   Radial Basis Function Based Neural Net   Radial Basis Function (RBF) neurons employ a specialized basis function to map an input pattern $\vec{X}$ to two soft levels 0 and 1. A pattern classifier with k classes usually has k basis functions, designed to map an unknown input vector to one of k classes. When a pattern falls in a class, its RBF function yields a value close to one. On the other hand when an input pattern does not fall in a class, the function returns a small value close to zero. Among the popular RBF functions, the Gaussian function is most common for its wide applications in science and engineering. Let X_c be the center of an RBF function. Then for any input vector X_c, we define the RBF function Y as given below:

$$Y = exp - \|(X - X_c)\|$$

where $\|.\|$ denotes an Euclidean norm.

A typical RBF neural net consists of two layers, the first layer being the RBF layer, and the last layer being realized by a perceptron neuron, the weights of which are determined by the perceptron learning algorithm. When an unknown input instance is supplied, the response of the first layered RBF neurons become close to zero for most of the neurons and close to one for one or fewer neurons. The weights generated by perceptron learning algorithm are later used to map an intermediate pattern into pattern class. A few research works employing Radial Basis Function Based Neural Network as a classifier for emotion recognition, include References 209–211.

1.5.1.3   Self-Organizing Feature Map Neural Net   In a Self-Organizing Feature Map (SOFM) Neural Net, we need to map input patterns onto a 2D array of neurons

based on the similarity of inputs with the patterns stored in individual neurons. The patterns stored by neurons have the same dimension as that of input patterns. These patterns are called weights of the respective neurons. A given input pattern is mapped onto a neuron with the shortest Euclidean distance. A neighborhood around the selected neuron is considered, and the weights of all the neurons in the neighborhood are adapted by the following equation:

$$W_{i_j}(t + 1) \leftarrow W_{i_j}(t) + \eta(X_k - W_{i_j}(t))$$

where $W_{i_j}(t)$ is the weight of neuron (i, j) in the neighborhood of the selected neuron in the 2D array at time t, X_k is the kth input vector, and η is the learning rate. After the weights are adapted, a new input vector is mapped onto the array by the distance criterion, and the process of neighborhood selection around the neuron and weight adaptation of the neurons in the neighborhood is repeated for all the inputs. The whole process of mapping input vectors onto a 2D array and weight adaptation of neurons is aimed at topological clustering of neurons, so that similar input vectors are mapped at close vicinity on the 2D array.

During the recognition phase, we need to retrieve one or more fields of a given vector, presuming that the remaining fields of the vector are known. Generally, the unknown vector has the same dimension to that of other input vectors used for weight adaptation. The unknown vector first is mapped onto a neuron in the 2D array based on the measure of minimum Euclidean distance between the input vector and all weight vectors, considering only the known fields of the unknown vector during Euclidean distance evaluation. The neuron having the best match, that is, having the smallest Euclidean distance between its weight vector and the unknown input vector is identified. The unknown fields of the input vector are retrieved from the corresponding fields of the weight vector of the selected neuron. SOFM can be used for emotion recognition from face [212], speech [213], EEG [214], as well as from gesture [215], for its high efficiency and accuracy as a classifier.

1.5.1.4 Support Vector Machine Classifiers Support Vector Machines have been successfully used for both linear and nonlinear classification. A linear SVM separates a set of data points into two classes with class labels +1 and −1. Let $X = [x_1 x_2 ... x_n]^{\mathrm{T}}$ be any point to be mapped into +1, −1 by a linear function $f(X, W, b)$, where $W = [w_1 w_2 ... w_n]$ is a weight vector and b is a bias term. Usually, $f(X, W, b) = \mathrm{Sign}(WX + b) = \mathrm{Sign}(\sum_i w_i x_i + b)$. Figure 1.3 illustrates classification of 2D data points. In 2D, the straight line that segregates the two pattern classes is usually called a hyperplane. Further, the data points that are situated at the margins of the two boundaries of the linear classifier are called support vectors. Figure 1.3 describes a support vector for a linear SVM.

Let us now select two points $X+$ and $X-$ as two support vectors such that for $X = X^+$, $WX^+ + b = +1$. Similarly, for $X = X^-$, $WX^- + b = -1$. Now, the separation between the two support vectors lying in the class +1 and class −1, called marginal width is given by

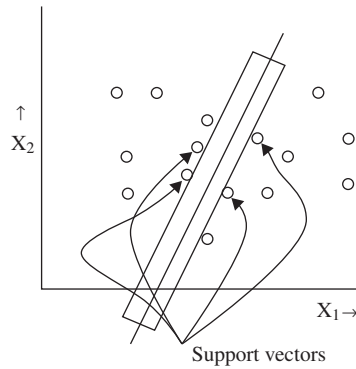$$M = (WX^+ + b) - (WX^- + b)/\|W\| = 2/\|W\|$$

FIGURE 1.3   Defining support vector for a linear SVM system.

The main objective in a linear SVM is to maximize M, that is, to minimize $\|W\|$, which is same as minimizing $1/2W^TW$. Thus, the linear SVM can be mathematically described by

$$\begin{aligned} &\text{Minimize } \phi(W) = 1/2W^TW \\ &\text{subject to } y_i(WX_i + b) \geq 1 \text{ for all } i \end{aligned} \qquad (1.3)$$

where y_i is either 1 or -1 depending on the class that X_i belongs to.

Here, the objective is to solve W and b that satisfies the above equation. The solution to the optimization problem is not given here due to space limitation. However, inquisitive readers can have it in the standard literature. Linear SVM has a wide range of applications in supervised classifiers. It is currently one of the best popular algorithms for pattern classification. Many researchers choose SVM [18, 24, 26, 216–219] as a classifier in emotion classification for its higher accuracy.

1.5.1.5 Learning Vector Quantization   In scalar quantization, the random occurrence of a variable x in a given range $[x_{\min}, x_{\max}]$ is quantized into few fixed levels. For example, suppose there are m quantization levels, uniformly spaced in $[x_{\min}, x_{\max}]$. So, the quantization step height is determined as $q = (x_{\max} - x_{\min})/m$. The kth quantization level has a value $x_k = x_{\min} + h.(k - 1)$. An analog signal x having a value greater than kth quantization level, but less than $(k + 1)$th quantization level is quantized to the kth level. This particular feature of the quantization process is called truncation. Sometimes, we use roundoff characteristics of the quantizer. In case of roundoff, an analog signal having a value less than $x_k + q/2$ but greater than x_k would be quantized to x_k. But if the analog signal $x > x_k + q/2$, but less than x_{k+1}, it will be quantized to x_{k+1}.

In vector quantization, vectors V of dimension n are quantized to fixed vectors V_i of the same dimension. If all components of V are close enough with respect to corresponding components of vector V_i, then V would be quantized to V_i. Usually, such quantization is widely used in data compression. In learning vector quantization

(LVQ), we use a two-layered neural net. The first layer is used for data reception, while the second is a competitive layer, where one of several neurons only fires, and the weights are reinforced using the input. Let X_i be the ith input vector, whose components are mapped to the neurons at the input layer of the neural net. Let us assume that there are p neurons in the second layer. Now let the weight vector of the neurons in the second layer be $W_1, W_2, \ldots, W_p$. Suppose

$$\|W_k - X_i\| <= \|W_j - X_i\| \text{ for all } j$$

Then we would adapt W_k by the following update rule:

$$W_k \leftarrow W_k + \eta(X_i - W_k)$$

where η is the learning rate between $(0, 1)$. The weights are thus reinforced by all the inputs. After the learning with all the input instances are over, we identify the unknown components of a given test instance vector by identifying the trained weight vector, with which the input instance has the smallest Euclidean distance, considering only the known components. The unknown fields of the selected weight vector are then used for subsequent applications. Researchers used LVQ as a classifier for emotion recognition from facial expression [220, 221] or speech [222].

1.5.2   Fuzzy Classifiers

Measurements obtained from facial expression, voice, and gesture/posture for emotion recognition often are found to be contaminated with various forms of uncertainty. For example, repeated measurements of facial, vocal, and bodily gestures of a subject experiencing the same emotion have wider variance. This is often referred to as an intrapersonal level of uncertainty. Further, when the measurements are taken from different subjects experiencing same/similar emotion, the variance in measurements is found to have a large value, causing interpersonal level of uncertainty. Classical type 1 fuzzy logic considers a single membership function to represent the uncertainty involved over a given measurement space, but fails to model the true spirit of intra- and interpersonal level of uncertainty [223]. Type 2 fuzzy sets provide an opportunity to represent both inter- and intrapersonal variations in uncertainty, and thus has immense scope in fuzzy classifiers, capable of correctly classifying emotions from measurements suffering from uncertainty.

In classical fuzzy rule–based classifiers, fuzzy rules are employed to map fuzzy encoded measurements into emotion classes with different degrees of certainty. Thus individual emotions support a given set of emotional features to different degrees, and naturally the class with the highest support is considered as the winner. Type 2 fuzzy rules on the other hand map a set of imprecise fuzzy encoded measurements obtained from different sources into emotion classes. The class offering maximum support to the measurement space is considered as the winning class.

An alternative approach to fuzzy pattern recognition is to cluster a set of features based on their similarity. Since measurements are noisy, data points lying on the

boundary of two classes can be categorized to both the classes with certain degrees. Fuzzy C-means clustering algorithm has widely been used as a basic tool of pattern clustering to overcome clustering of data points suffering from noisy measurements or incomplete specification of data dimensions. A few important research works employing fuzzy logic in classifiers include [60, 224, 225].

1.5.3 Hidden Markov Model Based Classifiers

Let X, Y, and Z be three random variables where they may take up any value from $x_1, x_2, \ldots, x_n$, $y_1, y_2, \ldots, y_m$, and $z_1, z_2, \ldots, z_r$, respectively. Suppose Y depends on X and Z depends on Y. So, we can represent the dependence relationship among X, Y, and Z by a directed graph, where X is a predecessor of Y and Y is a predecessor of Z. Now suppose, by experiments we have the conditional probabilities: $P(Y/X)$ and $P(Z/Y)$ for any X, Y, and Z. Now in a Markov process we consider the state-transition probabilities in one level only, that is, we consider $P(Z/Y, X) = P(Z/Y)$.

In an HMM, the probability of occurrence of a class is determined by a sequence of state transitions. For example, suppose if $X = x_2$ then $Y = y_3$, and if $Y = y_3$ then $Z = z_4$. Suppose, the sequence of state transitions from $X = x_2$ through $Y = y_3$ and $Z = z_4$ denotes class 1 of a pattern recognition problem. This state transition probability is $P(Z = z_4/Y = y_3) \cdot P(Y = y_3/X = x_2)$. Now, if $P(X = x_2)$ is the probability of occurrence of $X = x_2$, then $P(Z = z_4)$ following the sequence $X = x_2$ through $Y = y_3$ is given by $P(X = x_2) \cdot P(Y = y_3/X = x_2) \cdot P(Z = z_4/Y = y_3)$. Thus for all known sequences passing through X, Y, and Z, we know the probability of the pattern class.

Now, for an unknown sequence, we compute the state transition probability of the sequence, and check whether the probability of the sequence matches enough with the probability of the standard nearest sequence. If so, we consider the unknown sequence to lie in the class of the nearest matched sequence. HMM is vastly used in research for recognition of emotion from face [25], speech [226–228], and sometimes from both face and speech [229–231].

1.5.4 k-Nearest Neighbor Algorithm

k-Nearest Neighbor algorithm is a simple method to determine the class of an unknown pattern. It presumes that the given data points are pre-classified and the label of each class is known. Now, for an unknown pattern, represented by a data point X, the distance between X and all nearest neighbors Y_j of X are identified, and the count of Y_j lying in different classes is determined. The class having the largest number of nearest neighbors of X is declared as the class of the unknown data point X.

The distance metric selection is an important problem in the k-NN algorithm. Usually, Euclidean distance is used in most of the literature [232]. However, when the data points have a large dimension or the components are not scaled properly, the distance between any two data points sometimes is too large, and consequently the results of classification is not free from errors. The given set of data points therefore should be normalized by scaling d_{i_j} by $d_{i_j}/\text{Max}_j(d_{i_j})$, where d_{i_j} is the jth component of the ith data point d_i.

k-NN algorithm works fine for lower dimensional data, typically when dimension is less than 5. It has popularity in the pattern recognition community particularly for its simplicity. The complexity of the algorithm grows with an increase in the dimension of the data points. To avoid this increase in computational complexity, feature reduction algorithm is first employed to identify the independent components of the data points by using a feature reduction algorithm, and k-NN is applied on the data classes developed with reduced dimensionality. k-NN has been used in many research works as a classifier to recognize emotion [79, 90, 233].

1.5.5   Naïve Bayes Classifier

Suppose, there are n objects, each having a set of m features $f_1, f_2, \ldots, f_m$ and the objects are classified into k classes: $c_1, c_2, \ldots, c_k$ based on the similarity of their features/attributes. Now, by Bayes' conditional probability theory, we have the conditional probability

$$P(c_j \,|\, f_1, f_2, \ldots, f_m) = P(f_1, f_2, \ldots, f_m \,|\, c_j)P(c_j)/P(f_1, f_2, \ldots, f_m)$$

where the probabilities have their standard meaning. Now, for any j, the expression on the right-hand side contains a constant denominator: $P(f_1, f_2, \ldots, f_m)$. So, we instead of computing $P(f_1, f_2, \ldots, f_m)$, simply represent it by a scalar constant $1/A$. Therefore, the last result may be reconstructed as

$$P(c_j \,|\, f_1, f_2, \ldots, f_m) = A.P(f_1, f_2, \ldots, f_m \,|\, c_j)P(c_j) \tag{1.4}$$

Now, $P(f_1, f_2, \ldots, f_m \,|\, c_j)$ denotes the probability of the joint occurrence of $f_1, f_2, \ldots, f_m$, given the object class c_j. However, computation of $P(f_1, f_2, \ldots, f_m \,|\, c_j)$ is not simple. One simple assumption is to consider the features independent, that is,

$$P(f_1, f_2, \ldots, f_m \,|\, c_j) = P(f_1/c_i).P(f_2/c_i)...P(f_m/c_i) = \prod_{i=1}^{m} P(f_i/c_j) \tag{1.5}$$

Combining (1.3) and (1.4) we obtain,

$$P(c_j \,|\, f_1, f_2, \ldots, f_m) = A.P(f_1, f_2, \ldots, f_m \,|\, c_j)P(c_j) = A.\prod_{i=1}^{m} P(f_i/c_j).P(c_i)$$

Now, from experimental instances, if we can prepare a table $P(f_i/c_j)$ for all i, j, we can determine $P(c_j \,|\, f_1, f_2, \ldots, f_m)$ for all j, and identify the class k, for which

$$P(c_k \,|\, f_1, f_2, \ldots, f_m) >= P(c_j \,|\, f_1, f_2, \ldots, f_m) \text{ for all } j$$

The class k is the desired class for the given combination of features. This is how a Naïve Bayes classifier works. The current literature employing Naïve Bayes Classifier to recognize emotion includes References 234–236.

1.6   MULTIMODAL EMOTION RECOGNITION

Emotion of a subject is recognized from his or her facial expression, voice, gesture, and posture. The term modality refers to modes of recognition. Early research in emotion recognition stressed the importance of single modality. Single modality, however, cannot always provide sufficient information for emotion recognition. Current research on emotion recognition thus is targeted to multimodal information extraction to improve the reliability of the recognition process.

One common metric to represent the performance of the recognition process is to construct a confusion matrix, where each grid (i, j) of this matrix represents the percentage of emotional expression correctly classified in class j, when the actual emotion is i. Experiments undertaken on different modality of the same subjects reveal that the confusion matrices constructed for each modality have significant differences, indicating that the individual modality of classification is not highly reliable.

Multimodal fusion is to integrate all single modalities into a combined single representation. There are basically two types of fusion techniques that have been used in most of the literature to improve reliability in emotion recognition from multimodal information. They are (1) feature-level fusion and (2) decision-level fusion. Feature-level (early) fusion classifies emotion of subjects from their individual modalities in the first step, and then attempts to fuse the results of unimodal classification with an aim to reduce uncertainty in classification. One simple method of fusion is to check whether the individual classifiers for multiple modalities yield the same class for a subject. If so, the decision about the emotion class of the subject is certain. However, if the classifiers produce conflicting class labels for a subject, then the class label obtained from majority of the classifiers is honored. Decision-level (late) fusion feeds all measurements obtained from sensors employed for different modalities to a pretrained classifier, which in turn determines the class label of the unknown measurement.

In Reference 173, the authors applied both feature-level and decision-level fusion on face and bodily gestures and concluded that feature-level fusion achieves better performance than decision-level fusion. On the other hand, in Reference 172, the authors observed that the overall performance of the feature-level bimodal classifier and decision-level bimodal classifier is similar. But by analyzing with confusion matrix, they also observed that the recognition rate for each emotion type was totally different. In the decision-level bimodal classifier, the recognition rate of each emotion is increased (except happiness, which is decreased by 2%) compared to the facial expression classifier. Recognition rate of emotion anger and neutral are significantly increased in the feature-level bimodal classifier whereas recognition rate of happiness is decreased by 9% by using the same. Therefore, they concluded that "the best approach to fuse the modalities will depend on the application."

1.7   STIMULUS GENERATION FOR EMOTION AROUSAL

Elicitation of emotion usually is performed by presenting selective stimuli to the subjects. Selection of the modality of emotion elicitation and identification of the

right stimuli for a given modality are equally important. Among the well-known modalities of emotion elicitation, visual, audio, audiovisual, music are most popular. Traces of using ultrasonic signals for emotion elicitation are also found in the literature [237]. Proximity signals, such as touch of a mother to stop crying of babies, have also been used in the literature for emotion control.

Considering a particular modality of emotion elicitation, we can identify some fundamental basis for selection of stimulus. In Reference 60, the authors sought the opinion of several people to determine the candidate audiovisual stimulus, which seems to be most effective for elicitation of a particular emotion. To briefly outline their scheme, suppose we have n audiovisual stimuli presented before a group of m subjects. Each subject is asked to provide his or her feeling about possible percentage of elicitation of five emotions by the stimuli. After the score sheet containing the opinions of m subjects for five emotions excited by n stimuli is prepared, the sum of percentage support obtained from m subjects for each stimulus favoring individual emotion class is determined. The emotion class that receives highest support for each stimulus is selected for elicitation of that emotion.

Selection of stimuli and their modalities are important issues for arousal of emotion. The modalities include stimulation by audio, visual, audiovisual, consolation message, music, and dance. Among these modalities, significant research has been undertaken on audio [238], visual [239, 240], audiovisual [241–243] and music [244–248] based stimuli for emotion arousal. In Reference 244 the authors studied the effect of certain features in music as a qualitative measure of emotion arousal. They noticed that pitch range, tone quality, rhythmic regularity, and tempo are some of the important features capable of originating and regulating emotion. Some well-known music pieces are often used for generating particular emotion in subjects. Baumgartner *et al.* [240] selected some classical orchestral pieces like Gustav Holst: *Mars—The Bringer of War from the Planets*, Samuel Barber: *Adagio for Strings*, and Beethoven: *Symphony No. 6 (3rd mvt.)* for arousal of three basic emotions: fear, happiness, and sadness. The tones of Holst evoke fear, music of Barber evokes sadness, and the one by Beethoven evokes happiness.

Visual mode of excitation includes voiceless movie, particularly action movies describing thematic issues of a particular real-life context. The modality of video presentation, context, and climax keep the subject attracted toward the stimulus, and the context of the movie excites a specific faculty of the mind, arousing an emotion. For example, a set of consecutive pathetic scenes causes an arousal of sadness, while a scene of amusements causes relaxation/happiness. In Reference 240, the authors have shown 48 still pictures containing humans or human faces to 70 subjects for arousal of three basic emotions: fear, happiness, and sadness. They have selected pictures of a man attacking a woman with a knife or a man pointing a pistol to the viewer for arousing fear. Pictures of a man holding his smiling baby, laughing children playing on the beach, or athletes in a victory pose are used for making subjects happy. For arousing sadness, picture of a crying little boy standing in front of a destroyed house or a couple standing at a gravestone are shown.

Audiovisual modality of excitation has been found to very effective for control of emotional state-transitions. The onset time (i.e., the duration of stimulation after

which the brain responds) and offset time (i.e., the minimum time brain remains excited after removal of the stimulus) [31, 249] are important issues in audiovisual stimulation. These times are not always constant for a given subject. They are sensitive to both context and climax.

1.8 VALIDATION TECHNIQUES

After a system design is complete, we go for validation of the system performance using test samples. In case performance of the system is found satisfactory, the system design is certified for field applications. Otherwise, we should identify the free system variables/parameters, which are not direct inputs or outputs, for tuning. Machine learning and/or evolutionary algorithms are often used to tune the parameters of physical systems in order to improve the system performance with known test samples. In Reference 60, the authors have employed evolutionary algorithm to adapt the parameters of fuzzy membership functions to improve the classification accuracy of the fuzzy rule–based classifier. The classifier in that paper has been tested with facial features of subjects with known emotion classes.

A good system validation requires identifying right test instances taken from all classes of emotion with wider variance in feature space even for the same class. If the test instances are misclassified, we need to adapt system parameters. Selection of non-input/output parameters having a wider range of variation with a power of controlling the performance of the classification algorithm is also an important issue. Lastly, the evolutionary/learning algorithm to be incorporated should have the ability to tune system parameters with a motive of improving the performance index of the classifier. Usually system validation is done offline. So, no restriction is imposed on the convergence time of the validation algorithm. However, a fast algorithm is always preferred without allowing degradation in system performance.

In Reference 177, authors used leave-one-out cross-validation method for selecting the correct classifier between SFFS, FP, and a combination of SFFS and FP, called SFFS-FP. In each case, the feature vector per emotion to be classified was excluded from the dataset before the SFFS, FP, or SFFS-FP was run. The best set of features or best transform (determined from the training set) was then applied to the test data point to determine classification accuracy. They observed that SFFS-FP improves the performance and provides the highest recognition rate on eight classes of emotion.

Busso *et al.* [78], compares the performance of their proposed approach with the conventional classifier when there is a mismatch between the training and testing conditions. For this purpose, they separate the emotional databases into two language groups: English and German. One of these groups is used for training, and the other one for testing. The system is then trained and tested with different databases recorded in three different languages (English, German, and Spanish) to validate the robustness of the proposed approach. It has been seen that the performance of the conventional classifier without the neutral models decreases up to 17.9%. So they concluded that "using neutral models for emotional recognition is hypothesized to be more robust and, therefore, to generalize better than using a direct emotion classification approach."

1.8.1   Performance Metrics for Emotion Classification

Performance metrics are used to measure the performance of an emotion classifier. The simplest metric for performance measure is classification accuracy, which indicates the percentage of success in emotion classification from a given set of measurements of the subjects during the period of their emotion arousal. Classification accuracy cannot provide the subjective details about how many instances of emotional expressions falling under an emotion class are misclassified in all other emotion classes. For example, if out of 60 facial expressions, 30 appear in anger, 20 in fear, and 10 in anxiety classes, then we understand that the set of features is either incomplete or the emotional expressions are not sufficiently distinguishable. Confusion matrix offers the details of percentage misclassification of an emotion into other classes in percentage or absolute values, and so it is a popularly used metric to represent classification accuracy in emotion recognition problem.

Statistical tests are undertaken to compare the relative performance of emotion classifiers. Two types of tests are generally performed. The first type compares the performance of two classifiers, while the latter type determines ranks of individual classifiers based on some metric, usually classification accuracy. The merit of using these statistical tests is to ensure robustness in performance analysis. The tests require some measure of individual performance of a classifier tested with different databases, and ultimately determine the relative performance of the classifiers as obtained from all the databases together. Among the well-known statistical tests, McNemar's test and Friedman test [250, 251] need special mention. McNemar's test considers two classifier algorithms at a time, and identifies the relatively better between them based on a specific measure determined by the count of misclassified items by the individual algorithms. Friedman test considers all the classifier algorithms together and ranks them best on the individual performance of the algorithms.

1.9   SUMMARY

Pattern recognition at present is an established discipline of knowledge. A general pattern recognition problem comprises three or more steps: feature extraction, feature reduction, and classification. Emotion recognition falls under the basic category of pattern recognition problems. Thus, the basic steps of pattern recognition are equally applicable here. However, the main challenge in emotion recognition lies in extraction of essential features, and selecting a suitable classifier. The chapter provides a thorough review of the existing techniques of feature extraction and classification, commonly used in emotion recognition.

A review of the existing techniques of emotion recognition from a pattern classification perspective is given in the chapter. Special emphasis is given to modality-based recognition as for individual modalities the feature extraction and classification techniques have significant differences. An outline to different validation techniques for emotion classification is given here. In addition, the statistical techniques used to compare the relative performance in emotion classification are narrated at the end of the chapter.

On occasion single modality cannot correctly yield good accuracy. Multimodality becomes an important necessity under the context. The chapter provides an extensive discussion on multimodal emotion recognition. Stimulus generation for elicitation of emotion is an open problem, and very few works in this regard are available in the literature. The chapter provides a thorough overview on stimulus generation. Lastly, the chapter cites an exhaustive list of references [1–264] to emotion research, including both fundamental research and applications.

REFERENCES

[1] J. P. Marques de Sá, *Pattern Recognition: Concepts, Methods and Applications*, Springer, 2001.

[2] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2008.

[3] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.

[4] J. C. Bezdek and S. K. Pal, *Fuzzy Models for Pattern Recognition*, IEEE Press, New York, 1992.

[5] M. Kass, A. Witkin, and D. Terzopoulus, "Snakes: active contour models," *Int. J. Comput. Vis.*, 1(4): 321–331, 1988.

[6] A. Yuille, D. Cohen, and P. Halliman, "Extraction from faces using deformable templates," *Int. J. Comput. Vis.*, 8: 104–109, 1992.

[7] I. Craw, D. Tock, and A. Bennett, "Finding face features," in Proceedings of the Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19–22, 1992, pp. 92–96.

[8] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10): 1042–1052, 1993.

[9] M. Kirby and L. Sirovich, "Application of the Karhunen–Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1): 103–108, 1990.

[10] B. Moghaddam and A. Pentland, "Beyond Euclidean Eigenspaces: Bayesian Matching for Visual Recognition," in *Face Recognition: From Theories to Applications*, Springer, Berlin, 1998.

[11] B. Draper, K. Baek, M. S. Bartlett, and R. Beveridge, "Recognizing faces with PCA and ICA," *Comput. Vis. Image Underst.* (Special Issue on Face Recognition), 91: 115–137, 2003.

[12] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6): 780–788, 2002.

[13] B. Scholkopf, A. Smola, and K. M. Uller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.* 10(5): 1299–1319, 1998.

[14] M.-H. Yang, "Face recognition using kernel methods," in *Advances in Neural Information Processing Systems* (eds S. Becker, T. G. Ditterich, and Z. Ghahramani), 2002, pp. 215–220.

[15] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," in Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, April 14–16, 1998, pp. 396–401.

[16] I. Essa and A. Pentland, "Coding, analysis interpretation, recognition of facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 757–763, 1997.

[17] S. Kimura and M. Yachida, "Facial expression recognition and its degree estimation," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June 17–19, 1997, pp. 295–300.

[18] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin, "Automatically detecting pain in video through facial action units," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 41(3): 664–674, 2011.

[19] P. Ekman, "The argument and evidence about universals in facial expressions of emotion," in *Handbook of Social Psychophysiology* (eds H. Wagner and A. Manstead), John Wiley & Sons, 1989, pp. 143–164.

[20] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2): 97–115, 2001.

[21] D. J. Kim and Z. Bien, "Design of 'personalized' classifier using soft computing techniques for 'personalized' facial expression recognition," *IEEE Transactions on Fuzzy Systems*, 16(4): 874–885, 2008.

[22] Y. Huang, Y. Li, and N. Fan, "Robust symbolic dual-view facial expression recognition with skin wrinkles: local versus global approach," *IEEE Transactions on Multimedia*, 12(6): 536–543, 2010.

[23] H. Kobayashi and F. Hara, "The recognition of basic facial expressions by neural network," *Trans. Soc. Instrum. Contr. Eng.*, 29(1): 112–118, 1993.

[24] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 36(1): 96–105, 2006.

[25] T. Otsuka and J. Ohya, "Spotting segments displaying facial expression from image sequences using HMM," in Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, April 14–16, 1998, pp. 442–447.

[26] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

[27] P. Ekman, J. Hager, C. H. Methvin, and W. Irwin, *Ekman–Hager Facial Action Exemplars*, Human Interaction Laboratory, University of California, San Francisco, CA.

[28] P. Ekman and W. F. Friesen, "The repertoire of nonverbal behavioral categories—origins, usage, and coding," *Semiotica*, 1: 49–98, 1969.

[29] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in *Nebraska Symposium on Motivation* (ed J. Cole), University of Nebraska Press, 1972, pp. 207–283.

[30] P. Ekman and W. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, CA, 1978.

[31] P. Ekman and E. L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System*, Oxford University Press, Oxford, 1997.

[32] P. Ekman, *Face of Man: Universal Expression in a New Guinea Village*, Garland, New York, 1982.

[33] J. Cohn and P. Ekman, "Measuring facial action by manual coding, facial EMG, and automatic facial image analysis," in *Handbook of Nonverbal Behavior Research Methods in the Affective Sciences* (eds J. A. Harrigan, R. Rosenthal, and K. Scherer), Oxford University Press, New York, 2005, pp. 9–64.

[34] P. Ekman, T. S. Huang, T. J. Sejnowski, and J. C. Hager (eds), *Final Report to NSF of the Planning Workshop on Facial Expression Understanding*, Human Interaction Laboratory, University of California, San Francisco, 1993.

[35] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, 36: 253–263, 1999.

[36] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21: 974–989, 1999.

[37] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System*, The Human Face, Salt Lake City, UT, 2002.

[38] P. Ekman, "Darwin, deception, and facial expression," *Ann. New York Acad. Sci.*, 1000: 105–221, 2003.

[39] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12): 1424–1445, 2000.

[40] M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human–computer interaction," *Proceedings of the IEEE*, 91(9): 1370–1390, 2003.

[41] M. Pantic and L. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 34(3): 1449–1461, 2004.

[42] M. Pantic and I. Patras, "Dynamics of facial expressions—recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 36(2): 433–449, 2006.

[43] M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 42(1): 28–43, 2012.

[44] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn, "Spontaneous vs. posed facial behavior: automatic analysis of brow actions," in Proceedings of the ACM International Conference on Multimodal Interfaces, Banff, Canada, November 2–4, 2006, pp. 162–170.

[45] M. F. Valstar, M. Pantic, and H. Gunes, "A multimodal approach to automatic recognition of posed vs. spontaneous smiles," in Proceedings of the ACM International Conference on Multimodal Interfaces, Nagoya, Japan, November 12–15, 2007, pp. 38–45.

[46] M. Pantic and M. Bartlett, "Machine analysis of facial expressions," in *Face Recognition* (eds K. Delac and M. Grgic), I-Tech Education and Publishing, Vienna, Austria, 2007, pp. 377–416.

[47] M. K. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filters," *J. Am. Stat. Assoc.*, 94(446): 590–599, 1999.

[48] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in Proceedings of the IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, July 6, 2005, pp. 317–321.

[49] M. F. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the MMI facial expression database," in Proceedings of the International Conference on

Language Resources and Evaluation, Workshop on EMOTION, Valletta, Malta, May 17–23, 2010, pp. 317–321.

[50] I. Patras and M. Pantic, "Tracking deformable motion," in Proceedings of IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, Hawaii, October 10–12, 2005, pp. 1066–1071.

[51] I. Patras and M. Pantic, "Particle filtering with factorized likelihoods for tracking facial features," in Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, South Korea, May 17–19, 2004, pp. 97–102.

[52] Y. L. Tian, T. Kanade, and J. F. Cohn, *Handbook of Face Recognition*, Springer, New York, 2005.

[53] J. Lien, T. Kanade, J. Cohn, and C. Li, "Subtly different facial expression recognition and expression intensity estimation," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, June 23–25, 1998, pp. 853–859.

[54] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, March 26–30, 2000, pp. 46–53.

[55] J. F. Cohn and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles," *J. Wavelets, Multi-Resolution Inf. Process.*, 2(2): 121–132, 2004.

[56] J. F. Cohn, J. F. Reed, Z. Ambadar, J. Xiao, and T. Moriyama, "Automatic analysis and recognition of brow actions in spontaneous facial behavior," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 1, 2004, pp. 610–616.

[57] J. F. Cohn, "Foundations of human computing: facial expression and emotion," in Proceedings of the ACM International Conference on Multimodal Interfaces, Vol. 1, 2006, pp. 610–616.

[58] A. Chakraborty and A. Konar, *Emotional Intelligence: A Cybernetic Approach*, Springer, 2009.

[59] A. Chakraborty, A. Konar, and A. Chatterjee, "Controlling human moods by fuzzy logic," *Int. J. Math Sci.*, 6(3–4): 401–416, 2007.

[60] A. Chakraborty, A. Konar, U. Chakraborty, and A. Chatterjee, "Emotion recognition from facial expressions and its control using fuzzy logic," *IEEE Transactions on Systems, Man and Cybernetics, Part A (Systems and Humans)*, 39(4): 726–743, 2009.

[61] A. Halder, A. Konar, R. Mandal, P. Bhowmik, A. Chakraborty, and N. R. Pal, "General and interval type-2 fuzzy face-space approach to emotion recognition," in IEEE Transactions on Systems, Man and Cybernetics, Part A (Systems and Humans), 2011.

[62] A. Halder, R. Mandal, A. Konar, and L. C. Jain, "Fuzzy based hierarchical algorithm for template matching in emotional facial images," *J. Intell. Fuzzy Syst.*, 2012.

[63] B. Biswas, A. Konar, and A. K. Mukherjee, "Image matching with fuzzy moment descriptors," *Eng. Appl. Artif. Intell.*, 14: 43–49, 2001.

[64] A. Halder, P. Rakshit, S. Chakraborty, A. Konar, E. Kim, and A. K. Nagar, "Reducing uncertainty in interval type-2 fuzzy sets for qualitative improvement in emotion recognition from facial expressions," in Proceedings of the IEEE International Conference on Fuzzy Systems, Brisbane, Australia, June 10–15, 2012.

[65] A. Halder, R. Mandal, A. Chakraborty, A. Konar, and R. Janarthanan, "Application of general type-2 fuzzy set in emotion recognition from facial expression," in Joint International Conference on Fuzzy and Neural Network (FANCCO), Vishakhapatnam, India, 2011.

[66] A. Halder, A. Jati, G. Singh, A. Konar, A. Chakraborty, and R. Janarthanan, "Facial action point based emotion recognition by principal component analysis," in SocPros, Indian Institute of Technology, Roorkey, India, 2011.

[67] A. Konar, A. Chakraborty, A. Halder, R. Mandal, and R. Janarthanan, "Interval type-2 fuzzy model for emotion recognition from facial expression," in PerMIn, CDAC, Kolkata, India, 2012.

[68] A. Chakraborty, P. Bhowmik, S. Das, A. Halder, A. Konar, and A. K. Nagar, "Correlation between stimulated emotion extracted from EEG and its manifestation on facial expression," in IEEE Systems, Man and Cybernetics Conference, San Antonio, TX, 2009.

[69] S. Das, A. Halder, P. Bhowmik, A. Chakraborty, A. Konar, and A. K. Nagar, "Voice and facial expression based classification of emotion using linear support vector machine," in International Conference on Developments in eSystems Engineering (DESE), Abu Dhabi, UAE, 2009.

[70] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6): 681–685, 2001.

[71] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, 16(1): 172–187, 2007.

[72] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2): 153–158, 1997.

[73] P. Yang, Q. Liu, and D. N. Metaxas, "Boosting encoded dynamic features for facial expression recognition," *Pattern Recogn. Lett.*, 30(2): 132–139, 2009.

[74] J. Whitehill and C. Omlin, "Haar features for FACS AU recognition," in Proceedings of the Seventh IEEE International Conference on Automatic Face and Gesture Recognition, Southampton, UK, April 10–12, 2006, pp. 97–101.

[75] Y. Stylianou, "Modeling of speech signal for analysis purposes or mathematical modeling of jitter and shimmer", in eTERFACE, 2008.

[76] S. H. Chen and Y. R. Luo, "Speaker verification using MFCC and support vector machine," in International Multi Conference of Engineers and Computer Scientists (IMECS), Vol. 1, Hong Kong, 2009.

[77] K. R. Aida-Zade, C. Ardil, and S. S. Rustamov, "Investigation of combined use of MFCC and LPC features in speech recognition systems," *Int. J. Comput. Inf. Syst. Control Eng.*, 2(7): 6–12, 2008.

[78] C. Busso, S. Lee, and S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE Transactions on Speech and Audio Processing*, 17(4): 582–596, 2009.

[79] C.-M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing*, 13(2): 293–303, 2005.

[80] C. H. Wu and W. B. Liang, "Emotion recognition of affective speech based on multiple classifiers using acoustic–prosodic information and semantic labels," *IEEE Transactions on Affective Computing*, 2(1): 10–21, 2011.

[81] E. H. Kim, K. H. Hyun, S. H. Kim, and Y. K. Kwak, "Improved emotion recognition with a novel speaker-independent feature," *IEEE/ASME Transactions on Mechatronics*, 14(3): 317–325, 2009.

[82] E. Mower, M. J. Mataric, and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles," *IEEE Transactions on Speech and Audio Processing*, 19(5): 1057–1070, 2011.

[83] E. Mower, S. Lee, M. J. Mataric, and S. Narayanan, "Human perception of synthetic character emotions in the presence of conflicting and congruent vocal and facial expressions," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, NV, March 30–April 4, 2008, pp. 2201–2204.

[84] E. Mower, S. Lee, M. J. Mataric, and S. Narayanan, "Joint-processing of audio–visual signals in human perception of conflicting synthetic character emotions," in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Hannover, Germany, June 23–26, 2008, pp. 961–964.

[85] E. Mower, M. J. Mataric, and S. Narayanan, "Selection of emotionally salient audio–visual features for modeling human evaluations of synthetic character emotion displays," in Proceedings of IEEE International Symposium on Multimedia (ISM), Berkeley, CA, December 15–17, 2008.

[86] E. Mower, M. J. Mataric, and S. Narayanan, "Human perception of audio–visual synthetic character emotion expression in the presence of ambiguous and conflicting information," *IEEE Transactions on Multimedia*, 11(5): 843–855, 2009.

[87] C. Busso and S. S. Narayanan, "Interrelation between speech and facial gestures in emotional utterances: a single subject study," in IEEE Transactions on Speech and Audio Processing, Vol. 15, No. 8, 2007, pp. 1–16.

[88] S. Ananthakrishnan and S. Narayanan, "Automatic prosody labeling using acoustic, lexical, and syntactic evidence," *IEEE Transactions on Speech and Audio Language Processing*, 16(1): 216–228, 2008.

[89] C. Busso, S. Lee, and S. Narayanan, "Using neutral speech models for emotional speech analysis," in Proceedings of Interspeech'07 Eurospeech, Antwerp, Belgium, 2007, pp. 2225–2228.

[90] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions speech and multimodal information," in International Conference on Multimodal Interfaces, State College, PA, 2004, pp. 205–211.

[91] C. Busso and S. Narayanan, "Interplay between linguistic and affective goals in facial expression during emotional utterances," in Proceedings of the Seventh International Seminar Speech Production (ISSP'06), Ubatuba-SP, Brazil, 2006, pp. 549–556.

[92] C. Busso and S. Narayanan, "Joint analysis of the emotional fingerprint in the face and speech: a single subject study," in Proceedings of the IEEE International Workshop on Multimedia Signal Process (MMSP 2007), Chania, Crete, Greece, 2007, pp. 43–47.

[93] C. Busso and S. Narayanan, "Recording audio–visual emotional databases from actors: a closer look," in Proceedings of the Second International Workshop Emotion: Corpora for Research on Emotion and Affect, International Conference Language Resources and Evaluation (LREC 2008), Marrakech, Morocco, 2008, pp. 17–22.

[94] M. Grimm, K. Kroschel, E. Mower, and S. Narayanan, "Primitives-based evaluation and estimation of emotions in speech," *Speech Commun.*, 49: 787–800, 2007.

[95] C. H. Wu and G. L. Yan, "Speech act modeling and verification of spontaneous speech with disfluency in a spoken dialogue system," *IEEE Transactions on Speech and Audio Processing*, 13(3): 330–344, 2005.

[96] J. C. Lin, C. H. Wu, and W. L. Wei, "Error weighted semi-coupled hidden Markov model for audio–visual emotion recognition," *IEEE Transactions on Multimedia*, 14(1): 142–156, 2012.

[97] C. H. Wu, Z. J. Chuang, and Y. C. Lin, "Emotion recognition from text using semantic label and separable mixture model," *ACM Transactions on Asian Language Information Processing*, 5(2): 165–182, 2006.

[98] C.-H. Wu and Z.-J. Chuang, "Emotion recognition from speech using IG-based feature compensation," *Int. J. Comput. Linguist. Chin. Lang. Process.*, 12(1): 65–78, 2007.

[99] B. Schuller, "Recognizing affect from linguistic information in 3D continuous space," *IEEE Transactions on Affective Computing*, 2(4): 192–205, 2011.

[100] B. Schuller, B. Vlasenko, F. Eyben, M. Wollmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: variances and strategies," *IEEE Transactions on Affective Computing*, 1(2): 119–131, 2010.

[101] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004, pp. 17–21.

[102] B. Schuller, C. Hage, D. Schuller, and G. Rigoll, "Mister D.J., Cheer Me Up!': musical and textual features for automatic mood classification," *J. New Music Res.*, 39(1): 13–34, 2010.

[103] M. Schröder, R. Cowie, D. Heylen, M. Pantic, C. Pelachaud, and B. Schuller, "Towards responsive sensitive artificial listeners," in Proceedings of the Fourth International Workshop on Human-Computer Conversation, 2008.

[104] B. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," in Proceedings of the International Speech Communication Association, 2009, pp. 312–315.

[105] B. Schuller, R. Müller, M. Lang, and G. Rigoll, "Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles," in Proceedings of the International Conference on Speech Communication Association, 2005, pp. 805–808.

[106] A. Batliner, B. Schuller, S. Schaeffler, and S. Steidl, "Mothers, adults, children, pets-towards the acoustics of intimacy," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, pp. 4497–4500.

[107] B. Schuller, "Speaker, noise, and acoustic space adaptation for emotion recognition in the automotive environment," in 2008 ITG Conference on Voice Communication (SprachKommunikation), 2008, pp. 1–4.

[108] B. Schuller, G. Rigoll, S. Can, and H. Feussner, "Emotion sensitive speech control for human–robot interaction in minimal invasive surgery," in Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, 2008, pp. 453–458.

[109] B. Schuller, B. Vlasenko, D. Arsic, G. Rigoll, and A. Wendemuth, "Combining speech recognition and acoustic word emotion models for robust text-independent emotion recognition," in Proceedings of the International Conference on Multimedia and Expo, 2008.

[110] B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll, "On the influence of phonetic content variation for acoustic emotion recognition," in Proceedings of the Fourth IEEE

Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems, 2008.

[111] B. Schuller, B. Vlasenko, R. Minguez, G. Rigoll, and A. Wendemuth, "Comparing one and two-stage acoustic modeling in the recognition of emotion in speech," in Proceedings of the IEEE Workshop Automatic Speech Recognition and Understanding, 2007, pp. 596–600.

[112] B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll, "Combining frame and turn-level information for robust recognition of emotions within wpeech," in Proceedings of the INTERSPEECH, 2007, pp. 2249–2252.

[113] B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll, "Frame vs. turn-level: emotion recognition from speech considering static and dynamic processing," in Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ed. A. Paiva), 2007, pp. 139–147.

[114] A. Batliner, S. Steidl, B. Schuller, D. Seppi, K. Laskowski, T. Vogt, L. Devillers, L. Vidrascu, N. Amir, L. Kessous, and V. Aharonson, "Combining efforts for improving automatic classification of emotional user states," in Proceedings of the First International Conference on Language Technologies, 2006, pp. 240–245.

[115] M. Wöllmer, F. Eyben, S. Reiter, B. Schuller, C. Cox, E. Douglas-Cowie, and R. Cowie, "Abandoning emotion classes—towards continuous emotion recognition with modelling of long-range dependencies," in Proceedings of the Ninth Annual Conference of the International Speech Communication Association, 2008, pp. 597–600.

[116] G. Rigoll, R. Müller, and B. Schuller, "Speech emotion recognition exploiting acoustic and linguistic information sources," in Proceedings of the 10th International Conference on Speech and Computer, Vol. 1, 2005, pp. 61–67.

[117] E. H. Kim, K. H. Hyun, S. H. Kim and Y. K. Kwak, "Improved emotion recognition with a novel speaker-independent feature," *IEEE/ASME Transactions on Mechatronics*, 14(3): 317–325, 2009.

[118] L. Siegel, "A procedure for using pattern classification techniques to obtain a voiced/unvoiced classifier," *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP 27: 83–88, 1979.

[119] S. Ntalampiras and N. Fakotakis, "Modeling the temporal evolution of acoustic parameters for speech emotion recognition," *IEEE Transactions on Affective Computing*, 3(1): 116–125, 2012.

[120] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Transactions on Speech and Audio Processing*, 8(6): 708–716, 2000.

[121] D. Ververidis, C. Kotropoulos, and I. Pitas, "Automatic emotional speech classification," in Proceedings of the 29th International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 04), 2004, pp. 593–596.

[122] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: resources, features, and methods," *Speech Commun.*, 48(9): 1162–1181, 2006.

[123] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77(2): 257–286, 1989.

[124] R. Cowie and R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Commun.*, 40: 5–23, 2003.

[125] E. N. Luengo, I. Hernáez, and J. Sánchez, "Automatic emotion recognition using prosodic parameters," in Proceedings of the INTERSPEECH, 2005, pp. 493–496.

[126] S. G. Kooladugi, N. Kumar, and K. S. Rao, "Speech emotion recognition using segmental level prosodic analysis," in Proceedings of the International Conference on Devices and Communications, 2011, pp. 1–5.

[127] Y. Zhou, Y. Sun, J. Zhang, and Y. Yan, "Speech emotion recognition using both spectral and prosodic features," in International Conference on Information Engineering and Computer Science (ICIECS), Wuhan, December 19–20, 2009.

[128] B. Vlasenko and A. Wendemuth, "Tuning hidden Markov model for speech emotion recognition," in Proceedings of the DAGA, March 2007.

[129] D. Datcu and L. J. Rothkrantz, "The recognition of emotions from speech using gentleboost classifier. A comparison approach," in Proceedings of the International Conference on Computer Systems and Technologies, Vol. 1, 2006, pp. 1–6.

[130] D. Ververidis and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information," in Proceedings of the EUSIPCO, 2004, pp. 341–344.

[131] H. Meng, J. Pittermann, A. Pittermann, and W. Minker, "Combined speech-emotion recognition for spoken human-computer interfaces," in Proceedings of the IEEE International Conference on Signal Processing and Communications, 2007.

[132] D. Morrison, R. Wang, and L. C. De Silva, "Ensemble methods for spoken emotion recognition in call-centres," *Speech Commun.*, 49(2): 98–112, 2007.

[133] O. W. Kwon, K. Chan, J. Hao, and T. W. Lee, "Emotion recognition by speech signals," in Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH), 2003.

[134] A. Khasnobish, "Emotion recognition from stimulated EEG signals", M.E. Thesis, May 2012.

[135] G. Singh, A. Jati, A. Khasnobish, S. Bhattacharyya, A. Konar, D. N. Tibarewala, and R. Janarthanan, "Negative emotion recognition from stimulated EEG signals", in Third International Conference on Computing Communication & Networking Technologies (ICCCNT), July 26–28, 2012.

[136] S. Bhattacharyya, A. Sengupta, T. Chakraborty, D. Banerjee, A. Khasnobish, A. Konar, D. N. Tibarewala, and R. Janarthanan, "EEG controlled remote robotic system from motor imagery classification", in Third International Conference on Computing Communication & Networking Technologies (ICCCNT), July 26–28, 2012.

[137] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Transactions on Affective Computing*, 1(2): 81–97, 2010.

[138] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Transactions on Information Technology in Biomedicine*, 15(5): 737–746, 2011.

[139] Y. P, Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, 57(7): 1798–1806, 2010.

[140] J. J. B. Allen, J. A. Coan, and M. Nazarian, "Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion," *Biol. Psychol.*, 67(1/2): 183–218, 2004.

[141] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognit. Emot.*, 15(4): 487–500, 2001.

[142] W. Heller, "Neuropsychological mechanisms of individual differences in emotion, personality and arousal," *Neuropsychology*, 7: 476–489, 1993.

[143] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction, Foundations and Applications*, Springer, New York, 2006.

[144] R. Palaniappan, "Utilizing gamma band to improve mental task based brain-computer interface design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(3): 299–303, 2006.

[145] E. Altenmuller, K. Schurmann, V. K. Lim, and D. Parlitz, "Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns," *Neuropsychologia*, 40(13): 2242–2256, 2002.

[146] A. Choppin, "EEG-based human interface for disabled individuals: emotion expression with neural networks," Master's Thesis, Tokyo Institute of Technology, 2000.

[147] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," in Proceedings of the Second International Conference on Autonomous Robots and Agents, 2004, pp. 186–191.

[148] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: arousal evaluation using EEG and peripheral physiological signals," Technical report, University of Geneva, 2005.

[149] D. O. Bos. EEG-based emotion recognition: the influence of visual and auditory stimuli, 2006, http://hmi.ewi.utwente.nl/verslagen/capita-selecta/CS-Oude_Bos-Danny.pdf. Last accessed on Sept. 18, 2014.

[150] Z. Khalili and M. Moradi, "Emotion detection using brain and peripheral signals," in Proceedings of the Biomedical Engineering Conference, 2008, pp. 1–4.

[151] R. Horlings, D. Datcu, and L. J. M. Rothkrantz, "Emotion recognition using brain activity," in Proceedings of the International Conference on Computer Systems and Technologies, 2008, pp. 1–6.

[152] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidin and D. Hazry, "Lifting scheme for human emotion recognition using EEG," in Proceedings of the International Symposium on Information Technology, 2008, pp. 1–7.

[153] K. Schaaff and T. Schultz, "Towards an EEEG-based emotion recognizer for humanoid robots," in Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009, pp. 792–796.

[154] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE Transactions on Systems, Man and Cybernetics*, 37(3): 311–324, 2007.

[155] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, 2(2): 106–118, 2011.

[156] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *Int. J. Hum. Comput. Stud.*, 59(1–2): 213–225, 2003.

[157] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Affective Computing and Intelligent Interaction*, Volume 4738: Lecture Notes in Computer Science (eds A. Paiva, R. Prada, and R. W. Picard), Springer, Berlin, 2007, pp. 71–82.

[158] A. Camurri, B. Mazzarino, and G. Volpe, "Analysis of expressive gesture: the eyesweb expressive gesture processing library," in *Gesture-Based Communication in*

Human–Computer Interaction, Volume 2915: Lecture Notes in Computer Science (eds A. Camurri and G. Volpe), Springer, Heidelberg, 2004, pp. 460–467.

[159] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud, "Design and evaluation of expressive gesture synthesis for embodied conversational agents," in Proceedings of the Third International Joint Conference on Autonomous Agents Multi-Agent Systems, 2005, pp. 1095–1096.

[160] A. Ramamoorthy, N. Vaswani, S. Chaudhury, and S. Banerjee, "Recognition of dynamic hand gestures," *Pattern Recogn.*, 36: 2069–2081, 2003.

[161] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in Proceedings of the IEEE International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, June 1995, pp. 296–301.

[162] H. S. Yoon, J. Soh, Y. J. Bae, and H. S. Yang, "Hand gesture recognition using combined features of location, angle and velocity," *Pattern Recogn.*, 34: 1491–1501, 2001.

[163] J. Daugman, "Face and gesture recognition: an overview," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 675–676, 1997.

[164] M. De Meijer, "The contribution of general features of body movement to the attribution of emotions," *J. Nonverbal Behav.*, 13: 247–268, 1989.

[165] H. G. Wallbott, "Bodily expression of emotion," *Eur. J. Soc. Psychol.*, 28: 879–896, 1998.

[166] N. Bianchi-Berthouze and A. Kleinsmith, "A categorical approach to affective gesture recognition," *Connect. Sci.*, 15: 259–269, 2003.

[167] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *J. Netw. Comput. Appl.*, 30(4): 1334–1345, 2007.

[168] H. Gunes and M. Piccardi, "Creating and annotating affect data bases from face and body display: a contemporary survey," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2006, pp. 2426–2433.

[169] H. Gunes and M. Piccardi, "Observer annotation of affective display and evaluation of expressivity: face vs. face-and-body," in Proceedings of the HCSNet Workshop on Use Vision Human–Computer Interaction, 2006, pp. 35–42.

[170] Z. Zeng, J. Tu, M. Liu, T. S. Huang, B. Pianfetti, D. Roth, and S. Levinson, "Audio-visual affect recognition," *IEEE Transactions on Multimedia*, 9(2): 424–428, 2007.

[171] E. Mower, M. J. Matarić, and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles," *IEEE Transactions on Speech and Audio Processing*, 19(5): 1057–1070, 2011.

[172] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *International Conference on Multimodal Interfaces*, State College, PA, October 2004.

[173] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 39(1): 64–84, 2009.

[174] M. A. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *IEEE Transactions on Affective Computing*, 2(2): 92–105, 2011.

[175] G. Castellano, L. Kessous, and G. Caridakis, "Emotion recognition through multiple modalities: face, body gesture, speech," in *Affect and Emotion in Human–Computer*

Interaction, Volume 4868: Lecture Notes in Computer Science (eds C. Peter and R. Beale), Springer-Verlag, Berlin, 2008, pp. 92–103.

[176] A. Chakraborty, P. Bhowmik, S. Das, A. Halder, A. Konar, and A. K. Nagar, "Correlation between stimulated emotion extracted from EEG and its manifestation on facial expression," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, October 2009.

[177] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10): 1175–1191, 2001.

[178] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, 3(1): 2012.

[179] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," in 2nd International Conference on Autonomous Robots and Agents, Palmerston North, New Zealand, December 2004.

[180] J. Russell and J. Fernandez-Dols, *The Psychology of Facial Expression*, Cambridge University Press, New York, 1997.

[181] M. Cohen, *Perspectives on the Face*, Oxford University Press, Oxford, 2006.

[182] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, 1982.

[183] R. S. Lazarus and B. N. Lazarus, *Passion and Reason: Making Sense of Our Emotions*, Oxford University Press, 1996.

[184] S. Li and A. Jain, *Handbook of Face Recognition*, Springer, New York, 2005.

[185] J. Pittermann, A. Pittermann, and W. Minker, *Handling Emotions in Human–Computer Dialogues*, Springer, 2008.

[186] A. Young, *Face and Mind*, Oxford University Press, Oxford, 1998.

[187] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.

[188] H. Gunes and M. Piccardi, "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior," in Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), 2006.

[189] T. Sucontphunt, X. Yuan, Q. Li, and Z. Deng, "A novel visualization system for expressive facial motion data exploration," in IEEE Conference on Visualization Symposium, PacificVIS '08, 2008, pp. 103–109.

[190] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Speech and Audio Processing*, 16(2): 448–457, 2008.

[191] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in Proceedings of the European Conference on Computer Vision, Vol. 2, 1998, pp. 581–695.

[192] C. L. Huang and Y. M. Huang, "Facial expression recognition using model-based feature extraction and action parameters classification," *J Vis. Commun. Image Represent.*, 8(3): 278–290, 1997.

[193] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12): 1357–1362, 1999.

[194] W. Sato, T. Kochiyama, S. Yoshikawa, and M. Matsumura, "Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis," *Cogn. Neurosci. Neuropsychol.*, 12(4): 709–714, 2001.

[195] M. S. Bartlett and T. J. Sejnowski, "Independent components of face images: a representation for face recognition," in Proceedings of the 4th Annual Joint Symposium on Neural Computation, 1997.

[196] C. F. Chuang and F. Y. Shih, "Recognizing facial action units using independent component analysis and support vector machine," *Pattern Recogn.*, 39(9): 1795–1798, 2006.

[197] M. S. Bartlett and T. J. Sejnowski, "Viewpoint invariant face recognition using independent component analysis and attractor networks," *Neural Inf. Process. Syst. Nat. Synth.*, 9: 817–823, 1997.

[198] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, 13(6): 1450–1464, 2002.

[199] J. H. Holland, "Outline for a logical theory of adaptive systems," *J. ACM*, 9(3): 297–314, 1962.

[200] J. H. Holland, "Genetic algorithms and the optimal allocation of trials," *SIAM J. Comput.*, 2(2): 88–105, 1973.

[201] J. H. Holland, "Erratum: genetic algorithms and the optimal allocation of trials," *SIAM J. Comput.*, 3(4): 326, 1974.

[202] R. A. Fisher, *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford, 1930.

[203] R. Storn, "System design by constraint adaptation and differential evolution," *IEEE Transactions on Evolutionary Computation*, 3(1): 22–34, 1999.

[204] H. Kobayashi and F. Hara, "Facial interaction between animated 3D face robot and human beings," in Proceedings of the International Conference on Systems, Man, and Cybernetics, 1997, pp. 3732–3737.

[205] C. Padgett and G.W. Cottrell, "Representing face images forn emotion classification," in Proceedings of the Conference on Advances in Neural Information Processing Systems, 1996, pp. 894–900.

[206] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor wavelets-based facial expression recognition using multi-layer perceptron," in Proceedings of the International Conference on Automatic Face and Gesture Recognition, 1998, pp. 454–459.

[207] H. Kobayashi and F. Hara, "Recognition of six basic facial expressions and their strength by neural network," in IEEE International Workshop on Robot and Human Communication, 1992, pp. 381–386.

[208] J. Zhao and G. Kearney, "Classifying facial emotions by neural networks with fuzzy inputs," in Proceedings of the Conference on Neural Information Processing, Vol. 1, 1996, pp. 454–457.

[209] M. Rosenblum, Y. Yacoob, and L. Davis, "Human expression recognition from motion using a radial basis function network architecture," *IEEE Transactions on Neural Networks*, 7(5): 1121–1138, 1996.

[210] L. W. Chew, K. P. Seng, L. M. Ang, V. Ramakonar, and A. Gnanasegaran, "Audio-emotion recognition system using parallel classifiers and audio feature analyzer," in Third International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM), 2011, pp. 210–215.

[211] S. Zhang, X. Zhao, and B. Lei, "Spoken emotion recognition using radial basis function neural network," *Adv. Comput. Sci., Environ., Ecoinf., Edu.*, 214: 437–442, 2011.

[212] B. Takacs and H. Wechsler, "Locating facial features using SOFM," in Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 2, Conference B: Computer Vision and Image Processing, 1994.

[213] J. He and H. Leich, "Speech trajectory recognition in SOFM by using Bayes theorem," in International Symposium on Speech, Image Processing and Neural Networks, 1994, pp. 109–112.

[214] R. Khosrowabadi, H. C. Quek, A. Wahab, and K. K. Ang, "EEG-based emotion recognition using self-organizing map for boundary detection," in 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 4242–4245.

[215] G. Caridakis, K. Karpouzis, C. Pateritsas, A. Drosopoulos, A. Stafylopatis, and S. Kollias, "Hand trajectory based gesture recognition using self-organizing feature maps and markov models," in IEEE International Conference on Multimedia and Expo, 2008, pp. 1105–1008.

[216] T. Tabatabaei, S. Krishnan, and A. Guergachi, "Emotion recognition using novel speech signal features," in Proceedings of the IEEE International Symposium on Circuits and Systems, 2007, pp. 345–348.

[217] C. Ye, J. Liu, C. Chen, M. Song, and J. Bu, "Speech emotion classification on a riemannian manifold," in Advances in Multimedia Information Processing - PCM, 2008, pp. 61–69.

[218] O. W. Kwon, K. Chan, J. Hao, and T. Lee, "Emotion recognition by speech signals," in Proceedings of the 8th European Conference on Speech Communication and Technology, 2003, pp. 125–128.

[219] B. Schuller, R. Mller, M. Lang, and G. Rigoll, "Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles," in Proceedings of the Interspeech, 2005, pp. 805–809.

[220] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Eng. Appl. Artif. Intell.*, 1–9, 2007.

[221] C. Y. Chang, J. S. Tsai, C. J. Wang, and P. C. Chung, "Emotion recognition with consideration of facial expression and physiological signals," in IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 2009, pp. 278–283.

[222] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," in Proceedings of the 6th International Conference on Neural Information Processing (ICONIP '99), Vol. 2, 1999, pp. 495–501.

[223] J. M. Mendel and D. Wu, *Perceptual Computing*, IEEE Press and Wiley Publications, 2010.

[224] N. Amir and S. Ron, "Towards an automatic classification of emotions in speech," in Proceedings of the ICSLP, 1998, pp. 555–558.

[225] M. Grimm, E. Mower, K. Kroschel, and S. Narayanan, "Combining categorical and primitives-based emotion recognition" in Proceedings of the 14th European Signal Processing Conference, 2006.

[226] B. S. Kang, C. H. Han, S. T. Lee, D. H. Youn, and C. Lee, "Speaker dependent emotion recognition using speech signals," in Proceedings of the ICSLP, 2000, pp. 383–386.

[227] T. L. Nwe, F. S. Wei, and L. C. De Silva, "Speech-based emotion classification," in Proceedings of the TENCON, Vol. 1, 2001, pp. 297–301.

[228] T. S. Polzin, "Detecting verbal and non-verbal cues in the communications of emotions," Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 2000.

[229] L. C. De Silva and P. C. Ng, "Bimodal emotion recognition," in Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 332–335.

[230] Y. Yoshitomi, S. Kim, T. Kawano, and T. Kitazoe, "Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face," in Proceedings of the Ninth IEEE International Workshop on Robot and Human Interactive Communication, 2000, pp. 178–183.

[231] C. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, and S. Narayanan, "Emotion recognition based on phoneme classes" in Proceedings of the Interspeech, 2004, pp. 205–211.

[232] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.* 14(1): 1–37, 2008.

[233] S. Scherer, F. Schwenker, and G. Palm, "Classifier fusion for emotion recognition from speech," in Proceedings of the International Conference on Intelligent Environments, 2007, pp. 152–155.

[234] M. H. Yang, D. Roth, and N. Ahuja, "A SNoW-based face detector," in Proceedings of the Neural Information Processing Systems, Vol. 12, 2000, pp. 855–861.

[235] F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," in Proceedings of the International Conference on Spoken Language Processing, 1996, pp. 1970–1973.

[236] N. Sebe, M. S. Lew, I. Cohen, A. Garg, and T. S. Huang, "Emotion recognition using a Cauchy Naive Bayes classifier," in 16th International Conference on Pattern Recognition, Vol. 1, 2002, pp. 17–20.

[237] O. Lowery, Technical Report of Silent Sounds. Inc., 2004.

[238] R. Huber, A. Batliner, J. Buckow, E. Noth, V. Warnke, and H. Niemann, "Recognition of emotion in a realistic dialogue scenario," in Proceedings of the International Conference on Spoken Language Processing, 2000, pp. 665–668.

[239] J. D. Mayer, M. Dipaolo, and P. Salove, "Perceiving affective content in ambiguous visual stimuli: a component of emotional intelligence," *J Pers. Assess.*, 772–781, 1990.

[240] T. Baumgartner, M. Esslen, and L. Jäncke, "From emotion perception to emotion experience: emotions evoked by pictures and classical music," *Int. J. Psychophysiol.*, 34–43, 2005.

[241] A. Chakraborty and A. Konar, "Stability and chaos in cognitive reasoning," in *Reasoning in Intelligent Systems* (ed. J. Nakamatshu), World Scientific, 2010.

[242] M. Ghosh, A. Chakraborty, A. Konar, and A. Nagar, "Detection of chaos and limit cylces in the emotional dynamics from the facial expression of subjects stimulated with audio-visual movies," in *Computational Intelligence, Control and Computer Vision in Robotics and Automation* (ed. B. Subudhi), Narosa Publisher.

[243] I. Kanluan, M. Grimm, and K. Kroschel, "Audio-visual emotion recognition using an emotion space concept," in 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25–29, 2008.

[244] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. D. Friederici, and S. Koelsch, "Universal recognition of three basic emotions in music," *Curr. Biol.*, 19(7): 573–576, 2009.

[245] P. Eckerdal and B. Merker "'Music' and the 'action song' in infant development: an interpretation," in *Communicative Musicality. Exploring the Basis of Human Companionship* (eds S. Malloch and C. Trevarthen), Oxford University Press, Oxford, 2009, pp. 241–262.

[246] S. Koelsch, T. Fritz, D. Y. von Cramon, K. Muller, and A. D. Friederici, "Investigating emotion with music: an fMRI study." *Hum. Brain Mapp.*, 27: 239–250, 2006.

[247] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, 44: 293–304, 2007.

[248] P. N. Juslin, "Communicating emotion in music performance: a review and a theoretical framework," in *Music and Emotion: Theory and Research* (eds P. N. Juslin and J. A. Sloboda), Oxford University Press, New York, 2001, pp. 309–337.

[249] V. Bettadapura, "Face expression recognition and analysis: the state of the art", Technical Report, arXiv:1203.6722, April 2012.

[250] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, 7: 1–30, 2006.

[251] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, 10(7): 1895–1923, 1998.

[252] B. Abbound, F. Davoine, and M. Dang, "Facial expression recognition and synthesis based on an appearance model," *Signal Processing: Image Communication*, 19(8): 723–740, 2004.

[253] M. Wang, Y. Iwai, and M. Yachida, "Expression recognition from time-sequential facial images by use of expression change model," in Proceedings of the International Conference on Automatic Face and Gesture Recognition, 1998, pp. 324–329.

[254] R. Mandal, A. Halder, P. Bhowmik, A. Chakraborty, and A. Konar, "Uncertainty management in type-2 fuzzy face-space for emotion recognition," in Proceedings of the IEEE International Conference on Fuzzy Systems, Taiwan, China, 2011.

[255] R. Mandal, A. Halder, A. Konar, and A. K. Nagar "A fuzzy condition sensitive hierarchical algorithm for template matching in emotionally expressive facial images", in IPCV'11, Las Vegas, Nevada, July 18–21, 2011.

[256] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1): 34–58, 2002.

[257] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recogn.*, 25(1): 65–77, 1992.

[258] A. Konar, *Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of the Human Brain*, CRC Press LLC, January 2000, 788 pages, ISBN 0-8493-1385-6.

[259] A. Konar, *Computational Intelligence: Principles, Techniques and Applications*, Springer, Heidelberg, March 2005, 708 pages, ISBN 3-540-20898-4.

[260] A. Konar and L. C. Jain, *Cognitive Systems Engineering: A Distributed Computational Intelligence Approach*, Springer, London, July 2005, 350 pages, ISBN 10: 1-85233-975-6.

[261] A. Bhattacharya, A. Konar, and A. K. Mandal, *Parallel and Distributed Logic Programming: Towards the Design of a Framework for the Next Generation Database Machines*, Springer, 2006, 291 pages, ISBN 10: 3-540-33458-0.

[262] R. W. Picard, *Affective Computing*, MIT Media Laboratory Perceptual Computing Section, Cambridge, MA, Technical Report 321, November 1995.

[263] A. Halder, A. Chakraborty, A. Konar, and A. K. Nagar, "Computing with words model for emotion recognition by facial expression analysis using interval type-2 fuzzy sets," in FUZZ-IEEE, Hyderabad, India, 2013.

[264] M. J. Hasegawa, S. Levinson, and T. Zhang, "Children's emotion recognition in an intelligent tutoring scenario," in Proceedings of the Interspeech, 2004.

AUTHOR BIOGRAPHIES

Amit Konar received the B.E. degree from Bengal Engineering and Science University, Howrah, India, in 1983, and the M.E., M.Phil., and Ph.D. (Engineering) degrees from Jadavpur University, Kolkata, India, in 1985, 1988, and 1994, respectively.

In 2006, he was a visiting professor with the University of Missouri, St. Louis. He is currently a professor with the Department of Electronics and Telecommunication Engineering (ETCE), Jadavpur University, where he is the founding coordinator of the M.Tech. program in Intelligent Automation and Robotics. He has supervised fifteen Ph.D. theses. He has over 200 publications in international journals and conference proceedings.

He is the author of eight books, including two popular texts *Artificial Intelligence and Soft Computing* (CRC Press, 2000) and *Computational Intelligence: Principles, Techniques and Applications* (Springer, 2005). He serves as the Associate Editor of *IEEE Transactions on Systems, Man and Cybernetics, Part A* and *IEEE Transactions on Fuzzy Systems*. His research areas include the study of computational intelligence algorithms and their applications to the domains of electrical engineering and computer science. Specifically, he has worked on fuzzy sets and logic,' neurocomputing, evolutionary algorithms, Dempster–Shafer theory, and Kalman filtering, and has applied the principles of computational intelligence in image understanding, VLSI design, mobile robotics, pattern recognition, brain–computer interfacing, and computational biology. He was the recipient of All India Council for Technical Education (AICTE) accredited 1997–2000 Career Award for Young Teachers for his significant contribution in teaching and research.

Anisha Halder received the B.Tech. degree in Electronics and Communication Engineering from the Haldia Institute of Technology, Midnapore, India, and the M.E. degree in Control Engineering from Electronics and Telecommunication Engineering Department, Jadavpur University, Kolkata, India, in 2007 and 2009 respectively. She is currently pursuing her Ph.D. degree under the guidance of Prof. Amit Konar and Dr. Aruna Chakraborty at Jadavpur University.

Her principal research interests include artificial intelligence, pattern recognition, cognitive science, and human–computer interaction. She is the author of over 25 papers published in top international journals and conference proceedings.

Aruna Chakraborty received the M.A. degree in Cognitive Science and the Ph.D. degree in Emotional Intelligence and Human–Computer Interactions from Jadavpur University, Kolkata, India, in 2000 and 2005, respectively. She is currently an associate professor in the Department of Computer Science & Engineering, St. Thomas' College of Engineering & Technology, Kolkata. She is also a visiting faculty of Jadavpur University, where she offers graduate-level courses on intelligent automation and robotics, and cognitive science.

She has coauthored a book with Prof. A. Konar titled *Emotional Intelligence: A Cybernetic Approach* (Springer, Heidelberg, 2009). She serves as an editor to the *International Journal of Artificial Intelligence and Soft Computing*, Inderscience, UK. Her current research interests include artificial intelligence, emotion modeling, and their applications in next-generation human–machine interactive systems. She is a nature lover, and loves music and painting.