Contemporary Challenges for a Social Signal processing

Dr.T.KishoreKumar[1] and K.Sunilkumar[2]

[1]Department of Electronics and Communication Engineering,NIT Warangal,A.P,India
kishorefr@gmail.com
[2] Department of Electronics and Communication Engineering, VITS Hasanparthy,Warangal,A.P, India
sunil_veena10@yahoo.co.in

ABSTRACT

This paper provides a short overview of Social Signal Processing. The exploration of how we react to the world and interact with it and each other remains one of the greatest scientific challenges. Latest research trends in cognitive sciences argue that our common view of intelligence is too narrow, ignoring a crucial range of abilities that matter immensely for how people do in life. This range of abilities is called social intelligence and includes the ability to express and recognize social signals produced during social interactions like agreement, politeness, empathy, friendliness, conflict, etc., coupled with the ability to manage them in order to get along well with others while winning their cooperation. Social Signal Processing (SSP) is the new research domain that aims at understanding and modeling social interactions (human-science goals), and at providing computers with similar abilities in human-computer interaction scenarios (technological goals). SSP is in its infancy and the journey towards artificial social intelligence and socially-aware computing is still long, the paper outlines its future perspectives and some of its most promising applications.

KEYWORDS

Behavioral science, human computer interaction, emotion recognition.

1. INTRODUCTION

Social Signal Processing (SSP) is the new pioneering domain aimed at bringing social intelligence to computers. This is one of the multiple facets of human intelligence and can be thought of as the ability of dealing effectively with social interactions, whether this means to be accepted as leader in a working environment, to be an understanding parent, to be respected in a community, or to capture the attention of the audience. Since humans spend most of their life being involved in social interactions, social intelligence is definitely a key ability that can make the difference between success and failure in life.

Social signal processing aims at providing computers with the means of analyzing and adequately representing human social signals, which will allow them to adapt and properly function in social settings .Of course, the ultimate aim is not to create a digital dinner Companion, although for the lonesome scientist that may be a worthy goal in itself. Social signal processing promises to benefit a host of domains and makes many applications possible. For ambient intelligence, it means environments that are more responsive to the social context [1].For psychologists, it promises quantitative evaluation tools that can be used in coaching or diagnosis. For entertainment technology, more engaging games can be envisioned. And most importantly, new human-computer interaction challenges can be met by greatly increasing the sensitivity of the computer (or of the robot) to the interacting Person's emotional and mental state [2], [3].

Computer-mediated communication is now common place for most of us, and while several methods are developed to transmit social signals over its usually simplified channels (like emoticons), the richness of face-to-face communication and of social signals transmitted during real conversations is largely lacking. Even when the medium offers conditions similar to traditional communication media (for instance a phone interface for a travel agency), an automated reply system is at a disadvantage, and may be perceived as cold and un effective, instead of efficient and capable.

Bridging the social intelligence gap has two main aspects. Consider for a moment the automatic response system. To appear as human-like as possible, it should be able to analyze the incoming signals for their semantic and affective content, but it should also be able to produce appropriate affective responses in turn. The analysis and synthesis aspects allowed social signal processing researchers to create application settings with rigorous evaluation criteria right from the start, and ecological validity concerns introduced additional computational constraints to the already formidable challenge. Many issues are still open in the field, including the inherent uncertainty of machine detectable evidences of human behavior, multi scale temporal dynamics, and the appropriate psychological and cognitive theories that can provide useful concepts and models.

In this paper, we briefly describe this domain, its contemporary challenges. We refer the reader to previous surveys for an in-depth overview of the domain, as well as for historical insight into its development [4], [5]. We will focus here on the more recent developments, and present a short note of the field as it stands now.

2. SOCIAL SIGNALS

Poggi and D'Errico define a social signal as a communicative or informative signal that, either directly or indirectly, conveys information about social actions, social interactions, social emotions, social attitudes and social relationships [6].In [5], taxonomy is introduced for the analysis of social Signals. Verbal signals that are usually direct manifestations of communicative intent are accompanied by *nonverbal behavioural cues* that serve to convey information about emotion, personality, status, dominance, regulation, rapport, etc. in a given social context. These cues reside in different modalities (or *codes*), namely physical appearance, gestures and postures, face and gaze behaviour, vocal behaviour, and use of space and environment.

Nonverbal behavioural cues are at the core of social signal processing (SSP). They typically describe temporal muscular and physiological changes that occur over short time intervals. Some cues last for milliseconds and are therefore difficult to perceive, but still play a role: An example is the movement of orbicularis oculi muscle on the face, which can be used to distinguish real smiles from posed ones [7]. These cues are perceived by humans during communication (with remarkable accuracy) consciously or unconsciously, and can radically alter the interpretation of the situation: A slight muscle movement or an inflection in the voice may add sarcasm to an otherwise innocent comment.

Paul Ekman and Wallace Friesen, building on their research in the 60s, as well as on earlier work by Efron [8], have classified nonverbal cues according to their origin, usage, and coding [9].

The type of messages conveyed by such cues is:

- *Emblems:* Signs that have direct verbal translation and used consciously, like the gesture of cutting one's throat. They are employed as substitutes for verbal signals, or to emphasize them.
- *Illustrators:* Signs that emphasize speech by providing visualization of spatial and temporal aspects, like illustration of a timeline, pointing, or a sketch of an object drawn in the air.
- *Affect Displays*: Signs that display more intimate and personal states, like emotions shown on the face.
- *Regulators*: Signs that coordinate the timing of other signals during communication. Turn-taking cues and backchannel signals are of this type.
- *Adaptors*: Signs that originate from habits, either in a self-manipulative fashion (like wiping the lips with the tongue) or through manipulation of objects (like twirling a pen).

This categorization does not properly do justice to the rich vocal nonverbal behaviours relevant as social signals, like paralinguistic information and voice quality, although some non-linguistic vocalizations, silences, and turn-taking patterns would be considered regulators.

A few dimensions can arguably serve as relevant taxonomical distinctions in SSP. The temporal scale of signals is one such dimension, and it can span a range from milliseconds (as exemplified above with a muscle twitch) to minutes, hours, and much longer in case of behavioural habits. Another relevant distinction is individual vs. dyadic vs. group behaviour. Since we are evolutionary bound to produce social signals, we produce them even when we are alone, but dyadic interactions and multi-party interactions involve different phenomena, like cohesion [10] in groups and postural congruence in dyads [11].A third distinction pertains to the depth of analysis. We can classify a face into smiling/non-smiling classes, but we can go deeper and classify whether a smile is real or posed, or whether it carries hints of sarcasm or pity.

We will review recent work in SSP from two aspects here. We will use taxonomy in terms of signal channels, and report developments as per modality (i.e. faces and eyes, vocal behaviour, gestures, interaction geometry, and multimodal cues). There are social signals hidden in other modalities, for instance in appearance features (e.g. somatic types, make-up and clothing), but we will not describe these here. In applications we will take up a complementary application perspective, and list developments in several application areas.

While we have define here the domain of SSP as actual physical behaviour of humans (and of synthetic agents that primarily mimic them), human social signals are not restricted to the real world. Movement and interaction patterns of people traced via mobile phones, chatting and micro-blogging behaviour, connection formations over social networking platforms, and behaviours exhibited by avatars in virtual worlds all involve social signals, albeit of a different nature. The study of such virtual social signals goes by the name of *computational social science* [12].

3. SOCIAL SIGNALS ANALYSIS

3.1 Physical Appearance

To the best of our knowledge, the detection of people appearance has been addressed in relatively few works. These were never aimed at inferring social information and focused rather on biometric and surveillance applications. Several approaches have proposed measures of facial attractiveness based on symmetry and respect of canonical proportions in the geometry of face landmarks (eyes, nose tip, corners of mouth, brows, etc.) [13][14], while others have rather pointed on the adherence to "average" facial models. The modelling of the overall appearance of individuals (color of clothes, skin, hair, etc.) has been investigated for identification purposes in [15].

3.2 Face and eyes

Faces convey information about gender, age, and emotions of a person, which are valuable sources in social signal processing. Of these the most important area for SSP is facial expression analysis.

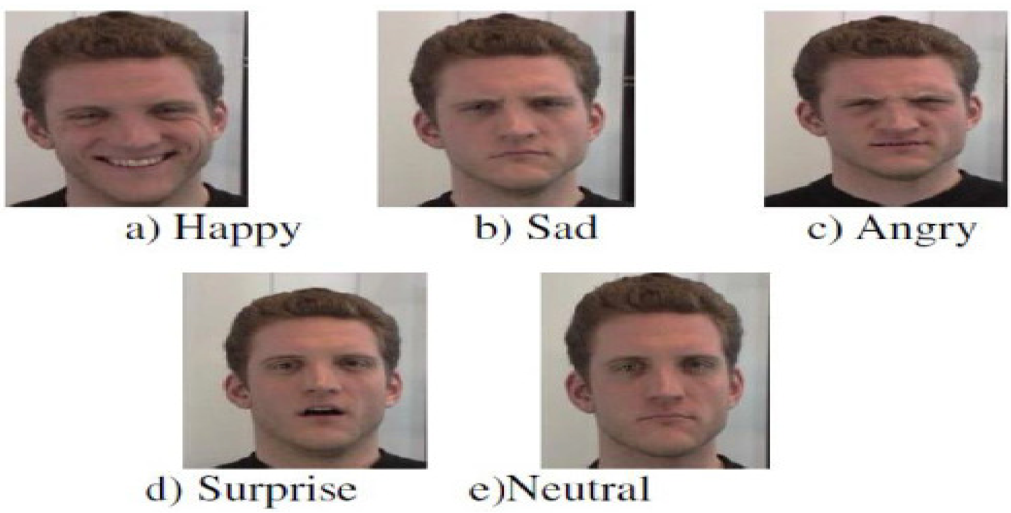

Fig. 1:Human Emotions

Human face reflects that how he/she feels or in which mood he/she is. Humans are capable of producing thousands of facial actions during communication that vary in complexity, intensity, and meaning. Emotion or intention is often communicated by subtle changes in one or several discrete features. Fig1 is showing human Emotions.

The facial expression analysis places emphasis on identifying Facial Actions (FACS), evaluation of expression in natural settings, as opposed to posed expressions and a more detailed analysis of the temporal evolution of expressions as opposed to analysis from static images. Methods of processing facial affect are extensively reviewed in [16]. As the complexity of the classification problem grows, there is greater need for incorporating domain-specific knowledge into the learning system. In a recent work on facial action detection, Zhu *et al.* achieve this by a smarter training set selection for subsequent learning through a dynamic cascade bidirectional bootstrapping scheme and report some of the best results so far in AU detection on the RU-FACS database [17].Face analysis is also used for performing mutual gaze following and joint attention actions. Joint attention is the ability of coordination of a common point of reference with the communicating party. This skill is investigated in [18] for interaction with a robot, and in [19] for interaction with a virtual agent. Gaze direction is important in face-mediated affect, as direct gaze communicates threat or friendliness and plays a role in the expression of joy and anger, whereas averted gaze facilitates avoidance-oriented expressions like fear and sadness.

One aspect of faces that received attention recently is the stereo typical judgments people base on face images. The appearance of a face can invoke (in an unjustified way) feelings of trust, warmth, confidence, etc. Alexander Todorov and his colleagues did an experiment in 2004, where they presented subjects with photographs of US general election candidates, for brief periods of time. From the ensuing competence judgments, they were able to predict election results with accuracy close to 70% [20].These findings demonstrate that facial appearance can act as a strong social signal, and even though the stereotypical judgments based on facial appearance do not have an objective basis (i.e. competent people do not necessarily look competent), they can be used to predict people's responses to them. Automatic analysis of faces in natural settings depends on accurate face registration and facial feature tracking, and these are active research topics.

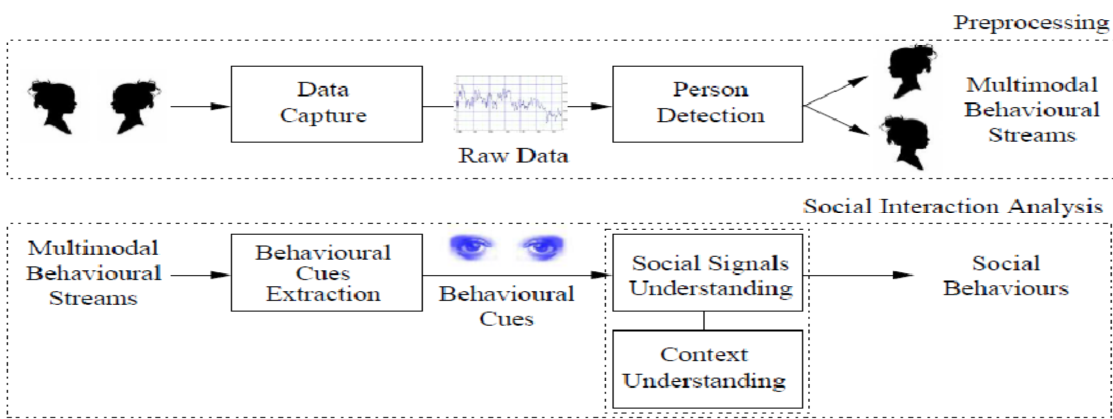

Fig 2: General scheme of an SSP approach

The problem of machine analysis of human social signals includes two major stages (see Figure 2): pre-processing and social interaction analysis. The pre-process*ing* includes data capture (the recording of the scene with multiple sensors) and detection of the people in the observed scene. In the case of microphones and cameras, the sensors most commonly applied, people detection corresponds to speaker diarization and face or full human figure [21] detection. The *social interaction analysis* includes the extraction of audio and/or visual behavioural cues displayed by people detected in the scene, the interpretation of this information in terms of social signals conveyed by the observed behavioural cues, the sensing of the context in which the scene is recorded, and classification of detected social signals into the target social-behaviour interpretative categories in a context-sensitive manner.

The pre-processing stage is based on technologies that have been extensively investigated in the recent years and are not specifically oriented to social interactions (e.g., speaker diarization can be performed for other purposes than the analysis of interactions), while the social interaction analysis stage is the actual *core problem* of SSP and it is still largely unexplored, as discussed in the rest of this section.

3.3 Gesture and Posture

Gesture recognition is an active research domain, but no attempts have been made, to the best of our knowledge, to Interpret gestures in terms of social information, with the exception of few efforts aimed at inferring affective states from gestures [22] . The most common approaches for gesture recognition start by detecting the different body parts (arms, legs, trunk, etc.) using features like the orientation of edge histograms, velocity features extracted with stereo cameras, or pixel colors.

Also automatic posture recognition has been addressed in few works, mostly aiming at surveillance [23] (using multi scale morphological method and Kalman motion estimation) and activity recognition (using an eigen space representation of human silhouettes obtained from Digital Cosine Transform coefficients). However, there are few works where the posture is recognized as a social signal, namely to estimate the interest level of children learning to use computers [24], to recognize the affective state of people [25], and the influence of culture on affective postures [26].

3.4 Vocal Behaviour

Nonverbal vocal behaviour accounts for roughly 50% of the total time in spontaneous conversations [27], thus it has been extensively investigated in speech processing, but only with the goal of improving speech recognition and synthesis systems [28]. In other words, no major efforts have been made, to our knowledge, to interpret nonverbal vocal behaviour in social terms. The five major components of vocal behaviour, namely voice quality, linguistic and non-linguistic vocalizations, silence and turn-taking patterns. The first corresponds to the prosody and accounts for *how* something. is said. The three main prosodic features, called the *Big Three*, are *pitch*, *tempo* and *energy*. The first is the frequency of oscillation of vocal folds during voice emission, the second relates to speaking rate and its variation, and the third is the energy carried by the vocal acoustic waves [28].The pitch is typically obtained by analyzing the Fourier transform of the speech signal from short intervals (in general 30 *ms*) 1. The tempo is measured through the rate of phonetically relevant events like vowels and syllables, or through the first spectral moment of the energy. The energy is a property of any digital signal and corresponds to the sum of the square values of the signal samples. In general the energy is extracted from short analysis windows (30 *ms* like the pitch) [28].

To the best of our knowledge, no efforts have been made to detect non-linguistic vocalizations, with the only exception of laughter for its ubiquitous presence in social interactions. The detection is typically performed by classifying vectors of common speech features like *Mel Frequency Cepstral Coefficients* [28] with models like Gaussian Mixture Models and Neural Networks. Recent approaches have shown that the detection performance can be dramatically improved using multimodal approaches based on both audio and visual features [29][30].

On the contrary, linguistic vocalizations have been extensively investigated to detect hesitations in spontaneous speech with the main purpose of improving speech recognition systems. The disfluencies are typically detected by mapping acoustic observations (e.g. pitch and energy) into classes of interest with classifiers like neural networks or Support Vector Machines. The detection of silence is one of the earliest tasks studied in speech analysis and robust algorithms, based on the distribution of the energy, have been developed since the earliest times of digital signal processing. The last important aspect of vocal behaviour, i.e. the turn taking, is typically a side-product of the speaker diarization, i.e. of the segmentation of speech recordings into single speaker segments .This allows to recognize who speaks with whom and to measure the nonverbal behaviour in correspondence of speaker transitions. Turn-taking has been used to model influence and dominance relationships [31].

3.5 Use of Space and Environment

Physical proximity information has been used in *reality mining* applications as a social cue accounting for the simple presence or absence of interaction between people [32][33]. These works use specially equipped cellular phones capable of sensing the presence of similar devices in the vicinity. The automatic detection of seating arrangements has been proposed as a cue for retrieving meeting recordings in [34]. Several approaches developed in computer surveillance to track people across public spaces can potentially be used to address the detection of social signals in the use of the space.

4. MAIN APPLICATIONS OF SOCIAL SIGNAL PROCESSING

A primary usage of SSP is in analyzing and evaluating interacting humans for certain aspects. This can serve different purposes. For instance, autism and social anxiety disorders cause problems in emitting and interpreting social signals. Psychologists, when they analyze such individuals, record long sessions of interaction and manually annotate these. Automatic analysis tools can reduce the effort of annotation significantly. The social interaction of humans is very rich, and recent research focuses on different aspects like dominance, pride, mirroring, agreement, etc. Each of these aspects can have many indicators. For instance in [35], audio-visual cues are used for detecting cases of agreement and disagreement during an interaction. Among agreement cues, the authors list head nods, listener smiles, eye brow raises, laughter, sideways leaning and mimicry, where as disagreement can be signaled by head shakes, ironic smiles, sarcastic cheek creases, nose flares, leg clamps, and many more. Indeed, the authors list almost 40 different ways of signaling disagreement, through head gestures, facial actions, body posture and actions, auditory cues, hand actions, and gaze. For a thorough understanding of a real interaction, these signals need to be captured.

Dominance is one of the signals that received a lot of attention. In [36] movement-based features from body and face, as well as mouth activity were analyzed in a classification framework to determine dominance in dyadic conversations. In [37], a small meeting scenario is considered, Where audio recordings are also available in addition to camera input. By using visual and audio cues, the authors demonstrate that audio cues are more successful than visual cues for establishing dominance, but the fusion of both improves over audio only.

The skills of a good tutor incorporate social skills to motivate, challenge, bolster, and intrigue students. One application of SSP is in creating computer systems for tutoring, which require implementing ways to emulate those skills. This means the student should be probed for signals of interest, boredom, curiosity, etc.

Another application is the assessment of the coach (or teacher), as these motivational devices need to be used properly and in a timely manner. Finally, the subject of coaching can be a social signal itself. The integration of real-time social signal processing into expert systems opens up new venues for this mature field. In this application nonverbal speech cues are extracted and used for assigning affective labels (absorbed, excited, interested, joyful, opposed, stressed, sure, thinking, unsure) to short speech segments, as well as for assessing the speech in terms of its perceived qualities (clear, competent, credible, dynamic, persuasive, pleasant), resulting in a novel and useful coaching scenario.

Building intelligent robots that can participate in conversations in a natural way, or to fulfill certain social roles in everyday environments is a great challenge. Among all computer systems, social robotics suffers most from a lack of social skills, as its aims are much more ambitious compared to other applications. These aims include childcare robots, health care robots, and service robots for domestic settings.

Apart from the capacity to analyze social signals, additional requirements for a socially responsive robot are primarily the ability to function in noisy environments, to process multimodal and temporal information in real time, and to produce correct signals at the correct time. This is also called “closing the affective loop,” and it is an issue for robots and virtual agents a like [40].

We should mention here the Robocup@Home initiative1, which is an attempt to make robots more sociable by integrating them into real domestic settings and by giving them simple tasks. The appearance and social behavior of the robot is judged by a jury in this challenge, in addition to sensory motor skills.

Embodied conversational agents (ECAs) require a number of capabilities that use non-verbal social signals for realistic interaction. These include initialization and termination of conversation sessions, turn-taking, and feedback functions [41]. The agent uses facial movements, head motions, and body movements to give these signals. Parametric models of body and limb motions are derived from actual interactions, and transferred to synthetic characters for realistic body and limb movements. For instance, gesture rate and performance can be adjusted to tune the appeared extroversion of a virtual agent [42].

During an interaction with a virtual agent, nonverbal cues can be very dominant. In particular, the sensitive artificial listener (SAL) technique proposes that it may be possible to give adequate responses based on such nonverbal cues, even if what the other party is saying is not understood [43].In [44] facial expressions are combined with movement cues obtained from the shoulder area, as well as with audio cues, to predict emotions in the valence-arousal space for an artificial listener, which monitors the interacting human for affective signals to give appropriate responses in real-time. This work also demonstrates that it is useful to learn correlations between valence and arousal.

Online worlds create novel social spaces where virtual avatars interact with each other. Game developers think of ways of enriching the expressiveness of the avatars, as well as natural ways of transmitting desired social signals from the controllers of the avatars to the actual virtual agent. Taken to the extreme, it is possible to induce affective responses in the real users through avatar interaction. An example application is presented in [45], where the users wear haptic interface devices to transmit social signals automatically through their avatars in Second Life. The HaptiHeart device conveys emotion related heart rates (through speakers positioned on the body), HaptiButterfly creates a fluttering in the stomach via vibration motors, HaptiShiver sends shivers down the spine through a cold airflow, and HaptiTickler induces joy by actually tickling the user in the ribs.

5. CONCLUSION

Social Signal Processing has the ambitious goal of bringing social intelligence in computers. The first results in this research domain have been sufficiently impressive to attract the praise of the technology and business communities. What is more important is that they have established a viable interface between human sciences and engineering - ,social interactions and behaviours, although complex and rooted in the deepest aspects of human psychology, can be analysed automatically with the help of computers. SSP brings together computer science, engineering and social sciences together in a unique way. There are many challenges, especially in creating systems that work on real-world data, and in integrating the numerous findings back into useful applications, but there is also great progress. Finally in this paper the author highlighted the importance of SSP further author discussed various social signals such as non verbal behavioral cues and its analysis, pertaining to physical appearance, face and eyes, gesture, posture and vocal behaviour.Finally elaborated various applications of SSP also its contemporary challenges.

REFERENCES

[1] E. Aarts and J. Encarnac¸ ˜ao, Eds., True Visions: The Emergence of Ambient Intelligence. Springer-Verlag, 2006.
[2] J. Crowley, J. Coutaz, and F. B´erard, "Perceptual user interfaces: thingsthat see," Communications of the ACM, vol. 43, no. 3, pp. 54–64, 2000.
[3] C. Breazeal, "Emotion and sociable humanoid robots," InternationalJournal of Human-Computer Studies, vol. 59, no. 1-2, pp. 119–155,2003.
[4] A. Pentland, "Social Signal Processing," IEEE Signal Processing Magazine,vol. 24, no. 4, pp. 108–111, 2007.
[5] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social Signal Processing:Survey of an emerging domain," Image and Vision Computing Journal,vol. 27, no. 12, pp. 1743–1759, 2009.
[6] I. Poggi and F. D'Errico, "Cognitive modelling of human social signals,"in International Workshop on Social Signal Processing, 2010, pp. 21–26.
[7] H. Dibeklioglu, R. Valenti, A. Salah, and T. Gevers, "Eyes do not lie: spontaneous versus posed smiles," in ACM International Conference on Multimedia, 2010, pp. 703–706.
[8] D. Efron, Gesture and environment. King's Crown Press, 1941.
[9] P. Ekman and W. Friesen, "The repertoire of nonverbal behavior:Categories, origins, usage, and coding," Semiotica, vol. 1, no. 1, pp.49–98, 1969.
[10] H. Hung and D. Gatica-Perez, "Estimating cohesion in small groups using audio-visual nonverbal behavior," IEEE Transactions on Multimedia, vol. 12, no. 6, pp. 563–575, 2010.
[11] T. Chartrand and J. Bargh, "The chameleon effect: The perception–behavior link and social interaction." Journal of personality and social psychology, vol. 76, no. 6, p. 893, 1999.
[12] D. Lazer, A. Pentland, L. Adamic, S. Aral, A. Barabasi, D. Brewer,N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King,M. Macy, D. Roy, and M. Van Alstyne, "Computational social science,"Science, vol. 323, pp. 721–723, 2009.
[13] P. Aarabi, D. Hughes, K. Mohajer, and M. Emami.The automatic measurement of facial beauty. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, pages 2644–2647,2001.
[14] Y. Eisenthal, G. Dror, and E. Ruppin. Facial attractiveness: Beauty and the machine. NeuralComputation, 18(1):119–142, 2005.
[15] T. Darrell, G. Gordon, M. Harville, and J. Woodfill Integrated person tracking using stereo, color, and pattern detection. International Journal of Computer Vision, 37(2):175–185, 2000.

[16] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31,no. 1, pp. 39–58, 2009.
[17] Y. Zhu, F. De la Torre, J. Cohn, and Y. Zhang, “Dynamic cascades with bidirectional bootstrapping for spontaneous facial action unit detection,” in IEEE Conference on Affective Computing and Intelligent Interaction, vol. II, 2009, pp. 1–8.
[18] Z. Yucel and A. Salah, “Resolution of focus of attention using gaze direction estimation and saliency computation,” in IEEE Conference on Affective Computing and Intelligent Interaction, vol. II, 2009.
[19] C. Peters, S. Asteriadis, and K. Karpouzis, “Investigating shared attention with a virtual agent using a gaze-based interface,” Journal on Multimodal User Interfaces, vol. 3, no. 1, pp. 119–130, 2010.
[20] R. Adams and R. Kleck, “Perceived gaze direction and the processing of facial displays of emotion,” Psychological Science, vol. 14, no. 6, pp. 644–647, 2003.
[21] T. Moeslund and E. Granum. A survey of computer vision-based human motion capture. Computer Vision and Image Understanding, 81(3):231–268, 2001.
[22] M. Pantic, A. Pentland, A. Nijholt, and T. Huang.Human-centred intelligent human-computer interaction (HCI2): How far are we from attaining it? International Journal of Autonomous and Adaptive Communications Systems, 1(2):168–187, 2008.
[23] Y. Li, S. Ma, and H. Lu. Human posture recognition using multi-scale morphological method and Kalman motion estimation. In Proceedings of International Conference on Pattern Recognition, pages 175–177,1998.
[24] S. Mota and R. Picard. Automated posture analysis for detecting learner’s interest level. In Proceedings of Conference on Computer Vision and Pattern Recognition, pages 49–56, 2003.
[25] R. De Silva and N. Bianchi-Berthouze. Modeling human affective postures: an information theoretic characterization of posture features. Journal of Computational Animation and Virtual World,15(3-4):269–276, 2004.
[26] A. Kleinsmith, R. De Silva, and N. Bianchi-Berthouze.Cross-cultural differences in recognizing affect from body posture. Interacting with Computers,18(6):1371–1389, 2006.
[27] N. Campbell. Conversational speech synthesis and the need for some laughter. IEEE Transactions on Speech and Language Processing, 14(4):1171–1178, 2006.
[28] X. Huang, A. Acero, and H. Hon. Spoken language processing. Prentice Hall, 2001.
[29] A. Ito, X. Wang, M. Suzuki, and S. Makino. Smile and laughter recognition using speech processing and face recognition from conversation video. In Proceedings of th International Conference on Cyberworlds, pages 437–444, 2005.
[30] S. Petridis and M. Pantic. Audiovisual discrimination between laughter and speech. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 5117–5121, 2008.
[31] J. Curhan and A. Pentland. Thin slices of negotiation:predicting outcomes from conversational dynamics within the first 5 minutes. Journal of AppliedPsychology, 92(3):802–811, 2007.
[32] N. Eagle and A. Pentland. Reality mining: sensing complex social signals. Journal of Personal and Ubiquitous Computing, 10(4):255–268, 2006.
[33] A. Pentland. Automatic mapping and modeling of human networks. Physica A, 378:59–67, 2007.
[34] A. Jaimes, K. Omura, T. Nagamine, and K. Hirata.Memory cues for meeting video retrieval. In Proceedings of Workshop on Continuous Archival and Retrieval of Personal Experiences, pages 74–85, 2004.
[35] K. Bousmalis, M. Mehu, and M. Pantic, “Spotting agreement and disagreement: A survey of nonverbal audiovisual cues and tools,” in International Conference on Affective Computing and Intelligent Interaction, vol. II, 2009, pp. 121–129.
[36] S. Escalera, O. Pujol, P. Radeva, J. Vitria, and M. Anguera, “Automatic detection of dominance and expected interest,” EURASIP Journal onAdvances in Signal Processing, vol. 2010, 2010.
[37] O. Aran and D. Gatica-Perez, “Fusing Audio-Visual Nonverbal Cues to Detect Dominant People in Group Conversations,” in International Conference on Pattern Recognition, 2010, pp. 3687–3690.
[38] M. Lepper, M. Woolverton, D. Mumme, and J. Gurtner, “Motivational techniques of expert human tutors: Lessons for the design of omputerbased tutors,” Computers as Cognitive Tools, pp. 75–105, 1993.

[39] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. McOwan, "Inter-ACT: an affective and contextually rich multimodal video corpus for studying interaction with robots," in ACM International Conference on Multimedia. ACM, 2010, pp. 1031–1034.

[40] "Affect recognition for interactive companions: challenges and design in real world scenarios," Journal on Multimodal User Interfaces, vol. 3, no. 1, pp. 89–98, 2010.

[41] J. Cassell, "Embodied conversational interface agents," Communicationsof the ACM, vol. 43, no. 4, pp. 70–78, 2000.

[42] M. Neff, Y. Wang, R. Abbott, and M. Walker, "Evaluating the effect of gesture and language on personality perception in conversational agents," in Intelligent Virtual Agents, 2010, pp. 222–235.

[43] E. Douglas-Cowie, R. Cowie, C. Cox, N. Amir, and D. Heylen, "The sensitive artificial listener: an induction technique for generating emotionally coloured conversation," in Workshop on Corpora for Research on Emotion and Affect, 2008.

[44] M. 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, in press.

[45] D. Tsetserukou, A. Neviarouskaya, H. Prendinger, M. Ishizuka, and S. Tachi, "iFeel IM: innovative real-time communication system with rich emotional and haptic channels," in International Conference on Human Factors in Computing Systems, 2010, pp. 3031–3036.

AUTHORS

Dr. T. Kishore Kumar received **Ph.D.** degree in the area of signal processing. Presently working as Associate Professor in Department of Electronics & Communication Engineering, National Institute of Technology (NIT), Warangal. He is faculty In-charge for Centre for Automation & Instrumentation, NITW.He has 17 years of experience in teaching. His areas ofinterest are signal processing and speech processing. At present he is guiding 4 PhD Scholars and 4 M.Tech Students. He has published 7 international Journals and 12 international conference papers during last five years. He attended different international conferences held at Czech Republic, Israel &Singapore. He has been deputed to ESIGELEC University,Rouen, Francefor taking assignment in Real Time Signal Processing. He has completed one in House R & D Project in Speech Processing. He has delivered expert lectures for RCI (DRDO) & ECIL Engineers.He is the Member of IEEE and also member of IETE. He is a Reviewer for TMH book on Signals and Systems. He is Member Board of Studiesfor variousinstitutes
Email: kishorefr@gmail.com

K. Sunil kumar received **M.Tech** degree in the area of VLSI DESIGN. Presently working as Asst. Professor in Department of Electronics & Communication Engineering,Vinuthna Institute of Technology and science, Warangal.Email: sunil_veena10@yahoo.co.in