From Smart to Personal Environment: Integrating Emotion Recognition into Smart Houses

Dmitrii Fedotov
Ulm University
Ulm, Germany
ITMO University
Saint Petersburg, Russia
dmitrii.fedotov@uni-ulm.de

Yuki Matsuda
Nara Institute of Science and Technology
Nara, Japan
Research Fellow of Japan Society for the Promotion of Science, Tokyo, Japan.
yukimat.jp@gmail.com

Wolfgang Minker
Ulm University
Ulm, Germany
wolfgang.minker@uni-ulm.de

***Abstract*—Recent advances in computational and sensing technologies allowed to incorporate different devices into a smart systems, making the ubiquitous or pervasive computing a hot topic for research and commercial projects. One technology, that can help the user to interact with invisible system representing smart environment is spoken dialogue system. Following the success in research on automatic speech recognition and natural language understanding, spoken dialogue systems have significantly improved themselves during the past decade and now bringing the communication between human and machine closer to natural level. Having user as a main subject, both system may benefit from explicit information about his current state and mood, adjusting their behaviour to the certain extent. In this paper we consider the combination of ubiquitous computing, spoken dialogue systems, and emotion recognition technologies, suggest possible ways of information flow, discuss future applications and potential problems. We find, that these technologies can be complementary to each other, increasing their flexibility, robustness and intelligibility when combined. We present the usage of such approach in a smart house environment, continuously tracking the state of the user, interacting with them in real time and reacting to mood changes.**

***Index Terms*—ubiquitous computing, emotion recognition, smart environments, spoken dialogue systems**

I. Introduction

Pervasive or ubiquitous computing is a concept and paradigm in information technologies, soft- and hardware engineering, that allows to spread computing to almost any device and made it available insensible and hidden, connecting devices into the complete network, embedded into person's everyday life.

Started around 50-60 years ago, personal computing has gone a long way, evolving from large and expensive machines to something, that can be literally called personal. Several decades ago computer became an explosive increase in popularity and availability, reaching the peak of sales in 2011. Later the focus was shifted to mobile computing, allowing people to have a personal computer in the pocket.

Further development of mobile and sensing technologies introduced wearable devices to a wide range of consumers. Having the sensors and computational power not only in a smartphone, but also on the hand of a person opened the perspectives to improve one's life quality, constantly monitoring his state. Nowadays even simple and inexpensive wearable devices can track the amount and timing of physical activity, sleep phases, etc. and give suggestions on having a better, healthier lifestyle. Some wearables contain more sensors and provide greater possibilities of tracking one's state, including heart rate, blood pressure, and skin temperature monitoring [1].

Making a step further from a desktop and mobile computing, ubiquitous computing eliminates the need of a user or operator to indeed make requests and give commands to a computing system. Such system is present in our lives, without any distraction or interruption. In some sense, following the Theory of Inventive Problem Solving [2], pervasive computing paradigm introduces new functionality without explicit definition of a physical object for it.

Extrapolating the trend of wider technology implementation, as well as increasing level of their invisibility and developments in making faster, more robust, widely spread and more stable connection, pervasive computing has great prospects for the nearest future.

A traditional way of communication between computer and person is giving the commands through a user interface, usually graphical, and controlling output in the same way. However, with a deeper implementation of technologies in our everyday life, it is demanded to make the communication more natural for a person. In real human-human interaction, people use speech to exchange information. It is also desired for the computer to understand the user naturally, without a set of necessary typical instructions, hence demand on spoken dialogue systems (SDS) has emerged.

SDSs started their way several decades ago, evolving faster with the development of technologies, those build a basis for them, such as speech recognition (ASR) and synthesis, natural language understanding (NLU), etc. Widest spread of SDSs through users is achieved with voice assistants in smartphones, such as Siri (Apple), Cortana (Microsoft), Google Assistant (Google) and Alisa (Yandex). These systems allow users to instantly retrieve useful information, as well as give simple

This research was funded by Ministry of Science and Higher Education of the Russian Federation (grant number 2.12795.2018/12.2), Japan Society for the Promotion of Science (JSPS) KAKENHI (grant number 16H01721) and German Academic Exchange Service (DAAD)

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

commands to a smartphone (set an alarm, call somebody from a contact list, plan appointment in a calendar) through a speech interface.

The same approach was applied to combine the ubiquitous computing technologies and spoken dialogue systems in the context of simple smart home concept: Alexa (Amazon), Google Home (Google), HomePod (Apple). These systems incorporate properties of both personal assistant and smart house: they can set an alarm, check the weather, turn off the light and increase the temperature in the room (if appropriate devices are connected to them).

The functionality of SDSs, described above, demands little intelligence from the system and the main condition, that has to be fulfilled, is the high quality of speech recognition. Nevertheless, this task is not solved completely yet, recent advances in speech recognition allowed it to be used with a high level of confidence and be integrated into commercial systems. Modern SDSs are able to communicate with users and sustain a conversation on a wide range of topics.

With the development of NLU approaches, the computer was trained to understand the meaning of phrases and sentences, i.e. the logical part of the exchange during a common interaction. One of the most important components that most SDSs are still lacking, is the one, that would allow them to understand the underlying information, appearing in a conversation: emotions, sarcasm, etc.

Emotion recognition has been a hot topic of research for a long period and a large amount of research was conducted. An automatic affect recognition can be beneficial in a variety of applications in areas of human-computer interaction (HCI) and human-human interaction (HHI). Emotional component in a HCI system allows it to perceive the emotional state of a speaker and adjust the response to increasing the quality of interaction. Nevertheless, the problem is far from being solved. Less than two decades ago, emotion recognition has left the laboratory conditions and faced real-world data and problems; such as cultural, linguistic and environmental differences [3], [4].

Emotion recognition usually deals with several modalities. Most popular and available of them are audio, visual and textual data. Recently more complex, physiological features were introduced for research community [5]. They followed the general trend of shifting from acted to natural, real world, in-the-wild data [6]–[8].

Incorporating data from wearable devices for emotion recognition, it is possible to seamlessly combine these technologies and provide constant monitoring of the user's emotional state. Similarly, it is possible to analyze textual data collected with SDS to detect emotions and use them in a SDS to improve the quality of interaction. Combining the data from both sources allows building a more flexible and robust system, supplementing data when several modalities are available.

From the other hand, feedback information, provided by an emotion recognition system can serve as an input data to both a smart environment and SDS, working together on an improvement of the user's mood.

In this paper, we present the concept of combining these three technologies together. The remainder of the paper follow is below: Section II introduces previous works done in related areas; Section III describes a general scheme of SDS, emotional categories and dimensions, and integration of emotion recognition component to the concept of SDS and smart environments; Section IV describes the possible ways to recognize emotions in the conditions of smart environment; Section V presents the prospects and future direction of development for such systems. Lastly, Section VI provides the conclusions.ons.

II. Related Work

In emotion recognition, audio- and/or visual-based approaches are popular fields in the field of dialogue systems and human-computer interaction [9]. In laboratory (indoor) conditions, existing audio-based emotion recognition systems which use a deep neural network have achieved great performance [10], [11]. Quck et al. proposed an audio-based emotion recognition system [12]. They built a dialogue system on mobile devices, and achieved around 60% recall score for four affective dimensions. Tarnowski et al. proposed an approach based on facial movements [13]. They obtained good classification accuracy of 73% for seven facial expressions. Moreover, they mentioned head movements (orientation) could significantly affect extracting facial expression features. Aiming at higher accuracies, bi-modal emotion recognition methods, combining audio and visual features, are also proposed [14], [15] and they achieved better accuracy (e.g., 91% for six emotion classes).

To infer the emotion of a person, the unconscious behaviour of humans may be a clue as well. Shapsough et al. described that emotions could be recognised by using typing behaviour on the smartphone [16]. This approach used a machine learning technique and induced high accuracy on emotion recognition. Resch et al. proposed the emotion collecting system for the urban planning, called *Urban Emotions* [17]. The wrist-type wearable devices and social media were used for emotion measurements.

Mikuckas et al. suggested a combination of emotion recognition and human computer interaction systems. Their method aims to recognize the stress level of an user using heart rate variability features derived from electrocardiogram.

Previous research on personalizing the smart environment, conducted by Pentland and Cloudhury [18], was concentrated on face recognition and user identification. They emphasized, that it is important for computers to be naturally integrated in human's lives to be ubiquitous. One examples of human-like and natural behaviour is to recognize and identify person by face or voice. Fernandez-Caballero et al. [19] used smart health environment to detect and regulate patient's emotions. They proposed a multimodal system with physiological signals, facial expressions and behavioural features. The final goal was to increase the quality of life and care. They regulated emotions and mood of a person with music and light, collecting feedback to estimate an effect. McKeown et al. [7] studied the emotional state of user talking to a dialogue system or its substitution, given that system have four

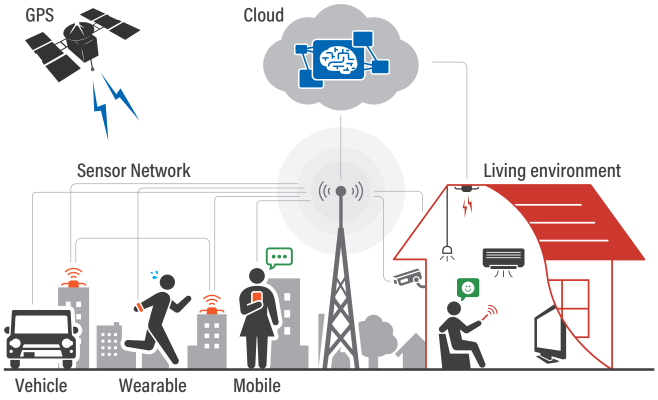

Fig. 1. Ubiquitous Computing Technology

different personalities: angry, calm, cheerful and pragmatic. The personalities of a system may be used to regulate the mood of a user, while having a conversation with him.

In our previous research, we concentrated on recognizing user's emotion and satisfaction level during a sightseeing tour [20], [21]. We found, that it can be achieved with audio, visual and behaviour features even in an outdoor in-the-wild conditions.

III. Proposed Integration Approach

In this paper we discuss the current state and perspectives of combining ubiquitous computing, spoken dialogue systems and emotion recognition. We assume, that these technologies can be complementary to each other and construct a powerful trio together.

A. Ubiquitous computing

Ubiquitous computing enhancing the role of the computer in our lives, making them less visible for the user at the same time [22]. Ubiquitous computing goes from standard personal or desktop computer paradigm, in which user has to find one and directly interact with it, to the usage of computational services anytime, anywhere (see Fig. 1). The ubiquitous computing system comprises of devices, that are able to collect the information; list of supported tasks, that can be performed by the system; infrastructure, that connects the devices to each other and allows information transition. A ubiquitous paradigm based on two main principles: ubiquity, i.e. availability of interaction between user and system at any time; and transparency: the system is smoothly integrated into the environment and do not distract the user.

B. Spoken Dialogue Systems

Spoken dialogue systems serve as a tool for convenient interaction between user and system, while performing information retrieval task, using services or just having a conversation (see Fig. 2). The system takes audio signal as an input at the first step. In order to work with an information, contained in this signal, it has to be transformed into text. To do it, an ASR component has to be implemented. When the information is converted into textul format, the system requires an understanding of the content. Therefore an NLU component should be introduced. At this stage, we already know, what user said and want to find the most appropriate action in our system to make a proper response. A dialogue management (DM) component is responsible for it. Often DM interacts with application-specific subsystems, that can address the information to databases or additional APIs. Now the system can already do the desired action, but it lacks the feedback to a user in order to inform him about the results of completing the task or request additional information if the decision cannot be made yet. To do so, the system needs to make an opposite of each previous step. Formal output has to be reformulated into a human understandable language with text generation component and then "pronounced" by the system with the help of speech synthesis component. Having these components in the system, it is possible to create a loop of dialogue and communicate with the system naturally.

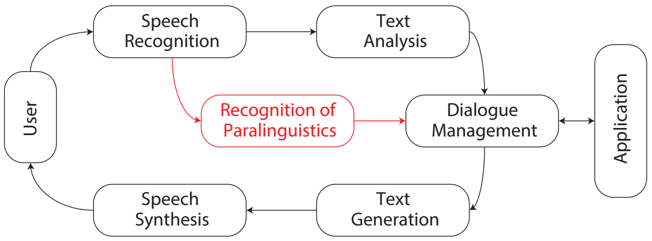

Fig. 2. Architecture of Extended Spoken Dialogue System

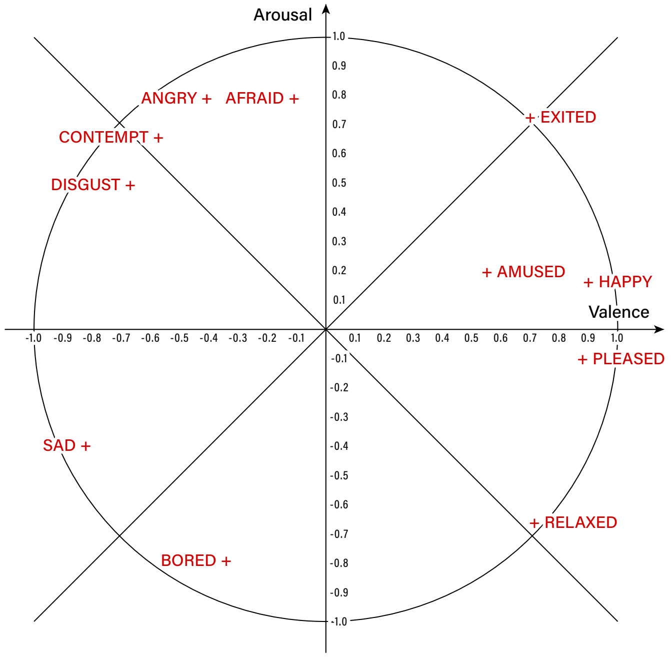

Fig. 3. Arousal-Valence affective space. Based on [23] and [24]

C. Emotion Recognition

Emotion recognition as a branch of research aims to detect an emotional state of a person at the present time. In order to do it, it takes features of some modality (audio, visual, textual, physiological, etc.) as an input and predicts the most probable emotion in categorical or dimensional scale. The range of predicting emotions may be constrained to the basic set: anger, disgust, fear, happiness, sadness, and surprise [25] or presented as a set of continuous values on affective scales (arousal, valence, dominance, etc.) [26]. Most of the research until the 2010s were concentrated on utterance-level emotion recognition and used categorical emotion representation. With a wider spread of continuously annotated corpora with natural emotions, affective scales became more popular, as it is not always possible to strictly define one of the basic emotions in real life conditions. Most popular scales are arousal (activation), which indicates the intensity of emotion; and valence, which represents the positiveness of emotion (see Fig. 3).

D. Combining Technologies

Technologies, discussed above can be combined into pairs and all together, allowing to benefit one from another. Introducing the features of first technology to the second, we can increase the flexibility and performance of it.

One can significantly improve the quality of a SDS, if the system understands not only the information, contained in speech, but also an emotional context. Two modalities can be used in such a system: audio - directly from received audio signals, and textual - the output of the ASR component. In Fig. 2 emotion recognition component is depicted in red. Taking an emotion or mood into account, SDS can adjust its behaviour, selecting appropriate words, phrases or even topics.

Smart environments can also benefit from awareness of the emotional state of a user. Incorporating smart devices, such as wearables, it may track the physiological features continuously, deriving useful information from them and building a general contextual picture of a particular user and group of users. With this knowledge, more comprehensive information may be collected from the environment.

Combining SDS and ubiquitous computing system may make it more transparent, i.e. user may communicate with the system in more natural way, excluding additional distraction and integrating the systems into user's life more smoothly.

IV. Spoken Dialogue Systems and Emotion Recognition in Smart Houses

When it comes to more detailed development of the combination of these systems in the context of smart houses, one should make several decision. Firstly, the amount and types of devices have to be chosen (for example see Fig. 5). To integrate any SDS into a ubiquitous computing system, it requires an audio input and output, i.e. microphones and speakers for communication between the user and SDS. Microphone ranges should cover the environment in order to provide ubiquity. To increase the robustness of emotion recognition, video modality can also be useful; therefore, one may equip the environment with video cameras. Obviously, in contrast to audio-based, performance of visual emotion recognition depends on the position and orientation of the user. It works properly with frontal face images and loses the performance, when it comes to tilts, shifts and turns. It may be hard to impossible to provide enough cameras to get a frontal image of the user at any time; therefore a compromise solution may be to analyze the typical user behaviour in the environment and install the cameras to cover the most popular places. For example it may be in front of the TV or kitchen stove. Additional modalities may be physiological features, collected with smart wearable devices. But in order to be able to analyze this type of data and predict emotions, an appropriate classification or regression system should be trained. To do this, one needs a corpus containing the mapping of features from same wearable devices to emotion of particular range or dimension.

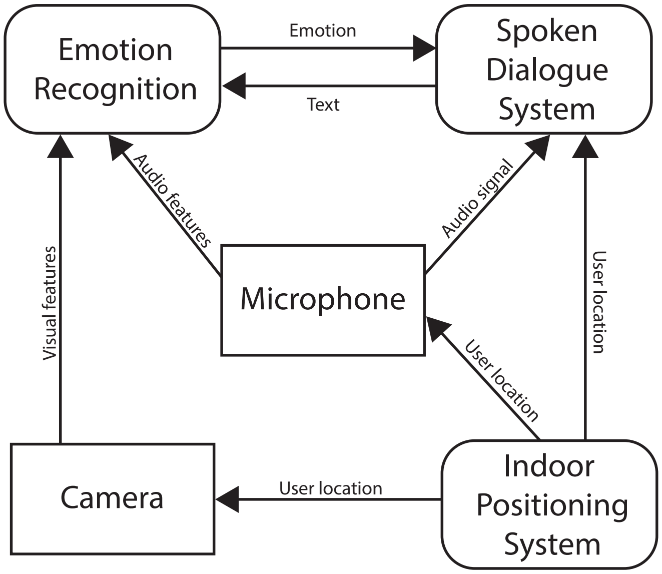

Fig. 4. Data flow in Smart House equipped with Spoken Dialogue System and Emotion Recognition System

Secondly, the data flow between subsystems should be designed. In this paper, we suggest the scheme described below (see Fig. 4). Our personal smart home system should track the user with position sensors. This information is used by microphones and cameras to record and by SDS to reply to the user based on his location in the house. For example, if a user changes the room, the system turns off the sensors and output devices there and turns them on in another room. Using the sensors located in the appropriate room, the system records data and sends it for processing to corresponding modules. SDS takes audio signals to perform speech recognition, as one step of its workflow (see. Fig 2). When speech is recognized and translated to a text, we can use it as an additional modality of emotion recognition module. Making the prediction based on the combination of three available modalities (audio, visual, textual), emotion recognition system sends a feedback to SDS, where it may be used to adjust the dialogue strategy.

Third, more concrete approaches of feature extraction should be defined. For example, from video stream action units (AUs) may be extracted. They are confirmed to be representative features for emotion recognition problem and are easy to extract in real time. AUs describe specific movements of facial muscles in accordance with the facial action coding system (FACS) [27], [28], e.g., AU-1 means an action of "raising up the inner brow." The following 17 AUs are usually used: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, 45. The real-time AUs detection can be performed with open source software OpenFace [29], [30]. The audio signal may be used in waveform directly or features may be extracted. One of the popular methods is to use low-level-descriptors (LLDs) as a set of basic features or extend it by including higher-level, more sophisticated features. As high dimensionality may increase the time for feature extraction and model performance, we suggest using LLDs as feature set [31]. They also can be extracted in real-time with open source software openSMILE [32]. When several people are living together and use the same dialogue system, a face recognition and speaker diarization is required before the feature extraction. With a proper performance level of audio and visual emotion recognition system, it is possible to just tune them a bit to adjust to a particular user. Trained systems don't require much computational power and therefore can process the data locally, keeping the information from cameras and microphones private.

Having the sensors installed and data communication system designed and set up, we can maintain continuous emotional state estimation and dialogue-type communication with smart house system on demand.

V. Future Research and Prospects

Information on the emotional state of a user brings any human-centered system to a new level. If a performance of an emotion recognition system reaches a certain degree of accuracy, a consecutive system may be built on top of it and use emotion not as a label, but as a feature. As human's emotion is often related to an environment, in which he is currently being, one can adjust, modify and correct particular aspects of it in the context of the smart house. Here, the simple techniques of emotion or mood regulation may be applied. For example, we can change the emotional state of the person to a certain extent via visual (color, light, etc.) and sound (music, silence, etc.) effects. It can be used in a silent and invisible system, that tracks human's mood and behaviour and suggests actions to improve one's lifestyle, giving useful advices.

Another method to regulate the emotion of a user is to do it through a dialogue. Psychologically reinforced behaviour of a system and argumentative strategies [33] along with emotion tracking, allow to sustain a conversation within certain limits and motivate, cheer up or relax the user if necessary.

Advances in the area of pervasive computing, wearable devices, dialogue systems, and emotion recognition will make it possible to increase the quality of such systems constantly and move forwards from smart to the personal environment, in which each person feels like home.

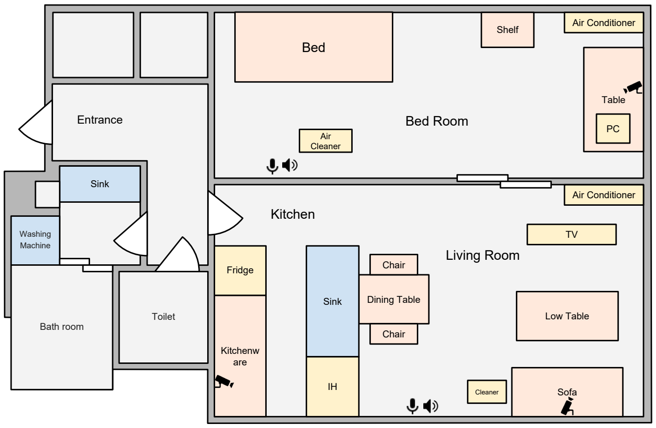

Fig. 5. Smart House Environment

VI. Conclusions

In this paper we discussed the current states and possibility of combining such research directions as ubiquitous computing, (spoken) dialogue systems and emotion recognition in the context of smart environment, namely, smart house. We found, that these technologies may be complementary to each other and benefit form fusion. Creating more complex, sophisticated and robust system, bringing these areas closer together, will boost the advancements in the field of pervasive computing, integrating it in an everyday life faster and seamlessly.

References

[1] Empatica Inc., "Empatica E4," https://www.empatica.com/research/e4/, (accessed: 26 Oct. 2018).

[2] G. Altshuller, G. Altov, and H. Altov, *And suddenly the inventor appeared: TRIZ, the theory of inventive problem solving*. Technical Innovation Center, Inc., 1996.

[3] N. Lim, "Cultural differences in emotion: differences in emotional arousal level between the east and the west," *Integrative medicine research*, vol. 5, no. 2, pp. 105–109, 2016.

[4] 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*, vol. 1, no. 2, pp. 119–131, 2010.

[5] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the recola multimodal corpus of remote collaborative and affective interactions," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–8.

[6] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, 2017, pp. 524–528.

[7] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, "The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 5–17, 2012.

[8] A. Metallinou, C. C. Lee, C. Busso, S. Carnicke, and S. Narayanan, "The USC CreativeIT database: A multimodal database of theatrical improvisation," *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, p. 55, 2010.

[9] C. H. Wu, J. C. Lin, and W. L. Wei, “Survey on audiovisual emotion recognition: databases, features, and data fusion strategies,” *APSIPA transactions on signal and information processing*, vol. 3, p. E12, 2014.

[10] K. Han, D. Yu, and I. Tashev, “Speech emotion recognition using deep neural network and extreme learning machine,” in *15th Annual Conference of the International Speech Communication Association (InterSpeech 2014)*, September 2014.

[11] H. Kaya, A. A. Karpov, and A. A. Salah, “Robust acoustic emotion recognition based on cascaded normalization and extreme learning machines,” in *Advances in Neural Networks - ISNN 2016*, 2016, pp. 115–123.

[12] W. Y. Quck, D. Y. Huang, W. Lin, H. Li, and M. Dong, “Mobile acoustic emotion recognition,” in *2016 IEEE Region 10 Conference (TENCON)*, Nov 2016, pp. 170–174.

[13] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Emotion recognition using facial expressions,” *Procedia Computer Science*, vol. 108, pp. 1175–1184, 2017.

[14] A. Subramaniam, V. Patel, A. Mishra, P. Balasubramanian, and A. Mittal, “Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features,” in *Computer Vision - ECCV 2016 Workshops*. Springer International Publishing, 2016, pp. 337–348.

[15] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, “End-to-end multimodal emotion recognition using deep neural networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, Dec 2017.

[16] S. Shapsough, A. Hesham, Y. Elkhorazaty, I. A. Zualkernan, and F. Aloul, “Emotion recognition using mobile phones,” in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Sept 2016, pp. 1–6.

[17] B. Resch, A. Summa, G. Sagl, P. Zeile, and J.-P. Exner, “Urban emotions – geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data,” in *Progress in Location-Based Services 2014*, 11 2014, pp. 199–212.

[18] A. Pentland and T. Choudhury, “Face recognition for smart environments,” *Computer*, vol. 33, no. 2, pp. 50–55, 2000.

[19] A. Fernandez-Caballero, A. Martínez-Rodrigo, J. M. Pastor, J. C. Castillo, E. Lozano-Monasor, M. T. López, R. Zangróniz, J. M. Latorre, and A. Fernández-Sotos, “Smart environment architecture for emotion detection and regulation,” *Journal of biomedical informatics*, vol. 64, pp. 55–73, 2016.

[20] Y. Matsuda, D. Fedotov, Y. Takahashi, Y. Arakawa, K. Yasumoto, and W. Minker, “Emotour: Estimating emotion and satisfaction of users based on behavioral cues and audiovisual data,” *Sensors*, vol. 18, no. 11, p. 3978, 2018.

[21] ——, “Emotour: Multimodal emotion recognition using physiological and audio-visual features,” in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, ser. UbiComp ’18. ACM, 2018, pp. 946–951.

[22] M. Weiser, “Some computer science issues in ubiquitous computing,” *Communications of the ACM*, vol. 36, no. 7, pp. 75–84, 1993.

[23] G. Paltoglou and M. Thelwall, “Seeing stars of valence and arousal in blog posts,” *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 116–123, 2013.

[24] J. A. Russell, “Core affect and the psychological construction of emotion.” *Psychological review*, vol. 110, no. 1, p. 145, 2003.

[25] P. Ekman, “Basic emotions,” *Handbook of cognition and emotion*, pp. 45–60, 1999.

[26] J. A. Russell, “A circumplex model of affect.” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[27] P. Ekman and W. V. Friesen, *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.

[28] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[29] T. Baltrušaitis, P. Robinson, and L. P. Morency, “Openface: An open source facial behavior analysis toolkit,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016, pp. 1–10.

[30] T. Baltrušaitis, “Openface,” https://github.com/TadasBaltrusaitis/OpenFace, 2017, (accessed: 15 Oct. 2018).

[31] B. W. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval, E. Marchi, and Y. Zhang, “The interspeech 2014 computational paralinguistics challenge: Cognitive & physical load,” in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014, pp. 427–431.

[32] F. Eyben, M. Wöllmer, and B. W. Schuller, “Opensmile: the munich versatile and fast open-source audio feature extractor,” in *Proceedings of the 18th ACM International Conference on Multimedia*, ser. MM ’10. ACM, 2010, pp. 1459–1462.

[33] N. Rach, S. Langhammer, W. Minker, and S. Ultes, “Utilizing argument mining techniques for argumentative dialogue systems,” in *Proceedings of the 9th International Workshop On Spoken Dialogue Systems (IWSDS)*, 2018.