Engagement with Artificial Intelligence through Natural Interaction Models

Sara (Salevati) Feldman
Simon Fraser University
Vancouver, Canada
sara_salevati@sfu.ca

Ozge Nilay Yalcin
Simon Fraser University
Vancouver, Canada
oyalcin@sfu.ca

Steve DiPaola
Simon Fraser University
Vancouver, Canada
sdipaola@sfu.ca

As Artificial Intelligence (AI) systems become more ubiquitous, what user experience design paradigms will be used by humans to impart their needs and intents to an AI system, in order to engage in a more social interaction? In our work, we look mainly at expression and creativity based systems, where the AI both attempts to model or understand/assist in processes of human expression and creativity. We therefore have designed and implemented a prototype system with more natural interaction modes for engagement with AI as well as other human computer interaction (HCI) where a more open natural communication stream is beneficial. Our proposed conversational agent system makes use of the affective signals from the gestural behaviour of the user and the semantic information from the speech input in order to generate a personalised, human-like conversation that is expressed in the visual and conversational output of the 3D virtual avatar system. We describe our system and two application spaces we are using it in – a care advisor / assistant for the elderly and an interactive creative assistant for uses to produce art forms.

Artificial Intelligence. Natural user interfaces. Voice systems. Expression systems. ChatBots.

1. INTRODUCTION

Due to the increase of natural user interfaces and untethered sensor devices, there is a corresponding requirement for computational models that can utilise interactive and affective user data in order to understand and emulate a more natural way for conversational communication. From an emulation standpoint, it is important to understand the mechanisms underlying the human to human multilayered semantic communication to achieve a more natural user experience. Humans tend to make use of gestures and expressions in a conversational setting in addition to the linguistic components that allow them to express more than the semantics of the utterances. This phenomenon is usually disregarded in the current automated conversational systems due to being computationally demanding and requiring a cognitive component to be able to model the complexity of the additional signals. With the advances in the current technology we are now closer to achieve more natural-like conversational systems. Gesture capture and recognition systems for video and sound input can be combined with output systems such as Artificial Intelligence (AI) based conversational tools and 3D modelling

systems in order to achieve human-level meaningful communication. This may allow the interaction to be more intuitive, open and fluent that can be more helpful in certain situations. In this work, we attempt to include the affective components from these input signals in order to generate a compatible and personalised character that can reflect some human-like qualities.

Given these goals, we overview our 3D conversational avatar system and describe its use in our two application spaces, stressing its use where AI systems are involved. Our first application space is CareAdvisor, for maintaining active and healthy aging in older adults through a multi-modular Personalised Virtual Coaching system. Here the natural communication system is better suited for the elderly, who are technologically less experienced, especially when used non-confrontationally and as an assistant conduit to health data from other less conversational devices. Our second application space is in the interactive art exhibition area, where our avatar system is able to converse with users in a more open way, compared to say forms and input systems, on issues of art and creativity. This allows for more open, intuitive conversation leading to an

appreciation of art (art education) as well as work with the computer AI system to produce unique works of art.

Our virtual avatar system can advise, assist, continuously learn and adapt based on user's behaviour by using multimodal affective interaction techniques. Compared to text, beeps, forms and images, our system of human style conversational communication works in a way that can be much better suited for human expression and creativity.

Besides a natural voice based component, our 3D conversational character agent uses several sensor systems that are able track the user's eye and facial analysis as well as body movements with full conversation AI system abilities similar to SIRI or Google Now. This allows for natural conservation dialoguing between the human and avatar system using Natural Language Processing (NLP), and an AI supported personality and topic knowledge base that allows the avatar to converse in an intelligent and natural way.

Our 3D conversational agent uses real-time high end (game level) 3D rendering and animation in the form of natural speech lip sync, facial expressions and hand, arm and body gestures. The system has frameworks that use sensor based communication systems which while still limited, allows the avatar system to perceive its environment and the user in front of it. This occurs via a still simple system of facial, gesture and movement sensors and recognition which are interpreted via semantic properties allowing for novel interactive conversational style communication models for health, education, art / museum exhibits.

Our main focus is to build a human computer interaction (HCI) experience that is more natural for the user, engaging, informative and worthwhile. Though this research we are looking beyond the screen and considering the relationship between digital and physical space through various channels of experiences such as gestural interfaces, social proximity, expressions, and intents. Social interaction does not only promote, but is a prerequisite for intellectual, personal and cultural development (Morrissey 2002).

2. TYPES OF NATURAL INTERACTION

The interaction models we will explore in this paper focus on nature human communication such as dialogue/speech with our intelligent virtual character as well as sensors that recognise human and facial gestures. Through this exploration we aim to reshape the current landscape of multimodal affective interaction and interfaces by fostering the adoption of novel UX and situation awareness

through an emotionally intelligent Artificial Intelligence (AI).

Multimodal interfaces may allow the system to be prone to errors as well as providing a more natural base for interaction. For an extensive survey of many multimodal interaction techniques (Jaime & Sebe 2007), affect detection (Calvo & D'Mello 2010, Zeng et al. 2009). Our goal is to focus on a natural, empathic, personalised and expressive communication model to allow for formation of emotional bonds that result in long-lasting relationships. We want our virtual character to respond efficiently, appropriately and adaptively to its users.

Our virtual 3D character system has avatars (see Figure 1) that are equipped with emotions that humans can read, however, it is not sufficient to only extract and understand emotions. AI needs to have the intelligence to extract the event related facial-voice-eyes-posture states, and to also generate emotions in a way that mirrors human ways of expressing emotions, for natural recognition of the emotions by the user. For our system we used a large available database for emotional faces beyond the seven Ekman expressions (Ekman 2004), and couple it with voice frequencies and pitch to validate the human perception of emotions.

Figure 1: *Virtual 3D characters.*

Emotional expressions are a basic component of human communication and interaction. The interaction between humans and computers will be more natural if computers are able to perceive and respond to human non-verbal communication such as emotions. Picard (1997), in his work Affective Computing describes four categories of affective computing: 1) Systems that cannot express and perceive affect 2) Systems that can express but cannot perceive affect 3) Systems that cannot

express but can perceive affect 4) The systems that can both perceive and express affect. Our aim is to continue to build out our agent system to be able to both perceive and express emotional cues in order to be able to engage in a natural communication. In this section, we explore some ways of obtaining emotional cues from full body gestures, facial gestures and speech content. Moreover, we will introduce our human-like avatar system in order to express emotions during communication. Figure 2 shows our avatar system, in a natural interaction setting. In the following sections we will cover the types of interactions and the technology behind these systems.
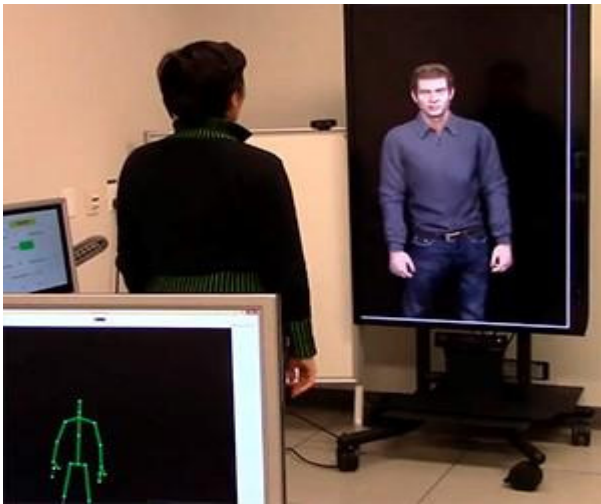

Figure 2: Natural Interaction with our virtual character with Kinect body sensor, camera for facial expression and mic for voice input.

2.1 Emotion recognition from speech signals

Speech, as being another substantial component of human interaction, includes both implicit and explicit forms of information. The explicit information contains the linguistic component, where the meaning and affect of the message can be extracted via careful analysis of the word sequence. This type of information is being investigated in the sentiment analysis research (Lehrer 1974, Turney 2003, Mohammad 2009).

Text analysis has been demonstrated to be a powerful tool to identify emotional states (Clavel 2013). The detection of emotions such as anger, joy, sadness, fear, surprise, and disgust have also been addressed (Bellegarda 2010, Mohammad 2009). However, spoken language is informal and provides information in an unstructured way so that developing tools to select and analyse sentiments, opinions, etc. is still a challenging topic (Justo 2014).

The implicit cues consists of a set of vocal cues (Murray & Arnott 1993) varying from basic prosodic

signals such as pitch, additional excitation signals due to the non-linear air flow in the vocal tract, vocal tract features such as formants, rate of speech and amplitude (Juslin & Scherer 2005, Ververidis & Kotropoulos 2006), prosodic features (Ben-David 2016) such as pitch loudness, voice quality and articulation (Girard and Cohn 2014), latency to speak, pauses. These approaches may also be combined in order to have a more natural and error-prone interaction (Lee & Naranyan 2005, Irastorza 2016).

Some works in the literature consider the combination of acoustic and language information for a more robust automatic recognition of a speaker's emotion (Schuller 2004, Lee et al. 2002) or combine semantic and acoustic features to design continuous dimensional speech affect recognition (Karadogan 2012).

2.2 Emotion recognition from face, eye & gaze analysis

The strong association between facial expressions and affective states have been studied extensively (Ekman & Oster 1977, Russell Bachorowski & Fernandez-Dols 2003). Although several approaches have been proposed to recognise human emotions based on facial expressions or speech, relatively limited work has been done to fuse the two modalities to improve accuracy and robustness of emotion recognition. In recent years, the research community begins to exploit more diverse signals to analyse human emotions. For example, (Busso 2004) show that emotion recognition improves considerably using facial expressions, speech and multimodal Information. Eye tracking has been the optimal technology for studying cognition and human emotion. Eye tracking device will collect information from the eyes to extract physiological measures that correlate to emotional states from eye-dynamics. This includes eye-gaze, which indicates the focus of attention, and shows the level of engagement and analyse the levels of emotion such as stress, interest, or sadness.

3. OUR SYSTEM

Our proposed project is intended to use the multimodal interaction techniques in a 3D full-body avatar system in order to achieve a more natural interaction where users will be more engaged and comfortable with interacting with a system. In this section, we will explain how the recognition systems were used in our prototype and talk about additional capabilities of our avatar system. We then explore our avatar system application within a practical domain of healthcare in a form of personalised virtual coaching system 'CareAdvisor'

and interactive creativity assistant system in the later sections.

3.1 Speech Input

For our prototype we used the linguistic characteristics of the speech input from the users. In order to support the verbal interaction, we propose to use a dialogue-based interaction style rather than unidirectional communication. To achieve this, we use an existing dialogue management application with the aid of an automatic speech recogniser (ASR). To be able to recognise the spoken input, we used a cloud based automatic speech recognition system Google ASR. After receiving the results with the highest probability, we pass these results to our dialogue management system. The dialogue manager is used to conduct an efficient dialogue within the boundaries of a pre-determined scenario. We are currently working on our prototype of the dialogue manager and in the meantime using API.AI (API.AI 2017) service to be able to manage the dialogue flow with the user.

The role of a dialogue manager is to understand the intent of the user and provide the most suitable reply depending on the current context in the dialogue flow. The manager processes the input received from the user and matches it with an intent. Intents are topic areas that can be used to control the flow of conversation by acting as a mapping between the user and the set of responses. Depending on the application area, the desired flow of conversation can be built into the dialogue manager. In our study, we will be providing two examples of application areas. For each of these areas a separate conversational flow is needed to be generated.

Furthermore, to be able to capture the affective properties of the provided spoken input, we used a sentiment analysis tool to contribute a richer interaction. In our work, we used Python NLTK Tool (Bird et al. 2009) with Naive Bayes Classifier and trained over sentiment dataset (Hutto & Gilbert 2014) which is consisted of sentence samples that has positive, negative and neutral sentences.

3.2 Gestural Input

For a face recognition software, we used OpenCV (Bradski 2000) for extracting Points of Interests (POIs) from the face image obtained from the real time video recordings. We used the extended Cohn-Kanade dataset for the emotion recognition which includes 486 sequences across 97 subjects (Lucey et al. 2010), which is the extended version of the original dataset (Kanade, Cohn & Titan 2000). Eight different categories for emotions are labelled in this dataset: angry, disgust, fear, happy,

sadness, surprise, contempt and neutral. For the initial model, we used four basic emotions derived from Russel's proposed circumplex model (1980) that categorises emotions by two axes of arousal and valence. We used four basic emotion types in each of the quadrants represented in Russel's model. These emotions are: happy, sad, fear and neutral/calm. Similarly, we have the same level of analysis for the user full body gestures using an overhead camera and a Microsoft Kinect motion capture system.
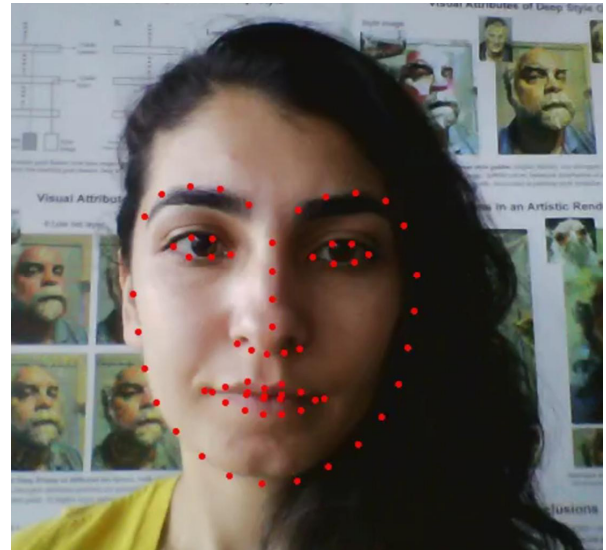

Figure 3: *Face recognition system for emotion tracking.*

Figure 3 shows an example of the points of interest (POI) in our face recognition system. Emotions are recognised by training the system on the relative positions of the POI over series of images.

3.3 Speech Output

We are able to personalise the avatar's voice either to the goal of the project or to user preference. The system uses two different voice systems that both have advantages and disadvantages. The Text to Speech (TTS) voice system has the disadvantage of being less natural sounding (although progress is being made especially with new Deep Neural Networks approaches) while have the advantage of being full programmable – that is any test that comes out of our AI dialogue system can be in real time spoken.

The other voice system uses real audio of pre-recorded speech – both these systems have automated lip-sync to them, but while this audio based system can have perfect inflection and emotion since it is created by a voice actor – it has the disadvantage of huge time and labour of pre-recording at times, hundreds of short sentences and of course the audio can not be manipulated

like the TTS system can. Either system has the controls to sync in a natural way with the facial and gesture animation of the avatar. The type of voice, and gestures are highly programmable and must be authored to create the impression of a personality type.

Our system can make use of both approaches, while the current TTS system can be used in a wider scope. Our project uses the built-in TTS and lip-synching capabilities of SmartBody (Thiebaux 2008) character animation platform. In the current setting, some conversations may sound mechanical and may need more adjustments.

3.4 Interactive 3D Output

Our system uses a fully body realistic game like 3D avatar with full facial and gesture parameterised animated control. We partner with the University of Southern California's open source effort called SmartBody (Thiebaux 2008) that uses an XML standard language called BML – Behavior markup language. BML allows use to send time based commands to the avatar system allowing use to programmatically control every aspect of facial and body gesture including conversational level human gesturing.

Figure 4: *Face expression, lip sync, natural breath control and social eye gaze are part of the system.*

Some reactive behavior such as shifting body posture, gaze or self-touching behavior as responses to the events happening in the environment and idle movement patterns during the lack of external input increases the quality of the interaction (Bernardet & DiPaola 2015, Bernardet et al. 2016). The consistency of these movements allows the system to generate a sense of personality in affective cognitive agents (Saberi, Bernardet & DiPaola 2014). Our system makes use of these bodily movements to contribute to a more natural interaction with the user. Besides body movements the system has full facial expression controls using the advanced Facial Action Coding (FACS) system (Ekman 2004) and

of course full lip sync (see Fig. 4). Recently our research group has added natural breathing and social eye gaze mechanism to the system.

3.5 Programmatic control

Building a virtual avatar system that interacts with humans through multimodal affective interaction techniques require a complex framework to coordinate the different techniques in a meaningful way. An example of achieving this type of coordination can be seen in previous work that controls gaze, bodily and facial gestures (Bernardet, Saberi & DiPaola 2016). The system includes the extension of such framework to include the necessary emotion recognition tools that provides important information on how to behave, as well as the dialogue component.

All these systems come together currently controlled by our AI dialog system which takes voice input questions from the user and both generates appropriate response, sending those responses through to the voice and gesture avatar system. At the same time, user facial and body gesture and positioning can be analysed to inform the system. We have research into advanced processes of human breathing and social eye gaze to make the conversational agent more natural and believable. All these processes are informed by psychology knowledge of how human's gesture, use emotion and have conversations.

4. A CASE STUDY: MEDICAL ADVISOR

Our first prototype, CareAdvisor, we explore new paradigms and platforms for maintaining active and healthy aging in older adults. CareAdvisor through emotional interactions and exchanges acts as a friend/companion that engages elders in sustaining health status (cognitive, social, physical activities), keeping them away from loneliness and enhancing quality of life. It acts as an integrative personalised virtual coaching system to provide guidance and care to older adults through empowerment and motivation in day-to-day activities without attention theft.

CareAdvisor uses remote non-intrusive technologies such as face analytics, adaptive spoken dialogue systems and natural language interfaces to extract physiological markers of emotional states in real-time for adaptive responses. This prototype system uses advances in holistic modeling of behavioral, computational, physical and social aspects to create an expressive virtual coach that is human like and emotionally responsive. This virtual coach through the power of narrative can advise and assist and continuously learn and adapt based on user's behavior and the information that it collects through various means.

CareAdvisor integrates seamlessly and effortlessly in user's life. The use of non-intrusive technology such as the humanistic virtual coach and the augmented everyday objects influences the user behavior without the necessity of their conscious attention. CareAdvisor acts as a friend by understanding the emotional status and the wishes of the user; and is able to establish a personal and long-lasting relationship with them. It understands their weaknesses and proposes actions and activities that will re-establish and maintain wellbeing.

Our methodology is iterative, based on participative design, relying on comprehensive consortium that excellence in persuasive technology, intelligent systems and HCI, Healthcare with domain knowledge of local older adults association.

The prototype project demonstration and validation phase with clearly-defined realistic use cases. In this early stage it focuses on evidence-based, user research and integration of intelligent user and context sensing methods through voice, eye and facial analysis, intelligent heuristics (complex interaction, user intention detection, distraction estimation, system decision), visual and spoken dialogue system, and system reaction capabilities. Through measurable end-user validation, will hope to perform additional user studies, with proposed methods and solutions to ensure usefulness, reliability, flexibility and robustness.

The effectiveness of the CareAdvisor system as well its accessibility will be assessed via co-design of our early prototype followed by a more evaluative trial. During the initial phase, the usability, user experience and acceptance of the early interfaces and prototypes will be evaluated through co-design with users. This will generate an iterative cycles of design refinements to be carried out in collaboration with the users and design team. The evaluation of the system as a whole will be accomplished in a longer-term study which includes scenarios walk-throughs preceded by a pilot trial and by several preparatory actions like system installation, recruitments, and protocol.

CareAdvisor prototype is socially innovative as it mindfully encourages and support activities of the aging user across areas of mental, physical and social well-being. It is contextually and emotionally aware of its users and provides advice that enhances social inclusion such as meeting a friend for a walk.

5. A CASE STUDY: INTERACTIVE CREATIVE ASSISTANCE

Our second case study is a 3D virtual agent that supports the creative process of individuals while engaging in a natural conversation and guiding the user to achieve its goals. The Virtual Assistant (VA) makes resources available for the individuals to receive further education in art, including art principles and history. An art library is made available to the users to browse and find information about their favourite pieces. Where necessary, connections to museums and art galleries will be made available online to allow the users to access more information.

Our virtual assistant helps individuals become more creative by enhancing their creative process through immersion. This is achieved by comparing the users' art design and work with a large collection of other pieces to provide a set of creative suggestions for the users to consider. The suggestions come from the inference as the output of creative models such as analogy models and conceptual blending.

The aim of the virtual assistant is to provide an interactive art education experience focusing primarily on the notion of active participation.

The virtual assistant through our AI based ePainterly (DiPaola 2014) system invites users to participate in activities related to a desired art theme, say Futurism or the colour palette and stroking style of Van Gogh, encouraging them to engage their imaginations through exploration of artist's creative process and the act of making an artwork in a specific style and mood. This interactivity acts as a catalyst for learning by providing new experiences and creative challenges, enticing users to further engage in the materials/ideas.

The Creative Assistant acts as a more natural communicative bridge to fully involve users with the presence of the artwork and the artist's intent through a participatory dialogue. It encourages users to explore a painting in depth and experiment by undertaking a personal journey and substantiating their personal narratives through emotional navigation in response to a selected artwork or art series.

The user through this dialogue and tour based guided interactive experience better understands and connects (compared to less natural web and book based systems) with the traditional painting or at least personally experience the palette, brush stroke style, and other aesthetic choices of that artist.

6. CONCLUSION

Autonomous smart agents or virtual personal assistants (VPAs) are on the rise to become more sophisticated, enhanced through advanced machine learning. VPAs such as Apple's Siri or Alexa are part of the ambient user experience, instead of interacting with buttons and menus users can speak to a smart app, which is an intelligent agent. Autonomous agents are on the verge of becoming a valuable asset in the workplace as well as for personal purposes. In the near future, intelligent agents will be able to provide the users with dynamic and contextual information and augment human activity. Whether on the Web or mobile devices, intelligent virtual agent (IVA) use is seeing rapid adoption.

In this paper, we have overviewed our prototype multilayers conversation agent system, in the hopes that others might learn from this early work and begin to use and build 3D fully embodied, expressive AI based conversation systems to allows users to have a more natural interaction through the affective components that captures the emotions of users and reflect consistent spoken and gestural behavior. These techniques support the perception of a system that has a personality, which can act dynamically in its environment, rather than a rule based rigid system that cannot answer users' needs. We propose that with the aid of these affective multimodal components, the richer interaction will enhance the experience of the user.

7. REFERENCES

Alm, C., Roth, D., and Sproat, R. (2005) Emotions from Text: Machine Learning for Text-based Emotion Prediction. In Proceedings of the Human Language Technology Conference and the 2005 Conference on Empirical Methods in Natural Language Processing, Vancouver, Canada, 6–8 October, pp. 579–586.

API.AI https://docs.api.ai/v25 (retrieved 10 Februray 2017).

Bellegarda, J. (2010) Emotion analysis using latent affective folding and embedding. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, California.

Ben-David A. and O. Bat-El. (2016) Developing phonology. In A. Bar-On & D. Ravid (eds.) *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*. Mouton De Gruyter, Berlin.

Bernardet, U. and DiPaola, S. (2015) A Framework for Exogenous and Endogenous Reflexive Behavior in Virtual Characters. In *International Conference on Intelligent Virtual Agents*, pp. 132–136, Springer. 2015.

Bernardet, U., Chollet, M., DiPaola, S., and Scherer, S. (2016) An Architecture for Biologically Grounded Real-time Reflexive Behavior. In *Intelligent Virtual Agents. Lecture Notes in Computer Science*. Springer.

Bernardet, U., Saberi, M., and DiPaola, S. (2016) Simulink Toolbox for Real-time Virtual Character Control. In *Intelligent Virtual Agents. Lecture Notes in Computer Science*. Springer.

Bird, S., Klein, E., and Loper, E. (2009) *Natural language processing with Python: analyzing text with the natural language toolkit.* O'Reilly Media Inc.

Bradski, G. (2000). OpenCV computer vision library. *Dr. Dobb's Journal of Software Tools.*

Busso C., Deng Z., Yildirim S., Bulut M., Lee C. M., Kazemzadeh A., Lee S., Neumann U., and Narayanan S. (2004) Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of Sixth International Conference on Multimodal Interfaces*, pp. 205–211.

Calvo, R. A. and D'Mello, S. (2010) Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1), pp. 18–37.

Clavel, C., Adda, G., Cailliau, F., Garnier-Rizet, M., Cavet, A., Chapuis, G., and Suignard, P. (2013) Spontaneous speech and opinion detection: mining call-centre transcripts. *Language Resources and Evaluation*, 47(4), pp. 1089–1125.

DiPaola, S. (2014) Computer Modelling Fine Art Painting using a Cognitive Correlative Heuristics Approach, Proc: Biologically Inspired Cognitive Architectures. MIT, MA, USA.

Ekman, P. and Oster, H. (1979) Facial expressions of emotion. *Annual Review Psychology*, 30, pp. 527–554.

Ekman, P. (2004). Emotional and conversational nonverbal signals. In *Language, knowledge, and representation*, pp. 39–50. Springer Netherlands.

Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., Hammal, Z., and Rosenwald, D. P. (2014). Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image and Vision Computing*, 32(10), pp. 641–647.

Hutto, C. J. and Gilbert, E. E. (2014) VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June.

Irastorza, J. and Torres, M. I. (2016) Identification of emotional shifts on customer services. Submitted to Inter-speech 2016.

Jaimes, A. and Sebe, N. (2007) Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1), pp. 116–134.

Juslin, P. N. and Scherer, K. R. (2005). Vocal expression of affect. In Harrigan, J., Rosenthal, R., Scherer, K. (eds). *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford University Press, Oxford, UK.

Justo, T. (2014) Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web, *Knowledge-Based Systems*, 69, pp. 124–133.

Kanade, T., Cohn, J. F., and Tian, Y. (2000) Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, pp. 46–53.

Karadoğan, S. G. and Larsen, J. (2012) Combining semantic and acoustic features for valence and arousal recognition in speech. In *3rd International Workshop on Cognitive Information Processing (CIP)*, pp. 1–6. IEEE.

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

Lee, C., Narayanan, S., and Pieraccini, R. (2002) Combining acoustic and language information for emotion recognition. In *Proceedings of ICSLP 2002*, pp. 873–876.

Lehrer, K. (1974) Knowledge. Oxford University Press, New York.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010) The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*. San Francisco, USA, pp. 94–101.

Mohammad, S. M., Dunne, C., and Dorr, B. (2009) Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP-2009)*, pp. 599–608, Singapore.

Morrissey, K. A. (2002) Pathways Among Objects and Museum Visitors. In S. G. Paris (ed.), *Perspectives on Object-Centered Learning in Museums*. Mahwah: Lawrence Erlbaum Associates.

Murray, I. R. and Arnott, J. L. (1993) Toward a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America,* 93(2), pp. 1097–1108.

Picard, R. W. (1997) *Affective Computing*, 252. Cambridge: MIT Press.

Russell J. A., Bachorowski J., and Fernandez-Dols J. (2003) Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54, pp. 329–349.

Russell, J. A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), pp. 1161–1178.

Saberi, M., Bernardet, U., and DiPaola, S. (2014) An Architecture for Personality-Based, Nonverbal Behavior in Affective Virtual Humanoid Character, *Procedia Computer Science. Bio Inspired Cognitive Architectures*, 41, pp. 204–211.

Schuller, B., Rigoll, G., and Lang, M. (2004) Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In *Proceedings of ICASSP 2004*, pp. 577–580.

Thiebaux, M., Marsella, S., Marshall, A. N., and Kallmann, M. (2008) Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems, Volume 1,* pp. 151–158. International Foundation for Autonomous Agents and Multiagent Systems.

Turney, P. and Littman, M. (2003) Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4), pp.315–346.

Ververidis, D. and Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9), 1162–1181.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Ttransactions on Pattern Analysis and Machine Intelligence*, 31(1), pp. 39–58.