

Survival at the Museum: A Cooperation Experiment with Emotionally Expressive Virtual Characters

Ilaria Torre
Trinity College Dublin
Dublin, Ireland
torrei@tcd.ie

Emma Carrigan
Trinity College Dublin
Dublin, Ireland
carrige@tcd.ie

Killian McCabe
ADAPT Research Centre
Dublin, Ireland
killianpmccabe@gmail.com

Rachel McDonnell
Trinity College Dublin
Dublin, Ireland
ramcdonn@scss.tcd.ie

Naomi Harte
Trinity College Dublin
Dublin, Ireland
nharte@tcd.ie

ABSTRACT

Correctly interpreting an interlocutor's emotional expression is paramount to a successful interaction. But what happens when one of the interlocutors is a machine? The facilitation of human-machine communication and cooperation is of growing importance as smartphones, autonomous cars, or social robots increasingly pervade human social spaces. Previous research has shown that emotionally expressive virtual characters generally elicit higher cooperation and trust than 'neutral' ones. Since emotional expressions are multi-modal, and given that virtual characters can be designed to our liking in all their components, would a mismatch in the emotion expressed in the face and voice influence people's cooperation with a virtual character? We developed a game where people had to cooperate with a virtual character in order to survive on the moon. The character's face and voice were designed to either smile or not, resulting in 4 conditions: smiling voice and face, neutral voice and face, smiling voice only (neutral face), smiling face only (neutral voice). The experiment was set up in a museum over the course of several weeks; we report preliminary results from over 500 visitors, showing that people tend to trust the virtual character in the mismatched condition with the smiling face and neutral voice more. This might be because the two channels express different aspects of an emotion, as previously suggested.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; • **Applied computing** → **Psychology**;

KEYWORDS

Trust; multi-modal emotional expression; smiling

ACM Reference Format:

Ilaria Torre, Emma Carrigan, Killian McCabe, Rachel McDonnell, and Naomi Harte. 2018. Survival at the Museum: A Cooperation Experiment with Emotionally Expressive Virtual Characters. In *2018 International Conference on Multimodal Interaction (ICMI '18)*, October 16–20, 2018, Boulder, CO, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3242969.3242984>

1 INTRODUCTION

Emotions are an essential aspect of social interaction, and as humans, we have evolved to accurately signal and perceive them in their different modalities, including facial cues [14], vocal cues [36], or bodily cues [21]. Emotional expressivity is also connected with positive personality traits, including cooperation and trust [4, 9, 27, 37], which are also fundamental to a functioning society [2]. While emotional expressions have been studied extensively in the context of human perception and behaviour, the literature on multimodal expression in machines is comparably scarce [7, 28]. However, as machines with 'faces', 'bodies' and 'voices' increasingly populate our social space, studying the effect of machine characteristics such as emotional expression on human decision making, are paramount to ensure cooperation between humans and machines. For example, would people rather trust a happy or sad navigation system? And what would be the valence of the visual and auditory channels?

Here, we present initial results from an experiment where participants cooperate with a virtual character expressing either a congruent multimodal emotional expression (smiling face and voice, or neutral face and voice) or an incongruent one (smiling face and neutral voice, or neutral voice and smiling face). The experiment was installed in a public setting in a museum; here we report preliminary results from over 500 visitors.

2 BACKGROUND

Emotional expressions in machines are generally well perceived. For example, Elkins and Derrick [15] found that people rated smiling Embodied Conversational Agents as more trustworthy than non-smiling ones; similarly, avatars displaying authentic smiles were trusted more in a trust game [22]. Also, robots exhibiting an emotional expression were "more fun to play with" [24] and trusted more [29]. However, most studies focused on only one channel of emotional expression [e.g. 24, 29], or considered a channel mismatch as 'confusing' or as an impairment [5, 26].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMI '18, October 16–20, 2018, Boulder, CO, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5692-3/18/10...\$15.00

<https://doi.org/10.1145/3242969.3242984>

Communication channels can be designed independently in a machine, and this allows, for example, to study which channel is attended most predominately in emotion perception [e.g. 33, 35]. These insights not only help studying the often overlooked ‘big picture’ of multimodal emotional expressivity, but they also have implications on the design of the machines and human-machine communication, even if this entails a humanly impossible channel incongruency. As our ultimate goal as Human-Machine Interaction researchers is that of making interactions with machines work, nothing prevents us from designing a machine that expresses a mismatched emotion, if there is reason to do so.

The multimodal nature of emotional expressions has been studied using mismatched stimuli in the past, with participants generally having to distinguish one of the six basic emotions [11] from either the visual or audio channel, or the two channels combined, for example by displaying a photograph of happy person paired with a happy or sad utterance [6]. Participants were more accurate in labelling the emotion in the congruent condition, and the face was found to be the predominant source of information regarding emotional expression in other experiments [16, 38]. For this reason, Fagel [16] suggested that the visual channel might express the valence of an emotion, and the audio channel the activation. For embodied agents, Creed and Beale [5] found that emotions were perceived more strongly when expressed in one channel, despite other channels presenting a contrasting expression.

However, most studies do not validate the emotion questionnaires beforehand, but assume that their labels will be understood in the same way by participants [3]. This approach risks missing out on subtle variabilities in the emotions continuum that a mismatched condition likely causes (cf. The McGurk Effect [31]), since people are typically ‘forced’ to choose out of a set of pre-defined emotion labels. The focus of these studies is also often the perception of human emotional expressions, thus using mismatched stimuli from a human source (e.g. a human face and a human voice [6]). However, humans cannot normally produce a certain emotion in the face and another in the voice. Therefore, rather than creating stimuli based on a particular emotion, we identified an emotional *expression*, the smiling, as the basis for the mismatch in our research.

Smiling is a universal [32], multimodal [10, 17, 39] emotional expression, which is also audible from the voice alone [1, 13]. While it does not have a univocal correspondence to a certain emotion, different emotions can be expressed through smiling [e.g. 8, 12]. As these are generally positive, smiling is also linked to other positive traits – including trustworthiness [37] – in a sort of ‘halo effect’ [23].

Here, we report preliminary results of an experiment aimed at studying the influence of mismatched audio and visual smiling on human-machine cooperation. Given the previous literature on the effect of smiling on person perception, we would hypothesise that a virtual character smiling in at least one channel will be trusted more than a character that presents a congruent neutral expression.

3 METHOD

3.1 Game design

We used a modified version of the ‘Lunar Survival Task’ originally devised by Hall and Watson [20], which allows to measure actual

cooperative behaviour, rather than subjective opinions [see 18, 19]. Participants imagined that they had crash-landed on the moon with a digital navigation assistant, and that they had to rank their only 6 intact items in order of importance for survival. These 6 items were chosen from a bigger list published by NASA [20]. The chosen items were: portable heating unit; milk tank; parachute; receiver transmitter; raft; nylon rope. After participants made their initial ranking, the virtual character suggested changing the position of some of the items: the items that participants put at position 1 (= most important), 2, 3, 4, 5, 6 (= least important) were moved to positions 5, 4, 3, 2, 1, 6, respectively. We varied the number of positions that the virtual character suggested changing to observe whether people would be more willing to accept small rather than big changes, and whether they would be more likely to move items upwards or downwards. This will be the subject of future analyses.

The virtual character gave a justification for each item, such that every item that was moved to position 1 or 2 was given a positive description, every item moved to position 3 or 4 a neutral description, and every item moved to position 5 or 6 a negative description. Participants’ final rankings provided an implicit measure of trust: if they deemed the virtual character to be trustworthy, they would be more likely to accept its suggestion to change an item position. While in general this acceptance could be influenced by other mechanisms, such as competence or dominance [25], in situations where risk is involved, such as a hypothetical survival scenario, trust is assumed to be an essential aspect for cooperation [e.g. 30].

3.2 Stimuli preparation

The virtual character was created using state-of-the-art computer graphics technology for modelling, animating, and rendering. First, the model, comprising over 250 scans of a real actor’s facial expressions, was created by a company called 3Lateral. These scans were then carefully combined into a controllable facial rig, which could then be driven by the motion capture. We recorded a different, Irish male actor in our motion capture studio at Trinity College Dublin using a 23-camera Vicon Vantage optical motion capture system for body motion capture and a Technoprops video-based head-mounted facial capture system. The actor was asked to read a set of pre-scripted sentences in neutral and smiling expressions. Audio was recorded using a wireless microphone attached to his face. The actor’s facial movements were then retargeted onto the model, using Faceware Tech software for the facial movement and inverse kinematics for the movement of the head. Finally, advanced shaders (e.g., subsurface scattering for the skin) were used to create the highly realistic appearance in Autodesk Maya 2018 software, as shown in Figure 1.

The recorded sentences comprised of a positive, a neutral, and a negative description for each item. For example, the positive sentence for the raft was: “The gas inside the raft could be used for propulsion, for moving faster”, the neutral: “We could use the raft for protection, but it won’t help us advance” and the negative: “I think the raft would be useless, since there are no water sources on the moon”. Thus, the actor read 18 different sentences in the neutral and smiling condition.

We chose smiling as an emotional expression because it is not univocally linked to one emotion, so we did not run a manipulation



Figure 1: Virtual character with a neutral (left) and smiling (right) facial expression

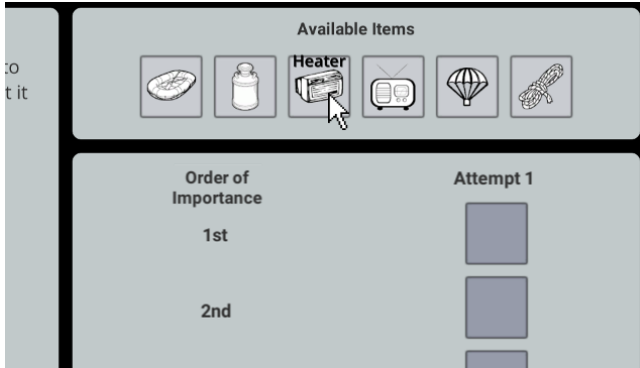


Figure 2: Screenshot from the game showing the 6 items and the beginning of the vertical grid for participants' ranking.

check on the emotion perceived from the smiling. We believe we were original in this sense, because rather than trying to categorise an emotion arbitrarily, we created a virtual character with an emotional expression and observed its effect on trust, regardless of the perceived emotion. However, as detailed later in section 4.2, participants were asked about the perceived happiness of the virtual character in the post-game questionnaire, and indeed the facial smiling was perceived to express this emotion often associated with smiling.

The individual recorded sentences were lip-synced with the facial animation – using the software Lightworks – to produce 4 different conditions: *congruent smiling* (smiling voice over smiling face), *congruent neutral* (neutral voice over neutral face), *smiling voice* (smiling voice over neutral face) and *smiling face* (neutral voice over smiling face). Thus, a total of 4 experimental conditions \times 18 sentences = 72 video clips were produced as stimuli for the experiment. The game was programmed in Unity version 2017.2.0.

3.3 Procedure

Participants were recruited among visitors of the 'FAKE' exhibition in the Dublin Science Gallery¹. Any visitor could play the game, but only the data of people who had given their written consent to participate in the study was collected. Participants were seated in front of a computer monitor which had previously been calibrated to the following settings: brightness 200 cd/m², white point 6500, display gamut 97% of sRGB; the character's video resolution was 1280 \times 720

¹Exhibition website: <https://dublin.sciencegallery.com/fake/>; last access: 2018-08-09

and the monitor resolution was 1920 \times 1080. Each participant was randomly assigned to one of the 4 emotional expression conditions. Participants were invited to wear a pair of good quality over-ear headphones attached to the experiment area. When they were ready to start, a screen showed them 6 images representing the 6 surviving items, which could be ranked by dragging and dropping them in a vertical grid with 6 available positions (Figure 2). When participants had completed their ranking, a prompt warned them that a video message from their virtual assistant was incoming. Then, the first video clip, where the virtual character re-ranked the item that participants had put at position 1, was played. The 6 videos corresponding to the 6 items were played one after the other, with the items being moved accordingly by the virtual character in a second vertical grid. After this ranking was finished, a third vertical grid appeared, and participants could re-arrange the items one last time. A final screen thanked participants for playing the game and asked them to fill out a questionnaire on the virtual character's perceived realism, appeal, eeriness, trustworthiness, knowledge, attractiveness, happiness and intelligence. The whole interaction lasted approximately 10 minutes.

3.4 Participants

The data from 504 museum visitors was collected. Of the people who filled out a demographics questionnaire, 234 identified as female and 195 as male; the majority were in the 26-35 age group (min = 13 years old); 14 rated their English fluency as 'basic', 129 as 'fluent', 39 as 'native-like' and 257 as 'native'.

4 RESULTS

4.1 Game data

To obtain a measure of participants' trust toward the virtual character, the difference between the virtual character's suggested item position, and the participant's second position for that item, was computed. Thus, a ranking difference of 0 indicates that participants completely accepted the suggested position, a difference of 1 indicates that participants moved their initial position closer to the virtual character's, and so on. The smaller this ranking difference, the higher the cooperation with the virtual character.

This ranking difference was used as the dependent variable in a cumulative link mixed model, with the virtual character's emotional expression as predictor and participant id as random effect. The emotional expression variable had 4 levels (one per condition), since we were interested in the effect of the two channels combined, rather than on the two individual channels. A comparison with a null model showed that emotional expression was a significant predictor for the ranking difference ($\chi^2(3) = 8.20, p = .04$). A post-hoc comparison of means using the Tukey HSD test showed that the *smiling face* condition attracted a smaller position difference – i.e., a higher trust – than the other conditions ($z = -2.09, p = .036$), as shown in Figure 3.

4.2 Survey

Participants rated the virtual character in terms of realism, appeal, eeriness, trustworthiness, knowledge, attractiveness, happiness and intelligence, on Likert scales ranging from 1 (= not at all) to 5 (= very much) (Figure 4). An ordinal regression model was fitted to

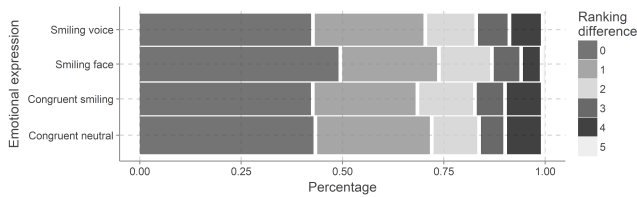


Figure 3: Ranking differences across the emotional expression conditions, for all participants and items.

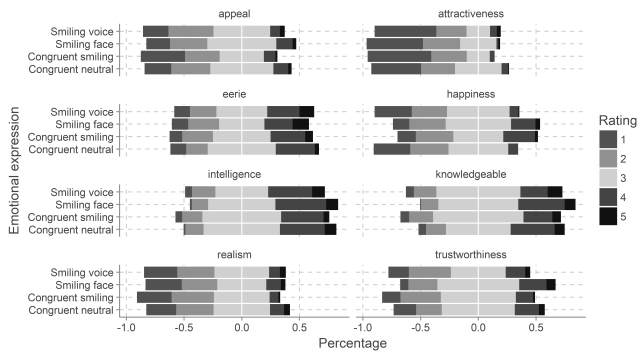


Figure 4: Summary of questionnaire responses.

each of these traits separately, using the virtual character’s emotional expression as a predictor. The emotional expression had an effect on the appeal, trustworthiness, knowledge, attractiveness and happiness traits. In particular, pairwise comparisons showed that people gave higher ratings of appeal for the character in the *congruent smiling* condition ($z = -2.71, p < .01$); higher ratings of trustworthiness in the *smiling face* condition ($z = 2.17, p = .03$); higher ratings of knowledge in the *congruent smiling* condition (marginal: $z = -1.88, p = .06$); higher ratings of attractiveness in the *congruent smiling* condition ($z = -2.29, p = .022$); and higher ratings of happiness in the *congruent smiling* ($z = 4.10, p < .001$) and *smiling face* ($z = 4.16, p < .001$) conditions. No other statistically significant differences were found.

5 DISCUSSION

The *smiling face* condition led to the lowest ranking difference in the cooperation game. This higher acceptance rate of the virtual character’s suggestions is surprising, as it indicates that people find this mismatched condition with a smiling face and non-smiling voice more trustworthy. This result suggests that the character’s smiling voice contains some features that were particularly distrusted in the context of the game. In fact, if the visual information had overridden auditory information in the character’s perception, people would have equally trusted the virtual character in the *congruent smiling* and *smiling face* conditions.

The different influence of the visual and auditory channel is also supported by the self-reported data from the questionnaire, where people rated the virtual character in the *smiling face* condition as more trustworthy, appealing, and knowledgeable. As Fagel [16] suggested, it is possible that the visual channel expresses the valence

of an emotion, and the audio channel the activation; this implies that participants trusted the ‘smilingness’ of the virtual character (as indicated by smiling in the face) but not how much of it there was (as indicated by a preference for a non-smiling audio). To test this dual emotional perception hypothesis, future studies could examine participants’ behaviour with a virtual character represented only by a voice (smiling or neutral) or by a face (smiling or neutral, with written text).

Another possible explanation for the result is that participants did not like the vocal smiling because they felt that such a ‘joyful’ voice was inappropriate for the dramatic context of a lunar survival quest. The fact that voice and face belonged to two different actors might have further accentuated a perceived mismatch [see e.g. 40].

However, the smiling voice might have had a negative impact only on some of the perceived virtual character’s traits, since people indicated that the *congruent smiling* condition was more appealing, knowledgeable, and attractive. A possible next step to test this interpretation would be to run the same experiment with the same virtual character, but a different voice.

It is also interesting to note that the smiling expression in the face (both in the *congruent smiling* and *smiling face* conditions) elicited higher happiness judgments, while the *smiling voice* condition did not. This suggests that the visual channel might have been enough to convey the smiling emotional expression in the virtual character.

In general, these initial results also suggest that both the audio and the visual channel seem to play a role in the perception of emotional expressions, either in a positive or negative sense [cf. 33, 34]. This has applications in the field of machine design, for example in cases where prioritisation of which expressive channel to build is required.

6 CONCLUSION

The experimental paradigm we employed allows collecting implicit data on cooperative behaviour. By manipulating the emotional expression channels in a virtual character, we were able to study whether people trusted and cooperated with one of these conditions more. The approach seems promising and shows a significant outcome for an unexpected combination of expressive modalities. Initial results suggest that people trusted a virtual character with a mismatched audio-video emotional expression more than one with a congruent expression.

ACKNOWLEDGMENTS

The research was funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713567, and by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2016) and is co-funded by the European Regional Development Fund. The second author received funding from the Science Foundation Ireland, Game Face (13/CDA/2135) project. We are grateful to the Science Gallery staff and visitors.

REFERENCES

- [1] Véronique Aubergé and Marie Cathiard. 2003. Can we hear the prosody of smile? *Speech Communication* 40, 1 (2003), 87–97.

- [2] Annette Baier. 1986. Trust and Antitrust. *Ethics* 96, 2 (1986), 231–260. <https://doi.org/10.1086/292745>
- [3] Russell Beale and Chris Creed. 2009. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies* 67, 9 (2009), 755–776.
- [4] R. Thomas Boone and Ross Buck. 2003. Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior* 27, 3 (2003), 163–182.
- [5] Chris Creed and Russell Beale. 2008. Psychological responses to simulated displays of mismatched emotional expressions. *Interacting with Computers* 20, 2 (2008), 225–239.
- [6] Beatrice De Gelder and Jean Vroomen. 2000. The perception of emotions by ear and by eye. *Cognition & Emotion* 14, 3 (2000), 289–311.
- [7] David DeSteno, Cynthia Breazeal, Robert H. Frank, David Pizarro, Jolie Baumann, Leah Dickens, and Jin Joo Lee. 2012. Detecting the Trustworthiness of Novel Partners in Economic Exchange. *Psychological Science* 23, 12 (2012), 1549–1556.
- [8] Amy Drahota, Alan Costall, and Vasudevi Reddy. 2008. The vocal communication of different kinds of smile. *Speech Communication* 50, 4 (2008), 278–287. <https://doi.org/10.1016/j.specom.2007.10.001>
- [9] Jennifer R. Dunn and Maurice E. Schweitzer. 2005. Feeling and believing: the influence of emotion on trust. *Journal of Personality and Social Psychology* 88, 5 (2005), 736.
- [10] Hedwig Eisenbarth and Georg W. Alpers. 2011. Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 4 (2011), 860.
- [11] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2 (1971), 124.
- [12] Paul Ekman and Wallace V. Friesen. 1982. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior* 6, 4 (1982), 238–252.
- [13] Kevin El Haddad, Ilaria Torre, Emer Gilmartin, Hüseyin Çakmak, Stéphane Dupont, Thierry Dutoit, and Nick Campbell. 2017. Introducing amus: The amused speech database. In *International Conference on Statistical Language and Speech Processing*. Springer, 229–240.
- [14] Hillary Anger Elfenbein and Nalini Ambady. 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin* 128, 2 (2002), 203.
- [15] Aaron C. Elkins and Douglas C. Derrick. 2013. The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation* 22, 5 (2013), 897–913.
- [16] Sascha Fagel. 2006. Emotional McGurk Effect. In *Proceedings of the International Conference on Speech Prosody*, Vol. 1. Dresden, Germany.
- [17] Sascha Fagel. 2009. Effects of Smiled Speech on Lips, Larynx and Acoustics. In *Proceedings of the Auditory Visual Speech Processing Conference (AVSP)*. Norwich, UK, 18–21.
- [18] Edward L. Glaeser, David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. 2000. Measuring trust. *Quarterly Journal of Economics* (2000), 811–846.
- [19] Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464.
- [20] Jay Hall and Wilfred Harvey Watson. 1970. The effects of a normative intervention on group decision-making performance. *Human Relations* 23, 4 (1970), 299–317.
- [21] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing* 4, 1 (2013), 15–33.
- [22] Eva Krumbhuber, Antony S. R. Manstead, Darren Cosker, Dave Marshall, Paul L. Rosin, and Arvid Kappas. 2007. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion* 7, 4 (2007), 730–735.
- [23] Sing Lau. 1982. The Effect of Smiling on Person Perception. *The Journal of Social Psychology* 117, 1 (1982), 63–67. <https://doi.org/10.1080/00224545.1982.9713408>
- [24] Iolanda Leite, André Pereira, Carlos Martinho, and Ana Paiva. 2008. Are emotional robots more fun to play with?. In *Robot and human interactive communication, 2008. RO-MAN 2008. The 17th IEEE international symposium on*. IEEE, 77–82.
- [25] Roy J. Lewicki, Edward C. Tomlinson, and Nicole Gillespie. 2006. Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management* 32, 6 (2006), 991–1022.
- [26] Benny Liebold and Peter Ohler. 2013. Multimodal emotion expressions of virtual agents, mimic and vocal emotion expressions and their effects on emotion recognition. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 405–410.
- [27] Robert B. Lount Jr. 2010. The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology* 98, 3 (2010), 420.
- [28] Gale Lucas, Giota Stratou, Shari Liebling, and Jonathan Gratch. 2016. Trust me: multimodal signals of trustworthiness. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 5–12.
- [29] Maya B. Mathur and David B. Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* 146 (2016), 22–32.
- [30] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. 20, 3 (1995), 709–734. <https://doi.org/10.2307/258792>
- [31] Harry McGurk and John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264, 5588 (1976), 746.
- [32] Marc Mehu and Robin I. M. Dunbar. 2008. Relationship between smiling and laughter in humans (*Homo sapiens*): Testing the power asymmetry hypothesis. *Folia Primatologica* 79, 5 (2008), 269–280.
- [33] Emily Mower, Maja J. Mataric, and Shrikanth Narayanan. 2009. Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. *IEEE Transactions on Multimedia* 11, 5 (2009), 843–855.
- [34] Maike Paetzel, Christopher Peters, Ingela Nyström, and Ginevra Castellano. 2016. Effects of multimodal cues on children’s perception of uncanniness in a social robot. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 297–301.
- [35] Christina Regenbogen, Daniel A. Schneider, Andreas Finkelmeyer, Nils Kohn, Birgit Derntl, Thilo Kellermann, Raquel E. Gur, Frank Schneider, and Ute Habel. 2012. The differential contribution of facial expressions, prosody, and speech content to empathy. *Cognition & Emotion* 26, 6 (2012), 995–1014.
- [36] Klaus R. Scherer, Rainer Banse, and Harald G. Wallbott. 2001. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural Psychology* 32, 1 (2001), 76–92.
- [37] Joanna Schug, David Matsumoto, Yutaka Horita, Toshio Yamagishi, and Kemberlee Bonnet. 2010. Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior* 31, 2 (2010), 87–94.
- [38] Sachiko Takagi, Saori Hiramatsu, Ken-ichi Tabei, and Akihiro Tanaka. 2015. Multisensory perception of the six basic emotions is modulated by attentional instruction and unattended modality. *Frontiers in Integrative Neuroscience* 9 (2015), 1. <https://doi.org/10.3389/fnint.2015.00001>
- [39] Ilaria Torre. 2014. Production and perception of smiling voice. In *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2013) Conference*. York, U. K.
- [40] Stephanie Walker, Vicki Bruce, and Claire O’Malley. 1995. Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics* 57, 8 (1995), 1124–1133.