Including Enthusiasm in Human–AI Communication

Anonymous Author(s)

ABSTRACT

Emotions are key components in conversations which is why virtual agents need to be able to detect and generate them. However, most virtual agents and communicative systems lack the ability to understand enthusiasm or generate multimodal enthusiastic communicative presentations. In this work, we argue for the importance of including enthusiasm in the design of human–AI collaboration and communication and review the existing datasets and models that can be used to bridge the gap in this area. We conclude with several proposals for ways that we can advance research to address this question in particular in the design of Virtual Agents. The data and analysis are publicly available [1].

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

datasets, enthusiasm, virtual agents, engagement, conversational expressions

ACM Reference Format:

Anonymous Author(s). 2021. Including Enthusiasm in Human–AI Communication. In *Fukuchiyama City '21: International Conference on Intelligent Virtual Agents, September 14–17, 2021, Fukuchiyama City, Kyoto, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/1122445.1122456

1 OVERVIEW

Human like communication is almost impossible without emotions. Although there has been extensive research in the area of emotionally intelligent AI systems, we have mostly overlooked enthusiasm or arousal.

The six basic emotions [16] are not sufficient to reflect the nuances that can occur in human communication. Conversational expressions, on the other hand, focus on all expressions that can occur during conversations. Examples of such expressions include clueless, annoyed, interested and various forms of smiling [7, 8, 10].

In this work, we want to call attention to enthusiasm as a conversational expression. Enthusiasm is one of the most desired traits in employees, co-workers, mentors, leaders, and teachers [3, 6, 11, 30, 34, 41]. Enthusiastic people are not only motivated and excited about a topic, they can also spark this excitement in their listeners and even move their audience to action [14]. Given that the ability to speak enthusiastically provides a clear advantage in human interactions when the goal is to engage the interlocutor emotionally, we should also aim to transfer this trait to virtual agents.

Although enthusiasm has been studied extensively in psychology, showing that students clearly benefit from enthusiastic teachers [6, 45] as well as companies do with enthusiastic leaders [25, 37], it is still unclear what exactly makes a person to be perceived as enthusiastic [22]. Most of the work on detecting enthusiasm automatically is based on using written human-to-human conversational dialogues [19, 40]. Limited work on enthusiastic virtual agents and robots exist. Liew et al. showed that virtual agents with enthusiastic voices improve learning outcomes in students [27–29], however, they used prerecorded voices from actors. Despite the limited work on detecting and generating enthusiastic behavior, the recent release of the first multimodal dataset on enthusiasm [1] is not only a chance to gain more understanding on enthusiasm, but also to create enthusiastic virtual agents.

In the following we will describe the opportunities for enthusiastic virtual agents as well as the challenges that need to be addressed in order to have an engaging conversation between humans and virtual agents. We will also present preliminary results using the enthusiasm dataset and how explainable AI techniques such as SHAP [31] can be used to understand what features are important for enthusiasm classification.

2 OPPORTUNITIES FOR ENTHUSIASTIC VIRTUAL AGENTS

There are several applications of virtual agents that could benefit from understanding and generating enthusiastic behavior. In the following, we want to focus on three specific use cases for enthusiastic virtual agents: a) teaching, b) coaching, and c) sales. All three applications have in common that they are not purely conversational. The essential component is to deliver a message in an engaging way. We believe that focusing on delivering messages enthusiastically is a feasible first step to improve existing virtual agents.

Teaching: Teacher enthusiasm has shown to improve students' performance in several studies [6, 22, 23, 45]. The same phenomena was also shown with virtual teachers using prerecorded enthusiastic voices [27, 29]. With the global shortage of teachers, pedagogical agents and robot teachers are gaining more importance [32] as it has been shown that both are accepted by students[12, 26]. One of the main advantages of virtual teachers is that they can adapt to the individual needs that students have [20]. Automatic generation of enthusiastic behavior during the delivery of teaching material would facilitate the access to enthusiastic pedagogical agents.

Coaching: Several virtual agents have been designed to coach humans and provide feedback. Zhao et al. developed a coach that provides feedback on elevator pitches in virtual reality [46]. El Kamali et al. created a conversational e-coach that promotes healthy lifestyles in older age [13] and Shameki et al. created a virtual coach

[1] www.anonymouslinkforsubmission.com

Fukuchiyama City '21, September 14–17, 2021, Fukuchiyama City, Kyoto, Japan

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

that helps patients that recover from spinal cord injuries adopt healthy behaviors [38]. The mentioned coaches have shown to be useful for the users. However, given that enthusiasm has the particular ability to move listeners to action, we believe that implementing enthusiastic behavior can improve the users' experience and their outcomes.

Sales: Virtual sales agents are gaining more interest as they could scale easily with the needs of online sales. Jusoh developed a conversational agent for online sales, combining recommendation and negotiation[21]. Matsumura et al. showed that a combination of a virtual sales agent and a robot that gives customers their purchases can be used as sales assistants [33]. By combining appealing arguments for a product or service with the non-verbal cues of enthusiasm, we believe that virtual agents can improve their convincingness and increase sales.

3 CHALLENGES

There are three main challenges that need to be addressed in order to use enthusiasm appropriately during conversations between humans and virtual agents.

Understanding human enthusiasm expression: Do all humans express enthusiasm in a similar way or are there cultural or even gender differences? Understanding how enthusiasm is expressed through facial expressions, voice, and linguistic content is the first step. The use of explainable AI techniques can help understand which features help models make good predictions [2]. Weber et al. used GradCam and layerwise relevance propagation to analyze image regions that make speakers be perceived as persuasive [43]. Although multimodal techniques do not exist yet, analysis of individual modalities is possible. Reverse engineering enthusiasm as Chen et al. have done for other conversational expressions [9] is another method that can help obtain a clear understanding of what is perceived as enthusiastic.

Responding to enthusiastic behavior: Once enthusiastic behavior is detected in humans, virtual agents should respond likewise with excitement and interest in learning more about the topic that fascinates the user. Mirroring and building rapport through social dialogue is an important component of user experience [17, 18, 35, 42]. Rapport has been shown to create interpersonal responsiveness and influence between two people which is beneficial in relationship of customers and employees or teachers and students[18]. The same is true for the relationship of humans and virtual agents.

Generating enthusiastic behavior: Data-driven methods can be used to generate enthusiastic behavior. Several architectures have been developed recently for facial reenactment such as Head2Head which are able to transfer facial expressions, gaze, and pose from a source actor to a target in a photo-realistic manner [24]. Style transfer methods have also been used to generate voices with different emotions, maintaining content [36, 47]. Rule-based methods can be used to implement simple findings of enthusiastic behavior, such as pitch variation or facial expressions [39].

4 PRELIMINARY ANALYSIS OF ENTHUSIASTIC HUMAN BEHAVIOUR

We performed a preliminary analysis using the enthusiasm dataset [1] to detect enthusiasm from speech and understand which acoustic parameters are most relevant for this task. The dataset is composed of 1126 sentences extracted from 113 different TED talks and is annotated as monotonous, normal, or enthusiastic speech.

We trained a Random Forest classifier using EGEMAPS features [15, 44] which contain 88 acoustic parameters that capture affective physiological changes in voice production. Our model performed with an F1-score of 77% using an 80/20 train-test split.

To better understand which features are important in the decision making of the model we performed SHAP analysis [31]. The features with the highest SHAP values (highest impact on the model output) were different statistical measures of pitch and loudness[2].

We performed further analysis using a one-way ANOVA to determine if pitch (F0) and loudness are independent from the enthusiasm level. Both p-values are >0.05, meaning that the enthusiasm labels depend on the acoustic features. In Fig. 1 (left) we can see that monotonous samples have a lower mean F0 than enthusiastic samples and that (right) monotonous samples have lower mean loudness than enthusiasm. These observations agree with the intuition that enthusiastic speakers speak louder and increase their pitch. Using these outcomes, it is possible to generate handcrafted voices with these characteristics [4].

We also performed two separate one-way ANOVAs on the Facial Action Units (AUs) detected with OpenFace [5] to evaluate the dependence of the mean and standard deviation of AUs with our labels. The AUs with p-value<0.05 are AU12 (lip corner puller), AU15 (lip corner depressor), AU17 (chin raiser), and AU26 (jaw drop). Analyzing the label distributions we observe that monotonous samples have more often low intensities for AU26 and very low standard deviation for AU12, compared to enthusiastic samples.

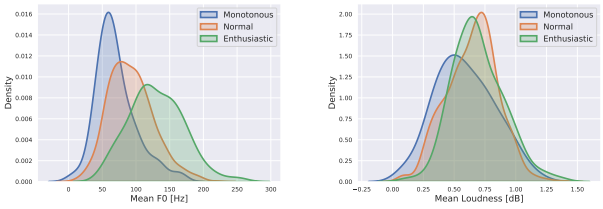

Figure 1: Label distribution for mean pitch (F0) and mean loudness. Left: enthusiastic samples have a higher mean F0 compared to monotonous samples. Right: monotonous speech tends to have lower mean loudness compared to enthusiastic speech.

5 CONCLUSION

In this paper we have presented the opportunities that enthusiasm as a conversational expression can bring to virtual agents, especially in teaching, coaching, and sales. With the recently released multimodal dataset on enthusiasm, it will be possible to start tackling several challenges on teaching virtual agents to understand and generate enthusiasm. The data and analysis are publicly available [3].

[2]Detailed description of the features are available at www.anonymouslinkforsubmission.com
[3]www.anonymouslinkforsubmission.com

REFERENCES

[1] Anonymous. 2021. Anonymized for submission.

[2] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115.

[3] Eve Ash. 2018. *The 15 traits of ideal co-workers: How many do you know?* https://www.smartcompany.com.au/people-human-resources/fifteen-traits-of-ideal-co-workers/

[4] Paolo Baggia, Paul Bagshaw, Michael Bodell, De Zhi Huang, Lou Xiaoyan, Scott McGlashan, Jianhua Tao, Yan Jun, Hu Fang, Yongguo Kang, et al. 2010. Speech synthesis markup language (SSML) version 1.1. *World Wide Web Consortium, Recommendation REC-speechsynthesis11-20100907* (2010).

[5] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1–10.

[6] Edward M Bettencourt, Maxwell H Gillett, Meredith Damien Gall, and Ray E Hull. 1983. Effects of teacher enthusiasm training on student on-task behavior and achievement. *American educational research journal* 20, 3 (1983), 435–450.

[7] Susana Castillo, Katharina Legde, and Douglas W Cunningham. 2018. The semantic space for motion-captured facial expressions. *Computer Animation and Virtual Worlds* 29, 3-4 (2018), e1823.

[8] Chaona Chen, Oliver GB Garrod, Robin AA Ince, Mary Ellen Foster, Philippe G Schyns, and Rachael E Jack. 2020. Building Culturally-Valid Dynamic Facial Expressions for a Conversational Virtual Agent Using Human Perception. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–3.

[9] Chaona Chen, Oliver GB Garrod, Jiayu Zhan, Jonas Beskow, Philippe G Schyns, and Rachael E Jack. 2018. Reverse engineering psychologically valid facial expressions of emotion into social robots. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 448–452.

[10] Douglas W Cunningham, Mario Kleiner, Christian Wallraven, and Heinrich H Bülthoff. 2005. Manipulating video sequences to determine the components of conversational facial expressions. *ACM Transactions on Applied Perception (TAP)* 2, 3 (2005), 251–269.

[11] Kevin Daum. 2021. *5 Desirable Traits of Great Employees.* https://www.inc.com/kevin-daum/5-tests-hiring-best-employees.html

[12] Bosede I Edwards, Idris O Muniru, Nosiba Khougali, Adrian D Cheok, and Rui Prada. 2018. A physically embodied robot teacher (pert) as a facilitator for peer learning. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.

[13] Mira El Kamali, Leonardo Angelini, Maurizio Caon, Giuseppe Andreoni, Omar Abou Khaled, and Elena Mugellini. 2018. Towards the NESTORE e-Coach: a tangible and embodied conversational agent for older adults. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 1656–1663.

[14] "enthusiasm". 2021. *Merriam-Webster.com.* https://www.merriam-webster.com

[15] Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2015. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE transactions on affective computing* 7, 2 (2015), 190–202.

[16] E Friesen and Paul Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3, 2 (1978), 5.

[17] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *International workshop on intelligent virtual agents*. Springer, 125–138.

[18] Dwayne D Gremler and Kevin P Gwinner. 2008. Rapport-building behaviors used by retail employees. *Journal of Retailing* 84, 3 (2008), 308–324.

[19] Michimasa Inaba, Fujio Toriumi, and Kenichiro Ishii. 2011. Automatic detection of "enthusiasm" in non-task-oriented dialogues using word co-occurrence. In *2011 IEEE Workshop on Affective Computational Intelligence (WACI)*. IEEE, 1–7.

[20] W Lewis Johnson, Jeff W Rickel, James C Lester, et al. 2000. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education* 11, 1 (2000), 47–78.

[21] Shaidah Jusoh. 2018. Intelligent conversational agent for online sales. In *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. IEEE, 1–4.

[22] Melanie M Keller, Thomas Goetz, Eva S Becker, Vinzenz Morger, and Lauren Hensley. 2014. Feeling and showing: A new conceptualization of dispositional teacher enthusiasm and its relation to students' interest. *Learning and Instruction* 33 (2014), 29–38.

[23] Melanie M Keller, Anita Woolfolk Hoy, Thomas Goetz, and Anne C Frenzel. 2016. Teacher enthusiasm: Reviewing and redefining a complex construct. *Educational Psychology Review* 28, 4 (2016), 743–769.

[24] Mohammad Rami Koujan, Michail Christos Doukas, Anastasios Roussos, and Stefanos Zafeiriou. 2020. Head2head: Video-based neural head synthesis. *arXiv preprint arXiv:2005.10954* (2020).

[25] David L Largent. 2016. Measuring and understanding team development by capturing self-assessed enthusiasm and skill levels. *ACM Transactions on Computing Education (TOCE)* 16, 2 (2016), 1–27.

[26] Jamy Li, René Kizilcec, Jeremy Bailenson, and Wendy Ju. 2016. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* 55 (2016), 1222–1230.

[27] Tze Wei Liew, Su-Mae Tan, and Chin Lay Gan. 2018. Interacting With Motivational Virtual Agent: The Effects of Message Framing and Regulatory Fit in an E-Learning Environment. In *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*. IEEE, 136–141.

[28] Tze Wei Liew, Su-Mae Tan, Teck Ming Tan, and Si Na Kew. 2020. Does speaker's voice enthusiasm affect social cue, cognitive load and transfer in multimedia learning? *Information and Learning Sciences* (2020).

[29] Tze Wei Liew, Nor Azan Mat Zin, and Noraidah Sahari. 2017. Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-centric Computing and Information Sciences* 7, 1 (2017), 9.

[30] Penny Loretto. 2019. *Eight Qualities of a Good Mentor.* https://www.thebalancecareers.com/qualities-of-a-good-mentor-1986663

[31] Scott Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874* (2017).

[32] Tetsuya Matsui and Seiji Yamada. 2019. The design method of the virtual teacher. In *Proceedings of the 7th International Conference on Human-Agent Interaction*. 97–101.

[33] Reo Matsumura, Masahiro Shiomi, and Norihiro Hagita. 2017. Does an animation character robot increase sales?. In *Proceedings of the 5th International Conference on Human Agent Interaction*. 479–482.

[34] Kelly Mitchell. 2018. *#1 Desired Trait? ENTHUSIASM!* http://www.socialjusticesolutions.org/2018/04/18/1-desired-trait-enthusiasm/

[35] Simon Provoost, Jeroen Ruwaard, Koen Neijenhuijs, Tibor Bosse, and Heleen Riper. 2018. Mood mirroring with an embodied virtual agent: a pilot study on the relationship between personalized visual feedback and adherence. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, 24–35.

[36] Kaizhi Qian, Yang Zhang, Shiyu Chang, Xuesong Yang, and Mark Hasegawa-Johnson. 2019. Autovc: Zero-shot voice style transfer with only autoencoder loss. In *International Conference on Machine Learning*. PMLR, 5210–5219.

[37] Birgitta Sandberg. 2007. Enthusiasm in the development of radical innovations. *Creativity and Innovation Management* 16, 3 (2007), 265–273.

[38] Ameneh Shamekhi, Ha Trinh, Timothy W Bickmore, Tamara R DeAngelis, Theresa Ellis, Bethlyn V Houlihan, and Nancy K Latham. 2016. A virtual self-care coach for individuals with spinal cord injury. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 327–328.

[39] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. 2021. A Conversational Agent Framework with Multi-modal Personality Expression. *ACM Transactions on Graphics (TOG)* 40, 1 (2021), 1–16.

[40] Ryoko Tokuhisa and Ryuta Terashima. 2006. Relationship between utterances and "enthusiasm" in non-task-oriented conversational dialogue. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*. 161–167.

[41] Pascal van Opzeeland. 2017. *The 8 Customer Service Skills And Traits You Should Look For.* https://www.userlike.com/en/blog/customer-service-skills-traits

[42] Isaac Wang and Jaime Ruiz. 2021. Examining the Use of Nonverbal Communication in Virtual Agents. *International Journal of Human–Computer Interaction* (2021), 1–26.

[43] Klaus Weber, Lukas Tinnes, Tobias Huber, Alexander Heimerl, Marc-Leon Reinecker, Eva Pohlen, and Elisabeth André. 2020. Towards demystifying subliminal persuasiveness: using XAI-techniques to highlight persuasive markers of public speeches. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*. Springer, 113–128.

[44] Wei Xue, Catia Cucchiarini, RWNM van Hout, and Helmer Strik. 2019. Acoustic correlates of speech intelligibility. The usability of the eGeMAPS feature set for atypical speech. (2019).

[45] Qin Zhang. 2014. Assessing the effects of instructor enthusiasm on classroom engagement, learning goal orientation, and academic self-efficacy. *Communication Teacher* 28, 1 (2014), 44–56.

[46] Zhenjie Zhao and Xiaojuan Ma. 2020. Situated Learning of Soft Skills with an Interactive Agent in Virtual Reality via Multimodal Feedback. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 25–27.

[47] Kun Zhou, Berrak Sisman, Rui Liu, and Haizhou Li. 2020. Seen and unseen emotional style transfer for voice conversion with a new emotional speech dataset. *arXiv preprint arXiv:2010.14794* (2020).