Towards Designing Enthusiastic AI Agents

Carla Viegas
Stella AI
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
cviegas@cs.cmu.edu

Malihe Alikhani
University of Pittsburgh
Pittsburgh, Pennsylvania, USA
malihe@pitt.edu

ABSTRACT

Immersive virtual worlds are increasingly being used for education, training, and entertainment, and virtual humans that can interact with human users in these worlds play many important roles. Understating the emotional constructs of the user and generating multimodal forms of communications that are aligned with the user's needs and input is key to designing AI agents. 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.

CCS CONCEPTS

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

KEYWORDS

datasets, enthusiasm, virtual agents, engagement, conversational expressions

ACM Reference Format:
Carla Viegas and Malihe Alikhani. 2021. Towards Designing Enthusiastic AI Agents. In *21th ACM International Conference on Intelligent Virtual Agents (IVA '21), September 14–17, 2021, Virtual Event, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3472306.3478366

1 OVERVIEW

We have recently observed a growing interest in the development of embodied agents that can understand user's emotional constructs. The fluent exchange of information and the display of cognitive and emotional states is essential for establishing and maintaining engagement. Studies have suggested that the six basic emotions [14] cannot best represent the emotional constructs that AI agents need to work with. Conversational expressions are more fine-grained and diverse. Examples of such expressions include clueless, annoyed, and interested [6, 7, 9]. 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 [2, 5, 10, 27, 31, 39]. 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 [12]. Given that the ability to speak enthusiastically provides a clear advantage in human interactions when the goal is to engage the interlocutor emotionally, we shall also aim to transfer it to virtual agents.

Although enthusiasm has been studied extensively in psychology, showing that students clearly benefit from enthusiastic teachers [5, 44] as well as companies do with enthusiastic leaders [23, 34], it is still unclear what exactly makes a person to be perceived as enthusiastic [19]. Most of the work on detecting enthusiasm automatically is based on using written human-to-human conversational dialogues [17, 38]. Limited work on enthusiastic virtual agents and robots exist. Liew et al. showed that virtual agents with enthusiastic voices improve learning outcomes in students [24–26], 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, called Entheos [40], 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 some preliminary analysis on the Entheos dataset[1] and discuss how explainable AI techniques such as SHAP [28] can be used to understand which features are important for enthusiasm classification.

2 OPPORTUNITIES FOR ENTHUSIASTIC VIRTUAL AGENTS

Although several applications of virtual agents could benefit from understanding and generating enthusiastic behavior, we will focus on three specific use cases: a) teaching, b) coaching, and c) sales. All three applications are not purely conversational, but focus on delivering a message in an engaging way, which we believe is a feasible first step to improving existing virtual agents.

Teaching: Enthusiasm has shown to improve students' performance not only with human teachers [5, 19, 20, 44] but also with virtual teachers using prerecorded enthusiastic voices [24, 26]. Given the increasing importance of pedagogical agents and robot teachers to overcome the global shortage of teachers [29], it is essential to automatically generate enthusiastic behavior during teaching to ensure students' performance and their engagement.

Coaching: Several virtual coaches have been designed to promote healthier behaviors targeting for ex. elderly people [11] or patients recovering from spinal cord injuries [36]. Virtual coaches

[1] https://github.com/clviegas/Entheos-Dataset

IVA '21, September 14–17, 2021, Virtual Event, Japan

ACM ISBN 978-1-4503-8619-7/21/09.
https://doi.org/10.1145/3472306.3478366

have also been successfully used in virtual reality, providing feedback on elevator pitches [45]. Although the mentioned coaches have shown to be useful for users, we believe that introducing enthusiastic behavior can improve the users' experience and their outcomes, given its particular ability to move listeners to action.

Sales: Virtual sales agents are gaining more interest as they are easily scalable with the needs of online sales. Virtual agents able to give recommendations and negotiate have been tested for online sales [18] as well as in combination with robots that give customers their purchases in stores [30]. Combining non-verbal cues with appealing arguments can improve the persuasiveness of virtual sales agents.

3 CHALLENGES

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 may vary depending on the user's age, gender and culture. The use of explainable AI techniques can help us better understand which features help models make better predictions in a diverse dataset [1]. Weber et al. used GradCam and layerwise relevance propagation to analyze image regions that make speakers be perceived as persuasive [42]. Reverse engineering enthusiasm as Chen et al. have done for other conversational expressions [8] is another method that can help obtain a better understanding of this culturally grounded emotional behaviour.

Once enthusiastic behavior is detected in users, 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 has improved the users experience [15, 16, 32, 41]. Rapport has been shown to create interpersonal responsiveness and influence between two people which is beneficial in the relationship of customers and employees or teachers and students[16]. The same is true for the relationship of humans and virtual agents. Generating multimodal enthusiastic behaviours can be explored through the new computational data-driven models that have been used for generating emotional presentations. 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 [21]. Style transfer methods have also been used to generate voices with different emotions, maintaining content [33, 46]. Rule-based methods can be used to implement simple findings of enthusiastic behavior, such as pitch variation or facial expressions [37].

4 ENTHEOS DATASET

Entheos is the first multimodal dataset for studying enthusiastic behaviors [40]. The dataset is composed of 1126 samples extracted from 113 different TED talks and is annotated as monotonous, normal, or enthusiastic speech (each with 123, 848, and 155 samples). Viegas et al. evaluated different labels and temporal granularity in a preliminary study [40] to describe enthusiastic speech, such as using the Public Speaking Competence Rubric [35] and vocal attributes on a sentence-level and entire talk. However, the highest inter-rater agreement for three annotators was obtained when using monotonous, normal, and enthusiastic labels on a sentence-level (Fleiss' kappa [22] = 0.82).

5 PRELIMINARY ANALYSIS OF ENTHUSIASTIC HUMAN BEHAVIOUR

We trained a Random Forest classifier using EGEMAPS features [13, 43] 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 [28]. The features with the highest SHAP values (highest impact on the model output) were different statistical measures of pitch and loudness. 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 [3].

We also performed two separate one-way ANOVAs on the Facial Action Units (AUs) detected with OpenFace [4] 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.

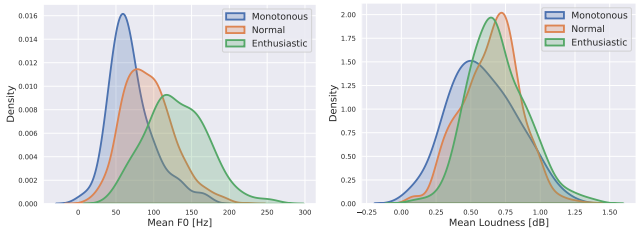

Figure 1: Label distribution for mean pitch (F0) and mean loudness. Left: enthusiastic samples have a higher mean F0 than monotonous samples. Right: monotonous speech tends to have lower mean loudness than enthusiastic speech.

6 CONCLUSION

Designing AI agents that can respond properly to more fine-grained emotional constructs can improve the quality of the interaction and increase the chances that the agent can achieve a common ground with the user in a more effective and human-like manner. We hope that the presented multimodal dataset Entheos will make future research on understanding and generating enthusiasm possible.

REFERENCES

[1] 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.

[2] 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/

[3] 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).

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

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

[13] 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.

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

[15] 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.

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

[17] 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.

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

[19] 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.

[20] 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.

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

[22] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* (1977), 159–174.

[23] 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.

[24] 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.

[25] 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).

[26] 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.

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

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

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

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

[35] Lisa M Schreiber, Gregory D Paul, and Lisa R Shibley. 2012. The development and test of the public speaking competence rubric. *Communication Education* 61, 3 (2012), 205–233.

[36] 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.

[37] 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.

[38] 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.

[39] 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

[40] Carla Viegas and Malihe Alikhani. 2021. Entheos: A Multimodal Dataset for Studying Enthusiasm. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)*.

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

[42] 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.

[43] 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).

[44] 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.

[45] 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.

[46] 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).