Spark Creativity by Speaking Enthusiastically

Communication Training using an E-Coach

Carla Viegas
TalkMeUp Inc.
Carnegie Mellon University
Pittsburgh, United States
cviegas@cs.cmu.com

Albert Lu
TalkMeUp Inc.
Carnegie Mellon University
Pittsburgh, United States

Annabel Su
TalkMeUp Inc.
Carnegie Mellon University
Pittsburgh, United States

Carter Strear
TalkMeUp Inc.
Carnegie Mellon University
Pittsburgh, United States

Yi Xu
TalkMeUp Inc.
Pittsburgh, United States

Albert Topdjian
TalkMeUp Inc.
Pittsburgh, United States

Daniel Limon
TalkMeUp Inc.
Pittsburgh, United States

JJ Xu
TalkMeUp Inc.
Pittsburgh, United States
jj@talkmeup.co

ABSTRACT

Enthusiasm in speech has a huge impact on listeners. Students of enthusiastic teachers show better performance. Leaders that are enthusiastic influence employee's innovative behavior and can also spark excitement in customers. We, at TalkMeUp, want to help people learn how to talk with enthusiasm in order to spark creativity among their listeners. In this work we want to present a multimodal speech analysis platform. We provide feedback on enthusiasm by analyzing eye contact, facial expressions, voice prosody, and text content.

CCS CONCEPTS

• **Human-centered computing** → **Visualization**; • **Computing methodologies** → **Feature selection**.

KEYWORDS

Dataset; Enthusiasm; Gaze Detection; Public Speaking; Multimodal E-Coach

ACM Reference Format:
Carla Viegas, Albert Lu, Annabel Su, Carter Strear, Yi Xu, Albert Topdjian, Daniel Limon, and JJ Xu. 2020. Spark Creativity by Speaking Enthusiastically: Communication Training using an E-Coach. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3382507.3421164

ICMI '20, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-7581-8/20/10.
https://doi.org/10.1145/3382507.3421164

1 INTRODUCTION

How we communicate can have a tremendous impact on our listeners, especially in teaching and leadership. Most of us have had at least one teacher that marked our career through their engagement and enthusiasm for their subject. Several studies have shown that enthusiastic teachers have students that perform better [1, 5]. Enthusiasm also has been shown to be one of the characteristics of leaders that drive employee's innovative behavior and even excitement in customers [2, 4].

Given the importance of enthusiasm in communication we developed an algorithm that takes into account that speaking has a multimodal nature. We analyze video, audio, and content and provide feedback to individual behaviors that contribute to enthusiasm. Users can record themselves through our online platform[1], receiving immediate feedback on their performance. By practicing several times, users are able to visualize their progress over several practice sessions and choose specific courses that target their weaknesses.

In the following, we will present the architecture behind our multimodal speech analysis platform, including an overview of the machine learning models and features employed, as well as the feedback users receive and how progress is shown. We conclude this work with a summary and outlook.

2 ARCHITECTURE

Our software analyzes data and provides written feedback in three stages: 1) data processing and classification, 2) conversion of classification results into performance values, and 3) translating performance results into meaningful written feedback that explains how to improve. In the following we will explain each stage visible in Fig. 1 in more detail.

[1] https://talkmeup.co

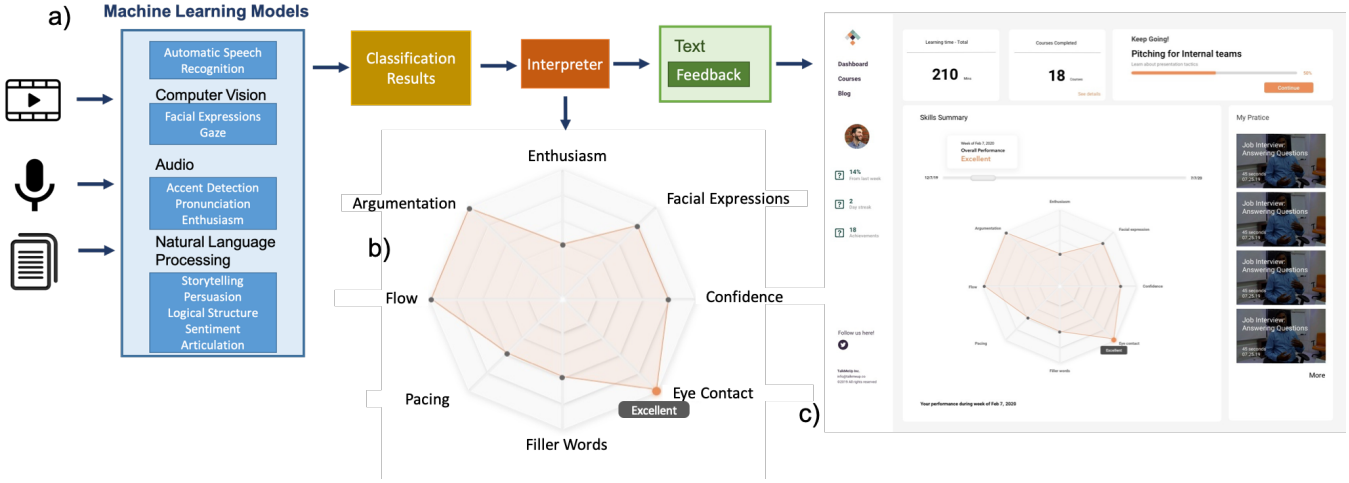

Figure 1: Overview of the backend architecture of our multimodal communication coach. In a) different state-of-the-art algorithms analyze data separately. The classification results are then fed into an interpreter which uses heuristic rules to create feedback for the user. In b) the performance in different categories is shown as a spiderweb. In c) the user's dashboard is visible where it is possible to select the practice session with a slider, showing in the spiderweb the performance.

2.1 Multimodal Classification

Given that communication is multimodal, we analyze different features over all three modalities (video, audio, text). We use a combination of deep neural networks, signal processing, computer vision, natural language processing, and machine learning to obtain classification results for different features. From video we evaluate facial expressions and eye contact. We use audio to evaluate pronunciation [3], enthusiasm, articulation quality, and filler word detection. For enthusiasm detection from speech, we created our own dataset composed by spoken sentences from TED talks labeled in three categories (monotonous, normal, enthusiastic). For text analysis, we use dictionary-based rules to evaluate storytelling, logical structure, and sentiment analysis. We trained a deep neural network to evaluate argument quality using the publicly available UKPConvArg1 corpus[2]. For each modality we implemented a microservice that is scalable to the usage. This allows us to analyze and provide feedback to the users in less than two seconds.

2.2 Numerical Interpretation and Feedback

The raw classifications are meaningless to users. For that reason, we implemented heuristics with support from expert communication coaches to interpret the results. We take the raw results and compute performance scores in the form of "excellent", "great", "good", "ok", and "keep practicing" for categories such as enthusiasm, argumentation, and more (see Fig. 1b). For each category and performance level we return written feedback with suggestions for improvement. As communication is multifaceted, we display the results in a spider web format that enables users to visualize and compare their performance in different categories at one view (Fig. 1b). In the main user dashboard (Fig. 1c) a slider is available that allows selecting training sessions over time to visualize how the performance changed in each session.

[2]https://github.com/UKPLab/ acl2016-convincing-arguments

3 CONCLUSION

We have presented a multimodal speech analysis platform that analyzes video recordings to provide feedback on different aspects of effective communication. We combine different state of the art techniques using video, audio, and text to obtain quantitative data on the performance by analyzing facial expressions, eye contact, argument structure, and more. We also collected data and trained a model able to detect enthusiasm levels from speech. In order to make the user experience more immersive we plan to develop a virtual agent able to analyze multimodal data and plan multimodal dialogue.

ACKNOWLEDGMENTS

We want to acknowledge the inspiration and support from the rest of TalkMeUp team, our board of directors and advisors, including Prof. Ron Placone from Carnegie Mellon University, Andrew Rabin and Bruce Gebhardt, Innovation Works, and Swartz Center for Entrepreneurship of Carnegie Mellon University.

REFERENCES

[1] 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. https://doi.org/10.3102/00028312020003435

[2] Jeroen P.j. De Jong and Deanne N. Den Hartog. 2007. How leaders influence employees innovative behaviour. *European Journal of Innovation Management* 10, 1 (2007), 41–64. https://doi.org/10.1108/14601060710720546

[3] Sandra Kanters, Catia Cucchiarini, and Helmer Strik. 2009. The Goodness of Pronunciation algorithm: a detailed performance study. *Speech Communication* (01 2009), 49–52.

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

[5] 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.