FILTWAM - A Framework for Online Game-based Communication Skills Training Using Webcams and Microphones for Enhancing Learner Support

Kiavash Bahreini, Rob Nadolski, Wen Qi, and Wim Westera

Center for Learning Sciences and Technologies, Open University of The Netherlands, Heerlen, The Netherlands

Abstract

This research provides an overarching framework embracing conceptual and technical frameworks for improving the online communication skills of lifelong learners. This overarching framework is called FILTWAM (Framework for Improving Learning Through Webcams And Microphones). We propose a novel web-based communication training approach, one which incorporates relevant and timely feedback based upon learner's facial expressions and verbalizations. This data is collected using webcams with their incorporated image and microphones with their sound waves, which can continuously and unobtrusively monitor and interpret learners' emotional behaviour into emotional states. The feedback generated from the webcams is expected to enhance learner's awareness of their own behaviour as well as to improve the alignment between their expressed behaviour and intended behaviour. Our approach emphasizes communication behaviour rather than communication content, as people mostly do not have problems with the "what" but with the "how" in expressing their message. We report on a small-scale proof of concept study that on the one hand exemplifies the practical application of our framework and on the other hand provides first evaluation results on that.

Introduction

We live in a knowledge society in which communication skills become more important as they used to be in the past. More jobs require more skilled people with respect to communication skills. The purpose of this research is to investigate novel training approaches for improving communication skills of everyone who has already finished their formal education. Communication skills are a lifelong affair for all members of our society. We propose an online web-based communication skills training framework (FILTWAM). This is not a replacement of the face-to-face training, but intends to offer a smooth setting for learners to improve their communication skills at their own pace, place, and time. FILTWAM uses devices, such as mobile phones, laptops, tablets, for learners' communication and comprises an affective computing tool with combining two modalities into a single system for face and voice emotion recognition.

This framework embraces both face emotion recognition and voice emotion recognition mechanisms. It recognizes six basic face and voice expressions (happy, sad, surprise, fear, angry, disgust) as well as neutral and provides adequate feedback to the learners. This study improves existing research (Chibulushi & Bourel, 2003; Ekman & Friesen, 1978; Kanade 1973; Petta, Pelachaud, & Cowie, 2011).

FILTWAM comprises an affective computing tool with combining two modalities into a single system for face and voice emotion recognition. The affective computing tool interpreters learners' emotional behaviour into emotional states in real-time. At the conceptual level, the rules engine manages game rules, pedagogy rules, and influences training content. It provides desire feedback based upon the rules. The technical framework represents different layers and the physical components.

Methods

We live in a knowledge society in which communication skills become more important as they used to be in the past. More jobs require more skilled people with respect to communication skills. The purpose of this research is to investigate novel training approaches for improving communication skills of everyone who has already finished their formal education. Communication skills are a lifelong affair for all members of our society. We propose an online web-based communication skills training framework (FILTWAM). This is not a replacement of the face-to-face training, but intends to offer a smooth setting for learners to improve their communication skills at their own pace, place, and time. FILTWAM uses devices, such as mobile phones, laptops, tablets, for learners' communication and comprises an affective computing tool with combining two modalities into a single system for face and voice emotion recognition.

This framework embraces both face emotion recognition and voice emotion recognition mechanisms. It recognizes six basic face and voice expressions (happy, sad, surprise, fear, angry, disgust) as well as neutral and provides adequate feedback to the learners. This study improves existing research (Chibulushi & Bourel, 2003; Ekman & Friesen, 1978; Kanade 1973; Petta, Pelachaud, & Cowie, 2011).

FILTWAM comprises an affective computing tool with combining two modalities into a single system for face and voice emotion recognition. The affective computing tool interpreters learners' emotional behaviour into emotional states in real-time. At the conceptual level, the rules engine manages game rules, pedagogy rules, and influences training content. It provides desire feedback based upon the rules. The technical framework represents different layers and the physical components.

Sixteen participants, all employees from the Centre for Learning Sciences and Technologies (CELSITEC) of Open University of the Netherlands (12 male, 4 female; age M=42.5, SD=10.9), volunteered to participate in this experiment. We used a straight forward simple feedback (red/green signal) to inform the learner whether the software detects the same 'emotion' as the participant was asked to 'mimic'. Five consecutive tasks were given to the participants: 

1. Train the database of the affective computing software,
2. Mimic the emotion that was presented through PowerPoint slides,
3. Mimic the seven face expressions two times,
4. Slides presented a text transcript (both sender and receiver) taken from a good-news conversation, participants were requested to read and speak aloud the sender 'idiot' of transcript.
5. All tasks were performed on a single Mac machine. The Mac screen was separated in two panels, left and right. The participants could watch their facial expressions in the affective computing software at the left panel, while they were performing the tasks using a PowerPoint file in the right panel.

The participants invited to participate in this experiment by an email. They were asked if they intend and like to improve their acting skills. They performed each individual session in about 20 minutes. They sat in a completely silent room with good lighting condition. Each participant was asked to read and sign the form before the session is started. The moderator of the session presented in the room, but with no intervention. All sessions were performed in about 325 minutes.

Conclusion

We propose a multimodal framework that in real-time interprets emotional behaviour into emotional states, is applied in educational settings, and is more precise for soft-skills training purposes. The results showed that the majority of the participants were able to accurately use the software; however they were not fully aware of their emotions to mimic them. They positively mentioned that they could easily fix the wrong emotions when looked at the reflections of their emotions in the mimic happy and neutral emotions, but they had a lot of problems to mimic other emotions. Almost all participants forgot the need for a new feedback solution to facilitate this process.