ABSTRACT
This paper describes how our FILTWAM software artifacts for face and voice emotion recognition will be used for assessing learners' progress and providing adequate feedback in an online game-based communication skills training. This constitutes an example of in-game assessment for mainly formative purposes. During this training, learners are requested to mimic specific emotions via a webcam and a microphone in which the software artifacts determine the adequacy of the mimicked emotion from either face and/or voice. Our previous studies have shown that these software artifacts are able to detect face and voice emotions in real-time and with sufficient reliability. In our current work, we present a software system architecture that unobtrusively monitors learners' behaviors in an online game-based approach and offers timely and relevant feedback based upon learner's face and voice expressions. Whereas emotion detection is often used for adapting learning content or learning tasks, our approach focuses on using emotions for guiding learners towards improved communication skills. Herein, learners need to have an opportunity of frequent guided practice in order to learn how to express the right emotion at the right time. We assume that this approach can address several issues with the current trainings in this area. We sketch the research design of our planned study that investigates the efficiency, effectiveness and enjoyableness of our approach. We conclude the paper by considering the challenges of this study.

Keywords
Formative assessment; communication skills; multimodal emotion recognition; serious gaming; software development; feedback provision.

1. INTRODUCTION
Communication skills are becoming more important in modern society. This causes a greater demand for communication skills training as it used to be in the past. Furthermore, there are still a lot of people that were educated in an area that communication was not that important in society as it is today [1]. Such people might have insufficient communication skills. Indeed, higher standards for communication skills - increasing the need for extended and varied practice - are applicable for all ages and foster new approaches towards communications skills trainings that are more suited for modern man. A flexible and online training program can offer a solution in which learners are able to practice a lot on a regular basis to enhance their communication skills. Nowadays, learners must attend specific courses that use face-to-face approach, turning them into quite inflexible training programs as far as freedom of place and time is concerned. The courses are costly and often demotivating because of the needed repeated practice. This frequent practice is inevitable for the learners in order to master the skill on a sufficient level. An additional problem is the shortage of trainers that can provide communication skills in face-to-face situations [2]. Serious games are games developed for educational purposes rather than entertainment. Such games seem adequate for addressing issues with online face-to-face trainings and to a certain extent also deal with the shortage of trainers as it can make a more effective use of their limited availability [3]. These games compared to the regular e-learning solutions are 1) engaging, 2) motivating, 3) user centric, 4) goal oriented, 5) more interactive, and 6) more personalized [4]. However, there is also an issue with many serious games with respect to their assessment of learning. They often do not reliably assess the learning [5]. Our approach intends to offer an effective training of communication skills while at the same time dealing with assessment of learning within the game. It is not a replacement of the face-to-face training, yet it offers much more flexibility and scalability.

We describe how our FILTWAM software artifacts for face and voice emotion recognition will be used for unobtrusive in-game assessment of learners' progress of learning. FILTWAM is integrated with a game-engine and is used for the development of a serious game for communication skills training. We call our serious game “Communication Advisor”. In this game, frequent feedback is provided for guiding learners towards improved communication skills (i.e., formative assessment [6], [7]). We assume that deploying the FILTWAM artifacts for multimodal emotion recognition can lead to better learning. Automated emotion recognition may compensate for the limited number of trainers that are available for the training of communication skills whereas a serious game is suggested because of its motivational strength for fostering learning in which frequent practice is needed for automatic skill mastery situations [2]. It is commonly acknowledged that emotions are important factors in any learning process, since it influences information processing, memory and performance [8]. Our previous research on face emotion recognition and voice emotion recognition has shown that it is possible to measure emotions from these two software artifacts with sufficient reliability in real-time [9, 10]. The easiest way and the most accessible equipment for gathering data for emotion recognition are webcams and microphones. There is valuable information inside face and voice expressions that can mirror affective aspects of learning, but that in the case of communication skills can also inform learners' progress towards their mastery. In our work, we focus on the latter usage of face and voice emotions, whereas most research deals with using emotions for adapting learning content or learning tasks. This insight has led to the research and development of affective tutoring systems [11]. Adequate communication is not only
about the ‘what’ (i.e., the content of the communication), but also about the ‘how’ (i.e., the way this content is delivered). Emotions need to be aligned with the message to have its intended effect. It is important for the learners to learn how to express the correct emotion at the right time. Feedback can guide the alignment between emotions and message and therefore expected to be significant in communication skills training. Also, feedback based on emotional states may enhance the learners’ awareness of their own behavior.

Communication Advisor allows learners to practice in so called conversation snippets. In each snippet the learner receives feedback. The feedback is based upon the detected learners' mimicked emotions in their expressed conversation part of the snippet. The content of the message has been chosen from text alternatives. Text alternatives are used so that the serious game can easily detect the content of the message in the snippet. Our approach proposes an example of unobtrusive in-game assessment for providing timely feedback to the learner. This assessment is meant for learning although the game could also be used as part of the setup for the assessment of learning in which case only limited or even no feedback would be given to the learner. It is assumed that deploying the FILTWAM artifacts for multimodal emotion recognition in Communication Advisor lead to better and more enjoyable learning of communication skills. In sum, to characterize the novelty of our work, we propose multimodal emotion recognition for assessment of learning in game-based communication training. In this paper, section 2 introduces the software system architecture. In section 3, we sketch the research design of our planned study that investigates the efficiency, effectiveness and enjoyment of Communication Advisor in which its current prototype is used to illustrate our approach in more detail. Section 4 discusses the challenges and provides few suggestions to conduct that study.

2. SOFTWARE SYSTEM ARCHITECTURE

We propose a loosely coupling system design of FILTWAM and EMERGO in our architecture. EMERGO is used as the game-based engine, and content manager of Communication Advisor. EMERGO is a methodology and open source toolkit for the development and delivery of serious games [12]. To connect Communication Advisor and EMERGO to FILTWAM, we followed the web service approach and developed a web service on both client and server sides of our architecture. The communication protocol between the game, EMERGO, and the software artifacts are established through a developed web service. Figure 1 demonstrates the components of the architecture and the relationships between them. We placed nine components in the data flow diagram: 1) Learner, 2) Browser, 3) EMERGO web service client, 4) Face emotion recognition software (FERS), 5) Voice emotion recognition software (VERS), 6) Real-time data file for FERS, 7) Real-time data file for VERS, 8) EMERGO serious game engine, and 9) EMERGO web service. The first three components are situated within the client side. The other six components have been placed in the server side. The learner opens the browser in the client side (number 1 in the figure 1) and launches the EMERGO serious game engine (number 2). The EMERGO serious game engine component deals with triggering specific feedback messages (i.e., content of feedback) of the game, manages game rules, and influences training content. The engine calls the EMERGO web service component (number 3) and triggers feedback based upon the rules. In the client side, the EMERGO web service client component that is already executed by the learner (number 4) is responsible to call FERS and VERS components (number 5 and 7). These two components generate a real-time data file for the FERS and for the VERS components (number 6 and 8). FERS and VERS do face emotion recognition and voice emotion recognition from the webcam input data that they receive from the learner. The EMERGO web service client uses both data files (number 9 and 10). It sends the real-time emotion data to the EMERGO web service (number 11) and this data will be sent through this component to the EMERGO serious game engine (number 12), then to the browser and the learner (number 14 and 15). As it is shown in figure 1, the learner will receive all feedback through a single browser on the client side.

Figure 1. Data flow of the software system architecture.

3. COMMUNICATION ADVISOR

We deploy Communication Advisor for skill-based learning using the EMERGO game engine. During development, the EMERGO method and toolkit is used. During design the EMERGO method is used. For testing and running the game, the EMERGO game engine is used. We offer the learner a varied set of audio-visual micro stories (tasks) in an online game-based setting. This approach intends to improve the alignment between the nonverbal behavior and the verbal message of the learner interaction snippets being part of such micro stories. An interaction snippet consists of (1) choosing an alternative (i.e., spoken text-message and emotion (which is either happy, sad, surprise, fear, disgust, angry, or neutral) from a small list of textalternatives, (2) speak the message with the chosen emotion in front of the webcam, (3) reaction from the conversation partner, and (4) presenting feedback as text format. The feedback is provided based on the message of the chosen text-alternative and
the shown emotion by the learner. Figure 2 illustrates the prototype of Communication Advisor when a task is presented to the learner.

![Communication Advisor prototype](image)

**Figure 2.** A screen capture of the Communication Advisor's prototype when a snippet is given to the learner.

### 3.1 Method

We follow the EMERGO approach for game design and development and use several game design guidelines from the literature [13, 14]. Ten micro stories (tasks) including one or more interaction snippets (twenty in total) will be offered to the learner within the game. Each snippet will be sequentially presented through webpages. The learner will see the following components on the screen (see figure 2): 1) a score counter, 2) a tasks’ list, 3) a number of completed snippets, 4) a recorded video of the conversation partner, 5) a reward component, 6) the text transcripts and the emotions, 7) an instruction text on how to proceed the game, 8) a button to proceed the game, 9) a button to redo the snippet 10) a message for the detected emotion of the face, and 11) a message for the detected emotion of the voice. The label of the detected emotion will be shown in green when the learner has presented the emotion from the chosen alternative correctly and will be shown in red when the learner did this incorrectly (see Figure 3). The reward mechanism will offer a prize when the learner expresses five out of twenty consecutive snippets correctly. The text transcripts and instructions for the micro stories will be selected from an existing OUNL training course [15] and a communication book [16]. Figure 3 presents an example of the feedback to the learner. The score counter component is increased by 5 and the completed snippets component is increased by 1 when the learner completes a snippet. The selected sentence and the emotion are also marked.

The entire game happens over the course of fifty minutes in an online virtual learning environment that is equipped with a microphone and a webcam. During our research, all learning sessions will be captured through an integrated webcam and a 1080HD external camera to capture and record the emotions of the participants. The interactions of the learner with the mouse and the keyboard on the computer screen will also be captured through Silverback usability testing software version 2.0. We examine five different versions of the game on its learning effectiveness, efficiency and attractiveness. The first version will only provide the feedback based on the chosen text alternative. The second and third versions will provide feedback on the combination of the chosen text alternative and one sensor (either webcam or microphone). The fourth version will provide feedback on the combination of the chosen text alternative and both sensors. Finally, the fifth version (i.e. control version) of the game does not provide any feedback. Some interaction snippets will be used as initial assessment for pre-skill measurement. Some will be used as summative assessment for post-skill measurement. In both pre-skill and post-skill measurement, no feedback and no reaction from the conversation partner will be shown to the learner. In addition, a pre-questionnaire (before the game) and a post-questionnaire (after the game) will collect other data that might be relevant for explaining individual differences in learning (e.g., attitude towards learning through games, motivation, and effort). We will especially examine the progress of each learner during the game play using assessment data from the sensors (recognized emotions) and chosen text alternatives within the micro stories. We will measure effectiveness of the learning progress after completing the game through comparing pre-skill measurement and post-skill measurement. The efficiency will be measured by measuring the study time. We measure enjoyableness of the learning by a questionnaire.

![Communication Advisor prototype](image)

**Figure 3.** A screen capture of the Communication Advisor’s prototype when the feedback is provided.

### 4. CHALLENGES AND DISCUSSION

In this article we described the integration of our FILTWAM software artifacts with an online serious game using the EMERGO platform and engine. The face and voice emotion recognition software and the game enable unobtrusive in-game assessment of learners’ progress are assumed to provide more adequate feedback than a game without those software artifacts. Both software artifacts are able to detect learner’s emotions for the purpose of providing timely feedback (i.e. instantly after the learner has pronounced the conversation snippet) within Communication Advisor. We focus on analyzing emotions for improving learner’s communication skills whereas previous
research on emotions in e-learning has mainly dealt with adapting learning content and tasks based on emotion detection. It is an interesting avenue for future research to combine both approaches. For Communication Advisor this would imply offering different learning tasks based on learners’ detected emotions during the game play (for example, easier or more difficult tasks). Speech recognition was left out in our approach in order to keep our research scope as simple as possible. There are still some issues with speech recognition (validity, reliability, performance) that might obscure our findings. However, our architecture can be easily extended with speech recognition. It is acknowledged that speech recognition would be a very valuable extension to our approach as the training situation is more aligned with the real situation. Such alignment is assumed to be important for transfer (i.e., application of the learning in other contexts). A technical challenge of our approach might be the performance of our game once more content and rules would be added. At this moment of writing we did follow a pretty straightforward approach with web services, but a stronger integration between the two software artifacts of FILTWAM and the EMERGO platform is needed if it occurs to be an issue to deliver the feedback timely.

5. ACKNOWLEDGMENTS
We would like to thank the Netherlands Laboratory for Lifelong Learning (NELLL) of the Open University of the Netherlands, which sponsors this research.

6. REFERENCES