Can You Help Me with My Pitch? Studying a Tool for Real-Time Automated Feedback

Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht, *Member, IEEE*

Abstract—In our pursue to study effective real-time feedback in Technology Enhanced Learning, we developed the Presentation Trainer, a tool designed to support the practice of nonverbal communication skills for public speaking. The tool tracks the user's voice and body to analyze her performance, and selects the type of real-time feedback to be presented. This paper describes an empirical study where we tested the effects of the Presentation Trainer's feedback on learners who used the tool while practicing for an elevator-pitch. Results from this study reveal that the feedback has a significant effect on the learners' motivation, confidence, self-awareness, and performance.

Index Terms—Computers and education, feedback, multimodal systems

✦

1 INTRODUCTION

FEEDBACK is one of the most significant interventions in learning [1]. The effects of this feedback depend on a variety of variables of the different dimensions the feedback can have [2]. For the particular case of public speaking, feedback is a key aspect for learning and developing the respective skills [3]. The effectiveness of this feedback depends on various variables. For example, feedback provided by a tutor in combination with feedback provided by peer students has proven to be more effective than feedback provided only by a tutor [4]. Regarding the timing of feedback, studies have shown that for aspects that can be corrected immediately, such as eye contact and body posture, immediate feedback is more effective than delayed feedback [5]. However, having tutors or peers providing us with feedback whenever we have time to practice is neither an affordable nor a feasible solution. To support this kind of learning outside of the traditional classroom setting, we developed the Presentation Trainer (PT). The PT is a sensor-based tool designed to support the development of basic public speaking skills, by providing learners with immediate feedback about different aspects of their nonverbal communication. In this article we report on a study on the impact that the feedback given by the PT had on participants training for an elevator-pitch[1].

2 BACKGROUND

Using computer systems to support learners with personalized feedback and instruction is a practice that has been around for several years now. A classical example of this type of systems is the LISP Tutor, which appeared in 1983

and was designed to give real-time feedback to learners of the LISP programming language [6]. The adaptation and feedback of computer-based learning support was based on user-modeling approaches taking into account performance information on tasks given in the learning environment. In recent years, advances in technology have made it possible to consider sensor information and interaction of users in context for learning support [7]. The coupling of sensors with multimodal computer interfaces made it possible to track and automatically analyze users' actions [8] and physiological states [9]. This led to the research and development of feedback systems able to support learners in a vast number of learning activities [10] that cover the cognitive, affective and psychomotor domain of learning [11].

In the study described in this article we investigate how tutor systems with sensing capabilities can support the development of nonverbal communication skills. Tutor systems presenting support for these skills have already been studied for scenarios such as job interviews [12], [13] and public speaking [14], [15], [16], [17]. These public speaking tutor systems use different approaches and focus on different aspects. The prototype in [14] used wearable technologies, e.g., the Google Glass, to present the user with feedback for specific nonverbal communication factors while giving a presentation. The study in [15] explored the use of an armband that provided haptic feedback to the user about the timing of her presentation whereas [16] designed an environment to help learners to overcome their public speaking anxiety by giving presentations in front of a virtual audience. In addition, this system is also able to make an assessment of some nonverbal aspects of the presentations. However, an evaluation of this assessment is not described in their work. In the work of [17] the authors explored a tool, which provided learners with immediate feedback about some nonverbal communication aspects while practicing their presentations. In addition, the tool was augmented with exercises focusing on specific elements. While the study revealed that the students were eager to use the system, it also showed that the dashboard interface employed was too difficult to follow while practicing for a presentation.

---

1. https://en.wikipedia.org/wiki/Elevator_pitch

• *The auhors are with the Open University of the Netherlands, Heerlen AT 6419, Netherlands. E-mail: {jan.schneider, dirk.boerner, peter.vanrosmalen, marcus.specht}@ou.nl.*
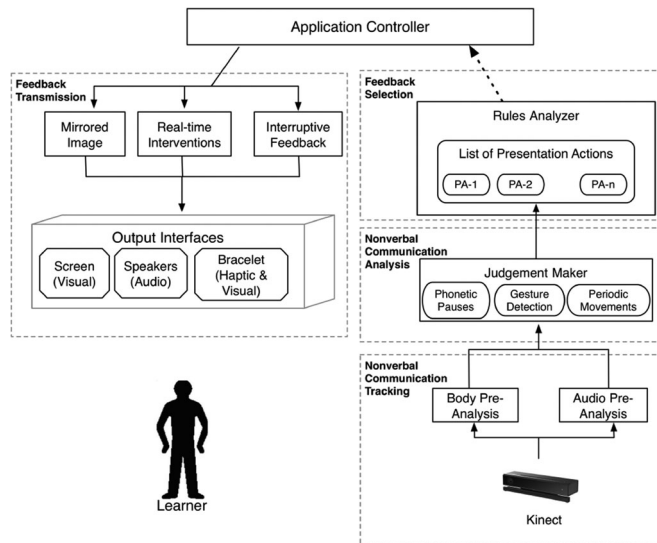
Fig. 1. Presentation trainer architecture.

To expand on the state of the art of tutor systems for public speaking and study the effects such a system can have on learners, we developed and evaluated the Presentation Trainer.

3 PRESENTATION TRAINER

The Presentation Trainer (PT) is a multimodal tool that supports learners with the self-practice of basic nonverbal communication skills for public speaking. It uses sensors to track the learners' body and voice to provide them with feedback about a set of basic nonverbal communication aspects for public speaking. Grounded on the results from related work, we decided to develop the PT based on the following assumptions:

- Immediate feedback is proven to be more effective for training nonverbal communication [5].
- The amount of cognitive load needed to practice for a presentation makes it difficult for the learner to pay attention to all the different elements simultaneously displayed on a dashboard interface. This makes dashboard interfaces far from ideal for immediate feedback for practicing public speaking skills [17].

Taking into account these two assumptions, the version of the PT described in this article has the capability to analyze the user's performance, and to select accordingly at most one nonverbal communication aspect to be presented as a feedback intervention.

3.1 Presentation Trainer Architecture

The PT has been developed in C# using the .NET framework 4.5. To capture the user's voice and body the current version of the PT uses the Microsoft Kinect for Windows V2 and its proprietary SDK. Its architecture shown in Fig. 1 has four main functionalities: Nonverbal Communication Tracking, Nonverbal Communication Analysis, Feedback Selection, and Feedback Transmission.

In order to track the user's nonverbal communication, the PT is constantly listening for new sensor data obtained from each sensor channel and stores this sensor data in their corresponding pre-Analysis object. The PT has a channel for

audio and a channel for the body of the user. Connected to these channels, it has an Audio pre-Analysis and a Body pre-Analysis object. The Audio pre-Analysis object has a 0.64 seconds long audio buffer that stores at a frequency of 16 kHz, the absolute volume values obtained from the microphones of the Microsoft Kinect for Windows V2. This object also contains a Boolean variable indicating whether the user is currently speaking. To infer whether the user is speaking or not the PT compares the average volume value of the buffer against isSpeaking threshold. If the average volume value is bigger than the threshold, the Boolean variable that indicates whether the user is speaking is set to true. The Body pre-Analysis object stores at a rate of 30 frames per second the current coordinates of the detected joints from the user's body. It also contains Boolean variables for all the postures that have been considered important for the analysis of the user's nonverbal communication for public speaking. If a posture is identified then its respective Boolean variable is set to true.

The JudgmentMaker object does the Nonverbal Communication Analysis. It analyses the data from the pre-Analysis objects in order to identify nonverbal communication mistakes or good practices. Whenever a specific mistake or good practice is identified the JudgementMaker creates a Presentation Action object and stores it on a list. If a Presentation Action is no longer identified, the JudgmentMaker removes it from the list.

The RulesAnalyzer is the object responsible for the Feedback Selection. It makes certain about the appropriate timing to present feedback. If timing is appropriate, it selects the oldest identified Presentation Action from the list of Presentation Actions, and triggers a feedback event about it. Whenever RulesAnalyzer identifies that the selected Presentation Action has been removed from the list, it triggers a correction event. The Application Controller of the PT receives the events and forwards them to the connected output channels that transmit the feedback to the users.

3.2 Presentation Trainer Feedback

The current version of the PT supports the training of basic public speaking skills, i.e., by providing learners with feedback about their use of pauses, voice volume, body posture, use of gestures, use of phonetic pauses, and steadiness in body posture. This set of nonverbal communication factors has been based on a synthesis of factors that according to public speaking manuals and courses affect the quality of a presentation [18], [19], [20], and we found them sufficient to study the feedback of the PT. Currently the PT is programmed to identify only mistakes in this set of factors. Whenever a mistake is identified, a Presentation Action is created. Next, we will discuss one by one each of the aspects we identify.

Good voice volume modulation in public speaking helps to communicate the message clearly and to keep the audience attention [19]. The PT uses the microphone from the Microsoft Kinect V2 to capture the voice volume of the learner. In the case where the PT perceives the voice volume as too soft or too loud, it creates a Voice Volume Presentation Action. To infer whether the speaking volume is too soft or too loud the PT follows the next procedure. It captures the sound through the use of a microphone at a rate of

16 kHz and stores the absolute volume values in a 0.64 seconds long buffer. Then it compares the average value of the buffer against three thresholds that can be manually set up during runtime to adapt to the acoustic needs of the room where the PT is being used. These thresholds are: speaking threshold, soft speaking threshold and loud speaking threshold. A soft volume Presentation Action is created when the average volume of the buffer is in between the speaking threshold and the soft speaking threshold. A high volume Presentation Action is created when the average volume is above the loud speaking threshold.

Using pauses correctly is a very important skill for public speaking [18]. The appropriate use of them allows the audience to take a breath when information is dense in content or emotion. Pauses also prepare the audience for the next subject, and are able to add some dramatic emphasis during the presentation. To identify a pause the PT has an isSpeaking volume threshold, volume values below that threshold are considered as silence. A pause is identified whenever the average value of the volume buffer remains below the isSpeaking threshold of a period longer than 0.25 seconds. Assessing the correct moment to pause during a presentation is highly dependent on its content e.g., pausing at the end of a sentence, after a rethorical question, etc. The aim of the PT's feedback regarding the use of pauses is to remind and make the learner aware about pausing while presenting, instead of pointing out the learner about the precise moment to pause. Therefore, whenever the PT does not detect a pause after the predefined time of 15 seconds, it raises a Presentation Action about pauses. We came up with the times of 0.25 seconds and 15 seconds by analyzing the average speaking time and pausing time of 15 different Ted Talks.

The speaker's body posture helps to convey confidence, openness and attentiveness towards the audience. To convey these attributes it is recommended to stand up with an open posture, straight, facing the audience, with the hands always visible inside of the acceptable box space and preferably above the hips [18]. The PT uses the Microsoft Kinect sensor V2 to track the learner's body. This body tracking presents the PT with the coordinates of the learner's joints. These coordinates are later used to infer the learner's body posture. Even when the learner stays still, these coordinates still flicker, however the flickering obtained by the Microsoft Kinect V2 is usually not big enough to interfere with the posture identification of the learner We apply a time threshold to improve the level of accuracy for posture identification, it helps to distinguish between postures and movements. The threshold is experimentally determined to be 0.3 seconds. This means that the PT recognizes a posture if the tracked body coordinates of the learner remain inside of some predefined posture values for a period longer than 0.3 seconds. Whenever the recognized posture violates the preset posture rules, the PT generates a body posture Presentation Action.

Hand gestures are a powerful tool to communicate your message in public speaking. They are able to enhance a speech by: painting vivid pictures in the listeners' minds, conveying the speaker's feelings and attitudes, enhance audience attentiveness and retention, dissipate nervous tension, etc. [20]. In the current version, the PT is not able to identify specific gestures. It is only able to recognize whether the learner uses gestures while speaking. To do this, the PT uses input of the Microsoft Kinect sensor V2 to get the coordinates of the learner's joints and keeps track of the angles between forearms and arms, and between arms and shoulder blades. The PT notices the decrement and increment of these tracked angles. If angles start increasing and stop decreasing, or the opposite way around, stop increasing and start decreasing a "pre-gesture" is identified. A gesture is recognized when the total increment or decrement of the "pre-gesture" angles are greater than 5 degree. This strategy of identifying gestures has proved to be very accurate. Because even with the constant flickering of body coordinates tracked by the Kinect sensor, the angle change between the tracked user's limbs is never greater than 5 degree when the user is not moving. Moreover, when using gestures the difference in angles are always far greater than 5 degree. A Presentation Action about gestures is created whenever the user is speaking, and no new gestures appear for a predefined time set to six seconds. We set the predefined time to six seconds because while tuning the PT, we observed that people who stay longer than six seconds without using gestures generally continue the presentation without using them at all, and that a gesture rarely takes longer than six seconds to be completed.

The phonetic pauses or filler sounds are all the "ehm", "hmm", "aah", etc. sounds made by the speaker. These sounds show hesitation, which is not a good practice for public speaking, therefore during the Toastmasters gatherings it is common to have an Ah-counter indicating the speakers how many times they have used a filler sound. The PT uses the speech recognition capabilities of the Microsoft Kinect V2 to recognize some of this filler sounds. The current accuracy for the PT to recognize these filler sounds is about 20 percent. This accuracy level is quite low, however we consider it satisfactory enough to remind users about this type of mistake. It is possible to increase this accuracy level but at the moment this would translate into the detection of false positives. Whenever the PT recognizes a filler sound it creates a phonetic pause Presentation Action.

By examining several presentations of novice speakers and interviewing teachers in public speaking, we identified that a common mistake that novice speakers make is to switch weight from one leg to the other, showing nervousness, a lack of confidence and giving the impression that they are dancing during their presentations. To track this behavior the PT uses the Microsoft Kinect sensor V2 to track the X and Z coordinates of the learner's hips. The PT uses a counter to take note of the number of swings from these coordinates, every 4 seconds this counter is reset. If the PT identifies 3 or more swings in 4 seconds, it creates a Presentation Action about staying grounded.

The PT stores all the current identified Presentation Actions on a list and deletes the ones that are no longer detected. This strategy of generating Presentation Actions whenever a rule is detected makes the PT scalable, allowing the inclusion of new type of "nonverbal mistakes" and "good practices" for updated versions of the tool.

Once the current Presentation Actions are identified, the PT is able to present the learner with feedback about their nonverbal communication. The amount of cognitive load [21] required from the learner while practicing for a presentation is quite high. The learner needs to know their topic
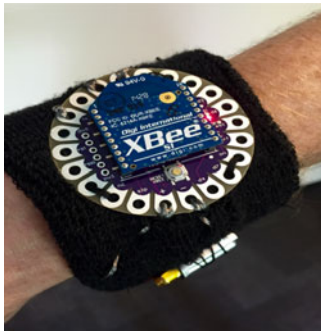
Fig. 2. Wristband used for haptic feedback.

(what to present and how to structure it) and how to deliver it (how to use their voice e.g., pitch, speed or volume, body, etc). Therefore we need to carefully design a feedback mechanism, that can actually help the learner to become aware of her nonverbal communication, adapt it, and use this increased awareness to improve her skills [17]. The main graphical interface of the PT constantly shows a mirrored image of the learner with the purpose to support the raise of self-awareness. When considered necessary, the PT transmits to the learner on top of this a feedback-instruction. To limit the cognitive load at most one feedback-instruction is given at a time. This feedback-instruction is currently transmitted through a visual and haptic channel, since research has shown that as the cognitive load increases more redundant multimodal communication is needed [22]. The visual feedback is displayed through the graphical interface of the PT on top of the mirrored image of the user. To transmit the feedback through the haptic channel, we developed a Feedback Wristband (FW) (Fig. 2).

This FW can be connected to the PT via Bluetooth. Whenever a feedback intervention is triggered, the PT sends a signal to the FW. At this time the FW produces a small vibration indicating the learner to pay attention to the screen, because a Presentation Action has been identified.

The procedure executed by the PT to provide the user with feedback is the following:

First the PT checks for the appropriate time to present feedback to the learner. A small user study conducted, indicated that a constant stream of feedback, even when it was only one type of feedback at the time, resulted in too much confusion for the users. Therefore, the current version of the PT waits at least six seconds after the last feedback stopped being shown, in order to consider it appropriate to give new feedback to the user.
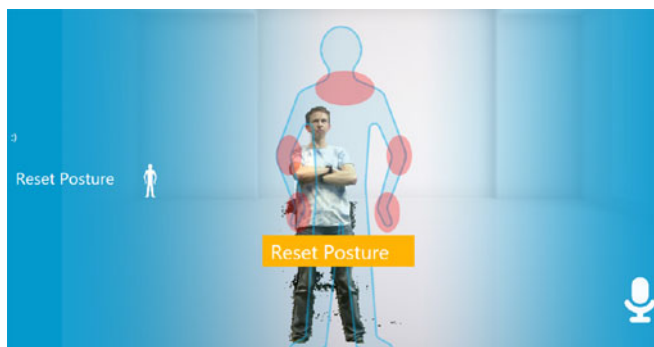

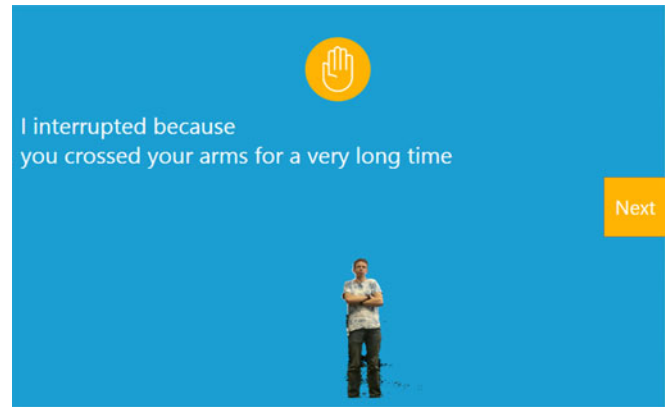Fig. 3. Immediate corrective feedback for crossing arms.

Fig. 4. Interruptive feedback for a long time of crossing arms.

Once the time to give feedback is appropriate, the PT looks at the list storing the current Presentation Actions. In the case the list is not empty, the PT selects the Presentation Action to be shown. Currently the PT uses a FIFO strategy to make this selection. This means that the selected Presentation Action is the one that has been for the longest time in the list. After making this selection, the PT decides whether the feedback should be corrective or interruptive. Corrective feedback indicates that a Presentation Action has been identified. This type of feedback produces in real-time a small vibration in the feedback wristband, and is visually displayed on top of the mirrored image of the learner. The feedback instruction displayed shows an icon and a short (maximum two words long) written instruction indicating how to correct the identified mistake (Fig. 3). This feedback icon remains on the screen until the mistake is corrected. When the mistake is corrected a check mark appears in the screen, the feedback icon and instruction fade away, and the appropriate time to show a new feedback starts its countdown.

In the case that a mistake is not corrected after 20 seconds or that a mistake has been repeated several times (currently set to 5) the PT presents the user with interruptive feedback. Interruptive feedback produces some vibration, a pause sound, stops the program, and displays on the screen the reason of the interruption (Fig. 4). The interface of the interruption offers the user the option to continue practicing the presentation receiving feedback on all nonverbal aspects, or only on the aspect that she was interrupted for, so that she can shift her focus on this specific skill.

4 PURPOSE OF THE STUDY

We conducted this empirical study with the purpose to evaluate the PT as an effective feedback tool for supporting the development of basic nonverbal communication skills for public speaking. To evaluate this effectiveness we explored the influences that the PT's feedback has in terms of learner's perception and learner's performance. For these two aspects we tested the following hypotheses:

Hypothesis 1. *The feedback of the PT will raise the learners' awareness of their nonverbal communication, increase their confidence on their skills, and increase their motivation to be trained.*

To test Hypothesis 1 we examined the learners' perception, i.e. we used two questionnaires, one for each phase of the experimental set-up.
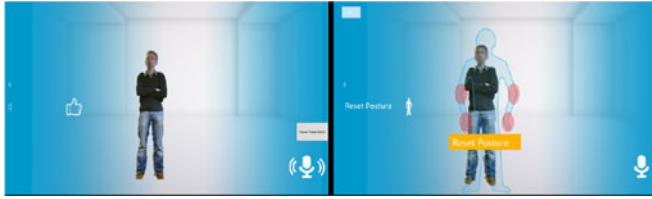
Fig. 5. Left: PT version for control group. Right: PT version for treatment group.

Hypothesis 1. *The feedback of the PT has a positive influence in the learners' performance. Learners who trained using the full version of the PT will perform better than learners who trained with the limited version of the PT.*

To test Hypothesis 2, we used the log files of the PT and analyzed the evolution of the performance scores from the pitches executed during the training sessions and the performance score of the final unsupported pitch.

5   METHOD

In this study we investigated the influence that the feedback of the PT has on learners practicing for an elevator-pitch. An elevator-pitch is a 30 to 120 seconds long speech that summarizes in lay terms what one does and why it is important. We deliberately chose the elevator pitch over other types of presentations because we consider it complex enough for participants to train their nonverbal communication, and short enough to fit in the time constrains of the experiment. To conduct this study we followed a quasi-experimental design [23] with a treatment and a control group, where the independent variable used was the feedback of the PT.

5.1   Participants

In this study we had a total number of 40 participants. Each group, the treatment and the control group, contained 9 female and 11 male participants. The age of the participants ranged between 24 to 62 years, with an average age of 42.6 years. All participants were professionals working at our university, with a similar western European cultural background. We recruited them by personally asking for their willingness to take part in our experiment. The criteria used to accommodate them in the treatment or control group was randomly based on the number of their experimental session. Participants from odd sessions (1st, 3rd, etc.) were assigned to the treatment group, and participants from even sessions (2nd, 4th, etc.) were assigned to the control group.

5.2   Apparatus and Materials

As an intervention tool we used two different versions of the PT. The control group used a limited version of the PT. This limited version of the PT did not provide the users with any feedback intervention; it only displayed the mirror image of the user (Fig. 5 Left). The treatment group used the full version of the PT described in previous chapters (Fig. 5 Right).

To measure the effects of the PT's feedback and test our hypothesis we used two questionnaires and the performance data logged by the PT. The first questionnaire is a user experience questionnaire containing three questions. These questions inquire the motivation, perceived amount of learning, and learning experience in comparison with a traditional
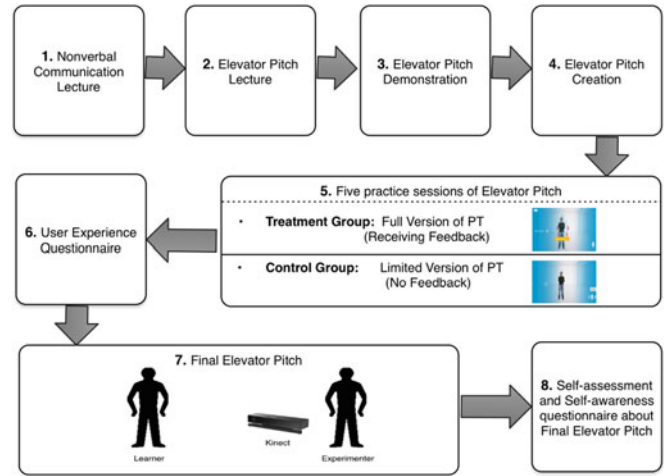

Fig. 6. Experimental procedure.

classroom setting. Answers to these questions are given on a Likert-type scale, whose ranking goes from 1 to 10.

The second questionnaire is a self-assessment and self-awareness questionnaire. This questionnaire contains eight items for self-assessment, six items for self-awareness, and one item about self-confidence. The items of self-assessment ask participants to provide ratings for the overall pitch, the overall nonverbal communication, body posture, use of voice, cadence, staying grounded and phonetic pauses. The ratings for these items are given in a Likert scale with values ranging from 1 to 5. The items about self-awareness ask participants to indicate the perceived amount of body posture, volume, gestures, phonetic pauses, cadence, and grounded mistakes performed during the pitch. Finally, the item of self-confidence asks participants to evaluate their self-confidence about their "elevator-pitch" skills. This evaluation is rated using a Likert scale whose values range from 1 to 5.

To measure the performance of the participants we used the log files generated by the PT during the training sessions and the final pitch. These log files include: the starting and ending time of each training session and pitch, all identified Presentation Actions (mistakes) together with their corresponding starting and ending timestamps, and all Feedback events with their corresponding timestamps.

5.3   Procedure

All experimental sessions in this study were individual and were performed following the procedure described in Fig. 6. Each session started with a five-minute lecture about nonverbal communication for public speaking. The nonverbal aspects taught in this lecture, were the same aspects as the ones tracked by the PT: body posture, use of hand gestures, voice volume, pauses, phonetic pauses, and ability to stay grounded. This lecture had two purposes. The first purpose is to assure that all participants had a similar baseline of basic knowledge about nonverbal communication for public speaking. It is important to clarify that the PT a tool designed to support the practice basic nonverbal communication skills for public speaking rather than a tool designed to teach basic nonverbal communication skills for public speaking. The second purpose of the lecture is to teach assure that people learned about these skills, before they start practicing.

Fig. 7. Training session setup with the PT giving feedback.

TABLE 2
Average Difference between Perceived Mistakes
and Mistakes Capture by the PT for the Final Pitch

Mistake type	Treatment Group	Control Group	t - test
Voice volume mistakes	M = 3.45, SD = 4.03	M = 4.15, SD = 4.5	t(38) = .53, p = .6
Posture mistakes	M = 1.25, SD = 1.07	M = 4.4, SD = 4.25	t(21) = 3.05, p < .01
Gestures mistakes	M = 1.55, SD = 2.44	M = 2.5, SD = 3.19	t(36) = 1.06, p = .3
Cadence mistakes	M = 1.5, SD = 1.3	M = 2.1, SD = 3.59	t(24) = .71, p = .49
Phonetic Pauses mistakes	M = 2.45, SD = 2.5	M = 2.2, SD = 2.01	t(36) = .35, p = .73
Grounded mistakes	M = 0.25, SD = .55	M = 1.7, SD = 2.78	t(20) = 2.29, p < .05

Once this lecture finished, participants had another five-minute lecture about elevator-pitches. In this lecture participants learned the basic elements needed to create their own elevator-pitch. This lecture finished with a live example of an elevator-pitch performed by the tutor. As soon as the lectures finished, participants had five minutes to create their own elevator-pitch. Participants were free to use any topic for their pitch.

On the next stage of the test participants practiced the delivery of their recently created pitch. Participants of both groups practiced the pitch in five consecutive training sessions. For these practice sessions participants stood between 1.5 m and 3 m in front of the Microsoft Kinect sensor and a 50 inches monitor displaying the interface of the PT (Fig. 7). Participants from the control group practiced their pitch using a limited version of the PT, only showing a mirrored image of the user, without the provision of any feedback. Participants from the treatment group practiced their pitch using the full version of the PT, receiving immediate and interruptive feedback when necessary.

After the fifth practice session the participants were asked to answer the user experience questionnaire. In the next phase of the experiment participants had to deliver a final elevator-pitch, without the assistance of the PT. To deliver this pitch participants stood 1.5 m to 3 m in front of the Microsoft Kinect V2 sensor. Just behind the Microsoft Kinect V2 sensor was the experimenter pretending to be the

TABLE 1
Average Results on the Three Dimensions of User
Experience Extracted from the Post Session
Questionnaire (Ratings from 1 to 10)

Dimension	Treatment Group	Control Group	t - test
Motivation *10 = very motivated	M = 7.89, SD = 2.05	M = 4.47, SD = 1.68	t(26) = 5.62, p < .01
Learning Perception *10 = learned a lot	M = 7.47, SD = 1.17	M = 6, SD = 2.42	t(35) = 2.38, p < .05
Practice using PT vs. Classroom *10 = much better than classroom	M = 6.94, SD = 1.93	M = 6.1, SD = 2.08	t(36) = 1.29, p = .2

audience of the pitch, and controlling the PT (starting and ending the pitch session). The PT in this final pitch was used only to log the performance of the participants. The participants were not able to see the interface. The experimental session ended by asking participants to fill in the self-assessment and self-awareness questionnaire about their final pitch.

6 RESULTS

The average of the answers from the 1st questionnaire regarding user experience are shown in Table 1. We used a heteroscedastic t-test to compare the difference between both groups. These results show that participants from the treatment group indicated to have felt significantly more motivated while practicing for their pitch than participants from the control group. The amount of perceived learning from participants of the treatment group was also significantly higher in comparison with the control group. Without significant differences between groups, both groups indicated a positive attitude towards the use of a PT-alike tool in comparison to learning in a traditional classroom setting.

In the final questionnaire the participants assessed their final pitch on Volume, Posture, Gesture, Cadence, Phonetic Pauses, Ability to stay grounded, Nonverbal communication in general, and overall pitch. The results indicated that participants from the treatment group self-assessed themselves with a higher rating than participants from the control group. However, when comparing the difference between both groups using a t-test only the self-assessment ratings for the ability to stay grounded showed significant results between both groups. Using a scale from 1 to 5, scores from the treatment group were $(M = 4, SD = 0.86)$ against $(M = 2.8, SD = 1.00)$ from control group; $t(37) = 4.06, p < .01$.

To evaluate the self-awareness level from participants we compared the difference between their perceived amount of mistakes during the final elevator-pitch and the amount of mistakes captured by the PT. The average amount of these differences is shown in Table 2. Results show that participants from the treatment group were better at making an educated guess about the amount of mistakes made during their last pitch for all the evaluated nonverbal categories with the exception of phonetic pauses.

TABLE 3
pTM Average for Each Training Session

Session	1	2	3	4	5
pTM Voice					
Treatment	.15	.09	.12	.07	.07
Control	.22	.27	.22	.22	.21
pTM Posture					
Treatment	.1	.05	.07	.04	.02
Control	.35	.29	.27	.26	.27
pTM Gesture					
Treatment	.22	.16	.15	.15	.14
Control	.27	.41	.37	.36	.4
pTM Pauses					
Treatment	.04	.01	.01	.003	.01
Control	.07	.14	.35	0.04	.09
pTM P. Pauses					
Treatment	.002	.002	.002	.002	.002
Control	.003	.003	.003	.003	.002
pTM Grounded					
Treatment	.008	.005	.007	.004	.004
Control	.014	.006	.01	.03	.03

We added the difference between measured and perceived mistakes for all of the trained categories to get the total difference. We compare this total difference between the treatment and the control group using a t-test. Results of this comparison show significant difference between both groups. The values for the total difference of mistakes for the treatment group were $(M = 10.45, SD = 8.06)$ and the values for the control group were $(M = 17.05, SD = 10.99); t(35) = 2.17, p < .05$. These results indicate that the feedback of the PT has a positive influence in the user's self-awareness.

Results also show significant differences between both groups in the confidence scores that participants assigned to their elevator-pitch skills. In a scale from 1 to 5 the treatment group scored $(M = 3.3, SD = 0.73)$ and the control group $(M = 2.75, SD = 0.85); t(37) = 2.19, p < .05)$. These results indicate that the feedback of the PT also has a positive impact in the user's confidence.

According to our criterion, the influence that Presentation Actions have in the quality of a presentation, depend on the percentage of time that they are being displayed throughout a presentation. For example, it is worse to speak too soft throughout the whole presentation, than to speak too soft on several occasions for short periods of time that in total last a fraction of the presentation. Therefore to assess the performance of the participants we used the percentage of Time in Mistake (pTM) for each type of Presentation Action. To measure the pTM we used the logged data generated by the PT on each of the sessions. For each session and each type Presentation Action logged, we added its duration and divided by the total time of the session.

TABLE 4
Total pTM Average for Each Training Session

Training Session	Treatment Group total pTM	Control Group total pTM	t-test
1	M = 0.51, SD = 0.48	M = 0.92, SD = 0.73	t(33) = 2.11, p < .05
2	M = 0.32, SD = 0.16	M = 1.11, SD = 0.95	t(20) = 3.71, p < .01
3	M = 0.35, SD = 0.28	M = 1.01, SD = 0.81	t(24) = 3.42, p < .01
4	M = 0.26, SD = 0.21	M = 0.92, SD = 0.75	t(22) = 3.78, p < .01
5	M = 0.25, SD = 0.19	M = 1.00, SD = 0.79	t(21) = 4.14, p < .01

The pTM average values for each of the training sessions on the different Presentation Actions are shown in Table 3. For all types of Presentation Actions and training sessions the pTM average values for the treatment were lower than the pTM average values for the control group. In the case of the Presentation Actions regarding voice volume, body posture, hand gestures, and correct use of pauses, the average pTM values for the treatment group decreased throughout the sessions in contrast with the average pTM values for the control group that remained stable. The average pTM values for phonetic pauses and ability to stay grounded were similar for both groups and remained stable throughout the sessions.

By adding the average values for all the Presentation Actions we can get the total pTM value for each session. The PT displays at maximum one corrective feedback at the time, nonetheless it still keeps track and logs all Presentation Actions, meaning that multiple mistakes can be tracked simultaneously, therefore the total pTM value can be larger than 1. The total pTM average values for every session are listed in Table 4. In order to calculate the significance of these results we used a heteroscedastic t-test.

From the first training session, there were already significant differences between both groups. These differences increased during the sessions. The average total pTM for the treatment group decreased throughout the sessions, while staying considerably stable for the control group. These results indicate that the feedback of the PT has a positive influence on the user's performance, and helps users to continue improving with practice.

We used the pTM of the tracked Presentation Actions to measure the performance for the final elevator-pitch. The average of these values are shown in Table 5. These results show that for all tracked nonverbal aspects with the exception of phonetic pauses, the pTM average values for the

TABLE 5
Total pTM Averages for the Final Elevator Pitch

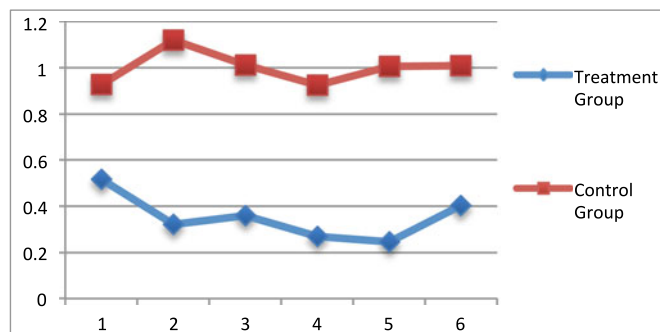|	Treatment group	Control group
Volume pTM	0.182	0.217
Posture pTM	0.032	0.188
Gestures pTM	0.158	0.541
Cadence pTM	0.022	0.036
P. Pauses pTM	0.0028	0.0027
Grounded pTM	0.007	0.234

Fig. 8. pTM average values for training sessions and final elevator pitch.

treatment group were lower than for the control group. In the case of phonetic pauses the measured performance of the control group was slightly better.

A heteroscedastic t-test was used to compare the difference of the total pTM values for both groups. There was a significant difference in the total pTM values for the treatment group ($M = 0.404$, $SD = 0.33$) and the total pTM values for the control group ($M = 1.01$, $SD = 1.05$) with $t(23) = 2.46$, $p < .05$. This shows a significant difference in the performance among participants from the control and treatment group for their final elevator-pitch. This shows that the PT's feedback received during the training sessions had a positive influence on the participants' performance during the final elevator-pitch.

By comparing the total average pTM of this final elevator-pitch against the training sessions (see Fig. 8), results show that the average pTM scores for the control group remained fluctuating in a range between 1.1 and 0.92 throughout the training sessions and the final elevator-pitch. This was in contrast to the case of the treatment group, where the pTM average values decreased throughout the sessions, and increased a bit for the final elevator-pitch. Nonetheless, this average value from the final elevator-pitch remained lower than the average pTM obtained from the first training sessions.

7 DISCUSSION

Background research showed the feasibility of using multimodal interfaces to support learning. Based on the knowledge obtained from this research we develop the PT. To go a step beyond a feasibility and usability study, and contribute to the state-of-the-art on multimodal systems for learning, in this study we explored the effects that the feedback of the PT has on learners practicing basic nonverbal public speaking skills. The effects can be arranged in two categories: learning perception and performance. When looking at the learning perception, participants show that the use of a tool such as the PT for learning compares relatively well in comparison to the educational practices occurring in traditional classroom settings. However, in contrast with our Hypothesis 1 the feedback of the PT was not the catalyzer for this result.

Results also indicate that the feedback of the PT has a positive influence in motivating learners to practice their speeches. This increase of motivation aligns to our Hypothesis 1 and can be explained by stating that the intervention performance feedback is an effective motivator for learners to achieve their goal [24]. The feedback of the PT helps learners to become aware of their performance, therefore motivates them to practice more.

When asking learners about their confidence on their elevator pitch skills, results indicate that the feedback of the PT has a positive influence in this confidence. This raise in confidence aligns with our Hypothesis 1, nonetheless differs with common practices in public speaking courses where in order to avoid hurting the confidence of the speaker, feedback is given using the sandwich technique [25]. In this sandwich technique weak points and mistakes made during the reviewed presentations are sandwiched in between the strengths of the speaker. This is in contrast with the PT that at its current stage only gives feedback about mistakes.

In self-regulated learning the strengths of involving learners as active participants in the assessment process is frequently discussed [26]. However, for the development of public speaking skills self-assessment has shown to be far less effective than the assessment coming from tutors [27], [28]. This can partially be explained by the lack or reflection that learners have about their performance [28]. Following the same line of reasoning the authors in [29] argument the relevance of external feedback in the development of academic skills. To explore whether the PT's feedback has a positive influence in the learners' self-assessment and self-awareness, in this study we compared the amount of mistakes that the participants reported to have make against the amount of mistakes tracked by the PT in the last elevator pitch. Results show a trend where the feedback of the PT helps learners to become better at identifying their own mistakes for all training areas on situations where the feedback of the PT is no longer present. This identification of own mistakes became particularly better for the trained areas that deal with the use of the body such as: body posture, hand gestures and staying grounded. By adding together the difference between perceived and identified mistakes from all the trained areas, results show significant differences between both groups, supporting our Hypothesis 1 stating that the feedback of the PT has a positive influence in the learner's self-awareness.

To test our Hypothesis 2 we analyzed the performance from training sessions showing learning was not a mere perception of the participants. Results reveal that the feedback of the PT has a positive influence in the performance of all the training areas except Phonetic Pauses. We have two reasons to explain the lack of influence regarding the use of Phonetic Pauses. The first reason has to do with the poor identification of them by the PT, on average only 20 percent of them are identified. Second reason deals with their timespan. Phonetic Pauses and their current feedback instruction have a duration that lasts only fractions of a second. We consider it difficult for users to correctly interpret and learn from these short time feedback instructions. To correct these types of short mistakes, research on different feedback strategies is suggested. For the rest of the training areas, which are better identified and have a longer timespan, the results of this study indicate that the feedback of the PT is effective during training sessions. This effectiveness is shown by the measured performance of the participants, where from the first session the treatment group already performed much better than the control group. Moreover results corroborate our Hypothesis 2 showing

that the performances of the treatment group kept improving considerably throughout the sessions, while the performances of the control group remained stable.

In alignment with Hypothesis 2 results indicate that the PT's influence regarding the learners' performance goes beyond the training sessions. The logged performances of the final elevator pitches revealed how participants from the treatment group on average performed better on all trained areas with the exception of Phonetic Pauses, than participants from the control group. This implies that the feedback of the PT received during the training sessions, significantly improved the overall performances from the last elevator pitches. These last performances were better than the performances from the first training sessions, but not as good as the performances from the following ones. We consider that more training sessions using the PT are required in order for learners to perform in their final pitch as good as in the training sessions. This highlights one of the limitations of our study, which is that the long-term usage and resulting learning effects of the PT were not tested. As discussed before, results showed that the feedback of the PT has a positive influence in motivating learner's to continue practicing. However, we assume that the novelty of this feedback played an important role in the learner's reported motivation. Therefore, to keep learners interested some other motivational strategies should be implemented in the PT.

Some other limitations in our study regard the capabilities of the PT to assess the quality of an elevator pitch or presentation. These limitations of assessment start with the fact that the quality of a presentation or a pitch highly depends on its content and not only on its delivery, and the PT is only able to interpret part of its nonverbal delivery. The PT has also limitations on what it can interpret from the sensor data. For example, the current version of the PT cannot distinguish between gestures used for emphasis, iconic gestures, or waving hands without any purpose. Luckily not using enough gestures while giving a presentation is a common mistake in public speaking in comparison with waving the hands without a purpose, which would be also considered as a mistake, but so far we have not identified a single case of someone portraying this behavior. One more limitation that the PT has in assessing the quality of a presentation regards the common consideration of public speaking as a performing art. The creativity and capacity of the speaker to impress the audience play a big role in the quality of a presentation. Therefore, experience speakers might deliberately break "rules" in order to create the desired impact on the audience. This is not a big limitation since breaking a rule to deliberately create an impact requires a certain degree of self-awareness. Results from this study indicate that the PT is a tool that supports learners with the increase of self-awareness and development of basic nonverbal communication skills for public speaking, therefore helping them to reach a competence level where they could make an educated decision on when to deliberately break a specific rule.

8   CONCLUSION AND FUTURE WORK

In this study we presented the Presentation Trainer, a system that with the use of sensors is capable to track, analyze and provide users with feedback that supports their learning in real-time. This learning support is achieved through a feedback mechanism that takes in consideration the learner's cognitive load required to perform the trained task while receiving, interpreting and adapting to the instructional feedback given by the system. In the scenario of using the PT as a supporting tool to practice for an elevator pitch, results of this study show that this type of automated feedback has a positive impact on:

- Increasing the learners' motivation to practice.
- Improving the ability of learners to identify their own mistakes in real time without the use of external feedback.
- Increasing the learners' confidence about their elevator pitch skills.
- Improving the learners' performance during and after the training sessions.

As discussed in the previous chapter, the PT has some limitations in terms of assessing the quality of a presentation. For future research it is important to investigate the relationship between the PT's assessment and human's assessment regarding the quality of a presentation. Moreover in order to improve the PT's assessment for future work we plan to investigate further how to assess the quality of a presentation based on nonverbal communication aspects, e.g., by conducting an expert study. The expected output of such a study is a more comprehensive and specific set of nonverbal communication rules for public speaking. Once implemented, this new and more comprehensive set of rules can improve the assessment capabilities of the PT. Another research gap that has not been addressed in this study, deals with the incorporation of a tool such as the PT in current educational practice and its long-term usage and learning effects. To address this gap we also plan to investigate its effects as a practice tool for learners following a public speaking course.

We consider wearable computing as an emerging trend with a lot of potential to influence learning. The current version of the PT makes use of a wristband to indicate learners about feedback events. One path of future research is to explore the usability and learning effects of this type of technologies in the contexts of the PT.

To conclude, the PT has shown to be a system able to interpret a small part of the user's natural nonverbal communication mechanism, and capable to communicate in real-time the results of this interpretation in such way that it has a positive impact on the learning process of the user. The PT has some limitations, as it cannot comprehend the content of a presentation and the provided feedback is simple and restricted to a limited set of basic rules. Thus, such a tool cannot substitute human tutors. Instead, the power of the tool relies on the ability to present opportunities for correct practice and rehearsal in cases where a human tutor is not available. This makes this multimodal sensor-based tool a valuable and effective addition to current educational practices.

REFERENCES

[1] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2007.

[2] E. H. Mory, "Feedback research revisited," *Handbook of Research on Educational Communications and Technology*, vol. 2. Berlin, Germany: Springer, 2004, pp. 745–783.

[3] S. Van Ginkel, J. Gulikers, H. Biemans, and M. Mulder, "Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education," *Educ. Res. Rev.*, vol. 14, pp. 62–80, 2015.

[4] V. W. Mitchell and C. Bakewell, "Learning without doing enhancing oral presentation skills through peer peview," *Manage. Learning*, vol. 26, no. 3, pp. 353–366, 1995

[5] P. E. King, M. J. Young, and R. R. Behnke, "Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions," *Commun. Educ.*, vol. 49, no. 4, pp. 365–374, 2000.

[6] A. T. Corbett and J. R. Anderson, "LISP intelligent tutoring system research in skill acquisition," in *Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches*, J. Larkin and R. Chabay, Eds. Englewood Cliffs, NJ, USA: Prentice-Hall Inc., 1992, pp. 73–110.

[7] A. Zimmermann, M. Specht, and A. Lorenz, "Personalization and context management," *User Modeling User-Adapted Interaction*, vol. 15, no. 3/4, pp. 275–302, 2005.

[8] H. Ghasemzadeh and R. Jafari, "Physical movement monitoring using body sensor networks: A phonological approach to construct spatial decision trees," *IEEE Trans. Ind. Informat.*, vol. 7, no. 1, pp. 66–77, Feb. 2011.

[9] K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," *Interactive Learning Environments*, vol. 24, pp. 1–16, 2014.

[10] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht, "Augmenting the senses: A review on sensor-based learning support," *Sensors*, vol. 15, no. 2, pp. 4097–4133, 2015.

[11] B. S. Bloom, M. B. Englehart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, "Taxonomy of educational objectives, the classification of educational goals" – *Handbook I: Cognitive Domain*. New York, NY, USA: McKay, 1956.

[12] T. Baur, I. Damian, P. Gebhard, K. Porayska-Pomsta, and E. André, "A job interview simulation: Social cue-based interaction with a virtual character," in *Proc. Int. Conf. Social Comput.*, 2013, pp. 220–227.

[13] M. E. Hoque, M. Courgeon, J. C. Martin, B. Mutlu, and R. W. Picard, "Mach: My automated conversation coach," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2013, pp. 697–706.

[14] I. Damian, C. S. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André, "Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 565–574.

[15] D. Tam, K. E. MacLean, J. McGrenere, and K. J. Kuchenbecker, "The design and field observation of a haptic notification system for timing awareness during oral presentations," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 1689–1698.

[16] L. Batrinca, G. Stratou, A. Shapiro, L. P. Morency, and S. Scherer, "Cicero-towards a multimodal virtual audience platform for public speaking training," in *Intelligent Virtual Agents*. Berlin, Germany: Springer, pp. 116–128, 2013.

[17] J. Schneider, D. Börner, P. Van Rosmalen, and M. Specht, "Stand tall and raise your voice! a study on the presentation trainer," in *Design for Teaching and Learning in a Networked World*. Berlin, Germany: Springer International Publishing, 2015, pp. 311–324.

[18] M. Bjerregaard and E. Compton, *Public Speaking Handbook*. Upper Saddle River, NJ, USA: Pearson, 2011. Retrieved December, 25, 2013.

[19] J. A. DeVito, *The Essential Elements of Public Speaking*. Upper Saddle River, NJ, USA: Pearson, 2014.

[20] Toastmasters International. "Gestures: Your body speaks." 2011. [Online]. Available: http://www.toastmasters.org/

[21] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning Instruction*, vol. 4, no. 4, pp. 295–312, 1994.

[22] N. Ruiz, F. Chen, and S. Oviatt, "Multimodal input," in *Multimodal Signal Processing: Theory and Applications for Human-Computer Interaction*. Amsterdam, Netherlands: Elsevier, 2009, pp. 231–255.

[23] J. Dinardo, "Natural experiments and quasi-natural experiments," *The New Palgrave Dictionary of Economics*. Basingstoke, U.K.: Palgrave Macmillan, pp. 856–859, 2008. doi:10.1057/9780230226203.1162. ISBN 978-0-333-78676-5.

[24] J. W. Jacobs and J. V. Dempsey, "Simulation and gaming: Fidelity, feedback, and motivation," *Interactive Instruction and Feedback*. Englewood Cliffs, NJ, USA: Educ. Technol. Publications, 1993, pp. 197–227.

[25] D. M. Docheff, "The feedback sandwich," *J. Physical Educ. Recreation Dance*, vol. 61, no. 9, pp. 17–18, 1990.

[26] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies Higher Educ.*, vol. 31, no. 2, pp. 199–218, 2006.

[27] L. De Grez, M. Valcke, and I. Roozen, "The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education," *Comput. Educ.*, vol. 53, no. 1, pp. 112–120, 2009.

[28] S. Van Ginkel, J. Gulikers, H. Biemans, and M. Mulder, "The impact of the feedback source on developing oral presentation competence," *Studies Higher Educ.*, pp. 1–15, 2015.

[29] R. Higgins, P. Hartley, and A. Skelton, "The conscientious consumer: Reconsidering the role of assessment feedback in student learning," *Studies Higher Educ.*, vol. 27, no. 1, pp. 53–64, 2002.

Jan Schneider received the master's degree in computer science from Jönköping University, Sweden, in 2007. On December 2013, he started his PhD studying the topic of sensor-based learning support at the Open University of the Netherlands in Heerlen, Netherlands. Previously, he worked as a Human Computer Interaction researcher with Hasselt University, Belgium. There, he worked on different research projects exploring different uses for multi-touch screens and mobile technologies.

Dirk Börner is an assistant professor in the Welten Institute, Open University of the Netherlands. With a background in computer science, he is conducting research on ubiquitous learning support. In particular, he focuses on the utilization of ambient information presentation in combination with mobile technology to support learners in authentic situations.

Peter van Rosmalen is an associate professor in the Welten Institute. He has been active in educational technology since the early eighties on a variety of topics including authoring tools, simulations, e-universities, computer supported cooperative learning, adaptive e-learning, peer support, learning networks, language technologies for learning, creativity, and serious games. His research centers on how to use technology to empower the learner or teacher with instruments to support or motivate their learning or working. His most recent research focuses on the potential of new forms of human computer interaction for learning and professional development.

Marcus Specht is a professor for advanced learning technologies in the Welten Institute, Open University of the Netherlands and director of the Learning Innovation Labs. He is currently involved in several national and international research projects on competence based life-long learning, personalized information support, and contextualized and mobile learning. His research focus is on mobile and contextualized learning technologies, learning network services, and social and immersive media for learning. He is a member of the IEEE.