Presentation Trainer, your Public Speaking Multimodal Coach

Jan Schneider, Dirk Börner, Peter van Rosmalen and Marcus Specht
Welten Institute, Open University of the Netherlands
Valkenburgerweg 177, 6419 AT Heerlen
+(31)45 576 2222
{jan.schneider, dirk.boerner, peter.vanrosmalen, marcus.specht}@ou.nl

ABSTRACT
The Presentation Trainer is a multimodal tool designed to support the practice of public speaking skills, by giving the user real-time feedback about different aspects of her nonverbal communication. It tracks the user's voice and body to interpret her current performance. Based on this performance the Presentation Trainer selects the type of intervention that will be presented as feedback to the user. This feedback mechanism has been designed taking in consideration the results from previous studies that show how difficult it is for learners to perceive and correctly interpret real-time feedback while practicing their speeches. In this paper we present the user experience evaluation of participants who used the Presentation Trainer to practice for an elevator pitch, showing that the feedback provided by the Presentation Trainer has a significant influence on learning.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI), Haptic I/O; K3.0 [Computers and Education]: general.

General Terms
Design, Experimentation, Human Factors.

Keywords
Immediate Feedback; Multimodal Interfaces; Sensors; Public Speaking; Presentation Training.

1. INTRODUCTION
"Practice does not make perfect. Only perfect practice makes perfect." Is a famous quote by Vince Lombardi, one of the most successful coaches in the history of Professional Football [12]. A key factor to achieve this "perfect practice" required for the development and improvement skills is feedback, which has also been identified as one of the most influential interventions in learning [13]. Having a human tutor providing us with high quality feedback whenever we have time to practice our skills is neither an affordable nor a feasible solution.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ICMI '15, November 09 - 13, 2015, Seattle, WA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3912-4/15/11…$15.00
DOI: http://dx.doi.org/10.1145/2818346.2830603

In our effort to study an affordable solution for this feedback availability challenge, we explored the topic 'public speaking skills'. Where we followed a design based research methodology [1] developing different prototypes of the *Presentation Trainer* (PT). The PT is an example of an automated feedback tool that tracks the learners' voice and body. It provides them with feedback about their nonverbal communication, with the purpose to support them with the development of their public speaking skills.

In this article we describe the current version of the PT, and present the user experience evaluation of a study, where participants had to prepare themselves for an elevator pitch. This study followed a quasi-experimental set-up where we explored the learning effects of the feedback provided by the PT.

2. Related Work
Research has proven that feedback provided by a tutor influences the development of public speaking skills [15] and that the magnitude of this influence depends on how this feedback is given to the learner. An important factor that has an effect in the development of these skills is the timing in which feedback is given. For aspects that can be corrected immediately, such as the nonverbal communication of the speaker, immediate feedback has proven to be far more efficient [16].

The nonverbal communication of a speaker is composed of multiple aspects such as: different qualities of the voice, use of gestures, postures, eye contact, facial expressions, etc. All these communication aspects are transmitted simultaneously to the audience; therefore our PT, in order to identify and give feedback about these communication aspects, needs to be a multimodal system.

Multimodal systems first appear in 1980 with the "Put That There" system [6], which opened the exploration of new human computer interactions designed to recognize naturally occurring forms of human language and behavior [19]. Hence, we can find this type of systems in different type of educational settings that range from the creation of collaborative diagrams through multimodal speech, pen and gesture interactions [2, 8]; to helping kids to learn biology through pen and audio interaction [10]. In the particular case of systems presenting some learning support through feedback, a technology that has been used for a vast number of learning applications, is the one of sensors. Sensor systems are able to track the learner's performance through one or multiple modalities, analyze this tracked data, and present the results of this analysis in the form of feedback [21].

Automated feedback has already been used to support nonverbal communication for scenarios such as job interviews [4, 14] and public speaking. The studies of tools designed to support public speaking, can be divided in two groups. The first group supports

the performance of the public speaker while giving a speech. Studies falling in this category researched how to present support by using augmented reality feedback for specific nonverbal communication factors [9], improving the voice quality of the presenter by using a smart synthesizer [17], improving the timing of a presentation by giving haptic feedback [24], etc. The second group studies how to support the training of public speaking skills. The tool studied in [3] has been designed to help learners to overcome their public speaking anxiety by giving presentations in front of a virtual audience; this system is also able to make an assessment of some nonverbal aspects of the presentations. The study in [22] showed a tool able to provide learners with some exercises designed to improve their nonverbal skills for public speaking and a dashboard interface giving immediate feedback about some nonverbal communication aspects while learners practice their presentations. This study also indicated that a dashboard interface is too difficult to follow while practicing for a presentation.

In order to come up with an effective tool to train public speaking skills, we decided to build upon the assumption that immediate feedback is proven to be more effective for training nonverbal communication [16], and that a dashboard interface to provide this feedback has been shown to be far from ideal [22]. Therefore the version of the PT described in this article has the capability analyze the user's performance, and to accordingly select at most one nonverbal communication aspect to be presented as feedback.

3. Presentation Trainer

In terms of likeability for face-to-face communication, the nonverbal communication is far more important than the words itself [18]. The Presentation Trainer (PT) is a tool designed for anyone who wants to practice and foster some basic nonverbal skills for a special type of face-to-face communication, which is public speaking. We developed the PT following a design based research methodology [1], in which we have iteratively designed, developed and tested different prototypes of the PT. In this chapter of the article we describe the current third version of the PT. This version of the PT supports the practice of basic public speaking skills by presenting trainees in public speaking feedback about the following nonverbal communication aspects: body posture, use of gestures, voice volume, use of pauses, use of phonetic pauses, and ability to stay grounded without shifting the weight from one foot to the other making movements that resemble the ones of dancing. This set of aspects does not cover all the nonverbal communication factors that influence a speech; however, these aspects are commonly mentioned in public speaking textbooks and manuals [5,11,25] and we consider them sufficient to help us studying the feedback mechanism of the PT. The following subsections of this article explain how the PT generates Presentation Actions about these aspects, and how these Presentation Actions are mapped into the feedback presented to the user.

3.1 Presentation Actions.

In this subsection of the article we describe the analysis done by the PT about different nonverbal communication aspects, and explain how it generates specific Presentation Actions about them. One of these nonverbal aspects is the body posture of the speaker. The body posture of the speaker is a tool that helps to convey confidence, openness and attentiveness towards the audience, in order to do so, it is recommended to stand up in an upright position facing the audience and with the hands inside of the acceptable box space; in front of the body without covering it, and preferably above the hips [5]. The PT uses the Microsoft Kinect sensor V2[1] to track the learner's body. This body tracking allows the PT to get the coordinates of different learner's joints. By knowing these coordinates the PT is able to infer the learner's body posture. Even when the learner stays still, these coordinates still flicker, however this flickering is usually not big enough to interfere with the posture identification. In order to improve this level of accuracy to a degree where the PT is able to detect all predefined postures without giving false positives, we added a time threshold to distinguish between postures and movements. Through some tuning we set this time threshold to 0.3 seconds. Meaning 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 the time threshold. Whenever the recognized posture violates the preset posture rules, the PT generates a body posture Presentation Action.

Hand gestures in public speaking enhance a speech in different ways such as: strengthening the audience's understanding of verbal messages, painting vivid pictures in the listeners' minds, conveying the speaker's feelings and attitudes, dissipate nervous tension, enhance audience attentiveness and retention, etc. [25]. The current version of the PT does not identify specific gestures. It only recognizes whether gestures are being used and gives feedback to the user whenever she has been speaking without using any gesture after a predefined amount of time. It recognizes if a gesture has been used, by using the input of the Microsoft Kinect V2 to capture the coordinates of the user's joints and keeping track of the angles between forearms and arms, and between arms and shoulder blades. The PT takes notice of the increment and decrement of all the angles. Whenever angles stop decreasing and start increasing, or the opposite way around, stop increasing and start decreasing a "pre-gesture" is identified. If the total amount of increment or decrement from the tracked angles is greater than 5° then a gesture is identified. The use of this strategy allowed us to identify new gestures accurately, without having to worry about the constant flickering of the body coordinates tracked by the Kinect sensor, because the angles between the tracked user's limbs remain stable when the user is not moving. If the user is speaking and no new gestures appear or the angles do not change sufficiently for a predefined time set to six seconds, a Presentation Action about gestures is created. The 5° of angle change and six seconds of speaking time have been obtained by tuning up the PT. We identified than when users move their hands and arms to make a gesture angles are much higher than 5°. Also identified that while not making any gestures the angles never changed more than 5°. We set the threshold to six seconds because while tuning the PT, we identified that a gesture rarely takes longer than six seconds, and that people who stays longer than six seconds without using gestures generally remains a whole presentation without using them.

Having a good voice volume modulation while public speaking is fundamental to transmit a clear message and keeping the audience attention [11]. The PT captures the sound through a microphone at a rate of 16 kHz and stores the absolute volume values in a buffer 0.64 seconds long. To interpret the voice volume of the learner the PT makes use of three thresholds that can be defined manually during runtime in order to adapt them to the acoustic needs of the room where the PT is used. These thresholds are: speaking threshold, soft speaking threshold, loud speaking threshold. The PT interprets silence when the average volume stored in the buffer is below the speaking threshold. It generates a soft volume

[1] https://www.microsoft.com/en-us/kinectforwindows/

Presentation Action whenever the volume of the buffer is in between the speaking threshold and the low volume threshold. Finally the PT generates a loud volume Presentation action whenever the average volume in the buffer is above the loud speaking threshold.

A foremost important skill for public speaking is the use of pauses [5]. When used correctly, pauses allow the audience to take a breathe when information is dense in content or emotion, create spaces for the audience to refocus on the given information, prepare the audience for the following subject, and can add dramatic emphasis during the presentation [25]. Whenever the average value of the sound buffer is below the speaking threshold longer than the predefined time of 0.25 seconds a pause is detected. In the case where no pauses are detected after the predefined time of 15 seconds, a Presentation Action about pauses is raised. We came up with these times of 0.25 seconds and 15 seconds after analyzing the average speaking time and pausing time in 15 different Ted Talks[2].

The filler sounds or phonetic pauses as we call them are all the "ehm", "hmm", "aah", etc. sounds that express hesitation. Showing hesitation is not a good practice while public speaking and therefore during the Toastmasters gatherings[3] 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 these filler sounds. The accuracy of recognizing these phonetic pauses is about 20%, which is quite low, however we found it satisfactory enough to remind learners about this type of mistakes. Whenever one of these phonetic pauses is recognized a Presentation Action about phonetic pauses is raised.

While conducting our research on the PT, examining several presentations and interviewing teachers in public speaking, we identified that a common mistake that students in this field usually do is to switch weight from one leg to the other, 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 user's hips. The PT has a counter that takes note of how many times these coordinates have increased and decreased in more degree than the predefined threshold. After four seconds this counter is reset. We came up with this threshold while tuning the PT. In the case where the counter reaches three or more swings in four seconds a Presentation Action about staying grounded is raised.

The PT stores all the raised Presentation Actions in a list and deletes them from it once they are not longer detected. This strategy of raising Presentation Actions whenever something is detected makes the PT scalable, making it possible to add new type of "nonverbal mistakes" or "good practices" for updated versions of the tool.

3.2 Interface and Feedback

The act of practicing for a presentation in itself already requires a lot of cognitive load from the learner [23]. If on top of this we want to give her feedback, we have to carefully design it, so that this feedback can actually help her to become aware of her nonverbal communication, adapt it, and use this increased awareness to improve her skills [22]. To support the increase of self-awareness the main graphical interface of the PT shows a mirrored image of the user. Taking in consideration the cognitive load of the learner, on top of this mirrored image, the PT presents the learner with at most one real-time feedback instruction at the time. This feedback instruction is transmitted through a visual and a haptic channel, since research has shown that as the cognitive load increases more redundant multimodal communication is needed [20]. Visual feedback is displayed on the graphical interface of the PT. The haptic feedback is communicated through the feedback wristband (Figure 1) that produces some vibration whenever a feedback event is triggered.

Figure 1. Wristband used to give haptic feedback.

In order to present the learner with a maximum amount of one feedback at the time, the PT executes the following procedure:

First it checks whether the time to give feedback is appropriate. In a small user study that we conducted we realized 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. Hence, the PT waits at least six seconds after the last feedback stopped being shown, in order to present the user with a new feedback.

When the time to give feedback is appropriate, the PT looks at the list of Presentation Actions and selects which one of them should be used to create a feedback event. So far we have tested two strategies to select the feedback to be displayed by the PT. The first one assigns scores to the different Presentation Actions according to their relevance. This relevance depends on the number of occasions that each Presentation Action has been on the list, amount of time inside of the list and user profiling [7]. The second strategy, which is the one that we used for this study, triggers the oldest Presentation Action stored in the list as a feedback event. Once the feedback event is triggered it keeps on being displayed until the mistake is corrected. When corrected a correction mark is shown to the user.

The PT selects the Presentation Action to be triggered as a feedback event and then decides whether it is presented as corrective or interruptive feedback. Corrective feedback indicates the user in real-time that a Presentation Action has been identified. It is presented to the user by displaying an icon and a short (maximum two words long) written statement indicating how to correct the identified mistake (Figure 2) together with a small vibration produced by the feedback wristband.

[2] https://www.ted.com/talks

[3] https://www.toastmasters.org/

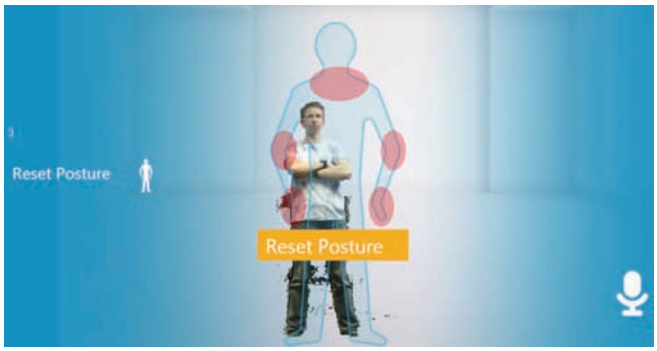

Figure 2. Immediate corrective feedback for crossing arms

Interruptive feedback is triggered whenever a mistake is considered to be severe. Examples of severe mistakes are: mistakes repeated several times, mistakes that stay for too long time without being corrected (currently a mistake repeated five times, or longer than 20 seconds without correction), and any type of predefined severe mistake. Interruptive feedback, produces some vibration, a pause sound, stops the program and displays the reason of the interruption (Figure 3). The interface of the interruption offers the user the possibility to continue practicing her presentation receiving feedback in all nonverbal aspects, or only on the aspect that she was interrupted for, so that she can focus on improving this specific skill.

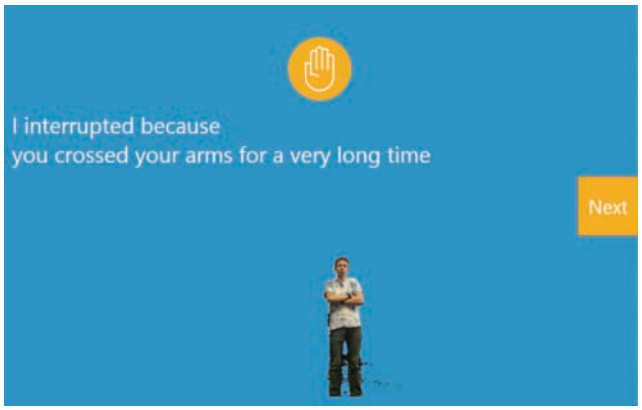

Figure 3. Interruptive feedback for crossing arms for a long time.

3.3 Presentation Trainer Architecture

The PT trainer has been developed in C# using the .NET framework 4.5[4]. In order to help people in their nonverbal skills, the PT needs to be able to track the user's voice and body; the current set-up of the PT does this tracking through the use of the Microsoft Kinect for Windows V2 sensor in conjunction with its SDK. Nevertheless its architecture shown in Figure 4 allows the inclusion of more sensor channels.

The PT receives the body and audio sensor data and stores it in the Body and Audio pre-Analysis objects respectively. The Body pre-Analysis object contains the current coordinates of the detected joints of the user's body and flags indicating whether certain postures rules, such as crossing arms are fired. The Audio pre-Analysis object contains the values of the audio buffer received from the Kinect microphones together with a flag indicating whether the user is currently speaking.

[4] http://www.microsoft.com/net

The JudgementMaker object uses the information stored in the Body and Audio pre-Analysis objects to create Presentation Actions. These Presentation-Actions are any type of nonverbal communication mistakes or good practices identified by the system. For example speaking for too long time without pausing, standing in an incorrect posture, etc.

A list of the current Presentation-Actions is passed to the RulesAnalyzer object, which according to the state of the program and performance of the user decides the feedback event to be triggered. The running operational state receives these feedback events and forwards them to the connected output channels.

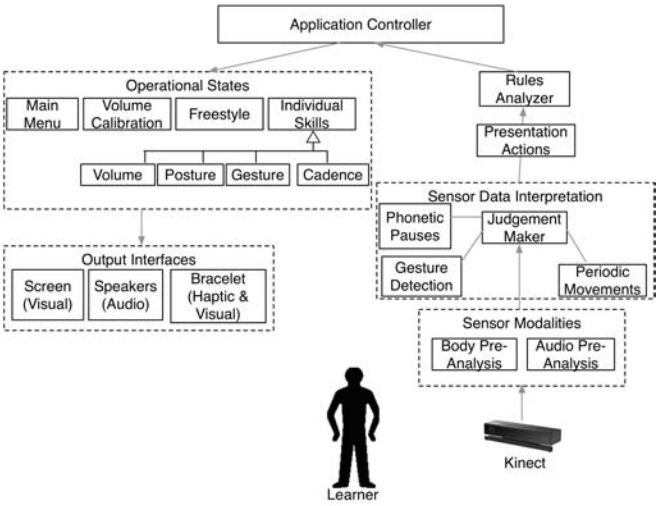

Figure 4. System Architecture.

4. Evaluation

In this study we investigated how the feedback provided by the PT has an effect on the user experience of learners practicing to give an elevator pitch. An elevator pitch is a 30 to 120 seconds long speech where one summarizes in lay terms what one does and why it is important[5]. In order to study this user experience we conducted a quasi-experiment with a treatment and a control group of participants.

4.1 Participants

For this experiment we had a total number of 40 participants. The control and the treatment group both had 11 males and 9 females. The average age of the participants was 42.6 with a range of 24-62. All participants were professionals working at our university, with a similar cultural background. We recruited them by personally asking for their willingness to take part in our experiment. Once they agreed to take part, they scheduled their experimental session themselves by selecting an available timeslot. The criteria used to accommodate them in the control or treatment 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 arriving from even sessions (2nd, 4th, etc.) were assigned to the control group.

4.2 Apparatus and Materials

To measure the user experience of participants we used a questionnaire designed to evaluate the user experience with

[5] https://en.wikipedia.org/wiki/Elevator_pitch accessed on August 2015

multimodal learning applications[6]. The questionnaire consists of ten questions. Each of the first six questions inquires a different dimension of the user experience with multimodal learning applications. This six dimensions are: naturalness of use, boredom, invasiveness, unlikelihood of use during free time, perceived learning, and comparison against traditional classroom learning. Answers of these six questions are provided through a rating from 1 to 10 in a Likert-type scale. In addition there is one multiple-choice question about the novelty of the system, and three open questions inquiring the participants' opinion about the system, recommendations, and insights while using the system. Besides the questionnaire the PT also created a log file for each training session. This log file includes: the starting and ending time of the training session, all identified Presentation Actions (mistakes) together with their corresponding starting and ending timestamps, and all Feedback events together also with their corresponding timestamps.

4.3 Experimental Setup

Each experimental session in this study was individual. The session started with a short, five-minute, lecture about nonverbal communication for public speaking. This lecture reviewed the aspects that the PT is able to track, which are: body posture, use of hand gestures, use of voice volume, pauses, phonetic pauses and ability to stay grounded. The reason for this lecture was to assure participants have a similar baseline of basic knowledge about nonverbal communication for public speaking.

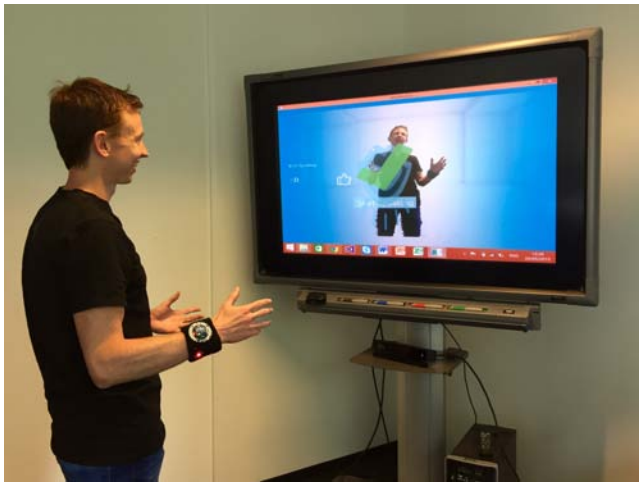

Figure 5. Training session setup with Presentation Trainer giving feedback.

Immediately after this lecture, participants had another short five-minute lecture about elevator pitches where they learned how to do an elevator pitch. Finishing the lecture the tutor showed the participants a live example of an elevator pitch. As soon as the lectures finished the participant had 5 minutes to create her own elevator pitch. Participants were free to use any topic they want for the pitch. Once the pitch was created, it was time to practice its delivery.

Participants of both groups had to practice the delivery of the pitch in five successive training sessions. Participants in the control group practiced using a version of the PT whose interface only shows a mirrored image of the user, in other words they did not receive any specific feedback. The treatment group practiced the pitch using the full version of the PT (Figure 5). They received, if necessary, both immediate and interruptive feedback.

After practicing the pitch for the fifth time, the participants were asked to answer a questionnaire about their user experience.

5. Results

The answers to the six dimensions regarding the user experience posted in the questionnaire are shown in Table 1. To calculate the significance of the results we used a t-test. A dimension that showed significant differences between the control and treatment group is the one of motivation. Participants from the control group felt significantly less motivated while using the PT scoring an average of 6.52 against an average of 3.05 scored by the treatment group. In terms of perceived learning the treatment group reported to learn significantly more with an average score of 7.47 against a 6.0. With no significant difference between groups, results indicate that participants found the interaction with the tool to be natural, not invasive, better for learning than the traditional classroom setting, and found it unlikely to use it during free time.

Table 1. Average results on the 6 dimensions of user experience extracted from the post session questionnaire (ratings from 1 to 10).

Dimension	Treatment Group	Control Group	P-Value
Naturalness *10=very natural	6.47	6.42	p = 0.92
Invasiveness *10=very invasive	2.68	3.84	p = 0.08
Boredom vs. Motivation *10=very bored	3.05	6.52	p <.01
Unlikelihood of free time use *10=very unlikely	6.05	6.05	p = 1
Learning Perception *10=learned a lot	7.47	6	p < .05
Practice using tool vs. Classroom *10=much better than classroom	6.94	6.1	p = 0.2

None of the participants from the treatment group had ever used an application similar to the PT, on the other hand one participant of the control group stated to have used a similar application, and two of them stated that maybe they have seen something similar.

When asking participants about the specific aspects learned while using the PT, most of the participants from the treatment group mentioned specific nonverbal communication aspects. The learning aspect most frequently mentioned by participants was the use of pauses; followed by the use of gestures, voice volume, posture, and finally phonetic pauses. On the other hand participants from the control group stated that they learned about self-awareness, staying calm and the importance of practice.

[6]http://www.sigmla.org/mla2015/ApplicationGuidelines.pdf accessed on August 2015

Twelve participants from the control group suggested improving the system by providing users with some real-time feedback about their performance. The suggestion for improvement from the treatment group include the use of more explicit icons for feedback, selection to practice one skill at the time, and giving a summary of the performance at the end of each training session.

The logged data generated by the PT show significant differences in the assessed performances of participants between both groups. To assess the performance of each training session, we looked at the proportion of time while the user is making a mistake (pTM), to get this value, first we added up the duration of all Presentation Actions (mistakes) because according to our criterion, one mistake that lasts for instance for an entire minute is much worse than five mistakes that in total have a duration of ten seconds.

Then we divided the total duration of the mistakes by the total time of the session. We performed this division because according to our criterion for instance 30 seconds of mistakes in a 45 seconds pitch is far worse than two minutes of mistakes in a 25 minutes long presentation. Even when the PT at maximum displays one corrective feedback at the time, it still keeps tracks and logs all Presentation Actions, meaning that the pTM value can be larger than 1, since multiple mistakes can happen simultaneously. The pTM average values for every session are listed in Table 2. In order to calculate the significance of these results we used a t-test.

Table 2. pTM average for each training session

Training Session	Treatment Group	Control Group	P-Value
1	0.51	0.92	$p < 0.05$
2	0.32	1.11	$p < 0.01$
3	0.35	1.01	$p < 0.01$
4	0.26	0.92	$p < 0.01$
5	0.25	1.002	$p < 0.01$

From the first training session, there were already significant differences between the groups with average values of 0.92 for the control and 0.51 for the treatment group. These differences (see Figure 6) increased during the sessions. The average pTM for the treatment group continued to decrease to an average of 0.25 for the fifth session, while staying considerably constant with an average of 1.002 on the fifth session for the control group.

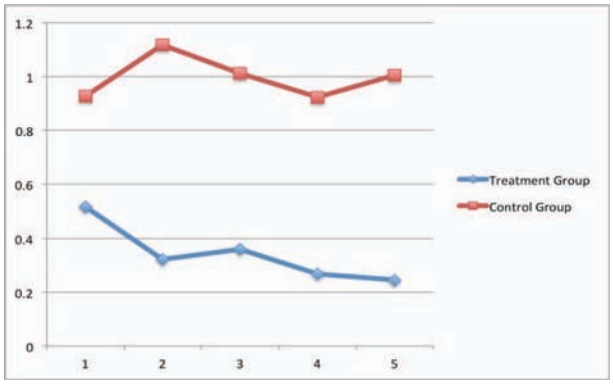

Figure 6. Reduction of ptm after 5 training sessions.

6. Discussion

Results and observations realized within this study showed that in general participants enjoyed using the PT to practice their communication skills, especially the version of the PT including feedback. Some participants stated during the firsts sessions how difficult it is to give a speech while paying attention to the feedback, however the general impression at the end of the training sessions was that it became stepwise much easier to use and that it is a very helpful tool for training nonverbal communication skills.

The answers to the questionnaire indicate that practicing with the PT is perceived by users as a better way to learn than the traditional classroom methods, particularly in the case where this practice includes feedback. This feedback apparently makes the user experience more natural and less invasive. It seems that participants that did not receive feedback and only saw their reflection, felt observed and became suspicious about what the system could be doing with all the tracked information. Participants receiving feedback experienced what the system was doing with the information and felt the interaction more natural and less invasive.

Results show how feedback can significantly motivate learners, however this feedback is not enough to motivate them to use such a tool during their free time. We have to take in consideration that none of the participants practiced public speaking as a hobby and that this version of the PT has been designed to be a learning supporting tool and not a game.

Participants that received feedback felt as if they have learned more during their practice sessions and listed concrete nonverbal communication aspects, when asked about the subjects learned while using the system. In contrast with participants from the control group, who reported to have learned more abstract subjects such as self-awareness, stay calm, etc. This difference in learning is not only shown in the reports of the participants, it has also been objectively corroborated by the performance measurements taken during the practice sessions, where from the first session the treatment group already performed much better than the control group. Moreover the performances of the treatment group kept improving considerably throughout the sessions, while the performances of control group remained stable.

A limitation of the PT is that its performance measurement mechanism cannot be directly translated to the assessment that a human would make about the quality of a presentation or an elevator pitch. 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 its delivery, and the PT is only able to interpret part of its nonverbal delivery. Additionally there are still limitations regarding what the sensors can perceive and what the PT can interpret out of the sensor data. In previous user tests conducted to tune the PT we learned that the identification of false positives should be avoided as much as possible. False positives lead users to believe that the system does not work properly and therefore users stop taking the feedback seriously. Thus it is better in some cases to reduce the accuracy of the system, as we did it with the phonetic pauses where the identification ration is only about 20%. Another current limitation of the PT in terms of interpreting data has to do with the interpretation of gestures; the PT cannot recognize the difference between iconic gestures, emphasis gestures, or waving hands without purpose. Fortunately waving hands without purpose is something that we have not seen

during the experiments, and are hardly seen throughout presentations.

Public speaking in some cases can be seen as a performing art, where the creativity and capacity of the speaker to impress the audience play a big role in the quality of a presentation. It is perfectly acceptable for a speaker to deliberately break any rule in order to create the desired impact in the audience. The PT is not capable to interpret between deliberately broken rules or mistakes by the presenter. However, we want to stress that the PT is designed to support the practice of basic nonverbal skills, by making learners aware of the identified Presentation Actions; and this study has proven its success to do so.

7. Future work and conclusions

The presented study identified that the feedback mechanism of the PT facilitates the improvement of basic nonverbal communication skills for public speaking. As a next step we plan to do an expert study to extract a rich set of nonverbal communication aspects and rules that have an effect on the quality of a presentation, and improve the PT based on those findings. In this expert study we also want to investigate how a tool such as the PT can become a complementary tool able to support the improvement of courses in public speaking.

In this context we also want to explore how the measured assessment of the PT compares to expert, peer and self-assessment, and study how these combination of assessments can support the development of public speaking skills.

To conclude, in this article we showed how the feedback mechanism implemented in the PT has supported users with the development of their public speaking skills by helping them to significantly improve their performance during training sessions. If we consider the PT as an example of a multimodal system; and consider public speaking as an example of an activity that requires a great deal of practice with effective feedback to become proficient at it, then this study points out a way in which multimodal systems are able to contribute in getting us one step closer to a perfect practice that makes perfect.

8. ACKNOWLEDGMENTS

The underlying research project is partly funded by the METALOGUE project. METALOGUE is a Seventh Framework Programme collaborative project funded by the European Commission, grant agreement number: 611073 (http://www.metalogue.eu).

9. REFERENCES

[1] Anderson, T., Shattuck, J.: Design-Based Research A Decade of Progress in Education Research? *Educational Researcher*, 41(1) (2012), pp. 16-25

[2] Barthelmess, P. Kaiser, E., Huang, X, Demirdjian, D. Distributed pointing for multimodal collaboration over sketched diagrams. *Proceedings of the Seventh International Conference on Multimodal Interfaces*, ACM: (New York 2005), pp. 10-17.

[3] Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., Scherer, S. Cicero - towards a multimodal virtual audience platform for public speaking training. *In Proc. IVA*, vol. 8108. Springer, 2013, 116–128.

[4] Baur, T., Damian, I., Gebhard, P., Porayska-Pomsta, K., Andre, E. A job interview simulation: Social ´ cue-based interaction with a virtual character. *In Proc. SocialCom. IEEE*, (2013), pp. 220–227.

[5] Bjerregaard, M., Compton, E. Public Speaking Handbook. Snow College, Supplement for Public Speaking (2011)

[6] Bolt, R. A. Put-that-there: Voice and gesture at the graphics interface. *Computer Graphics*, 14 (3) (1980), pp. 262-270.

[7] Brouwers, J. Implementation and Testing of an Improved Automated Presentation Trainer. Master Thesis, Maastricht University, Faculty of Humanities and Sciences, Department of Knowledge Engineering, Aug-Sept. 2015.

[8] Cohen, P.R., McGee, D. (2004) Tangible multimodal interfaces for safety critical applications. *Communications of the ACM*, vol. 47 (January 2004) 41-46.

[9] Damian I., Tan C.S.S., Baur T., Schöning J., Luyten K., André E. Augmenting Social Interactions: Realtime Behavioural Feedback using Social Signal Processing Techniques. *In CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* (Seul 2015), pp. 565-574.

[10] Darves, C., Oviatt, S. Talking to digital fish: Designing effective conversational interfaces for educational software. *In Evaluating Embodied Conversational Agents*, Z. Ruttkay, C. Pelachaud, Eds., (Dordrecht 2004), pp. 271-292.

[11] DeVito J.A. *The Essential Elements of Public Speaking*. Pearson (2014)

[12] ESPN Greatest coaches in NFL History. http://espn.go.com/nfl/story/_/page/greatestcoach1/greatest-coaches-nfl-history-vince-lombardi . Accessed July 2015.

[13] Hattie, J., & Timperley, H., The power of feedback. *Review of Educational Research* (2007), pp. 81–112

[14] Hoque, M. E., Courgeon, M., Martin, J., Mutlu, B., and Picard, R. W. Mach: My automated conversation coach. *In Proc. UbiComp* (2013).

[15] Kerby, D. & Romine, J.: Develop Oral Presentation Skills Through Accounting Curriculum Design and Course-Embedded Assessment. *Journal of Education for Business*. Vol. 85, Iss. 3. (2009).

[16] King, P., Young, M., & Behnke, R. Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions. *Communication Education*, 49(4) (2000), pp. 365-374.

[17] Li, X., and Rekimoto, J. Smartvoice: a presentation support system for overcoming the language barrier. *In Proc. CHI'14, ACM* (2014), pp. 1563–1570.

[18] Mehrabian, A. Nonverbal communication. Oxford, UK: Aldine-Atherton, 1972.

[19] Oviatt, S. Multimodal Interfaces. The human-computer interaction handbook. 3rd edition Lawrence Erlbaum: New Jersey, 2012.

[20] Ruiz, N., Chen, F. and Oviatt S., "Multimodal Input", *in Multimodal Signal Processing: Theory and Applications for Human-Computer Interaction.* Edited by Thiran, J.P., Marques, F. and Bourlard, H., Academic Press, 2010, Chapter 12, pp. 231-255.

[21] Schneider, J., Börner, D., van Rosmalen, P., Specht, M. Augmenting the senses: a review on sensor-based learning support. *Sensors* **15**(2) (February 2015), pp. 4097–4133.

[22] Schneider, J., Börner, D., van Rosmalen, P., Specht, M. Stand Tall and Raise your Voice! A Study on the Presentation Trainer. *In EC-TEL'15 Proceedings of the tenth European Conference on Technology enhanced learning* (Toledo 2015).

[23] Sweller, J. Cognitive Load Theory, learning difficulty, and instructional design. *Learning and Instruction* 4 (4) (1994), pp. 295–312.

[24] Tam, D., MacLean, K., McGrenere, J., and K. J. Kuchenbecker, K. J. The design and field observation of a haptic notification system for timing awareness during oral presentations. *In Proc. CHI '13*. ACM, 2013.

[25] Toastmasters International. Gestures: your body speaks. http://www.toastmasters.org/ (2011)