Feedback Design in Multimodal Dialogue Systems

Peter van Rosmalen[1], Dirk Börner[1], Jan Schneider[1], Olga Petukhova[2] and Joy van Helvert[3]

[1]*Welten Institute, Open University of the Netherlands, P. O. Box 2960, Heerlen, The Netherlands*
[2]*Universitaet des Saarlandes, Saarbrucken, Germany*
[3]*University of Essex, Colchester, U.K.*

Keywords: Feedback, Sensors, Instructional Design, Multimodal Dialogue, Reflection Support.

Abstract: This paper discusses the design and development of the instructional aspects of a multimodal dialogue system to train youth parliament members' presentation and debating skills. Real-time, in-action feedback informs learners on the fly how they perform key skills and enables them to adapt instantly. About-action feedback informs learners after finishing a task how they perform key skills and enables them to monitor their progress and adapt accordingly in subsequent tasks. In- and about-action feedback together support the enhancement of the learners' metacognitive skills, such as self-monitoring, self-regulation and self-reflection thus reflect in- and about action. We discusses the theoretical considerations of the feedback, the type of data available and different ways to analyse and combine them, the timing of feedback and, finally, provide an instructional design blueprint giving a global outline of a set of tasks with stepwise increasing complexity and the feedback proposed. We conclude with the results of the first experiment with the system focussing on non-verbal communication skills.

1 INTRODUCTION

The variety of interfaces used for interaction in environments is rapidly growing. Interfaces increasingly use one or more modes of interaction resembling natural communication by using input and output modalities such as speech, text, gesture, facial expressions, movement detection or pointing devices. While there is experience in education with systems e.g. using written language for interaction (Nye, Graesser and Hu, 2014) and motion sensors (Triantafyllou, Timcenko, and Triantafyllidis, 2014), there is limited experience in using other or more modalities at the same time to support interaction for learning. The increasing computable power and miniaturization, however, opens up numerous new application scenarios in education; for example, using sensors to provide input about learners, between learners or between learner(s) and the environment they explore.

In this paper we will discuss the design of the METALOGUE multimodal dialogue system to train debating skills. Whereas the argumentative elements of debating have received ample attention as a means to enhance learning (e.g. D'Souza, 2013), learning all aspects of debating has received less attention. Giving an interactive presentation, i.e. a presentation including an argumentation, is a complex task. A trainee needs not only to master the content (i.e. what to present, how to structure their presentation and which strategy to use in the closing argumentation) but also other modalities (Trimboli and Walker, 1987), such as voice (i.e. how to control and use their voice e.g. pitch, speed or volume) and body language (i.e. how to control and use their body e.g. arms, hands or align their body). Additionally, the trainee has also to be continuously aware of the effects of their arguments, voice and use of their body language towards their audience or opponents and therefore monitor, reflect and adapt when necessary (metacognitive aspects). There are numerous materials such as seminars, courses, books and magazines that can help us to develop our debating skills, however, it is difficult to obtain sufficient practice.

This paper focuses on the instructional aspects of an eventually fully automated multimodal dialogue system which will provide individualised debate training; in particular, it considers the task and feedback design. The modalities included are speech, gestures and movement. Personal traits and social aspects (e.g. stage fright) involved have been neglected. The envisioned system focuses on the support of the initial (private) training phase, while

providing only minimal support during the actual public performance. Furthermore this initial training aims to convey basic, generally accepted debating rules that can be processed by the system, rather than supporting the development of distinct personal communication skills.

In the next section we will start by introducing the design aspects, and then continue by explaining the instructional design blueprint. Finally, we conclude by describing the results of the first system experiment focussing on non-verbal communication skills.

2 DESIGN ASPECTS

The use of a multimodal dialogue system for educational purposes has to address a number of different perspectives. In this section we start with a discussion of the theoretical aspects of the instructional design. Next, we will discuss the feedback options available taking into account both the educational aspects such as usefulness and timing, and the technical aspects i.e. the data available from the sensors and the different ways to analyse and combine them. We conclude with the instructional design blueprint derived.

2.1 Instructional Design

Giving an interactive presentation is a complex task. The design, therefore, has to pay specific attention not to overload the learner (Sweller, 1994), while at the same time the tasks will have to be sufficiently challenging and, at the end, meet the full complexity required. To assure that tasks are sufficiently inspiring Kiili et al (2012) suggest to take into account in particular sense of control, clear goals, the challenge-skill relation, and, finally, feedback. Feedback is one of the most powerful interventions in learning (Hattie and Timperley, 2007). According to some authors (Nicol and Macfarlane-Dick, 2006), the most beneficial thing tutors can do to students is to provide them feedback that allows them to improve their learning. Common practice in education and training is to give feedback after a task has been performed. However, depending on the task, the type and content of the feedback and the availability of a (virtual) tutor, feedback may also be given while performing a task. Schön (1983) coined the notions of reflection-in-action (reflection on behaviour as it happens, so as to optimize the immediately following action) and reflection-about-action (reflection after the event, to review, analyse,

and evaluate the situation, so as to gain insight for improved practice in future).

In current educational design practice there is a growing interest in using whole-tasks models. Whole-tasks models aim to assist students in integrating knowledge, skills and attitudes into coherent wholes, to facilitate transfer of learning. As part of this they take into account how to balance the load of the learner, make the tasks sufficiently challenging and how to give feedback. The 4C-ID model is a whole-tasks instructional design model that has been widely researched and applied in course and curriculum design (Van Merriënboer and Kirschner, 2013). Recently also for the design of serious games (Van Rosmalen et al., 2014), since the key elements of the 4C-ID instructional design model (i.e. authentic tasks, task classes which take into account levels and variation, the distinction between supportive and procedural information and the extra practice of selected part-tasks) fit well with game (design) practice. For the same reasons, it fits well with the instructional design of METALOGUE where the users have to stepwise understand and learn how to present and argue working with realistic, engaging tasks adjusted to their personal needs in terms of complexity levels, and if necessary, have the option to practice selected types of subtasks.

2.2 Data and Feedback

Three types of sensor specific data (Figure 1) will serve as input for the system: (1) speech signals from multiple sources (wearable microphones and headsets for each dialogue participant and all-around microphone placed between participants); (2) visible movements tracking signals from Kinect and Myo sensors capturing body movements and facial expressions; and (3) video signal captured by the camera that records the whole dialogue training session (also includes sound).
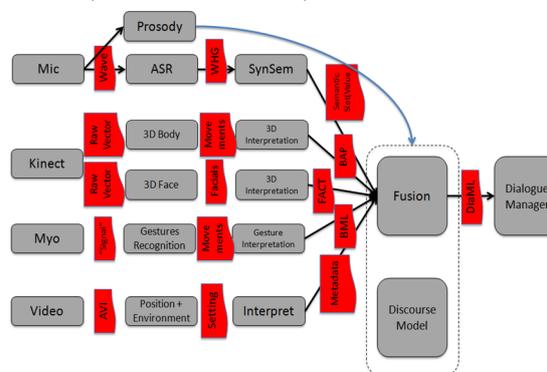

Figure 1: METALOGUE workflow and formats of data.

Speech signals will serve as input for 2 types of further processing. Automatic Speech Recognition should answer the question: 'What was said?'. Prosodic analysis should answer the question: 'How was it said?'. The latter is mostly concerned with generating feedback relating to voice quality aspects such as speech rate, volume, emphasis and pausing. Moreover, prosodic analysis is important to identify participant's emotional state, for instance nervousness level, and degree of uncertainty (e.g. hesitation phases using speaking rate (speech speed) and pausing).

The visible movements will serve as input to the analysis of body language. It includes aspects such as gaze (re-) direction, head movement and orientation, facial expressions, hand and arm gestures, posture shifts and body orientation.

2.2.1 Semantics of the Data

Gaze shows the focus of attention of the dialogue participant. Gaze is also an important signal of liking and disliking, and of power and status (Argyle, 1994). Gaze is also used to ensure contact between participants. Instructions for good debating and presentational skills include recommendation on keeping eye-contact with your opponent.

Head movements and head orientation are the basic forms of signalling understanding, agreement and approval, or failure (Duncan, 1970). Head movements are also used to indicate aspects of information structure or to express a cognitive state, e.g. uncertainty or hesitation. Heylen (2006) noticed that head movements may have a clear semantic value, and may mark interpersonal goals and attitudes.

Hand and arm gestures have been studied extensively, especially for their relation to the semantic content of an utterance (e.g. Kendon, 2004). The beginnings of gesticulations have been observed to mark turn-initial acts (Petukhova, 2005). So-called beat gestures are often used by the speaker to signal most important parts of their verbal message, e.g. to emphasise/accent new important information. Guidelines for good debating and negation style include several recommendations based on long-standing traditions and observations such as "Keep hands out of your pockets" or "Do not cross/fold your arms".

Posture shifts are movements or position shifts of the trunk of a participant, such as leaning forward, reclining, or turning away from the current speaker. Posture shifts occur in combination with changes in topic or mode of participation (e.g. Scheflen,1964).

In debating posture and overall body orientation plays an important role. Debating guidelines talk about confidence posture such as "Keep legs aligned with your shoulders" or "Turn body towards the opponent".

Facial expressions are important for expressing emotional reactions, such as happiness, surprise, fear, sadness, anger and disgust or contempt (Argyle, 1994). Emotions will be analysed in combination with verbal and prosodic components. Moreover, face can also display a state of cognitive processing, e.g. disbelief or lack of understanding.

In debates, performance is often judged on three main criteria, i.e. argument content, organization and delivery (http://www.wikihow.com/Debate). Delivery is about how the debater speaks. Good debaters should give a strong impression that they truly believe in what they say. To express authority the debater needs not only to use his voice and body but also support his arguments with statistics, facts and figures, including personal experience or experience from the real life experience of others. Likability is about showing respect and friendliness. In summarised, there are 5 global aspects to be considered: Audibility, Engagement, Conviction, Authority and Likability (AECAL).

Nevertheless, debate is about argumentation, the planning and preparation involving arguments as a general conclusion, supported by reason(s) and evidence. Good debaters use discourse markers and dialogue announcement acts such as "I will talk in favour of ... Because ... Since international research shows...". The debaters' way of structuring arguments are analysed using a recently proposed argumentation scheme (Peldszus and Stede, 2013). The scheme is based on detecting proponents' and opponents' moves in a basic debating situation. In addition to argument structure annotation, links between premises and conclusions, as well as rebutting and undercutting links, are annotated with discourse relations as defined in Rhetorical Structure Theory (Mann and Thompson, 1988) extended with relations from Discourse Penn TreeBank corpus (Prasad et al., 2008).

Finally, a pragmatic analysis takes care of the overall perspective. This type of analysis is based on identifying speaker's intentions in terms of dialogue acts as specified in ISO 24617-2 (www.iso.org). This taxonomy distinguishes the following core dimensions, addressing information about:
- the domain or task (Task);
- feedback on communicative behaviour of the speaker (Auto-feedback) or other interlocutors (Allo-feedback);

- managing difficulties in the speaker's contributions (Own-Communication Management) or those of other interlocutors (Partner-Communication Management);
- the speaker's need for time to continue the dialogue (Time Management);
- about who should have the next turn (Turn Management);
- the way the speaker is planning to structure the dialogue, introducing, changing or closing the topic (Dialogue Structuring);
- the information motivated by social conventions (Social Obligations Management);
- and one optional dimension, addressing establishing and maintaining contact (Contact Management).

2.2.2 Feedback

Drawing on Schön's (1983) distinction between reflection in-action and reflection about-action, we distinguish here between in- and about-action in terms of learner feedback in the context of METALOGUE. In-action or immediate feedback is potentially powerful but in order to be effective, it should be (Hattie and Timperley, 2007; Engeser and Rheinberg, 2008; Coninx, Kreijns and Jochems, 2013):

- specific and goal oriented, i.e. focus on key aspects of the learner's interaction so they become aware of strong or weak points, comprehend their meaning, and adjust their behaviour accordingly;
- clear, unambiguous and not requiring complex reasoning about its cause and how to respond;
- concise, i.e. short so they are minimally disruptive;
- predictable, i.e. the type of feedback should be known/agreed upon in advance.

Taking these guidelines into consideration, the in-action feedback will concentrate on aspects of argument delivery, i.e. aspects of voice quality and visible movements (non-verbal behaviour), which are relatively straightforward to understand and to respond to. Aspects related to argument content and argument organisation will only be implicitly addressed through the discourse constructed in the METALOGUE system. Consequently, in-action feedback will mainly concentrate on promoting awareness. The feedback should enable the learner to become aware of their strong and weak points and their development. For the learner this would imply that they will come to understand which aspects are

of relevance and, ultimately, be able to recognise these aspects in their performance or the performance of others.

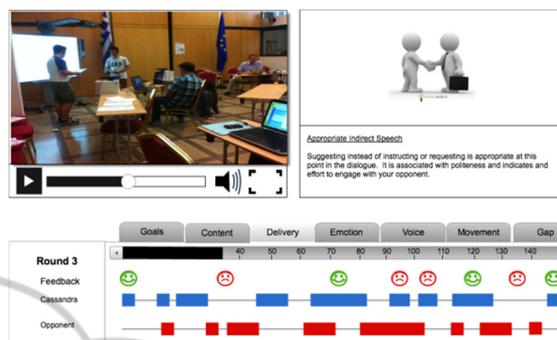Figure 2: Screen mock-up: about-action delivery feedback.

In contrast, the about-action feedback will mainly support reflection. Closely connected with awareness, reflection goes one step beyond. The feedback should enable the learner directly after a debate performance to review, analyse, and evaluate the situation, to gain insight for improved practice in the future. Here, for the learner the ultimate goal is to train their self-monitoring, self-regulation and self-reflection. For the learner this would imply that as they practice through their tasks in a number of rounds that they stepwise seamlessly are able to adjust their performance with respect to their own utterances and behaviour and their opponent's. The about-action feedback will build upon the in-action feedback providing valuable insight based on aggregations of the in-action feedback and feedback based on the semantics of the verbal contents and dialogue act use. Together, about-action feedback (Figure 2) will be structured within the following partly related categories:

- *Goals*. The status of the goal to be achieved, progress and distractions. The goal will have two qualities, one related to the objective of the dialogue and one related to the (meta-)cognitive aspects of dialogue (i.e. the ability of the learner to anticipate on their 'opponent' and adapt accordingly.
- *Content and organisation*. An integrative perspective on the use of argument, reason and evidence. It will build on an analysis of the verbal part of the discourse.
- *Delivery*. Delivery will focus on individual and integrative (AECAL) aspects of how the speaker speaks.
- *Emotion*. Given the importance of the awareness and appreciation of the emotional state of the user and opponent special attention will be given on the emotional state of the participants.

● *Voice*. Aligned with the in-action feedback, voice aspects will be aggregated, analysed and commented upon.

● *Movements*. Aligned with the in-action feedback, movements aspects will be aggregated, analysed and commented upon.

2.3 Instructional Design Blueprint

The basis of the instructional design is the skill to be trained. The skill "debating" (and its associated knowledge and/or attitudes) can be elaborated in the following skills hierarchy (Figure 3).
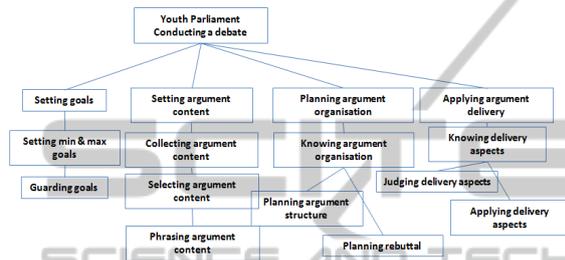

Figure 3: Skills hierarchy: "conducting a debate".

The METALOGUE system will be delivered in 3 rounds: an initial pilot, a second pilot and the final system including a fully automated dialogue. The instructional design aligns with the incremental design of the system. The need of a stepwise increase of complexity of the tasks to be mastered fits with the stepwise increase of the complexity of the system.

Given its complexity, learning to debate has to be carefully designed. For a trainee the challenge is not to master one of the skills but to apply all required skills simultaneously. Focussing on the arguments easily leads to a lack of attention to delivery aspects or vice versa. The trainee, therefore, will from the beginning practise on debating with tasks that integrate all skills required. The tasks will be combined in 3 task classes. In the first task class the trainee will get acquainted with debating, however, focussing on just a few specific aspects and within a relatively easy debating context. In the second task class the set of aspects to be trained upon will be expanded and the debate task will be more complex. At the final level, the trainee will mainly receive integrated feedback within a realistic debating context. Table 1 gives an overview of the final level. It describes the context and it indicates the feedback available indicating the type and amount of debating aspects to be mastered. Learners are expected to be sufficiently fluent at one level

before moving on to the next level. Given the large amount of possible feedback, it is expected that the feedback will be limited to a selection based on user preferences or priority rules related to e.g. seriousness of an error or chances of improvement.

Based on the task complexity aspects discussed below there are three task classes with each a number of tasks, supportive information and criteria to be matched. Adaptation will be possible by adapting the sequence and amount of tasks based on the performance of the learner. The assumption is that in the final setting, the training of the learner will follow through the tasks of each of the three task classes, based on their individual performance, in one or more sessions with in each session a separate round for each individual task. Below the tasks, supportive information and criteria for task class 1 are described.

Tasks. In the first task class the trainee will get acquainted with debating. The trainee will, however, only have to focus on a limited number of specific aspects i.e. voice volume, confident posture, time usage and overall performance. On the first two aspects in-action feedback will be given. The debating itself will be relatively simple e.g. a position statement and one argument exchange. Additionally, the trainees will familiarise themselves with the system with the help of "present yourself and discuss one interest" warming-up task. Examples of tasks in task class 1 are:

- Task 1a. Observe an expert debate video of approximately 3 minutes.
- Task 1b. Observe and assess a video of a 'standard' debate of approximately 3 minutes.
- Task 1c. Prepare and present yourself and discuss one interest
- Task 1d. Prepare and present your position on the topic "ban smoking" and debate

Supportive Information. An introduction on how to prepare, structure and deliver a debate will be provided. Special attention is given to the aspects, which are introduced at this level. How and why to use one's voice and how and why to show a confident posture and an appropriate use of time. Additionally, the trainee will get an introduction to the system.

Criteria for the Tasks. The main criteria to judge debating skills are generally accepted and connected to the skills distinguished in the skills hierarchy (Figure 3). They focus on content, argument structure and presentation and the ability of the trainee to set and guard their goals. Unfortunately, the criteria used in current practice are mostly

general and only qualitative. For instance they focus on posture in general ("appears confident") and are rated with qualitative assessments (such as e.g. poor, fair good or excellent) without a clear objective measurement procedure. At this stage, we therefore do not always have a simple way to translate the METALOGUE measurements to meaningful judgements or scores. Meaningful in this case means in line with and/or similar to a human qualitative assessment. For instance, translating a 'voice too low for 30 seconds' measurement to an summative judgement such as 'your use of voice volume is insufficient, sufficient or good' or alternatively to a formative judgement 'your use of voice volume is: not yet appropriate, sometimes appropriate, regularly appropriate, often appropriate or always appropriate'. As the system develops we will have to incrementally develop system output that provides meaningful formative or summative judgement by comparing and relating system measurements to human assessors (e.g.: Turnitin "Grade Anything: Presentations" http://vimeo.com/88075526? autoplay=true) both for single aspects such as "voice volume", and integrated aspects such as "authority", "likeability" or "overall dialogue performance", which are based on combinations of aspects.

Table 1: Task context and aspects to be mastered. In *italics* aspects on which also in-action feedback will be given.

Task Complexity	Level 3
Context Topic	Full topic. Number of argument exchanges is decided by participants
Context Opposition	Agreeable and disagreeable opponent
Context Length	max 10 min.
Goals	Indicator: - overall dialogue performance - target achievement
Contents and organisation	Visualisation Argument – Reason – Evidence use
Delivery overall	Visualisation AECAL Relative speaking time and turn time
Delivery voice	*Voice volume* *Speaking cadence* + Overall visualisation voice aspects
Delivery body language	*Confident posture* *Hands and arms usage* + Overall visualisation body language aspects
Emotion	Visualisation Emotions – Response pairs

3 FIRST EXPERIMENT

Given its complexity the METALOGUE system will be developed in three consecutive rounds with stepwise increasing functionality. While already at this time a global instructional design is available, many details are depending of the actual technical achievements and the usefulness of the design proposed. The latter in particular has to be confirmed in practice by the main stakeholders i.e. learners and teachers. The final selection of aspects to give feedback upon will be based on the use stakeholders preferences (youth parliament trainers), balance between voice and movement aspects, achieved preciseness of the aspects proposed and whether it can be mediated to the user in an understandable way. To this end in parallel with the three development rounds a series of pilots has been planned and a number of smaller experiments to validate the design on specific elements. In line with this, in our work towards the instructional designs for real-time in-action feedback we developed a prototype application called the Presentation Trainer. The application was developed with the purpose to study a model for immediate feedback and instruction in the context of one aspect of debating i.e. the initial presentation. The application utilises different sensor information to analyse aspects of nonverbal communication, such as body posture, body movements, voice volume and speaking cadence. The results of this analysis are then presented as feedback and instruction to the user. In the context of METALOGUE and the envisioned meta-cognitive real-time feedback, the application aims to ensure the situational awareness of the presenter by providing real-time feedback on the actual performance.

3.1 Presentation Trainer

The Presentation Trainer is a software prototype designed to support the development of nonverbal communication aspects for public speaking, by presenting immediate feedback about them to the user. The nonverbal communication aspects currently analysed by the Presentation trainer are: body posture, body movements, voice volume and speaking cadence.

Voice Analysis. To track the user's voice the Presentation Trainer uses the integrated microphone of the computer together with the Minim audio library (http://code.compartmental.net/tools/minim/). By analysing the volume input retrieved from the microphone it is possible to give instruction to the

user regarding her voice volume, voice modulation and speaking cadence. Speaking loud during a presentation is good to capture the attention of the audience, give emphasis and clear instructions. Speaking at a low volume during a presentation can be useful to grab the attention of the audience while giving personal opinions, sharing secrets and talk about an aside point. Nevertheless talking at a high or low volume for an extended period of time makes it difficult for the audience to follow the presentation (DeVito, 2014). Therefore the Presentation Trainer gives feedback to the user when the volume of her voice has been too loud, too low or has not been modulated for an extended period of time. In order to do this voice analysis the Presentation Trainer makes use of four different volume thresholds regarding the volume value received from the microphone. These thresholds can be set in running time according to the setting where the Presentation Trainer is being used.
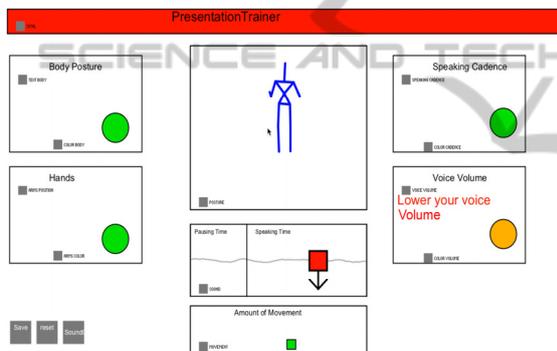

Figure 4: Presentation Trainer interface.

Body Language Analysis. The Presentation Trainer uses the Microsoft Kinect sensor (www.xbox.com/en-US/kinect) in conjunction with the OpenNI SDK (www.openni.org) to track the body of the user. This fusion allows the creation of a skeleton representation of the user's body. With the use of this skeleton representation, the Presentation Trainer is able to analyse the user's body posture and movements in order to give her feedback and instructions about it. While speaking to an audience it is important to project confidence, openness and attentiveness towards the audience. The body posture of the speaker is a tool to convey those qualities. Therefore 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, above the hips, and without the arms being completely extended (Bjerregaard, and Compton, 2011). To make it

possible for the Presentation Trainer to give feedback regarding the user's body posture we predefined some postures that should be avoided while giving a public presentation if one wants to convey confidence, openness and attentiveness. These postures are: arms crossed, legs crossed, hands below the hips, hands behind the body and hunchback position. The skeleton representation of the learner's body is compared against those postures and when a match is presented, the posture mistake is fired.

3.2 First User Study

The purpose of this first study (Schneider et al, 2014) was to explore the users' acceptance of the Presentation Trainer i.e. in particular the type of feedback provided and the timing of the feedback during the presentation itself. Before doing the user test, we introduced the prototype in a meeting where we explained the tool and its purposes. At the end of the presentation we let the audience give their feedback and impressions about the tool. After the presentation six participants volunteered for the user test.

The test consisted on giving a short presentation while using the Presentation Trainer as an immediate feedback training tool. In the experiment the participant were requested to give their presentation at a distance of approximately 2.5 m in front of the Microsoft Kinect and 2 computer screens. One of the screens displayed the Presentation Trainer (Figure 4), the other the slides that had to be presented. The people inside of the room during the test were the participant and the examiners. The test started by showing the participant a comic story containing 6 pictures and asking her to give a short presentation about it. Once the participant saw all the pictures and acknowledged being ready, (s)he started with the presentation. During the presentation, the Presentation Trainer was tracking the participant and displaying immediate feedback and instruction about the nonverbal communication.

After the presentation, participants were asked to fill in a System Usability Scale (SUS) (Brooke, 1996) questionnaire, followed by an interview. During the interview we showed the user interface of the Presentation Trainer to the participants and asked them questions to find out which components of the interface were the most used, helpful and interesting. We also asked questions on their general opinion about the Presentation Trainer and what they would like to get from it in the future.

3.3 Results and Discussion

Six participants took part on the study, half of them female and half of them male. The age of the participants ranged from 24 to 40 years old. The working experience of all of them is in the field of learning or computer sciences. Moreover, as part of their work, they have to perform public presentations a couple of times a year. The amount of participants are in line with the recommended amount of participants for this type of study (Nielsen and Landauer, 1993). The average scores for the SUS were: 67.5 for SUS, 77.1 for learnability, and 65.1 for usability (Lewis and Sauro, 2009). These relatively high SUS scores align with the enthusiasm expressed by the participants during the interview session towards using a tool like the Presentation Trainer to prepare for their presentations.

All participants concluded that the most observed element of the interface during the presentation was the Skeleton Feedback module and the second most observed was the Voice Feedback module. The coloured circles were observed but participants did not know how to change their behaviour based on them. The users had not observed the displayed texts with instructions. Some participants suggested using icons instead of text to give the instructions. Participants remarked about the overload of information required to give a presentation and the need to be aware of all the feedback at the same time. Nevertheless, after using the tool they all stated their enthusiasm towards the immediate feedback.

Observations during the user tests showed that though this version of the Presentation Trainer was only partially successful, the overall outcome did meet our expectations. In our instructional design blueprint we deliberately introduce a set of tasks in each task class to stepwise increase the complexity. In this experiment we knowingly did ask the user only to do one presentation to get their very first opinion on the system. Not being prepared for giving a presentation, regardless of its simplicity, confirmed to be a fairly complex task. It consumed most of the participants' attention; hence only a small percentage of their attention was paid on the Presentation Trainer. By examining the different feedback representations used during the tests, we identified that the ones continuously reflecting the actions of the participants', such as the skeleton and the voice feedback, were the easiest ones to be understood and followed during the presentation. As a result in our next prototype we will focus on further simplifying (iconizing) the representation of the feedback and to introduce part-task practice to train specific attention points.

4 CONCLUSIONS

The rising availability of sensors has created the space to design, develop and create tools to support learning and to give users in-action and about-action feedback on their performance. System design and instructional design do not normally go hand in hand. The latter commonly follows the first. However, with the practically unlimited amount of data we choose to elaborate the instructional design already at an early stage in our project to complement the technical design. In this paper, as a result, we discussed the design and development of the instructional aspects of a multimodal dialogue system to train presentation and debating skills. We outlined our instructional design taking into account the instructional design requirements for the task at hand and the data and their semantics available through the sensors used. The use of sensors enables us to propose a combination of in-action, immediate feedback and about-action feedback. Real-time, in-action feedback informs learners on the fly how they perform key skills and enables them to adapt instantly. About-action feedback informs learners after finishing a task how they perform key skills and enables them to monitor their progress and adapt accordingly in subsequent tasks. In- and about-action feedback together support the enhancement of the learners' metacognitive skills, such as self-monitoring, self-regulation and self-reflection thus reflection in- and about action. We discussed the practical challenge and our on-going work to select and develop criteria based on the sensor input, which are useful and in line with human judgment. Finally, we presented the result of our first experiment which demonstrated the technical feasibility of our approach and also indicated that, overall, users in principle accept the approach followed. In the forthcoming period we will continue to expand the system and support the development with experiments and pilots.

ACKNOWLEDGEMENTS

REFERENCES

Argyle, M., 1994. *The psychology of interpersonal behaviour.* Penguin Books, London.

Bjerregaard, M. and Compton, E., 2011. *Public Speaking Handbook.* Snow College, Suppliment for Public Speaking.

Brooke, J., 1996. SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland. *Usability Evaluation in Industry.* London: Taylor and Francis.

Coninx, N., Kreijns, K. and Jochems, W., 2013. The use of keywords for delivering immediate performance feedback on teacher competence development. *European Journal of Teacher Education*, 36(2), 164–182.

DeVito J.A., 2014. *The Essential Elements of Public Speaking.* Pearson.

D'Souza, C., 2013. Debating: a catalyst to enhance learning skills and competencies. *Education + Training*, Vol. 55 Iss 6 pp. 538 – 549.

Duncan, S., 1970. Towards a grammar for floor apportionment: A system approach to face-to-face interaction. In *Proceedings of the Second Annual Environmental Design Research Association Conference*, pages 225–236, Philadelphia.

Engeser, S. and Rheinberg, F., 2008. Flow, performance and moderators of challenge-skill balance. *Motiv Emot* 32: 158-172.

Hattie, J., and Timperley, H., 2007. The power of feedback. *Review of Educational Research*, 77(1), 81–112.

Heylen, D., 2006. Head gestures, gaze and the principles of conversational structure. *International journal of Humanoid Robotics*, 3(3):241–267.

Kendon, A., 2004. *Gesture: visible action as utterance.* Cambridge University Press, Cambridge.

Kiili, K., De Freitas, S., Arnab S. and Lainemac, T., 2012 The Design Principles for Flow Experience in Educational Games. *Procedia Computer Science 15 (2012) 78 – 91. Virtual Worlds for Serious Applications* (VS-GAMES'12).

Lewis, J. R., and Sauro, J., 2009 The Factor Structure of the System Usability Scale. *Proceedings of the first Human Centered Design Conference HCD* 2009, San Diego, 2009.

Mann W. and Thompson, S., 1988. *Rhetorical Structure Theory* http://www.cis.upenn.edu/~nenkova/Courses/cis700-2/rst.pdf.

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

Nielsen, J. and Landauer, T., K., 1993 A mathematical model of the finding of usability problems. *Proceedings of ACM INTERCHI'93 Conference,* pp. 206-213.

Nye, B.D., Graesser, A.C. and Hu. X., 2014. AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal Artificial Intelligence Education*, 24, pp 427-469.

Peldszus, A. and Stede M. 2013. From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence* (IJCINI) 7(1), 1–31.

Petukhova, V., 2005. *Multidimensional interaction of multimodal dialogue acts in meetings.* MA thesis,. Tilburg University.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. and Webber, B., (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC). Marrakech, Morocco.

Scheflen, A., 1964. The significance of posture in communication systems. *Psychiatry*, 17:316–331.

Schneider, J., Börner, D., Van Rosmalen, P. and Specht, M., 2014. Presentation Trainer: A Toolkit for Learning Non-verbal Public Speaking Skills. In *Proceedings of the 9th European Conference on Technology Enhanced Learning, EC-TEL 2014. Open Learning and Teaching in Educational Communities, Lecture Notes in Computer Science* Volume 8719, 2014, pp 522-525. Springer International Publishing.

Schön, D. A., 1983. *The Reflective Practitioner: How Professionals Think in Action.* (T. Smith, Ed.) (p. 374). Basic Books.

Sweller, J., 1994. "Cognitive Load Theory, learning difficulty, and instructional design". *Learning and Instruction* 4 (4): 295–312.

Triantafyllou, E., Timcenko, O., and Triantafyllidis, G., 2014. Reflections on Students' Projects with Motion Sensor Technologies in a Problem-Based Learning Environment. *Proceedings of the 8th European Conference on Games Based Learning ECGBL 2014.* Berlin, 2014.

Trimboli, A., and Walker, M. B., 1987. Nonverbal dominance in the communication of affect: A myth? *Journal of Nonverbal Behavior*, 11(3) pp. 180–190.

Van Merriënboer, J. J. G. and Kirschner, P. A., 2013. *Ten Steps to Complex Learning* (2nd Rev. Ed.). New York: Routledge.

Van Rosmalen, P., Boyle, E.A., Van der Baaren, J., Kärki, A.I. and Del Blanco Aguado, A., 2014. A case study on the design and development of mini-games for research methods and statistics. *EAI Endorsed Transactions on Game Based Learning* 14(3): e5.