D3.1 Instructional Designs for Real-time Feedback

metalogue

METALOGUE: Deliverable D3.1

Instructional Designs for Real-time Feedback

The METALOGUE Consortium

October 2014

Version N°: V1.0

Main Authors:

P. van Rosmalen, D. Börner, J. Schneider (OUNL),

V. Petukhova (UdS) and J. van Helvert (UESSEX)

Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development

Project ref. no.	ICT – 611073
Project title	METALOGUE – Multiperspective Multimodal Dialogue: dialogue system with metacognitive abilities

Document status	**Final version**
Contractual date of delivery	31 October 2014
Actual date of delivery	21 November 2014
Document number	D3.1
Deliverable title	Instructional Designs for Real-time Feedback
Dissemination level	RE
Type	Report
Number of pages	65
WP contributing to the deliverable	WP 3
WP / Task responsible	OUNL / OUNL
Contributing partners	OUNL, UdS, UESSEX
Reviewers	Fokie Cnossen, RUG
Author(s)	Peter van Rosmalen, Dirk Börner, Jan Schneider, Olga Petukhova & Joy van Helvert
EC Project Officer	Pierre-Paul Sondag
Keywords	Feedback, Sensors, Instructional Design, Multimodal Dialogue

Table of Contents

1. Executive Summary

The main objective of METALOGUE is to produce a multimodal dialogue system that is able to implement an interactive behaviour that seems natural to users and is flexible enough to exploit the full potential of multimodal interaction. The METALOGUE system will be arranged in the context of educational use-case scenarios, i.e. for training active citizens (Youth Parliament) and call centre employees. This deliverable describes the intended real-time feedback and reflection in-action support to support the training. Real-time feedback informs learners how they perform key skills and enables them to monitor their progress and thus reflect in-action. This deliverable examines the theoretical considerations of reflection in-action, what type of data is available and should be used, the timing and type of real-time feedback and, finally, concludes with an instructional design blueprint giving a global outline of a set of tasks with stepwise increasing complexity and the feedback proposed.

This deliverable is structured as follows. Chapter 3 explains the 4C-ID instructional design model, which is particularly suited to design complex learning. In addition, the main theoretical considerations underlying feedback and immediate feedback are reviewed, i.e. Schön's model of reflective practice as well as the concepts of Situational Awareness, Cognitive Load, and Flow Theory. In chapter 4 the available data, i.e. speech signals from multiple sources, visible movements tracking signals capturing body movements and facial expressions, and video signal captured by the camera that records the whole dialogue training session, is described. Based on this, feedback categories are derived, i.e. *goals* (the status of the goal to be achieved, progress and distractions), *content and organisation* (an integrative perspective on the use of argument, reason and evidence), *delivery* (an integrative perspective of how the speaker speaks), *emotion* (the emotional state of the user and opponent), *voice quality,* and finally *movements (*non-verbal behaviour*)*. The provided immediate, real-time feedback (i.e. feedback on behaviour as it happens, so as to optimize the immediately following action, e.g. argument delivery) will concentrate on the latter two aspects, which are relatively straightforward to understand and respond upon.

Chapter 5 discusses in detail two case studies. The first study, the Presentation Trainer, was developed with the purpose to study a model for immediate feedback and instruction for public speaking. It presents feedback and instruction to the user regarding aspects of their nonverbal communication, i.e. voice and body language. In the second study, the Feedback Cubes were developed prototypically in an attempt to research and develop a balanced ambient way to provide real-time feedback.

Finally, the last chapter summarises with an instructional design blueprint. It starts with a skills hierarchy of "conducting a debate" including an overview of which feedback for reflection in-action *and* about-action will be given in the three consecutive versions of METALOGUE and in the three task-classes designed for the trainee, thus aligning the METALOGUE incremental development with the instructional design. Finally, it describes the tasks of each task level and discusses how the main criteria to judge debating skills will be derived based on the feedback categories discussed in chapter 4.

2. Introduction

The main objective of METALOGUE is to produce a multimodal dialogue system that is able to implement an interactive behaviour that seems natural to users and is flexible enough to exploit the full potential of multimodal interaction. It will be achieved by understanding, controlling and manipulating the system's own and users' cognitive processes. The METALOGUE system will be deployed in particular in the context of an educational use-case scenario, i.e. in social educational contexts for training young entrepreneurs and active citizens (Youth Parliament). In addition the transfer to a second educational use case scenario, i.e. a business education context for training call centre employees to successfully handle their customers, will be explored.

An important aspect therefore of the METALOGUE project is the development and implementation of the instructional design of the educational dialogue to enable to train self-monitoring, self-regulation and self-reflection. The main goal of this work package is the development and implementation of the instructional design of the educational dialogue. The work package starts from the scenarios described in WP1 and the specified data points to be collected. As they are developed it will take into account the cognitive models from WP2. This work package will define the adaptive and personalized learning support for real-time feedback and reflection in-action support, reflection about-action and learning analytics, multi-perspective instructional designs, as well as strategic feedback based on the cognitive modelling.

The aim of this deliverable D3.1 is to describe the intended *real-time feedback and reflection in-action support*. In conclusion a global instructional design blueprint is defined based on real-time feedback loops. The provided real-time feedback informs learners how they perform key skills and enables them to monitor their progress. Giving an interactive presentation, i.e. a presentation including an argumentation, is a complex task. A trainee needs to master both content aspects (i.e. what to present, how to structure their presentation and which argument to use in the closing argumentation) and other modalities, such as voice aspects (i.e. how to control and use their voice e.g. pitch, speed or volume) and body language aspects (i.e. how to control and use their body e.g. arms, hands or align their body). At the same time 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). Similar, also the call centre trainee has to master content aspects as well as other modalities and has to be aware of the effects of interactions with the customer and therefore continuously monitor, reflect and adapt when necessary. In the real world all interactions will happen at once and at full scale, i.e. for a trainee it is 'sink or swim'. The METALOGUE system, however, should be able to moderate and adapt tasks and support to the level that it fits a trainee while assuring that the task at hand is motivating, realistic and not too easy nor too complex (Instructional Design) and at the same time the feedback does not interfere too much (Cognitive Load, Flow Principle) with the task performed. Moreover, the feedback should not be merely used to correct but also to help the trainee reflecting on their actions (model of reflective practice), thus becoming aware of how they can steer their interactions (Situational Awareness).

In this deliverable we will one by one discuss the theoretical considerations, pointed to above, that will guide the design of the development and implementation of the instructional design of the educational dialogue. The focus of this deliverable is real-time feedback and reflection in-action support. Closely connected, and preferably read together, is D3.2 which focuses on Learning analytics and reflection about-action support. In chapter 3 we will start with a discussion of the 4C-ID instructional design model which should enable a design taking into account the requirements for adaptivity and the requirements raised by Schön’s model of reflective practice, Situational Awareness, Cognitive Load, and Flow Theory. In chapter 4 we will give an overview of the data available for real-time feedback and discuss the use for both reflection in-action and about-action (thus partly already introducing D3.2). Annex 1-3 complete this chapter with a tabular overview of all possible feedbacks foreseen. In chapter 5, we discuss two studies, which explored the consequences of the theories discussed in chapter 3 for the design of real-time feedback. The studies focussed on what, how and when to present immediate feedback. Finally, in the last chapter we will outline an instructional design blueprint taking into account the background discussed in chapter 3, the available data (chapter 4) and the preliminary application example studies carried out in the context of this deliverable (chapter 5).

3. Background

3.1 Instructional Designs: 4C-ID Model

The instructional model in METALOGUE will be based on the four component instructional design model (4C-ID model) (Van Merriënboer, 1997) which aims at teaching complex skills. The model is accompanied by a comprehensive description of how to design in line with the model (Van Merriënboer & Kirschner, 2013) and is based on extensive research of the acquisition of complex skills or professional competencies. The design principles incorporated in the model are directed at promoting transfer of what learners learned by focussing on whole authentic tasks integrating knowledge, skills and attitudes.

Giving an interactive presentation, i.e. a presentation including an argumentation or a deal with a customer call, is a complex task. A trainee needs to master both content and organisation aspects (i.e. what to present, how to structure it and which argument to use) and delivery aspects such as voice quality (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). At the same time a trainee has also to be continuously aware of the effects of their interaction and (metacognitive aspects) monitor, reflect and adapt when necessary. Moreover, the METALOGUE system will put a demand on a trainee. Common practice in education and training is to give feedback after a task has been performed. METALOGUE, however, will also provide real-time feedback during task execution. The design, therefore, has to pay specific attention not to overload the learner, while at the same time the tasks will have to be sufficiently challenging and at the end meet the full complexity required.

The underlying assumption of the 4C-ID model (Four Component Instructional Design) is that complex learning can be designed with the help of four interrelated components (Van Merriënboer, & Kirschner, 2013):

1. Learning tasks. Authentic, whole tasks preferably based on real-life tasks and organised in task classes with variation and increasing complexity.
2. Supportive information. Information that is supportive to the non-recurrent aspects of the tasks and explains how a domain is organised.
3. Procedural information. Information that is prerequisite to the recurrent aspects of tasks and instructs how to perform the routine aspects of a task. This information is available just-in-time and typically, stepwise will fade out when experience increases.
4. Part-task practice. Additional practice for routine aspects of learning tasks that require a high level of automation.

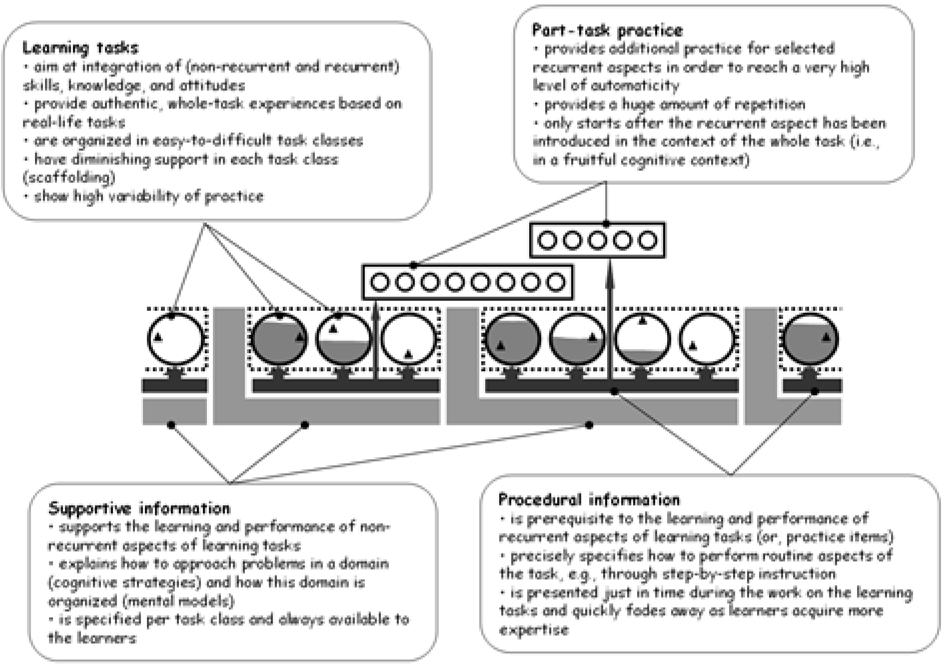


Figure 3.1 The Four Component Instructional Design Model (Van Merriënboer, & Kirschner, 2013).

The 4C-ID-model assumes that all human knowledge is stored in cognitive schemata and, is supported by cognitive load theory (Sweller, 1994; see section 3.3.2), a cognitive architecture with a working memory with a very limited capacity when dealing with novel information, as well as an effectively unlimited long term memory, holding cognitive schemata. The instructional interventions that are included in the 4CID model are directed at managing the cognitive load in working memory, enhancing schema construction and facilitating schema automation. The 4C-ID model, typically, is and has been applied for course and curriculum design (Hoogveld, Janssen-Noordman & Van Merriënboer, 2011). Recently, the model has also been applied for the design of serious games, 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 proposed practice to automation of selected part-tasks) fit well with game (design) practice (Huang, W. D., & Johnson, 2009; Lukosch, Van Bussel & Meijer, 2012; Enfield, 2012). 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 in a youth parliament setting or to successfully handle their customers working with realistic, engaging tasks adjusted to the user on the appropriate level of complexity, and if necessary, the option to practice selected subtasks.

3.2 Reflection Processes

The essential aim of an instructional design for real-time feedback in the context of METALOGUE is the support of a reflection process for the participants, i.e. the partners involved in the dialogue. Reflection as a theoretical concept has been widely discussed in the literature. Just recently Verpoorten (2012, p16) discussed the concept in great detail in the context of his dissertation on “Reflection amplifiers in self-regulated learning”:

> *“Reflection is an influential factor of learning both in regular classrooms and in eLearning settings (Heargraves, 2005; Higgins, 2011). Meta-analyses enduringly rank reflective practice among the strongest levers for learning (Hattie, 2009; Higgins, Kokotsaki, & Coe, 2011; Lai, 2011; Marzano, 1998; Wang, Haertel, & Walberg, 1990). As a typically human negotiation process between the self and the experience of the world, reflection is not just an “add-on” to instruction, but an essential component of a deep approach to learning (Marton, Dall’Alba, & Beaty, 1993). Its practice before, during and after action helps gradually develop learners’ awareness of what supports or hampers a consistent orchestration of the various dimensions of their learning, so that they can evolve into expert learners (Ertmer & Newby, 1996)”.*

Besides highlighting the importance of reflection on learning, the author (Verpoorten, 2012, p. 18) also explore how the concepts are interwoven:

> *“The confluence of experience (action) and thought (reflection) creates learning (e.g., Freire, 1973; Kolb, 1984). Learning is both an active and a reflective process. It is difficult to extricate one from the other since they operate often in “parallel processing” (Burns, Dimock, & Martinez, 2000). Furthermore reflection interacts often subconsciously in the midst of doing (Koriat & Levy-Sadot, 2000)”.*

One of the main outcomes of this exploration is the argument that “in order to foster pupils’ development as learners, it is useful, [...] to trigger and externalize reflection.” This is in line with the aim of the METALOGUE system to support the development of metacognitive skills, i.e. providing instructional designs for real-time feedback in order to inform learners how they perform, and enables them to monitor their progress. Verpoorten (2012, p. 20-21) also outlines the relation between reflection and meta-cognition:

> *“According to common sense, reflection lies somewhere around the notion of learning and thinking. People reflect in order to learn. Reflection is therefore practised for the sake of considering an object in more details (Amulya, 2004; Bengtsson, 1995; Moon, 2001). [...] Beyond the intuitive grasp, reflection turns quickly into a complex construct. The notion of reflection is akin to constructs like meta-cognitive instruction and development (Gama, 2004), learning to learn (Hoskins & Fredriksson, 2008), learning about learning (Watkins, 2001), learning/study skills (Hattie, Biggs, & Purdie, 1996; Higgins, Baumfield, & Hall, 2007; Tabberer, 1984), self- regulated learning (Isaacson & Fujita, 2006; Ridley, Schutz, Glanz, & Weinstein, 1992) and, more recently, situation awareness (Salmon et al., 2007). This proximity has lead to a variety of*

different interpretations and understandings of the word "reflection" among educational researchers and practitioners (Zeichner, 1984). This ill-defined nature of reflection has triggered fierce incriminations (Eraut, 2002; Ixer, 1999). Despite this invigorating criticism, reflection is a term which is often used in education and it is difficult to deny any legitimacy to it. References to a self-reflective consciousness can be traced as far back as Socrates' "inner voice".

The idea of a self-reflective mind has been given a new impetus by Flavell (1979), who attempts to generate a formal model of meta-cognition in the realm of educational psychology, and by Schön (1983, 1987) who grants a major importance to reflection, in his effort to elucidate the inner working of professional practice and learning organisations.

The notions of meta-cognition and reflection are strongly interwoven, if not overlapping or interchangeable (Georghiades, 2004; Scharp, 2008)."

Finally the author (Verpoorten, 2012, p. 21) defines reflection as "an active process of witnessing one's own learning experience and evaluating its different aspects. Reflection is considered as a means by which learners can build and evolve a mental model of the learning process they are committed to and of their position inside this process (Seel, Al-Diban, & Blumschein, 2002), so that appropriate directions and actions can be procured."

Schön's model of reflective practice

Schön (1983) defines reflective practice as the practice by which professionals become aware of their implicit knowledge base and learn from their experience. He coins 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). Within the METALOGUE project we refer to this basic distinction between reflection-in-action and reflection-about-action. While this deliverable focuses on an instructional design for reflection-in-action, deliverable D3.2 will focus on the reflection-about-action aspect.

Situational Awareness

Beside the discussed reflective practice, another aspect is important to consider when supporting reflection-in-action with real-time feedback, namely the situational awareness of the professional or learner. Endsley (2000) defines situation awareness as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". Following this definition the author presents three levels of situational awareness that can be used for classification, namely perception, comprehension, and projection. Perception is related to situational cues and important or needed information, comprehension relates to how people integrate combined pieces of information and evaluate their relevance, and finally projection relates to how people are able to forecast future events and situations as well as their dynamics. In the context of METALOGUE the concept allows to model and discuss the emerging feedback loop (see Figure 3.2) when enriching the dialogue setting with real-time feedback.

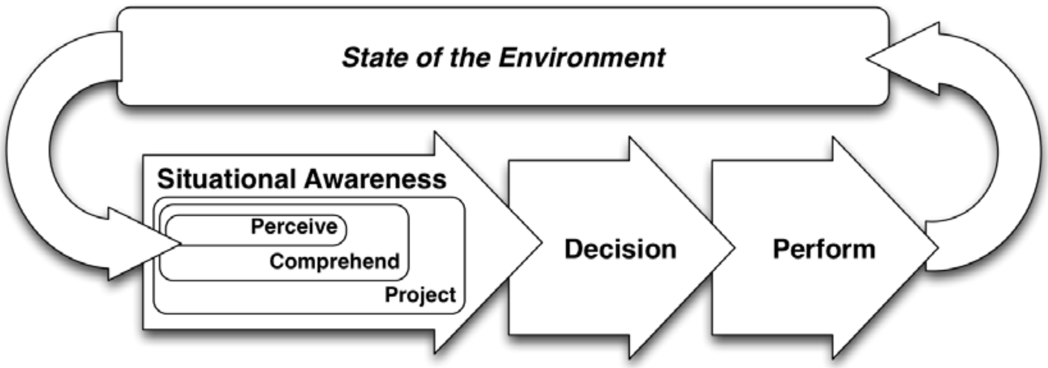


Figure 3.2 Situational Awareness Feedback Loop (Endsley, 2000)

3.3 Feedback[1]

Feedback is one of the most powerful interventions in learning (Hattie & Timperley, 2007). According to some authors (Nicol & Macfarlane-Dick, 2006), the most beneficial thing tutors can do to students is to provide them feedback that allows them to improve their learning. Moreover, high quality feedback is a requirement for formative assessment (Gedye, 2010). Therefore we decided to analyze the type of feedback given by sensor-based platforms. Feedback in this study is defined as the information about a person's behaviour or performance of a task, which is used as a basis for improvement (Oxford Dictionaries, 2014). By focusing on the learning support this review aims to analyze how the design patterns of the prototype align to the effective feedback framework (Hattie & Timperley, 2007).

According to the work of Hattie & Timperley (2007) effective feedback gives answers to the following questions: "where am I going?", "how am I going?" and "where to next?". The question "where am I going?" refers to the learner's goals; goals produce persistence at task performance in the face of obstacles, and support the resumption of disrupted tasks in the presence of more attractive alternatives (Bargh et al., 2001). The answer to "how am I going?" provides information relative to a task or performance goal of the user. Finally the answer to "where to next?" shows the learner the next steps to take towards the completion of his goal. Implementing the answers to these questions on a computerized system is not a straightforward task. In order to answer the question of "where am I going?" first it is important to know the goals of the user. The challenge comes in reminding the user about these goals and presenting the user with feedback on how the current task and performance aligns to the goals. The work of Goetz (2011) has suggested that by presenting the user with evidence of his current behaviour together with the consequences allows the user to

[1] *This section is part of a journal submission of a METALOGUE literature review on sensor-based learning support.*

perceive an alignment between his performance and goals. Sensors can be used as tools to collect this evidence. Presenting this evidence and the potential consequences is something that can be implemented on a sensor-based platform.

In order to answer, “how am I going?”, the performance of the user needs to be tracked, and this performance has to be compared against some rules. Through the dimension of feedback complexity described by Mory (2004) this answer can be presented to the user. This dimension identifies 5 different levels of feedback, including no feedback, simple verification, correct response, elaborated feedback, and try again feedback.

The implementation to the answer of “where to next?” has two basic requirements. First a map with all steps to achieve the learner’s goal is required. Second it is important to identify the current position of the learner on this map. The measuring and analysis qualities of sensor-based platforms seem suitable to identify the current position of the learner inside on the learning map. Moreover, having sensor-based platforms make use of system adaptation techniques such as direct guidance, content-based filtering (Brusilovsky, 2004), and self-adaptation through feedback loops (Brun et al., 2009), opens the possibility for them to present the learner with a personalized learning map.

Feedback research variables of interest

The concept of providing (instructional) feedback was based on an extensive research review in this area by Mory (2004). While her review is not focused specifically on computer-mediated feedback, the general feedback research variables of interest presented are also applicable for studying the interaction between learners and ambient displays. These variables are information content and load, referred to as complexity, timing, error analysis, learning outcome, and motivation. The author differentiates several levels of complexity such as simple verification, try-again feedback, or elaborated feedback. The timing of the feedback can be immediate or delayed, while errors can be analysed, if at all, in a corrective or confirmatory manner. The learning outcome again can target several levels, including declarative knowledge or concept learning and even higher-level outcomes, such as rule learning, problem solving, cognitive strategies, psychomotor skills, or attitude learning. In addition, feedback can have effects on a motivational level, e.g. in relation to self-efficacy and task expectancy, triggered by goal or performance discrepancy, or exposed by causal attributions.

3.4 Other Theoretical Considerations

3.4.1 Cognitive Load Theory

Cognitive Load Theory (CLT) (Sweller, 1994) aims to explain under what conditions learning will be optimal and gives guidelines for instructional design. The central notion in this theory is the limited capacity of people’s working memory. People are only able to deal with a certain amount of cognitive load.

CLT distinguishes three types of cognitive load: intrinsic, germane and extraneous load. Intrinsic load is a result of the inherent level of difficulty associated with the specific assignment. In layman's terms: intrinsic load is the amount of thinking that is required for the activity. Germane load describes the load needed for the processing, construction and automation of schemas. A good example of schema construction and application is calculus & mathematics. As adults we don't have to think about what the result is of 3+2. We immediately know that the answer is 5. However, these rules have to be learned during early childhood in order to apply them in more complex settings. Extraneous load is caused by the way the information is presented. Design decisions have therefore a direct impact on the amount of extraneous load.

In order to store information in long term memory, it has to be processed by our working memory which is limited. Hence, it is important for instructional designers to reduce extraneous load as this will free resources for dealing with intrinsic and germane load. Learners are then able to encode information into long term memory. If working memory is overloaded, learning will become ineffective.

3.4.2 Flow Theory

"The flow experience, or the state of total involvement in an activity that requires complete concentration" (Csikszentmihalyi, 1999) was originally studied in the domain of psychology within the context of happiness. It roots back to the more general question why people are highly involved in activities without obvious external rewards. Based on the work of Csikszentmihalyi (1975) and Rheinberg (2008), Engeser and Rheinberg (2008) summarise flow state as follows: "(1) A balance between perception of one's skills and the perception of difficulty of the activity (task demand). In this state of balance, one feels both optimally challenged and confident that everything is under control. (2) The activity has coherence, contains no contradictory demands, and provides clear, unambiguous feedback. (3) The activity seems to be guided by an inner logic. (4) A high degree of concentration on the activity due to undivided attention to a limited stimulus field. (5) A change in one's experience of time. (6) The self and the activity are not separated, leading to a merging of the self and the activity and the loss of self-consciousness". Not surprisingly, since the playing experience in serious games of the learners is crucial in their engagement and therewith the achievement of their learning goals, flow is also studied in the context of serious games (GALA Network of Excellence: D2.4 "Flow for Serious Games", to appear) i.e. which aspects of serious games contribute to flow experience of the user and how. Kiili et al (2012) distinguish sense of control, clear goals and challenge-skill dimensions of flow as the most important aspects. Other dimensions of importance are a rewarding experience and feedback. Notably, Kiili et al argue and connect flow with cognitive load. The use of the artefact should be effortless and easily learned, i.e. reducing extraneous load as much as possible. The learning tasks themselves should be engaging, i.e. not effortless. They should ask for a realistic effort imposing a germane cognitive load essential for knowledge construction. Users will be more willing to use effort to accomplish an engaging task.

The learning tasks should offer sufficient variety in differences in complexity to be engaging. The in-action, immediate feedback should be (Hattie & Timperley, 2007; Engeser & Rheinberg 2008; Coninx, Kreijns & Jochems, 2013) specific and goal oriented, i.e. focus on the aspects of the interaction so that the learners become aware and in combination with the about-action feedback comprehend their meaning and use them accordingly; clear, i.e. not ambiguous so there are no interpretation problems; concise, i.e. short so they are as little disruptive as possible; and predictable, i.e. the type of feedback should be known/agreed upon in advance.

4. Types of Data and Feedback

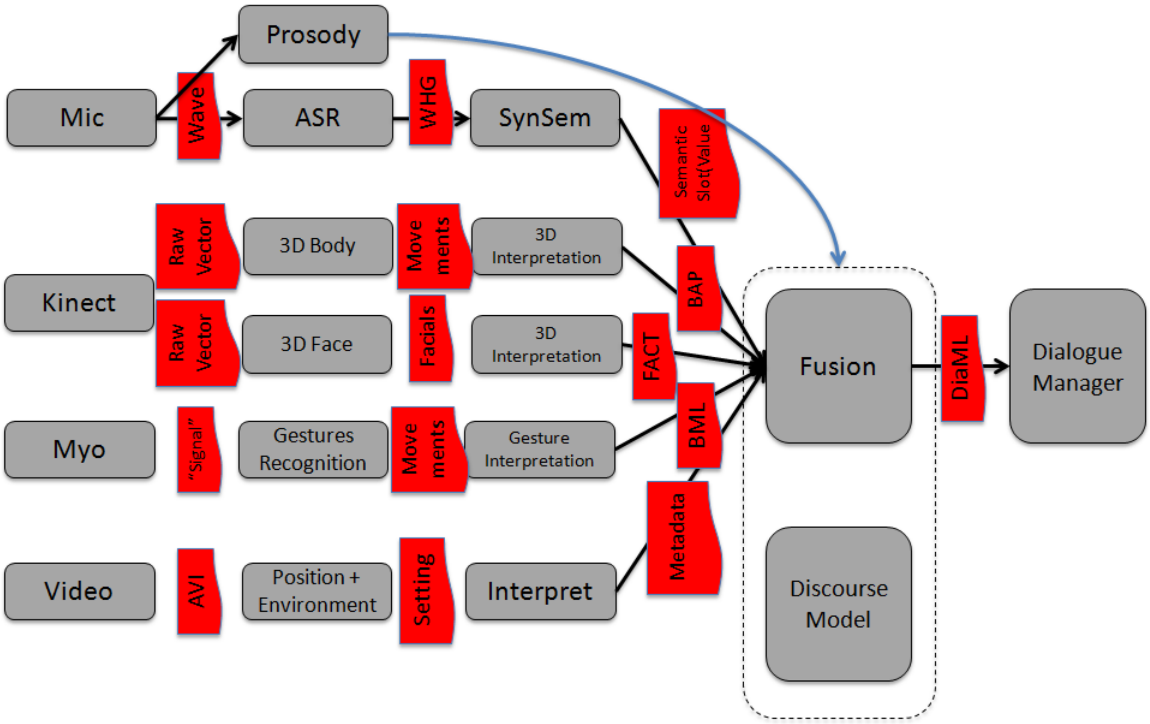


Figure 4.1 METALOGUE processing workflow and formats of data

4.1 Types of data

Sensor-specific input, such as captured from microphones, Kinect and Myo sensors and video cameras, is described in details in D1.1 which concerns the overall METALOGUE system architecture. Figure 4.1 above depicts the METALOGUE processing workflow and formats of data stream for each module input and output. For this deliverable we specify in more detail what raw data is collected in data collection experiments (see D1.2, D1.3 and D1.5) and elaborate how this data will be used for in-action (real-time) and about-action feedback generation by the METALOGUE system. There basically are 3 types of sensor specific data that 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).

4.1.1 Speech signals

Speech signals originating from all types of microphones used are encoded in wav files (see format details in D1.5) which will serve as input for 2 types of further processing: (1) Automatic Speech Recognition (ASR), described in length in D1.1, generates as output Word Hypothesis Graph (WHG) that is input for further syntactic and semantic analysis and for discourse model update (should answer the question: What was said?) (2) Prosodic Analysis (should answer the question: How it was said?). The latter is mostly concerned with

(1) quantitative and qualitative acoustic voice analysis, such as spectrogram, energy and pitch (fundamental frequency) and speech durational and temporal analysis, such as segmentation and speaking rate but also temporal regions of pitch accents. Prosodic features encoding is one of the topics of Deliverable 1.6. Results of prosodic analysis are important input for the system to generate feedback concerning **voice quality** that will include feedback on the following phenomena:

- Speech rate (fast; slow; adequate tempo)
- Volume (loud; soft; adequate loudness)
- Emphasis (flat intonation; uneven/unbalanced intonation; correct ratio/balance of accented/stressed, and unaccented/unstressed segments)
- Pausing (too long silences within segments, e.g. > 500ms; no pausing before new/important information; no silence/pausing at all)

Moreover, prosodic analysis is important to identify participant’s emotional state, e.g. nervousness level, and degree of uncertainty, e.g. hesitation phases using speaking rate (speech speed) and pausing.

4.1.2 Visible movements

Body language is an important modality to consider in debating and negotiation. This component will employ a Kinect sensor - it includes a camera for a video feed, an infrared projector and a sensor for 3D positioning (see D1.1 and D1.5 for more details). The following aspects of body language will be captured and analysed in METALOGUE:

- Gaze (re-) direction
- Head movement and head orientation
- Facial expressions
- Hand and arm gestures
- Posture shifts
- Body orientation

4.1.2.1 Semantics of visible movements

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. For example, if two people of different power or status meet, the low-power person looks at the other much more as he listens than as he talks, while there is no such difference for the high-power individual (Argyle, 1994). Gaze is also used to ensure contact between participants, for example, the speaker looking at an addressee signals that he is interested in his attention, wanting him to be involved. For this purpose so-called ‘mutual gaze’ is used, where people are looking at each other for some time. Participants break ‘mutual gaze’ when they close the interaction. 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. Head nods, shakes, turns, and jerks have been distinguished as actions performed by listeners to provide speakers with feedback on their message (Duncan, 1970). It has also been suggested that these head movements are responses to head movements of speakers, who may use this as a means to request feedback (McClave, 2001). Feedback functions of head movements can thus interact with turn management functions. Hadar et al.(1984) investigated whether it is likely that head movements are used for the latter purpose. They reported that the vast majority of head movements (89 out of 99) were performed by speakers rather than by listeners. Most of the speaker's head movements were located around initiations of speech after breaks between either syntactic clauses or turns. They concluded that speakers use head movements both to mark syntactic boundaries and to regulate the process of turn-taking.

Head movements are also used to indicate aspects of information structure, e.g. to mark alternatives, or contrast; 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 (see e.g. Kendon, 2004; McNeill, 1992; Ekman and Friesen (1981). Hand and arm gestures may also have interactive functions, especially, when aligned with speech in such a way that they are finished before the end of the turn. Stopping to gesticulate can be recognized by the hand dropping into a resting position, or the relaxation of a tensed hand position. These movements can therefore serve as a signal that the turn will soon end. Since co-speech gestures can make clear that a speaker is not about to finish talking, their presence can signal a Turn Keep function (Duncan, 1970). 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:

i. Keep hands out of your pockets
ii. Do not fiddle with your hair, nails, other body parts or objects in your hands (e.g. cue cards or clicking pen) or in your environment (e.g. tap on table); in other words, avoid all adaptors (also called manipulators) like rubbing your face, touching your nose, etc.
iii. Keep gesticulation calm (no fast abrupt movements)
iv. Avoid pointing gestures and if you need to point to something or emphasise something use open palm up gesture with all fingers together
v. 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), Condon and Osgton (1971), Erickson (1975), Hirsch (1989)). Cassell et al. (2001) found that both turn boundaries and discourse segment boundaries had an influence on the occurrence of posture shifts. Posture shifts occur more frequently, and tend to be more energetic, at discourse unit boundaries than within discourse units. Also, participants were shown to be five times more likely to show posture shifts at a turn boundary than within a turn. When a participant simultaneously starts a new turn and a new discourse unit, this is marked with a posture shift ten times more often than when a participant starts a new turn within the same discourse unit. As such, posture shifts may be more related to discourse structure than to turn management.

In debating and/or negotiations, or when presenting, posture and overall body orientation plays an important role. Debating guidelines talk about confidence posture:

- Keep legs aligned with your shoulders
- Your feet approximately 10-15 cm apart
- Distribute your weight equally on both legs
- Keep shoulders slightly back
- Turn body towards the opponent
- Never cross your legs
- Do not press down your weight on one hip

Facial expressions are the most complex signals of all the above mentioned. Face has 43 muscles identified. They all contribute to generate a facial expression of a certain type. Parts of face that are normally analysed as important contributors to certain facial expressions or actions are forehead (e.g. constricted or relaxed), eyebrows (e.g. raised or lowered), eyes (e.g. narrowed or widened), nose (e.g. wrinkled), cheeks (e.g. raised), and lips (e.g. corner pulled). Facial expressions are important for expressing *emotional* reactions, such as happiness, surprise, fear, sadness, anger and disgust or contempt (Argyle, 1994). These six basic emotions are found in all cultures. Emotions as complex signals will be analysed in METALOGUE in combinations with verbal and prosodic components.

Moreover, face can also display a state of *cognitive processing*, e.g. disbelief or lack of understanding.

4.1.3 Semantics of verbal contributions and pragmatics of multimodal input

In debates, debater's performance is often judged on three main criteria: (1) argument content; (2) argument organization and (3) argument delivery[2]. So far, what is discussed in 4.1.2, such as tone of voice, speech rate, body language, emotions, etc , can be used to evaluate the later criterion – delivery. To recap, delivery is about how the debater speaks: confident, near-native pronunciation, tone, pace, posture, gesture and eye contact. There are 5 things to be considered: **A**udibility, **E**ngagement, **C**onviction, **A**uthority and **L**ikability (AECAL). Good debaters should give a strong impression that they truly believe what they say. To express authority the debater needs not only use his voice and body but also support his arguments with statistics, facts and figures, but also personal experience or experience from real life of other people. Likability is about showing respect and friendliness.

Nevertheless, debate is about argumentation. Argumentation is the planning and preparation involving **argument** as a general conclusion, supported by **reason**(-s) and **evidence**. This structure is often called ARE[3]:

A = Argument (e.g. Marijuana should be legalized)
R = Reason (e.g. It does not harm a human body)

E = Evidence (e.g. According to recent research reported in Harm Reduction Journal, May 9 2006, frequent marijuana use is unlikely to be neurotoxic to the normal development of adolescent brain)

Good debaters are distinguished by concise clear connected by implicitly signalled structure of those, e.g. by discourse markers and dialogue announcement acts. For example, 'I will talk in favour of ... Because ... Since international research shows...'

[2] See 'How to Debate' rules: http://www.wikihow.com/Debate

[3] *See* http://www.slideshare.net/Cherye/advanced-debating-techniques

A well-known technique for structuring arguments is ‘**Chunking’**:

1. Chunk up – abstract overall principle. For example, `*We live in a society that allows us to use things that do not harm us. Marijuana does not harm. It should be legalized’*.
2. Chunk down – example from real life. For example, `*Do you know that Barack Obama, Bill Gates, William Shakespeare and Albert Einstein have all used marijuana? These people seem perfectly normal to me.’*
3. Chunk sideways - analogy, e.g. compare use of marijuana with use of alcohol

The METALOGUE trainee’s performance will be judged based on criteria defined in Table 4.2. Debaters’ way of **structuring arguments** will be analysed and annotated. The most recently proposed argumentation scheme of Peldszus and Stede (2013) will be used. The scheme is based on detecting proponent’s and opponent’s moves in a basic debating situation. The authors distinguish between basic elements of an argument which consists of non-empty set of premises and a conclusion. There are different support links between premises and a conclusion, such as *linked support* where two or more premises together support one conclusion; *multiple support* where two or more premises independently support one conclusion; *serial support* where one premise is a support for another premise which on its turn supports a conclusion; and *example support* where a premise provides an example for a conclusion.

Further, arguments can be either *attacked* by the opponent, *anticipated* by the proponent (temporal role switch proponent vs opponent, e.g. express awareness of exceptions), or *counter-attacked* by either the proponent or the opponent. There are two possible ways to attack an argument: (1) to present an argument against conclusion or its premise (*rebutting*) and (2) to diminish their supporting force (*undercutting*), see Peldszus and Stede, 2013.

In addition to argument structure annotation, links between premises and conclusions, as well as rebutting and undercutting links will be 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). The following relations are currently considered in METALOGUE: Elaborate, Exemplify, Justify, Motivate, Explain, Cause, Condition, Restatement, Concession, Alternative, Exception and List. This set will be potentially modified in order to better fit the METALOGUE data.

Machine learning algorithms will be trained in order to build classifier(-s) to detect argument units, its internal structure and type of relations between premises and conclusions.

The pragmatic analysis, in our view, brings all discussed in Section 4.1 together. This type of analysis is based on identifying speaker’s intentions in terms of **dialogue acts** as specified in ISO 24617-2 (also see D1.1). The ISO 24617-2 taxonomy distinguishes 9 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 1 optional dimension addressing establishing and maintaining contact (*Contact Management*).

38 domain-specific speaker's intentions are identified like Turn Grab or Turn Keep, Stalling or Feedback Elicitation, etc., and 44 general purpose intentions like Request, Agreement, Confirmation, Suggestion, Offer, etc. (see full specification http://dit.uvt.nl/#iso_24617-2 and Deliverable 4.1). There are feedback and functional dependence links, and rhetorical relations between segments and dialogue acts identified. Additionally, there is a set of qualifiers defined in order to better describe dialogue participant's behaviour in terms of (1) speaker's sentiments towards the addressee, side-participants, towards what he/she is saying or towards things that he/she intends to do; (2) the strength or weakness of certain speaker's assumptions and beliefs; and (3) the physical and emotional abilities and state of a dialogue participant.

Thus, a dialogue act specification includes

- pointers to stretches of performed speaker's behaviour as discussed above (either verbal input from ASR with additional prosodic analysis attached to it or visible movements, or, which is more often, both, in case of multimodal segments),
- representation of semantic content (what the segment is about, e.g. in terms of predicate-argument structure),
- identified communicative function (read speaker's intention) and
- links referring to previous segments or dialogue acts in a dialogue history (e.g. rhetorical links, but also functional and feedback dependence links).

The output is represented in DiAML (XML-based) as illustrated in D1.1.

4.2 Types of Feedback

Following Figure 4.1 in the previous section, we discussed the METALOGUE processing workflow and formats of the data stream for each module input and output, and elaborated how this data will become available for feedback generation by the METALOGUE system. In this section we introduce how we aim to use the available feedback to create and offer in-action and about-action feedback (Figure 4.2, see also D3.2) and provide 3 tables summarizing the indicators for dialogue acts, voice and body language.

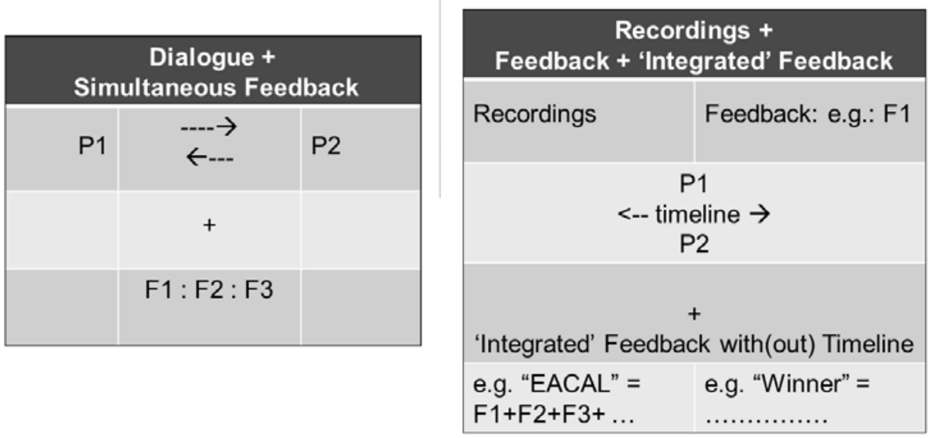


Figure 4.2 Left (D3.1): In-action feedback "F1", "F2", "F3" aligned with the ongoing dialogue. Right (D3.2): About-action Feedback a combination of Recordings and Feedback either on one aspect e.g. a voice aspect or on a combination of aspects e.g. "AECAL".

As discussed in chapter 3, an interactive presentation, i.e. a presentation including an argumentation or a deal with a customer call, is a complex task. A trainee needs to master both content aspects (e.g. what to present and how to structure it), delivery aspects (e.g. how to control and use their voice and their body) at the same time a trainee has also continuously to be aware of the effects of their interaction and (metacognitive aspects) monitor, reflect and adapt when necessary contents and delivery. Whereas immediate feedback is powerful, in order to be successful the in-action, immediate feedback should be (Hattie & Timperley, 2007; Engeser & Rheinberg 2008; Coninx, Kreijns & Jochems, 2012):

- *specific and goal oriented*, i.e. focus on key aspects of their interaction so that the learners become aware and in combination with the about-action feedback comprehend their meaning and use them accordingly;
- *clear,* i.e. not ambiguous so there are no interpretation problems about its meaning or requiring complex reasoning about its cause and how to respond to it;
- *concise,* i.e. short so they are as little disruptive as possible;
- *predictable,* i.e. the type of feedback should be known/agreed upon in advance.

Therefore the in-action, immediate feedback (i.e. feedback on behaviour as it happens, so as to optimize the immediately following action) 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 respond upon. Aspects related to argument content and argument organisation will be only implicitly addressed through the discourse constructed in the METALOGUE system. The in-action aspects to be used will be based upon the set described in Annex 1 (Voice quality aspects) and Annex 2 (Non-verbal behaviour aspects). The final selection of aspects will be based on use case preference (call centre or youth parliament), balance between voice and movement aspects, fit with the about-action feedback, achieved preciseness of the aspects proposed and whether it can be mediated to the user in an understandable way and, if necessary, selected on their usefulness through small experiments (c.f. chapter 5).

The about-action feedback (i.e. feedback after the event, to review, analyse, and evaluate the situation, so as to gain insight for improved practice in the future) will build upon the in-action feedback and give feedback based on aggregations of the in-action feedback and feedback based on the semantics of the verbal contents and dialogue act use (annex 3 Dialogue act use aspects). Together, about-action feedback use the following partly related and interconnected 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 (c.f. WP2 agent and user model)).
- *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 give aspects of and an integrative perspective of how the speaker speaks (AECAL).
- *Emotion*. Given the importance of the awareness and appreciation of the emotional state of the user and opponent special attention (depending on the achieved recognition preciseness) 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.
- *Gap*. Finally, the user will be enabled to define (see D3.2 section 3.2.2) their individual points of reflection, so called gap moments. Gap moments are personal moments of struggle, angst or uncertainty or success; moments where sense cannot immediately be made.

Similar to the in-action feedback, the feedback given will not be exhaustive but be based upon use case preference (call centre or youth parliament), balance between voice and movement aspects, fit with the about-action feedback, achieved preciseness of the aspects proposed and whether it can be mediated to the user in an understandable way and, if necessary, selected on their usefulness through small experiments.

5. Application Examples

In our work towards instructional designs for real-time feedback we explored the concepts introduced in section 3 developing two application examples. The *Presentation Trainer* application was developed with the purpose to study a model for immediate feedback and instruction for public speaking. 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 metacognitive real-time feedback, the application aims to ensure the situational awareness (as discussed in section 3.2) of the presenter by providing real-time feedback on the actual performance. Some aspects, such as the cognitive load (as discussed in section 3.4.1) for the presenter still need to be researched in depth. An initial evaluation points to a potential overload of the user, which needs to be more balanced in order to keep up the user's flow (as discussed in section 3.4.2).

The second application example, the *Feedback Cubes*, was developed prototypically in an attempt to research and develop a balanced ambient way to provide real-time feedback. The prototypes utilise the embodied interaction principles of tangible interaction and ambient displays to support different learning scenarios. In the context of METALOGUE the prototypes will be used either to facilitate the interaction of the participants with the dialogue feedback system or to indicate real-time feedback to the dialogue partners in an ambient way. Especially the second option again tries to ensure the situational awareness of the participants, keeping the cognitive load constant and the participants in their flow.

5.1 Presentation Trainer[4]

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.

Background

Feedback is one of the most influential learning tools, thus learners' achievements both positive and negative vastly depend on it (Hattie & Timperley, 2007). The means to present feedback vary greatly and several dimensions of feedback have been identified. One of these dimensions refers to the timing of feedback, which can be delayed or immediate (Mory, 2004). Most of the studies conducted comparing both types of feedback concluded that for most learning situations the impact of immediate feedback is more positive, since delayed impact tends to delay the acquisition of needed information (Mory, 2004). A

[4] This section was based on a submission to the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2015)

challenge for immediate feedback relies on the implementation of it, as it requires personal tutors to be constantly evaluating the learner. However, the currently increasing accessibility to sensors (Swan, 2012) has led to a vast research of tutoring systems able to proportionate immediate feedback.

The technique of using sensors to track the learner's current state or behaviour in order to provide them with immediate feedback has already been used since the late 1970s. In 1978 sensor-learning support was used to treat Idiopathic Bladder Instability. The changes in bladder pressure were translated into auditory and visual stimuli, making patients aware of them (Cardozo, Abrams, Stanton & Feneley, 1978). In those early stages biofeedback has also been used for teaching people how to relax (Burish & Jenkins, 1992) and how to reduce migraine (Gauthier, Bois & Allaire, 1981). An early finding about these types of tutor systems was that feedback should be consistent and should either always be presented or not presented at all. Partial feedback just increases the confusion in learners (Morley, 1979).

Lately, the pursue of studying new automatic tracking recognition techniques using sensors has led to an exploration of different learning fields which can be supported by immediate feedback tutoring applications. The field of learning sports is one that has received vast amount of support by these applications. Research on immediate feedback applications has already been conducted in sports such as cross-country running (Vales-Alonso, López-Matencio, Gonzalez-Castaño, Navarro-Hellín, Baños-Guirao, et. al., 2010), Karate (Takahata, Shiraki, Sakane & Takebayashi, 2004), rowing (Baca & Kornfeind, 2006), snowboarding (Spelmezan & Borchers, 2008), and Taekwondo (Kwon, & Gross, 2005). In these cases the immediate feedback is presented letting learners know about their current performance during the practice of one specific technique at the time. Besides sports, immediate feedback sensor-systems have also been studied in physical rehabilitation (Brunelli, Farella, Rocchi, Dozza, Chiari & Benini, 2006), treatment of Parkinson disease (Paradiso, Morris, Benbasat, & Asmussen, 2004), and treatment of attention deficit disorder (Linden, Habib & Radojevic, 1996).

Recently, research has also been conducted in developing applications designed to support the training of nonverbal communication skills. This is the case of the MACH (My Automated Conversation Coach) software, which aims to help users on improving their nonverbal communication skills for conversations, such as job interviews, by giving feedback on their facial expressions and voice tone (Hoque & Picard, 2014). In order to further elaborate on the research of applications providing immediate feedback for learning support, specifically in the area of training nonverbal public speaking skills, we developed the Presentation Trainer.

5.1.1 Presentation Trainer Application

The Presentation Trainer was developed with the purpose to study a model for immediate feedback and instruction for public speaking. It presents feedback and instruction to the user regarding aspects of her nonverbal communication such as voice and body language.

Voice Analysis

To track the user's voice the Presentation Trainer uses the integrated microphone of the computer together with the Minim audio library[5]. By analyzing 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. Values below the silence threshold are considered as silence. Once silence is detected the pausing timer starts to tick. Whenever the pausing timer reaches a certain time the "long pause mistake" is fired. The other volume thresholds defined are the low volume and high volume threshold. In case these volume levels are reached, their corresponding timers start ticking. The voice volume modulation is calculated by subtracting the minimum volume tracked from the highest volume tracked, both of them retrieved in a predefined amount of time. Pauses are considered as a stop while speaking. When used correctly, pauses allow the audience to take a breather 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. Too short pauses do not allow the audience to digest the message, and too long pauses can leave the audience wondering when the presenter will start to speak again, creating moments of awkwardness during the speech. Therefore mastering the use of pauses is an important skill for public speaking (Bjerregaard & Compton, 2011). To help with the improvement of this skill, the presentation trainer gives feedback to the user about the proper use of pauses. Whenever the volume level captured by the microphone is below the silence threshold the speaking timer is reset and the pausing timer is started. The opposite happens when the volume level is above the silence threshold. If one of these timers ticks a longer time than their corresponding long pausing or long speaking time, then their corresponding "pausing too long" or "speaking too long" mistake is fired. The long pausing time and long speaking time can be defined manually; their default values have been obtained by adding 0.5 seconds to the average speaking or pausing time from 10 different

[5] Minim: http://code.compartmental.net/tools/minim/

analyzed Ted talks[6].

Body Language Analysis

The Presentation Trainer uses the Microsoft Kinect sensor[7] in conjunction with the OpenNI SDK[8] to track and 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, M. & 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.

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. (Toastmasters International, 2011) The current version of the Presentation Trainer does not identify specific gestures; nevertheless it gives feedback to the user whenever she is not using any gesture for a certain amount of time. In order to identify whether the user has been gesturing or not, the Presentation Trainer calculates the amount of movement of the user's hands.

Freestyle Mode

This mode offers the main functionality of the Presentation Trainer. Just by standing in front of the Microsoft Kinect and speaking, the user will start to receive immediate feedback and instruction about her nonverbal communication for public speaking. We called this the Freestyle Mode; because in this mode users are not restricted to perform different tasks.

The interface for the Freestyle Mode contains 6 modules: Posture, Movements, Voice Histogram, Enhanced Mirror, Speaking Cadence, and Voice Volume. The posture module is responsible for giving instruction about the posture, including the posture of arms, hands, legs and the body. In order to reduce the amount of possible instructions that one can get about her posture, such as: uncross your arms, straighten up, look forward, don't hide your

[6] http://www.ted.com/.
[7] Microsoft Kinect: http://www.xbox.com/en-US/kinect
[8] OpenNI: http://www.openni.org

left hand behind your body, etc.; we decided to let the system tell the user to Reset Posture. The Reset Posture is a posture used by many public speakers, where they stand straight, facing the audience, with their legs uncrossed, their hands in front of their body, above their hips, and letting the fingers of the right hand touch the fingers of the left one, as shown in Figure 5.1.

Figure 5.1 Example of Reset Posture

The Movements module provides feedback when the user has stayed still for a large period of time, and it instructs the user to use more hand gestures.

The Voice Histogram module shows with vertical bars the average voice volume that was capture at a specific moment of time. The newer values appear on the right side of the histogram and with time move to the left. This module also shows two horizontal lines, which represent the value for the low and high voice volume. The aim of this module is to visualize the voice modulation and pauses performed by the user for the last 8 seconds. When the user's pauses are between 1.5 to 2.5 seconds long, the colour of the bars fade to green, showing the user the appropriate time to start talking again. After the 2.5 seconds the bars fade into red colour indicating the user that the pause has become too long.

The Enhanced Mirror shows a reflected image of the user, and in red colour it adds an overlay of the user's limbs that at the moment are in a wrong position. It also shows in green an overlay of the user skeleton for positive reinforcement while standing on the Reset Position and pausing, or leaning in and talking at a low volume.

The Speaking Cadence module provides feedback when the user has been speaking without pauses for a long period of time, or when the user's pause has become too long. This module instructs users when they should stop or start speaking.

The Voice Volume module provides feedback when the user speaks too loud, too soft or without modulating her voice volume. The instruction given by the module depends on the mistake of the user.

For immediate feedback the use of keywords have shown to be more effective than the use ad hoc explanations (Coninx, Kreijns & Jochems, 2013), therefore the text instructions from the feedback modules have to be short, i.e. maximum two word phrases such as: Reset Posture, Move, Pause, Speak, Raise Volume, Lower Volume, and Modulate Volume.

The Presentation Trainer gives some positive reinforcement by informing the user about things performed correctly while presenting such as:

- Returning to the Reset Posture while delivering a pause.
- Delivering a pause, which is between 1.5 to 2.5 seconds long.
- Talking at a low volume while leaning in.

This positive reinforcement is represented by a distinctive sound that is played when users performed one of the previously described actions.

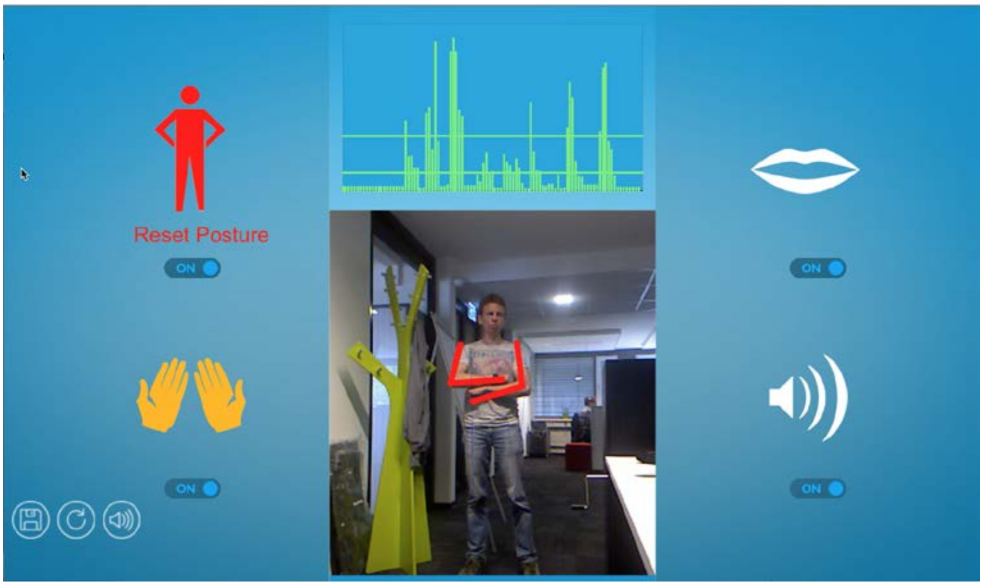


Figure 5.2 Top Left: The posture Module activated and instructing the user to Reset Posture. Bottom Left: Movement Module semi activated. Top Centre: Voice Histogram Module indicating the user that is a perfect time to start talking again. Bottom Centre: enhanced mirror module indicating that the user has his arms in a wrong position. Top Right: Speaking Cadence Module. Bottom Right: Voice Volume Module.

Exercise Mode

The four-component instructional design (4C-ID) model (Van Merriënboer, 1997) is a model design to promote complex learning, which is defined as the integration of knowledge, skills and attitudes; coordinating qualitatively different integral skills, transferring what is learned during the classroom to real life situations (Van Merriënboer & Kirschner, 2013). In order to develop complex skills, instructional design models usually divide the complex skill into sub-skills and teach these sub-skills separately. The 4C-ID encourages a holistic approach where these sub-skills are taught in a context where learners can understand the relevance of these sub-skills in the whole task. The 4 components of the 4C-ID model are:

- Learning tasks that help learners to develop an integrated knowledge base through a process of inductive learning.
- Supportive information, which is specific for a task, is always available and explains how problems in the domain should be approached.
- Procedural information, which appears just in time and explains the next step to follow.
- Task practice, which are design for the learner to acquire a high level of automaticity in routine tasks. (Van Merriënboer & Kirschner, 2013).

Since developing nonverbal public speaking skills has shown to be a complex learning task, we decided to follow the 4C-ID model in the Presentation Trainer, i.e. a Freestyle Mode to practice a variety of complete tasks and the Exercise Mode for part-task practice. This mode fractionates the task of developing nonverbal public speaking skills into different exercises or learning tasks. Each exercise is designed to train the user in a specific aspect or sub-skill of her nonverbal communication. Each exercise is presented with its explanation and relevance for public speaking, in order to present the user with supportive information in a holistic manner. Each exercise presents the user with procedural information. This information instructs her about the current step in need to be done to correctly continue with exercise. Each exercise provides the user with immediate feedback about her current performance (Figure 5.3)

The exercises developed so far are: reset posture, voice volume, hands gesticulation, pause control, leaning in while speaking soft, and questions and answers section. The *reset posture exercise* should get the user acquainted with the reset posture, a posture that allows them to be perceived open and attentive towards the audience. In the exercise the user is explained how well-trained presenters, commonly after using their hands to express their ideas, always return to the reset posture. As feedback the Presentation Trainer shows the mirrored image of the user with its overlaid tracked skeleton. Whenever the user uses the reset posture, the overlaid skeleton changes colour from blue to green and a distinctive sound is played letting the user know that the exercise was performed correctly. While standing in the reset posture, the Presentation Trainer informs the user to stay for some moments in that position before starting to move again. Once the user start moving for some moments the system informs her to return to the reset posture again.

The *voice volume exercise* intends to make the user aware of the importance of modulating the voice volume while public speaking and helps her practice speaking at different volumes. During this exercise the user has to first speak at a loud volume for 3 seconds, then speak at a low volume for 5 seconds, and finally speak using a loud and a low volume for 5 more seconds. During this exercise the immediate feedback of the Presentation Trainer shows a volume histogram to the user together with a high volume line. Values surpassing this line are considered as speaking with a high volume and values under this line are considered as speaking with soft volume. The trainer also displays the current instruction together with a timer showing the user for how much longer she still needs to speak at the instructed volume.

The *hand gesticulation exercise* aims at training the use of hand gestures that are inside of the acceptable box space, while teaching the importance of using these types of gestures. The feedback presented by the Presentation Trainer shows a mirror image of the user together with an overlaid skeleton. In this exercise targets that are inside of the acceptable box space appear in the screen and users have to gesticulate and reach them. After touching a target, they have to return to the Reset Posture in order for the next target to appear. The trainer also informs the user whether is time to reach for the target or to stay on the Reset Posture.

The *pause control exercise* has the purpose to teach and train users about the proper way to use pauses while public speaking. In this exercise the user has to speak for some seconds and then make a pause that is at least 2 seconds long. The immediate feedback in this exercise shows the volume histogram and a timer indicating the time that the user needs to wait before starting to talk again. If the user talks in between the 2 seconds of the pause, the pausing timer restarts again. Once the 2 seconds pause are over, a distinctive sound is played and the procedural instruction displayed in the screen changes from Pause to Speak.

In the previous exercises the purpose was to either train the body language or the voice. In the *leaning in while speaking soft exercise* we explore the combination of training both aspects at the same time. The technique of leaning in and speaking soft can be used in public speaking to help you to connect at a personal level with the audience while sharing some secrets or personal opinions (Devito, 2014). So in this exercise we wanted users to learn the importance of this technique while practicing it. The feedback shown in this exercise shows the volume histogram and the overlaid skeleton. When the user leans in the colour of the skeleton turns from blue to green. Once the user speaks in this position at a low volume (histogram values below the high volume line) for a couple of moments, a distinctive sound is played indicating that the exercise was performed correctly. After performing the exercise correctly the Presentation Trainer instructs the user to return to the Reset Position in order to repeat the exercise.

The purpose of the *questions exercise* is to recapitulate the lessons from the previous exercise in a question session fashion for public speaking. During this exercise the user has to stand in the Reset Posture waiting for the Presentation Trainer to ask a question related to the nonverbal skills learned. An example question is: “What are the benefits of pauses while public speaking?” Once the user answered the question, she has to stay quietly in the reset position waiting for the next one. The immediate feedback in this exercise shows the mirror image with the skeleton overlay and the volume histogram. The instructions displayed are “wait in the reset position”, and “answer the question”.

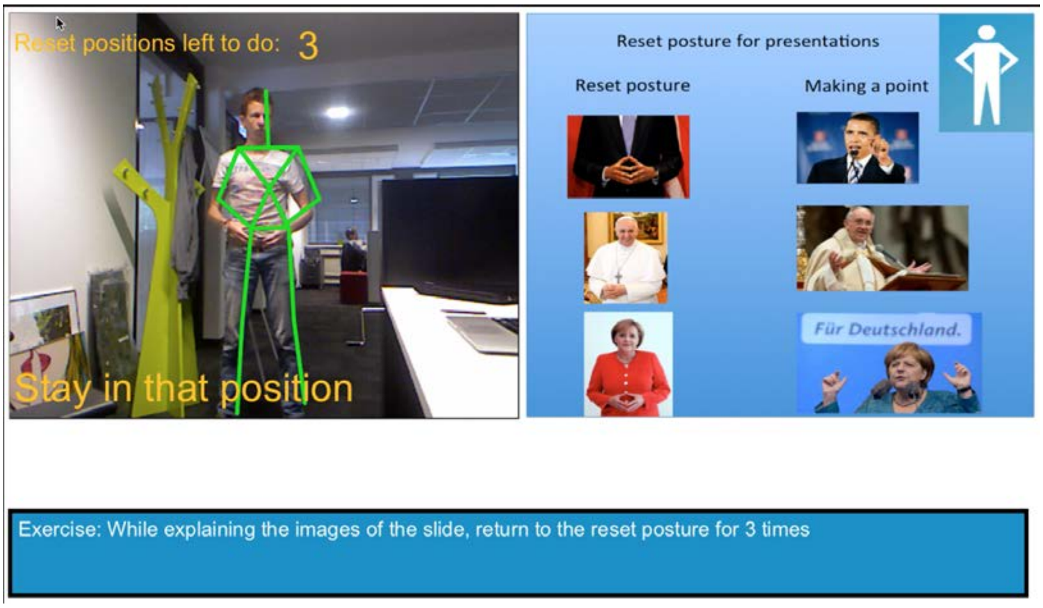


Figure 5.3 Interface of Reset Posture exercise. Top Left: Enhanced mirror showing a green skeleton overlay, because the user is on Reset Posture. Top Right: Sample slide to use for the exercise. Bottom: Exercise instruction. The yellow text shows the instructions for current state of the exercise.

System Architecture

The Presentation Trainer was developed in Processing 2.1, an open source JAVA-based programming language. It has an OpenGL integration that allows fast graphic manipulation making it suitable for 2D and 3D programs.

The software architecture of the Presentation Trainer (Figure 5.4) has 4 main components: the Sensor Objects, the Feedback Objects, the Exercise Objects, and the Main Controller. The Sensor Objects are all derived from the SensorObject class. Each instance of these objects is bounded to a specific sensor, for example the microphone of the computer. The function of these objects is to retrieve the sensor data, and analyse this according to the predefined rules. The result of this analysis is stored as the current state of the object.

The Feedback Objects are derived from the FeedbackObject class. Each of these objects is linked to a set of Sensor Objects. By retrieving the current state of their linked Sensor Objects, the Feedback objects are able to give feedback and instruction about the current performance of the user. This feedback and instruction can be in the form of playing sounds, displaying images or text on the screen, or sending messages to other actuators.

For the inclusion of the Exercise Mode we created the Exercise Objects, which derive from the ExerciseObject class, which in turn derives from the FeedbackClass. Exercise Objects use the elements of the FeedbackClass to track and to give immediate feedback about the user's performance, for example informing the user whether she is speaking loud enough, standing in the correct position, etc. Each Exercise Object contains different states. For example, the hand gesticulation exercise has two states: reaching for target and waiting state. During the reaching target state, the system presents feedback on how the user is reaching for the target, and during the waiting state the system analyses whether the user stays in the Reset Posture. Each Exercise Object also contains: an explanation of the exercise including its relevance in public speaking; procedural instructions which are shown according to the state of the exercise; and a slide showing sample information for the user to present while doing the exercise.

The Main Controller is responsible for the instantiation of each of the Sensor Feedback and Exercise Objects. It also has control over the application loop, which executes the information retrieval and analysis of the Sensor Objects, the output produced by the Feedback Objects in the case for the Freestyle mode, or the current Exercise when running on the Exercise Mode. The controller is also responsible for logging the states of the Sensor and Feedback Objects for a posterior analysis of the user's performance.

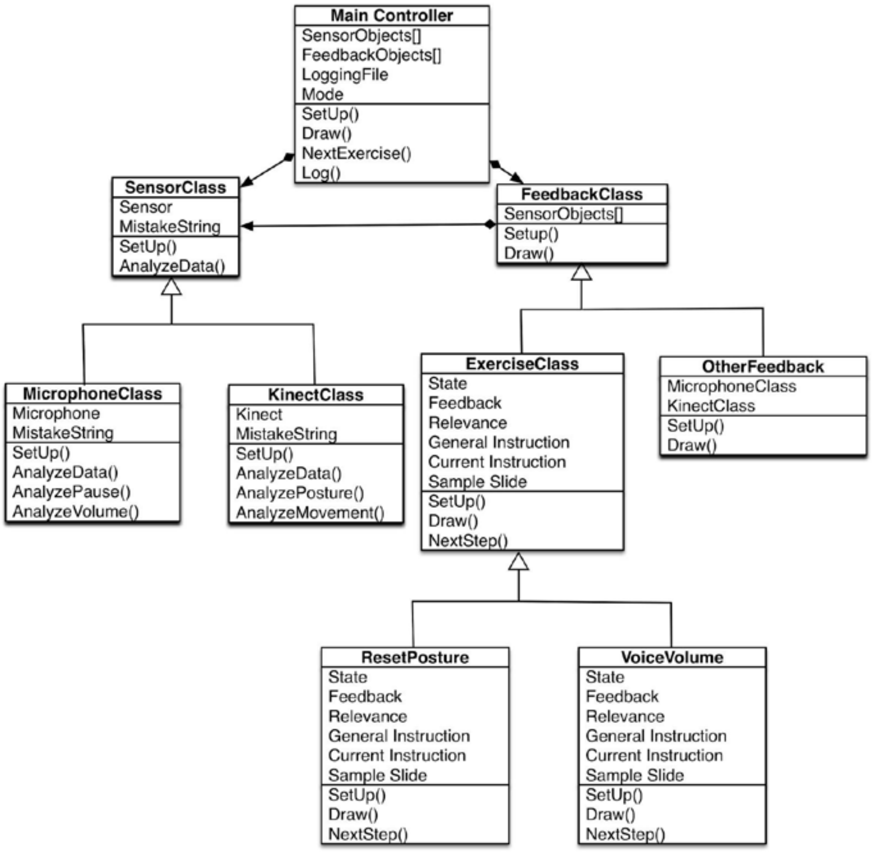


Figure 5.4 Class diagram of the Presentation Trainer

5.1.2 First User Study

The purpose of this first study was to explore the users' acceptance of the *Presentation Trainer* and to identify the first educational challenges towards building an immediate feedback tool designed to support users with the development of their public speaking skills. This section includes an explanation of the first version of the *Presentation Trainer's* output interface, a description of the setup used for our first user study, and a report of our findings.

Preparation

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.

User Test

The test consisted on giving a short presentation while using the Presentation Trainer as an immediate feedback training tool. The experimental setup sketched in figure 5.5 shows the participant standing 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, 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.

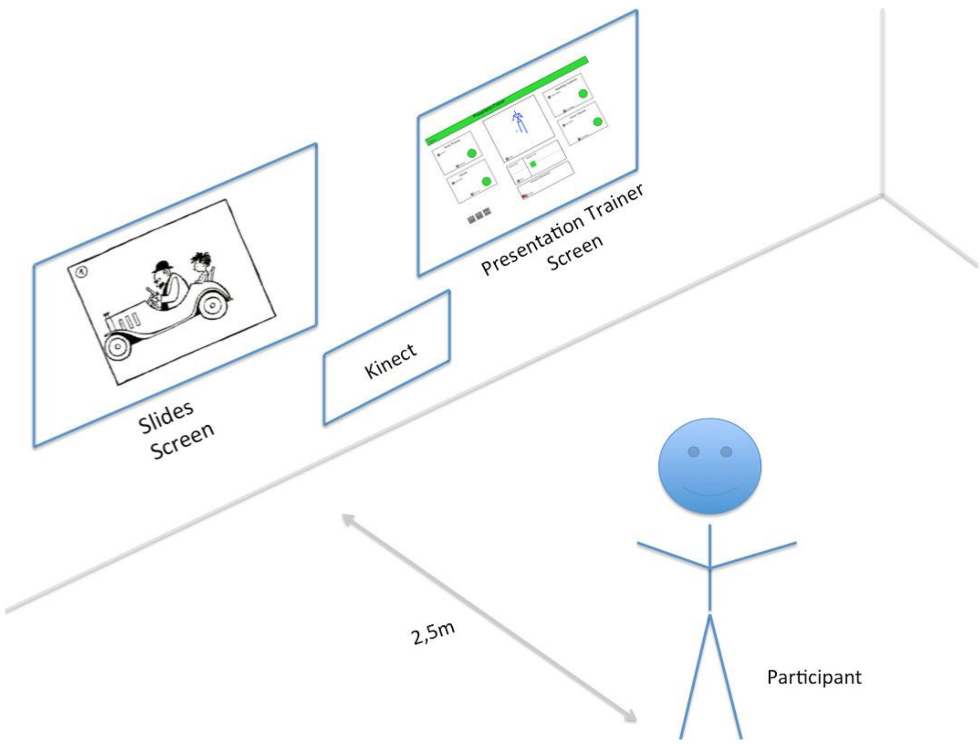


Figure 5.5 The setup sketch of the first user test

Results

In total we had 6 participants for this part of the study. The background of the participants was either in learning or computer sciences. The average scores for the SUS questionnaire were: 67.5 for SUS, 65.1 for usability, and 77.1 for learnability.

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 be aware of all the feedback at the same time. Therefore it was suggested to use a learning strategy focusing on giving feedback on one aspect of the trained skills only at the time. During the public demonstration of the Presentation Trainer, most participants showed scepticism about the immediate feedback. Thus, they suggested to use the tracking capabilities of the Presentation Trainer to show the users' performance and mistakes after the presentation, with the purpose of supporting their learning more effectively and allowing them to reflect about their performance. Nevertheless, after using the tool they all stated their enthusiasm towards the immediate feedback.

Evaluation of first results

Participants in the user tests showed great enthusiasm towards the Presentation Trainer. The remarks about the immediate feedback received were positive and participants liked the idea of using a similar tool to train for their presentations. However, observations executed during the user tests showed that the purpose of the Presentation Trainer has only been partially accomplished. Participants did not always adapt their behaviour, even when the Presentation Trainer was suggesting them to do so. We attribute this lack of response from the participants towards the feedback given by the trainer, mainly due to the amount of cognitive load. Not being prepared for giving a presentation, regardless of its simplicity proved to be a fairly complex task, consuming 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. Semaphores captivated the users' attention but its information was not enough to let them know how to adapt their behaviour. Finally participants did not perceive the instructional text.

5.1.3 Second User Study

In this study we wanted to explore the users' impressions and interactions towards the new features of the Presentation Trainer. These new features include the implementation of our approach to tackle the challenges arisen from the first study and the inclusion of positive reinforcement on the interface. This section describes the new features of the

Presentation Trainer and includes the description and results of our second one. Prior to the test participants were asked to prepare a short introductory presentation. For that we sent the participants a template of six slides including topics such as origin, profession, personal hobbies and movies that they find interesting. Once the participant arrived to the test, we asked her to sign up a form of consent that explaining that the recordings done during the test were going to be used for study purposes. After that we gave her a briefing about the purpose of the Presentation Trainer and explained her tasks to do during the test.

The user tests consisted of two phases, the Exercises phase and the Presentation phase. During the Exercises phase, the participant used the Exercise Mode of the Presentation Trainer. For this the participant stood at approximately 2.5m in front of the Microsoft Kinect sensor and a 27-inch display displaying the Presentation Trainer interface. During this phase of the test, the Presentation Trainer was running on the Exercise mode and asked the participant to perform 6 different exercises: reset posture, voice volume training, hand gesticulation, pause control, leaning in while speaking soft, and questions. On the Presentation phase, the participant stood 2.5m in front of the Microsoft Kinect sensor, a 27-inch display showing the Presentation Trainer interface on Freestyle Mode and another screen showing her prepared introductory presentation. For this phase the participant had 6 minutes to give her presentation while receiving feedback from the freestyle mode of the presentation trainer.

After finishing with the two phases of the test the participant was asked to fill in a SUS questionnaire. After filling in the questionnaire the examiner conducted an interview with her. During the interview the examiner asked the participant about her impressions of the Presentation Trainer; her opinions about the Exercise Mode; her opinions about the use positive reinforcement together with corrective feedback in contrast to showing only corrective feedback; the added value that the tool brings in contrast of preparing for a presentation in front of a mirror without any tool giving feedback; finally, the examiner also asked about additional comments and suggestions. The tests and the interviews were recorded in order to allow a proper analysis of the results.

Results

For this study we had 5 participants, whose background was from either learning or computer science. The results of the SUS questionnaire gave the Presentation Trainer an average score of 57.0 in the SUS, 59.4 in usability and 47.5 in learnability.

All participants during the interviews stated that the load of information for giving a presentation and using the system for the first time is too much to handle, therefore they all indicated that would need some time to get used to the tool before being able to assimilate the feedback and instructions provided by it. All participants liked the idea of having positive reinforcement; nevertheless 3 of them were not able to make sense of the sounds played when this positive reinforcement was given. All participants pointed out that they found the Presentation Trainer as a useful tool and that they would like to use it in order to prepare for their upcoming presentations, remarking the significance of receiving objective immediate feedback while preparing for them. With regard to the Exercise Mode, all participants found this mode necessary to develop their nonverbal skills. Still they remarked on the necessity on making a tutorial for each of the exercises. Two participants suggested on making the feedback more explicit feedback while performing the exercise, indicating whether the exercise is being performed correctly or not.

During this user study, while observing the participants it was possible to identify that the feedback provided by the Presentation Trainer was having some effect on the user's behaviour. However, this change of behaviour was not always the desired one. For example when the Move More caption was presented, users tend to shook their body a little or took one or two small steps to the sides, instead of gesticulating more with their arms while speaking. It was also observed that all participants mastered the Reset Posture by using it throughout their introductory presentation, providing us an indicator that some learning might have happened while using the system. While the Presentation Trainer proved to be robust enough during the tests, there are still some technical problems that once tackled will improve the user experience of the tool. The leaning in position is to similar as the hunchback position and it can give mixed signals to the user: on the one hand with positive reinforcement indicating that they are about to use a predefined public speaking strategy, and on the other hand instructing the user to stand straight.

5.1.4 Discussion and Conclusion

The scores obtained in this second study were considerably less high than in the previous study, leading us to think that the older version of the trainer was indeed more usable and learnable than the new version. However, in contrast with the previous test where participants did not change their behaviour while receiving feedback from the trainer, in this second test using the new version they did modify their behaviour when feedback and instruction was given to them. Also we found it very encouraging observing participants using the Reset Posture, leading us to infer that some learning took place while using the Presentation Trainer. These observations indicated that the trainer became more usable and learnable even when users perceived the opposite. We attribute this change in perception of the participants towards the usability of the tool to different factors. The first one is that with the new version participants were able to identify their mistakes with the feedback given by the trainer but did not know how to correct them at the moment, leading them to realize that they need some time to learn how to correctly use the system. The second factor deals with the expectations of the participants towards the Presentation Trainer. We explain these increasing expectations to the fact that in this second version the interface the tested tool no longer looks like a prototype anymore, leading participants to think that it should work as a commercial system and not as an experimental prototype.

The high cognitive load is still an important issue that needs to be tackled. Observing that users did perceive the feedback and try to adapt to it was very encouraging, nevertheless users still need to learn in real time how to change their behaviour adding a new level of complexity to the task. In order to reduce this complexity we plan to add a learning module to the Presentation Trainer for explaining users the meaning of the feedback and how to react to it before they start their training. With this module we will try to reduce the cognitive load and misleading interpretations of feedback of the user.

The Exercise Mode looks really promising; there is big room for improvement. Feedback showing whether the exercise is being performed correctly or incorrectly should become more explicit. Each exercise needs a tutorial on how to perform it because screen instructions are not explicit enough. Exercises need to be well designed so that users are able to integrate all the skills learned during a presentation. For example, as shown during the tests, participants learned how to gesticulate, but not how to talk and gesticulate at the same time. This was reflected during their presentations where none of them talked and used hand gesticulations at the same time.

Conclusion

The rising availability of sensors has created the space to design, develop and explore tools able to provide users with immediate feedback on their performance. This work presents the findings of our ongoing studies on the Presentation Trainer, a novel tool that tracks and analyses the body posture and voice volume of users in order to provide them with feedback and instruction with the purpose to train and develop their nonverbal public speaking skills. Reducing the amount of cognitive load required from the users is still an issue, that needs to be solved. Besides the users' positive views towards the Presentation Trainer, our studies have also shown that users were able to receive its feedback while presenting; furthermore the first indicators that learning took place while using the tool were also revealed.

5.2 Feedback Cubes[9]

Following similar principles, i.e. "*tangible bits*" by Ishii and Ullmer (1997) and "*ambient displays*" by Wisneski et al. (1998), both paved the way for embodied interaction. Tangible interfaces are considered as more natural and intuitive than traditional types of interfaces (Dourish, 2001). Ambient systems are subtle and non-intrusive means for interfacing people with peripheral information. With all these characteristics in mind, both concepts offer great potential to support learning scenarios in various ways.

Also related research work in this domain supports that. A review on tangibles for learning by O'Malley and Fraser (2004) concluded that tangibles bring physical activity and active manipulation to the forefront of learning, i.e. they reduce the learners' cognitive load in order to enable learners to allocate the resources relevant for the task. With similar conclusions Börner et al. (2013) just recently reviewed ambient displays for learning and highlighted their potential to support learning implicitly by raising, enhancing, or supporting awareness, changing behaviour, giving feedback, providing assistance and guidance, or just by

[9] This work was based on a work-in-progress paper at the 9th International conference on Tangible, Embedded and Embodied Interaction (TEI 2015)

presenting information.

The presented research and development in this paper aims to utilize both embodied interaction principles to support learning scenarios. With the underlying idea to combine both approaches in a flexible and easy-to-use system, the main focus lies on the exploration and formative evaluation of a respective prototypical system design called the *Feedback Cube*.

5.2.1 Prototypical System Design

To study the support of learning scenarios with tangible interfaces and ambient displays, respective prototypes were developed. The design process followed a system design approach, i.e. putting the actual system in the centre of the process and arranging a set of components to create the desired design solution (Saffer, 2007). For the envisioned system, the ability to detect motion, provide visual and auditive cues, and communicate wirelessly were considered as most important. Based on these criteria, the components were chosen and assembled. Figure 5.6 shows all parts and materials used as well as an assembled prototype.

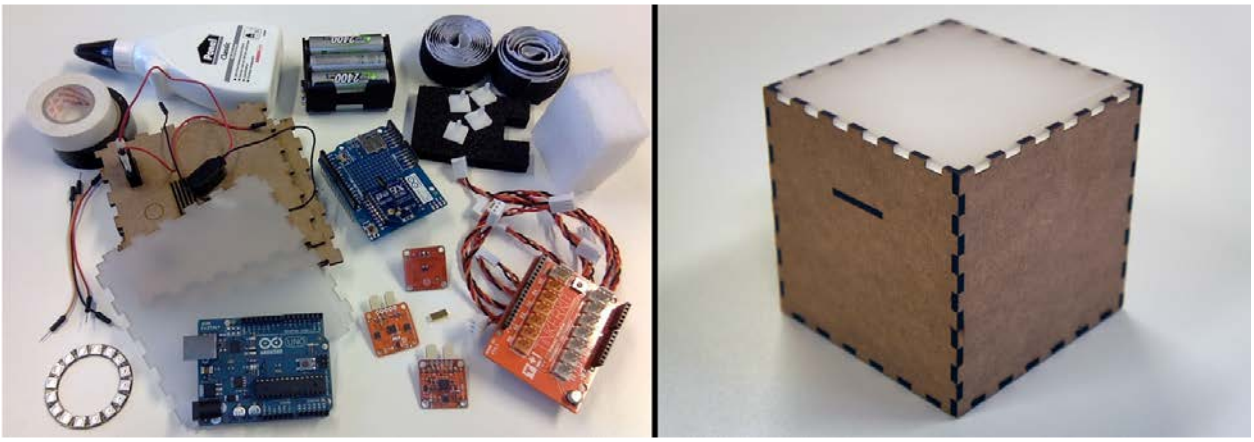


Figure 5.6 Hardware parts and additional material used (left) to assemble Feedback Cube prototype (right)

Hardware Platform

For the prototypical system design a cubic shape has been chosen. As solid three-dimensional objects, cubes represent familiar physical structures that can be utilized for tangible manipulation, spatial interaction, or expressive representation as characterized in Horneker and Buur's framework of tangible interaction (2006).

The exterior of the cube prototypes was made from high-density fibreboard and semi-transparent Plexiglas, whereas five sides of the cube are opaque and only the top is semi-transparent. The interior comprises a set of various sensor, actuator, and communication components as well as the necessary hardware to operate them. The cubes have an edge length of 100mm, so that all components fit in, while still ensuring a reasonable size for tangible interaction. The hardware operating the prototypes is based on the open-source electronics platform Arduino (http://www.arduino.cc). The main components are an Arduino

Uno microcontroller board, an Arduino WiFi/Wireless SD shield, and a TinkerKit (http://www.tinkerkit.com) Sensor shield. The wireless shield enables the microcontroller to connect and communicate wirelessly either via wireless or mesh networks. The integrated micro-SD card slot can be used to read and store files. The sensor shield provides an easy-to-use hub to connect sensor and actuator components directly to the microcontroller. All hardware components can be powered either by a built-in rechargeable battery or tethered via the integrated USB interface.

Sensors and Actuators

The hardware platform is enhanced with various sensor and actuator components. The sensors are used to detect changes in the environment, providing input to the prototypical system. The actuators are used to act upon the detected changes, providing a system output.

As sensor components the prototypes include a TinkerKit Accelerometer, Gyroscope, and Hall sensor. The three-axis accelerometer measures acceleration and can be used to detect movement. The two-axis gyroscope measures orientation and can be used to detect movement and rotation. The measured output of the accelerometer and the gyroscope were combined to emulate an inertial measurement unit. This allows more accurate measurement of the prototypes' inclination relative to the ground. Finally the hall sensor measures changes in the surrounding magnetic field, which can be used for instance to calculate the distance to a nearby magnet.

As visual actuator the prototypes include an Adafruit (http://www.adafruit.com) NeoPixel ring mounted below the semi-transparent side of the cube. The ring consists of 16 RGB LEDs that can be individually addressed via their built-in microcontroller. Using the available Arduino library the colour and brightness of each LED can be controlled. Furthermore the prototypes include a 12-Watt mini speaker that can be used as auditive actuator.

5.2.2 System Characteristics

The specified form factor, used hardware platform, and chosen sensor and actuator components enable different possibilities to use the *Feedback Cubes*. In general the prototypes can either facilitate some kind of interaction with users or objects (interaction facilitator), indicate feedback information to users or the immediate surrounding (feedback indicator), or do both at the same time.

Interaction Facilitator

The interaction facilitator concept as illustrated in Figure 5.7 is characterized by the prototypes' sensor components. The used accelerometer and gyroscope are able to detect movement on the x-, y-, and z-axis and rotation on the x-, and y-axis (five degrees of freedom). More specifically moving the prototype along the axes, i.e. forward, backward, left, right, up, and down, as well as rotating the prototype along the axes, i.e. rolling and tilting, can be detected. The used hall sensor is able to detect changes in the surrounding magnetic field and thus the presence of magnetic objects, e.g. other prototypes can be detected.

Other means of facilitating interaction are the communication components. The prototypes have built-in serial communication facilities that can be used, e.g. via the available USB-to-serial converter. Furthermore the used WiFi shield enables the prototypes to take client and/or server roles for communication within wireless networks. Alternatively the used wireless shield enables the prototypes to support point-to-point, point-to-multipoint, and peer-to-peer mesh network topologies with other prototypes.

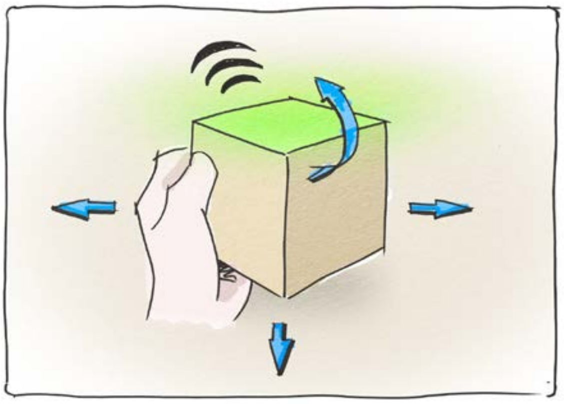


Figure 5.7 Interaction facilitator concept

Feedback Indicator

The feedback indicator concept as illustrated in Figure 5.8 is characterized by the prototypes' actuator components. The used LEDs are capable of displaying the full RGB colour space with 16777216 colours at 256 brightness levels. All 16 RGB LEDs on the ring can be controlled individually, which allows programming various visual patterns and effects, such as fading, blinking, or colour transitions.

The used mini speaker can produce sounds in response to the electrical signal input delivered by the microcontroller. Programmatically manipulating the signal input allows creating various audio patterns and effects, such as playing single tones, complex melodies, or even encoded audio files. When using the integrated storage capabilities, it is also possible to create an accessible music and sound effect library.

5.2.3 Formative Study

To explore usability issues of the *Feedback Cubes* a formative study was conducted with a group of 8 participants. Therefore the general idea and the basic functionality of the prototypes were briefly introduced to the group. In a first round the characteristics of the interaction facilitator concept were highlighted and each participant had the chance to examine the prototypes and test the respective functionality. The participants were then asked to fill in an all-positive version of the System Usability Scale (SUS) (Sauro, 2011), focusing their ratings solely on the interaction facilitator concept. In a second round the procedure was repeated once again for the feedback indicator concept.

The interaction facilitator concept received a mean score of 71.9 (SD = 10.3). The feedback indicator concept received a mean score of 69.1 (SD = 15.8). When comparing both scores to other hardware systems (Bangor, 2009), the interaction facilitator concept scored higher than 52.3% of the other systems with a “C” grade at an acceptable level, which can be described as “Good”. The feedback indicator concept scored higher than 41.6% of the other systems with a “D” grade at a marginal level, which can be described as “Ok”.

The results show above average ratings for both concepts with room for further improvements. The interaction facilitator concept scored higher than the feedback indicator concept. The participants stated that the tangible interaction is much more intuitive, while the ambient display principle requires an additional mapping to make sense of the provided information. It can be assumed that this changes once the mapping is clearly defined.

After evaluating the perceived usability of the single concepts, the whole group was asked to capture their general impressions about the prototypes using a modified electronic version of the Product Reaction Cards originally developed by Benedek and Miner (2002). The group was asked to agree on 6 cards with words that described their experience with the prototypes best and comment on their selection.

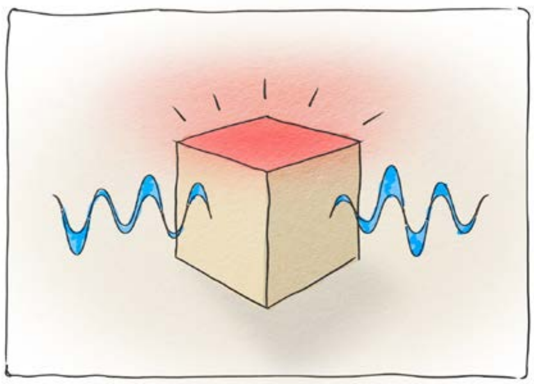


Figure 5.8 Feedback indicator concept

The group agreed on the following selection of words: *engaging*, *straight forward*, *customizable*, *responsive*, *fragile*, and *familiar*. The group commented that the prototypes are *familiar* in a sense that form and function are evident without creating additional obstacles or distraction. However, the prototypes’ design was also characterized as *fragile* and several improvements were suggested to make it more robust. Besides that, the prototypes were characterized as *engaging* due to the fact that specific user interactions and reactions are encouraged. The interactive and especially the indicator functions were characterized as *straight forward* and *responsive* with the potential to implement feedback and direct interaction mechanisms. The group also commented that the prototypes are *customizable* in a sense that various individual and collaborative scenarios could be supported.

5.2.4 Discussion and Conclusions

Following the formative study the participants were asked to think about and discuss specific learning scenarios that could be supported. This discussion, the given system characteristics, as well as the formative results, helped to outline the following application scenarios. Based on the introduced interaction facilitator concept the *Feedback Cubes* could for instance support memorization tasks by enabling users to easily relate visual and spatial information. Under the assumption that using the tangible interactive prototypes facilitates retention, the prototypes would augment the task. The users would receive visual instructions for a randomized sequence of moves that they have to repeat with the prototypes. After each correct repetition one more random move could be added to the sequence. In a collaborative group setting the interaction capabilities of the prototypes could also be used to moderate an ongoing discussion session, e.g. if a common agreement is needed at the end of the session. Each opposing party would receive a *Feedback Cube* and could confirm or decline arguments by tilting the prototype left and right. Whenever the parties agree on the same argument the prototypes should be moved closer to each other, reaching the final agreement when both prototypes touch each other.

Other application scenarios based on the feedback indicator concept could support the users' individual or group performance by increasing the awareness on certain indicators. The assumption would be that this increased awareness triggers reflection and eventually provokes users to adapt their behaviour accordingly. In this context the ambient display functionality of the prototypes would be used to provide this feedback. In a collaborative setting with several groups, each group could receive one *Feedback Cube* that indicates for instance the (externally measured) general speech volume within the group. Whenever the volume gets too loud (and thus prevents the exchange of arguments in the discussion) the prototype makes the group aware of that through visual or auditive feedback. In a similar manner the prototypes could also be used as personal peripheral displays. For instance in combination with a time/task management application the *Feedback Cube* could indicate the timing or completion of certain tasks. In combination with an activity tracking application the prototype could provide an overview of individual performance parameters or patterns.

In conclusion the presented system design of tangible interactive ambient display prototypes allows utilizing embodied interaction principles to support learning scenarios. A formative study underpinned the prototypes' potential to facilitate interaction and/or indicate feedback. Especially the interactive capabilities were considered as functional and encouraging, while several improvements were suggested. Finally the gathered insights informed application scenarios of the *Feedback Cubes* in a learning context, which will be implemented as proof of concept for further evaluation in a next step.

6. Instructional Design Blueprint

Following the theoretical background discussed in chapter 3, the data and feedback categories considered in chapter 4 and the initial experiments described in chapter 5, in this chapter we will outline the instructional design blueprint for METALOGUE following the 4C-ID model introduced in section 3.1. In this section we will first describe the skills hierarchy connected to the task at hand and the task complexity aspects divided over three levels. In the next section we will explain how the design aligns with the METALOGUE incremental development. In the final section, will outline for each of the task classes proposed a set of task varying in difficulty and perspective, the required supportive information and, finally, we will discuss how we plan to derive assessment criteria for the tasks to be practiced.

Conducting a debate[10], i.e. a presentation including an argumentation, is a complex task. The skill to be mastered is in brief “convincingly present, argue and respond about a current hot issue”. For this, a trainee needs to have knowledge and skills about both argument content and structure aspects (e.g. what to present, how to use and structure their arguments, how to rebut, what and how to close the argument) and delivery aspects (e.g. how to use their voice e.g. pitch, speed or volume, body etc). On top of this, the trainee has to be continuously aware of the effects of their debating inputs and guard their goals by monitoring the level of agreement, not only content wise but also how they and their opponents respond and reflect and adapt accordingly when necessary.

The skill (and its associated knowledge and/or attitude) required to perform this task adequately can be divided in four skills (figure 6.1):

- ‘Setting argument content’: search, select and phrase the relevant content;
- ‘Planning argument organisation’: organise content, arguments, counter-arguments and objections;
- ‘Applying argument delivery’: present the content taking into account delivery aspects;
- ‘Setting goals’: set and guard the desired target with regard to the aim of the dialogue (e.g. pass a proposal with as few changes possible) and the ability of the learner to anticipate on their ‘opponent’ and adapt accordingly to achieve their goal ‘at best’.

[10] *For the Call Centre scenario a similar instructional design will be developed.*

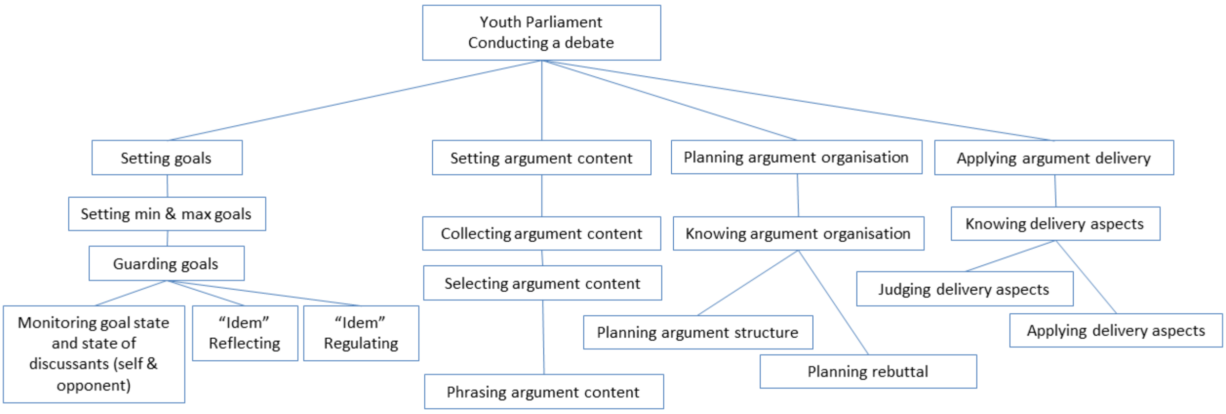


Figure 6.1 Skills hierarchy Youth Parliament "conducting a debate"

Given the complexity of debating, 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. For a trainee focussing on the arguments to be used 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 (c.f. the levels in table 6.1). 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 more complex. At the final level, the trainee will mainly receive integrated feedback within a realistic debating context. Only if necessary, the trainee should zoom into the constituting aspects of the feedback.

Table 6.1 Task complexity aspects: type of topic and types of aspects to be mastered. In *italic* aspects on which in-action feedback will be given.

Task Complexity / level		**Level 1**	**Level 2**	**Level 3**
Context	**Topic**	Simple e.g. present yourself and discuss your interest (to get to know the system) or a position statement with e.g. just one argument exchange	Full topic. Limited number of arguments or argument exchanges	Full topic. Number of arguments or argument exchanges depend on the participants (or max 10 minutes)
	Opposition	Agreeable opponent	Agreeable & disagreeable opponent	Agreeable & disagreeable opponent
	Length	3-4 minutes	5 minutes	5-10 minutes
Set & guard goals		Indicator: - overall dialogue performance (based on the available data)	Indicator: - overall dialogue performance (based on the available data)	Indicator: - overall dialogue performance - target achievement
Contents and organisation		-	Indicator/visualisation Argument use	Indicator/visualisation Argument – Reason – Evidence use
Delivery	**voice**	*Voice volume*	*Voice volume* *Speaking cadence*	+ Overall Indicator/visualisation voice aspects
	body language	*Confident posture*	*Confident posture* *Hands & arms usage*	+ Overall Indicator/visualisation body language aspects
	other	Relative speaking time; Relative turn time	Relative speaking time; Relative turn time DA: Communicative behaviour: Politeness	Indicator/visualisation AECAL Relative speaking time; Relative turn time
Emotion		-	Indicator/visualisation One Emotion – Response pair	Indicator/visualisation selected Emotions – Response pairs

6.1 METALOGUE development alignment

The METALOGUE system will be delivered in 3 rounds: an initial tutoring pilot, a second pilot and the final dialogue system. The instructional design outlined below aims at aligning 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. Final choices and details of the design will be added as the system and its design develops. For the initial tutoring pilot it will imply that the learner will be able to sample two types of in-action feedback i.e. feedback on one aspect of their voice quality (e.g. 'voice volume') and one aspect of their body language (e.g. 'confident posture'); and at least one type of about-action feedback (c.f. also D3.2), i.e. a time line overview of the selected voice aspect and one or two overall performance indicators (i.e. a score relative to other learners or relative to the dialogue) e.g. relative total speaking time or length of turn overview. As a result the learner can start learning with the system and will have an overview of the system's functionality straight from the beginning and subsequently can be questioned about strength and weakness or asked for suggestions (c.f. also D6.1). Table 6.1 gives an indicative overview how the METALOGUE development will align with the instructional design. It describes the type of topic, from simple to complex, and it indicates the METALOGUE feedback available i.e. indicating the type and amount of debating aspects to be mastered at a given level. Learners are expected to be sufficiently fluent at a 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 (c.f. also section 4.2).

6.2 Instructional Design

Based on the task complexity aspects discussed above the design below outlines three task classes with each a number of tasks, supportive information and how the criteria will be developed. Adaptation will be possible by adapting the sequence and amount of tasks based on the performance of the learner. *The details of the tasks will be decided upon in close collaboration with the pilot sites as the system develops*. 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.

Task Class Level 1

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.

Task 1a. Observe an expert debate video of approximately 3 minutes.

The video is shown “annotated” with the in-action feedback aspects and concludes with a tour of the about-action feedback.

The task closes with a reflection (together with a tutor[11]) on the criteria and feedback examined and the impact, if any, observed on the debate.

Task 1b. Observe and assess a video of a ‘standard’ debate of approximately 3 minutes.

A video is shown to the trainee of a video of debate. The trainee should observe and assess the initiating debater. The observing should result in an assessment of good and bad performance based on the aspects as defined in task 1a. The in-action assessment is done on paper with the help of a scoring form or with the help of pre-defined interface with e.g. buttons.

The task closes with a reflection on the aspects assessed in comparison to the system’s assessment and the impact, if any, observed on the debate.

Task 1c. Prepare and present yourself and discuss one interest

The trainee is asked to prepare and perform a “present yourself and discuss one interest”. The presentations are about 1 minute each. The discussion should be approximately 2 minutes. The trainee receives in-action and about-action feedback. The system will show the selected in-action feedback task while the task is performed.

The task closes with a reflection on the aspects assessed in comparison to the system’s assessment and the impact, if any, observed on the debate.

Task 1d. Prepare and present your position on the topic "ban smoking" and debate

The trainee is asked to prepare and perform a debate on "ban smoking" (in favour) i.e. prepare a position statement and (counter) arguments for e.g. one exchange of arguments. The position statements should be about 1 minute each. The exchange should be approximately 2 minutes, the opponent will give a mild opposition. The trainee receives in-action and about-action feedback. The system will show the selected in-action feedback while the task is performed.

The task closes with a reflection on the aspects assessed in comparison to the system’s assessment and the impact, if any, observed on the debate.

Possible variations: Task 1a and 1b can be repeated with similar videos. Task 1c can be repeated with different topics. Task 1d can be repeated with a change of role between pro-contra, give a prepared position and arguments instead of having the trainee prepare; or by asking for a set of arguments with pre-specified constraints.

Supportive information. An introduction (or links to relevant resources) of the preparation of a debate, the structure of a debate and the delivery of a debate. 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 overview of the METALOGUE system, what to expect, how it operates and its interface.

[11] *Depending of the stage of development the tutor or opponent can be artificial or human.*

Task Class Level 2

In the second task class the set of aspects to be trained upon will be expanded and the debate task will become more complex. The trainee is expected to generally know the METALOGUE system and the general principles of debating. The aspects to be trained upon will be introduced and explained one by one (c.f. table 6.1). The trainee will both have to be able to debate with a relatively agreeable and a strongly disagreeable opponent. The trainees will start with familiarising themselves with the aspects required by observing and assessing a video of a debate.

Task 2a. Observe and assess a video of a debate of 3 - 5 minutes.

A video is shown to the trainee of a video of debate with a strongly opposing opponent. The trainee should observe and assess the aspects as explained in the introduction (see supportive information). The in-action assessment is done on paper with the help of a scoring form or with the help of pre-defined interface with e.g. buttons. The trainee will report their about-action feedback with the help of a template.

The task closes with a reflection on the aspects assessed in comparison to the system's assessment and the impact, if any, observed on the debate.

Task 2b. Prepare and present your position on the topic "smoking in public places" and debate

The trainee is asked to prepare and perform a debate on "smoking in public places " (in favour). For this the trainee will *receive* a draft with a position statement and a set of (counter) arguments. The position statements should be about 1 minute each. The exchange should be at least three rounds. The opponent will give a mild opposition. The trainee receives in-action and about-action feedback. The system will show the selected in-action feedback while the task is performed.

The task closes with a reflection on the aspects assessed in comparison to the system's assessment and the impact, if any, observed on the debate.

Task 2c. Prepare and present your position on the topic "ban smoking" and debate

The trainee is asked to prepare and perform a debate on "ban smoking" (against) i.e. prepare their own position statement and (counter) arguments. The position statements should be about 1 minute each. The exchange should be at least three rounds. The opponent will give strong opposition. The trainee receives in-action and about-action feedback. The system will show the selected in-action feedback task while the task is performed.

The task closes with a reflection on the aspects assessed in comparison to the system's assessment and the impact, if any, observed on the debate.

Task 2 Part task practice. Observe and exercise standard voice and posture practice.

The part task practice task allows the trainee to practice a pre-defined set of dialogue moves with regard to voice and posture either as a specific move or in response to a given situation, if the trainee has difficulty to adequately perform with regard to these aspects (see section 5.1, figure 5.3 for an example).

Possible variations: Tasks can be repeated with stressing different feedback attention points and/or with different topics.

Supportive information An introduction (or links to relevant resources) of the preparation of a debate, the structure of a debate and the delivery of a debate. Special attention is given to the aspects which are introduced at this level and should be mastered. Additionally, the trainee will get an overview of the elements of the METALOGUE system added at level 2, what to expect, how it operates and its interface.

Task Class Level 3

The focus at this level is to monitor and adjust to the flow of the debate i.e. to the opponent and to the progress with regard to one's goals. At the final level, the trainee will receive integrated feedback within a realistic debating context. If necessary, the trainee can zoom into the constituting aspects of the feedback.

Task 3a. Prepare and present your position on the topic "smoking regulation and youth" and debate

The trainee is asked to prepare and perform a debate on "smoking regulation and youth" (pro) i.e. prepare their own position statement and (counter) arguments. The position statements should be about 1 minute each. The completion of the argument exchange is controlled by the debaters themselves (or max 10 minutes i.e. time expired). The opposition is mild. The trainee receives in-action and about-action feedback. The system will show the selected in-action feedback task while the task is performed.

The task closes with a reflection on the aspects assessed in comparison to the system's assessment and the impact, if any, observed on the debate.

Task 3b. Prepare and present your position on the topic "ban smoking" and debate

The trainee is asked to prepare and perform a debate on "ban smoking" (against) i.e. prepare their own position statement and (counter) arguments. The position statements should be about 1 minute each. The completion of the argument exchange is controlled by the debaters themselves (or max 10 minutes i.e. time expired). The opposition is strong. The trainee receives in-action and about-action feedback. The system will show the selected in-action feedback task while the task is performed.

The task closes with a reflection on the aspects assessed in comparison to the system's assessment and the impact, if any, observed on the debate.

Possible variations: The trainee will receive the real-time feedback on all criteria in red-green signals only (e.g. with the help of a feedback cube). The trainee will have additional instructions on the focus or the position statement and arguments or will receive a prepared position and arguments.

Part task practice. Recognise and classify arguments or emotions and dealing with them.

The part task practice task allows the trainee to monitor and assess a set of video-taped dialogue moves to learn to recognise and respond to arguments or emotions of the opponent. The set consists of a number of good and bad examples of short videos (max 1 minute each) The assessment is done with the help of a paper form (or similar with an overlay on top of the video).

Possible variations: Tasks can be repeated with stressing different feedback attention points and/or with different topics.

Supportive information. An introduction is given to the trainee about the importance to have an overall awareness of the debate, i.e. to be continuously aware of the target to be achieved and to be aware of the opponent: what is relevant to look at, how that may influence the debate and its outcome and how one may adapt to that. Additionally, the trainee will get an overview of the elements of the METALOGUE system added at level 3, what to expect, how it operates and its interface. In contrast to level 1 and 2, in level 3 the METALOGUE system will organise its feedback around integrated aspects.

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 6.1). They focus on content, argument structure and presentation and the ability of the trainee to set and guard their goals. Table 6.1 gives a first indication what criteria will be used. Unfortunately, the criteria used 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 withand/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 (for an example see: 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 6.2 gives a simplified example of the criteria (METALOGUE aware rubrics) of judgement of a single aspect "voice volume".

Table 6.2 Simplified example of criteria (not exhaustive) to judge a single aspect.

	Not yet	Sometimes	Regularly	Often	Always
uses voice volume appropriately	Adequate Normal < 60% or > 90% (see Annex 1)	Adequate Normal < 65% or Max Loud > 4% Max Soft > 4%	Adequate Normal between 65-75%	Max 4% Loud Max 4% Soft Adequate Normal between 65-75%	Max 2% Loud Max 2% Soft Adequate Normal between 75-85%

References

Amulya, J. (2004). What is Reflective Practice? Cambridge, MA: MIT - Center for Reflective Practice. Retrieved from http://www.itslifejimbutnotasweknowit.org.uk/files/whatisreflectivepractice.pdf

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

Baca, A., & Kornfeind, P. (2006) Rapid Feedback Systems for Elite Sports Training. Microprocessors, 70 – 76.

Bangor, A., Kortum, P.T. and Miller, J.T. (2009) Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies 4, 3, 114-123.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A. Y., Barndollar, K., & Troetschel, R. (2001) The automated will: Nonconscious activation and pursuit of behavioral goals. Journal of Personality and Social Psychology, 81, 1014-1027

Benedek, J. and Miner T. (2002) Measuring Desirability: New methods for evaluating desirability in a usability lab setting. In Proceedings of Usability Professionals Association 2003, UXPA, 8-12.

Bengtsson, J. (1995). What is Reflection? On reflection in the teaching profession and teacher education. Teachers and Teaching, 1(1), 23-32.

Börner, D., Kalz, M., and Specht, M. (2013) Beyond the channel: A literature review on ambient displays for learning. Computers & Education 60, 1, 426–435.

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

Brun, Y., Marzo Serugendo, G., Gacek, C., Giese, H., Kienle, H., Litoiu, M., Shaw, M. (2009). Engineering self-adaptive systems through feedback loops. In B. H. C. Cheng, R. Lemos, H. Giese, P. Inverardi, & J. Magee, (Eds.), Software Engineering for Self-Adaptive Systems (Vol. 5525, pp. 48–70). Berlin, Heidelberg: Springer Berlin Heidelberg.

Brunelli, D., Farella, E., Rocchi, L., Dozza, M., Chiari, L., & Benini, L. (2006) Bio-feedback System for Rehabilitation Based on a Wireless Body Area Network. Processing, 531-536.

Brusilovsky, P. (2004). Adaptive navigation support: From adaptive hypermedia to the adaptive web and beyond. PsychNology Journal, 2, 7–23.

Burish, T. G., & Jenkins, R. A. (1992) Effectiveness of Biofeedback and Relaxation Training in Reducing the Side Effects of Cancer Chemotherapy. Health (San Francisco). 17–23.

Burns, M, Dimock, V., & Martinez, D. (2000). Action + reflection = learning – Paulo Freire and education for critical consciousness. Tap into learning, (3)2, 1-4. Retrieved from http://www.sedl.org/pubs/tapinto/v3n2.pdf

Cardozo, L. D., Abrams, P. D., Stanton, S. L., & Feneley, R. C. L. (1978) Idiopathic Bladder Instability Treated by Biofeedback. Review Literature And Arts Of The Americas, 521–523.

Cassell, J., Nakano, Y., and Bickmore, T. (2001). Non-verbal cues for discourse structure. In Proceedings of Association for Computational Linguistics Annual Conference (ACL), pages 106–115.

Condon, W. and Osgton, W. (1971). Speech and body motion synchrony of the speaker-hearer. In Horton, D. and Jenkins, J., editors, The perception of language, pages 150–184. Academic Press, New York.

Coninx, N., Kreijns, K., & 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.

Csikszentmihalyi, M. (1975) Beyond boredom and anxiety. Jossey-Bass, San Francisco.

Csikszentmihalyi, M. (1999) If we are so rich, why aren't we happy? in: The American psychologist vol:54 iss:10 pp 821 -827

Dourish, P. (2001) Where the action is: the foundations of embodied interaction. MIT Press.

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.

Ekman, P. and Friesen, W. (1981). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica, 1:49–98.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), Situation Awareness Analysis and Measurement. Mahwah, NJ: Lawrence Erlbaum Associates.

Enfield, J. (2012) Designing an educational game with ten steps to complex learning. Doctoral Dissertation

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

Eraut, M. (2002). Menus for Choosy Diners. Teachers and Teaching, 8(3), 371-379.

Erickson, F. (1975). One function of proxemic shifts in face-to-face interaction. In Kendon, A., Harris, R., and Key, M., editors, Organization of behavior in face-to-face interaction, pages 175–187. Mouton, Den Haag.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. Instructional Science, 24(1), 1-24.

Flavell, J. H. (1979). Metacognition and cognitive monitoring : A new area of cognitive- developmental inquiry. American Psychologist, 3, 906-911.

Freire, P. (1973). Education for Critical Consciousness. New York: Seabury Press.

GALA Network of Excellence (to appear). D2.4 “Flow for Serious Games”

Gama, C. (2004). Integrating Metacognition Instruction in Interactive Learning Environments. (Doctoral thesis, University of Sussex, Brighton, UK). Retrieved from http://homes.dcc.ufba.br/~claudiag/thesis/Thesis_Gama.pdf

Gauthier, J., Bois, R., & Allaire, D. (1981) Evaluation of Skin Temperature Biofeedback Training at Two Different Sites for Migraine. Reading.

Gedye, S. (2010). Formative assessment and feedback: a review. Planet, (23), 40–45.

Georghiades, P. (2004). From the general to the situated: three decades of metacognition. International Journal of Science Education, 26(3), 365-383.

Goetz, T. (2011). Harnessing the power of feedback loops. Wired magazine. Available online: http://www.wired.com/2011/06/ff_feedbackloop/all (accessed February 2014).

Hadar, U., Steiner, T., Grant, E., and Rose, F. (1984). The timing of shifts of head postures during conversations. Human Movement Science, 3:237–245.

Hattie, J. (2009). Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement. London: Routledge.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. Review of Educational Research, 66(2), 99-136. Hatton,

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

Heargraves, D. (2005). About Learning - Report of the Learning Working Group. London: Demos.

Retrieved from http://www.demos.co.uk/publications/aboutlearning

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

Higgins, S., Baumfield, V., & Hall, E. (2007). Learning skills and the development of learning capabilities. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. Higgins,

Higgins, S., Kokotsaki, D. & Coe, R.J. (2011). Toolkit of Strategies to Improve Learning: Summary for Schools Spending the Pupil Premium. London: Sutton Trust. Retrieved from http://www.educationendowmentfoundation.org.uk/toolkit

Higgins, D. (2011). Why reflect? Recognising the link between learning and reflection. Reflective Practice, 12(5), 583-584.

Hirsch, R. (1989). Argumentation, information, and interaction: Studies in face-to-face interactive argumentation under differing turn-taking conditions. Department of Linguistics, University of Göteborg.

Hoogveld, A. W. M., Janssen-Noordman, A., & Van Merriënboer, J. J. G. (Red.) (2011). Innovatief onderwijs in de praktijk: Toepassingen van het 4C-ID-model. Groningen: Noordhoff Uitgevers.

Hoque, M. E., & Picard, R. W. (2014) Rich Nonverbal Sensing Technology for Automated Social Skills Training. Computer, 47(4), 28–35.

Hornecker, E. and Buur, J. (2006) Getting a grip on tangible interaction. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'06), ACM Press, 437–446.

Hoskins, B., & Fredriksson, U. (2008). Learning to Learn: What is it and can it be measured? (JRC Scientific and Technical Report EUR 23432 EN). Brussels: European Commission, Joint Research Centre. Retrieved from http://publications.jrc.ec.europa.eu/repository

Huang, W.D., Johnson, T. (2009) Instructional game design using cognitive load theory. In: Ferdig, R. (ed.) Handbook of Research on Effective Electronic Gaming in Education, Hershey, PA, Information Science, pp. 1143–1165

Isaacson, R., & Fujita, F. (2006). Metacognitive knowledge monitoring and self-regulated learn- ing: Academic success and reflections on learning. Journal of the Scholarship of Teaching and Learning, 6(1), 39-55.

Ishii, H. and Ullmer, B. (1997) Tangible bits: towards seamless interfaces between people, bits and atoms. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'97), ACM Press, 234-241.

Ixer, G. (1999). There's No Such Thing As Reflection. British Journal of Social Work, 29(4), 513-527.

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

Kiilia, K., De Freitas, S., Arnab S., & 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).

Kolb, D. A. (1984). Experiential Learning. Englewood Cliffs, NJ: Prentice-Hall.

Koriat, A., & Levy-Sadot, R. (2000). Conscious and Unconscious Metacognition: A Rejoinder. Consciousness and Cognition, 9(2), 193-202.

Kwon, D. Y., & Gross, M. (2005) Combining Body Sensors and Visual Sensors for Motion Training. Training, 94–101.

Lai, E.R. (2011). Metacognition. A Literature Review. San Antonio, TX, USA: Pearson Research. Retrieved from http://www.pearsonassessments.com

Linden, M., Habib, T., & Radojevic, V. (1996) A Controlled Study of the Effects of EEG Biofeedback on Cognition and Behavior of Children with Attention Deficit Disorder and Learning Disabilities I, 35–49.

Lukosch, H., Van Bussel, R., Meijer, S. (2012) A game design framework for vocational education. Int. J. Soc. Hum. Sci. 6(1), 453–457

Mann W. & Thomson, S. (1988) Rhetorical Structure Theory http://www.cis.upenn.edu/~nenkova/Courses/cis700-2/rst.pdf

Marton, F., Dall'Alba, G., & Beaty, E. (1993). Conceptions of Learning. International Journal of Educational Research, 19(3), 277–300.

Marzano, R. J. (1998). A theory-based meta-analysis of research on instruction. Aurora, CO: Mid-continent Regional Educational Laboratory.

McClave, E. (2001). Linguistic functions of head movements in the context of speech. Journal of Pragmatics, 32:855–878.

McNeill, D. (1992). Hand and Mind: what gestures reveal about thought.The University of Chicago Press, Chicago and London.

Moon, J. (2001). Reflection in Higher Education Learning. York: Higher Education Academy - LTSN Generic Centre. Retrieved from http://www.heacademy.ac.uk

Morley, S. (1979) Partial Reinforcement in Human Biofeedback Learning ~ 221–227.

Mory, E. H. (2004). Feedback research revisited. In Handbook of research on educational communications, 745–784.

Nicol D. & 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

O'Malley, C. and Fraser, D. S. (2014) Literature Review in Learning with Tangible Technologies. 2004. Oxford Dictionaries. Oxford University Press

Paradiso, J. A., Morris, S. J., Benbasat, A. Y., & Asmussen, E. (2004) Interactive Therapy with Instrumented Footwear. Music Therapy, 1341–1343.

Peldszus, A. and M. Stede (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.

Rheinberg, F. (2008). Intrinsic motivation and flow-experience. In H. Heckhausen & J. Heckhausen (Eds.), Motivation and action (pp. 323–348). Cambridge, UK: Cambridge University Press.

Ridley, D., Schutz, P., Glanz, R., & Weinstein, C. E. (1992). Self-regulated learning: the interactive influence of metacognitive awareness and goal-setting. Journal of Experimental Education, 60(4), 293-306.

Saffer, D. (2007) Designing for Interaction: Creating Innovative Applications and Devices. New Riders.

Salmon, P., Stanton, N., Jenkins, D., Walker, G., Young, M., & Aujla, A. (2007). What Really Is Going on? Review, Critique and Extension of Situation Awareness Theory. Engineering Psychology and Cognitive Ergonomics, 4562, 407-416.

Sauro. J. and Lewis, J. (2011) When Designing Usability Questionnaires, Does It Hurt to Be Positive? In Proceedings of the ACM SIGCHI Conference in Human Factors in Computing Systems (CHI 2011), ACM Press, 2215-2224.

Scharp, K. (2008). Locke's theory of reflection. British Journal for the History of Philosophy, 16(1), 25-63.

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

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

Schön, D. (1987). Educating the Reflective Practitioner. San Francisco: Jossey- Bass

Seel, N., Al-Diban, S., & Blumschein, P. (2002). Mental Models & Instructional Planning. In M. Spector, J. M., & T. Anderson (Eds.), Integrated and holistic perspectives on learning, instruction and technology: understanding complexity (pp. 129-158). Dordrecht, Boston: Kluwer Academic Publishers.

Spelmezan, D., & Borchers, J. (2008) Real-time Snowboard Training System. Computing, 3327–3332.

Swan, M. (2012) Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. Journal of Sensor and Actuator Networks, 217–253

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

Tabberer, R. (1984). Introducing study skills at 16 to 19. Educational Research, 26(1), 1-6.

Takahata, M., Shiraki, K., Sakane, Y., & Takebayashi, Y. (2004) Sound Feedback for Powerful Karate Training. Technology, 13–18.

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

Vales-Alonso J, López-Matencio P, Gonzalez-Castaño FJ, Navarro-Hellín H, Baños-Guirao PJ, Pérez-Martínez FJ, Martínez-Álvarez RP, González-Jiménez D, Gil-Castiñeira F, Duro-Fernández R. (2010) Ambient Intelligence Systems for Personalized Sport Training. Sensors. 2359-2385.

Van Merriënboer, J. J. G. (1997) Training complex cognitive skills: A four-component instructional design model for technical training. Englewood Cliffs, NJ: Educational Technology Publications..

Van Merriënboer, J. J. G., & Kirschner, P. A. (2013) Ten steps to complex learning (2nd Rev. Ed.). New York: Routledge.

Verpoorten, D. (2012). Reflection amplifiers in self-regulated learning. Retrieved from http://dspace.ou.nl/handle/1820/4560

Wang, M., Haertel, G., & Walberg, H. (1990). What Influences Learning? A Content Analysis of Review Literature. Journal of Educational Research, 84, 30-43.

Watkins, C. (2001). Learning about Learning Enhances Performance. London: Institute of Edu- cation, University of London. Retrieved from http://eprints.ioe.ac.uk

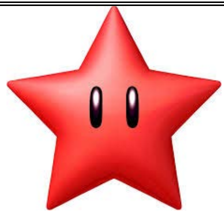
Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., and Yarin, P. (1998) Ambient displays: turning architectural space into an interface between people and digital information. In Cooperative buildings: Integrating information, organization, and architecture, Springer, 22–32

Zeichner, K. (1994). Research on teacher thinking and different views of reflective practice in teaching and teacher education. In I. Carlgren, G. Handal, & S. Vaage (Eds.), Teacher minds and actions: Research on teachers' thinking and practice (pp. 9-27). London: Falmer Press.

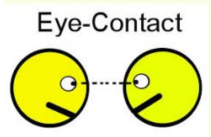
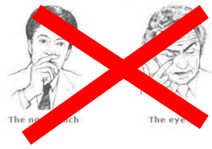
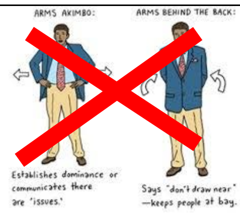
Annexes

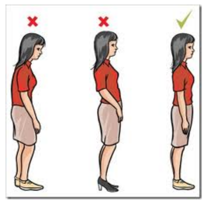
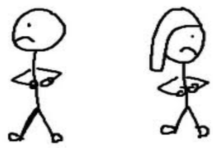
Annex 1. Table Voice Quality aspects

Interaction phenomena	Values (red – problems or failure; green – adequate use)	Indicators/Parameters	Display: Feedback visualization
Voice quality			
Speech rate	Fast Slow Adequate	Measured in number of syllables per second/ To estimate concrete values for these characteristic (e.g. speech rate range what is the adequate value)	
Tone of voice	Hard Normal Soft	The fundamental frequency will be calculated for every 0.1 sec duration using OpenSMILE and analysed to detect the tone of voice.	
Volume	Loud Soft Adequate	Measured in decibels/ 60-65dB normal conversation at 3 meters 30 decibels, equivalent to quiet conversation, soft whisper 5dB – noticeable perceived change in volume Very loud – 78db Loud: 72 Raised: 66 Normal: 60 Relaxed: 54	 

Emphasis	flat intonation; uneven/unbalanced intonation; correct ratio/balance of accented/stressed, and unaccented/unstressed segments	One way to measure it is in ratio of voiced/unvoiced frames (?) Another to estimate pitch range	
Pausing	too long silences within segments; no pausing before new/important information; no pausing at all (both within and between segments within a speaking turn)	Measured in milliseconds: within a segment silence should not exceed 500ms Absence of pause before pitch accented tokens Talking non-stop without any pausing	 Or 

Annex 2. Table Non-verbal behaviour (body language) aspects

Non-verbal behaviour **(body language)**			
Interaction phenomena	Values (red – problems or failure; green – adequate use)	Indicators/Parameters	Display: Feedback vizualization
Eye-contact	'mutual gaze' before starting an interaction No direct eye-contact between participants	Participants looks into the Kinect camera in front of him/her If otherwise	
	Extensive amount of gaze aversion while listening	Ratio direct/averted gaze when listening	
	Averted gaze on accented (important) segments (or parts of it) when speaking	No direct (not looking into Kinect camera) when producing important parts ('emphasis phases' above)	
Head movements	Turning away from the partner or down	Signals from Kinect sensor: head position, trajectory of movement	
Hand and arm gestures	Hands in the pockets	Signals from Kinect sensor: hands position	
	Adators (e.g. rubbing, touching body parts, face, hair or fiddling with objects in hands or in the direct environment)	Signals from Kinect sensor: hands position	
	Crossing your arms	Signals from Kinect sensor: hands position	
	Finger pointing gestures	Signals from Kinect sensor: hands shape and trajectory of arm movement	

	Abrupt, fast gesticulation	Signals from Kinect sensor: hands/arms positions and speed of movements	
	Beat gestures while producing important parts of verbal message	Signals from Kinect sensor and prosodic analysis output: timing of pitch accents, pausing and arm/hand movement	
Overall posture and posture shifts	Confidence posture before and at the very beginning of interaction	Signals from Kinect sensor: as described above	 
	Turning away from the partner	Signals from Kinect sensor: overall trunk orientation	
Facial expressions	Expression of anger	Forehead muscles constricted; eyebrows lowered; yes narrowed; lips corner down	
	Expression of uncertainty	Lip-pout, lip-compression + one corner up possibly accompanied by head waggles (sideways movements) Lips half-opened eyes widened + eyebrows raised (puzzled look) Puzzlement is also displayed by curving the mouth downward, lowering the eyebrows and eyelids, dropping the jaw, and constructing the forehead muscles	 

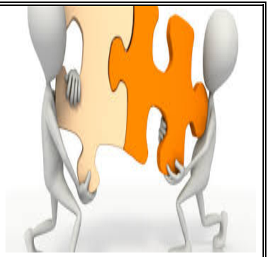
	Expression of disgust/dislike	Forehead muscles constricted, eyebrows lowered, averted gaze (often down), lips corners down, chin pouted	
	Friendly face	An expression in which the corners of the mouth curve upward, and the outer corners of the eyes crinkle into crow's-feet	

Annex 3 Table Dialogue act use ('speaker's intentions') aspects

Dialogue act use ('speaker's intentions')		**Indicators/ definition/examples**	**Feedback Display**
Social obligation acts	Use indirect speech acts for politeness purposes	Suggest instead of Request or Instruct; Use Indirect requests instead of direct request Use Check questions instead of Informs	
	Use social conventions	Apologize for misunderstandings, errors; Thank for offers, for answering questions, for good debate Greet in the beginning Farewell at closing the debate Stay friendly	
Task (and Task Management)	argument short to the point	Grammatical (parsable) sentence	
	Say why are you talking about	Reason, justification	
	Provide evidence	Use figures, statistics, personal experience	

	Do not forget rebuttal	Address directly opponent's arguments, criticize arguments not a person	
Auto/Allo Feedback			
Negative feedback	Signal misunderstandings explicitly	**S**peakers perception of the previous utterance was successful, but he encountered a problem in trying to assign an interpretation to the utterance (for example, Speaker was unable to make sense of the semantic content).	 
	Signal the partner's misinterpretation of your messages	S believes that **A**ddressee's interpretation of S's previous utterance was unsuccessful	
	Interruption when signalling negative feedback	Partner has not release or assigned the turn Turn release: S wants to make the turn available to any participant	
Number of **positive** feedback/back-channelling and strategies	No or little backchannels provided; or wrong timing	Backchannel form: verbally 'ok', 'right', 'mmhm' ' uhu'; nonverbally - head nods Backchannel relevant places: when partner is looking at you, has finished a chunk, is about to proceed while may pause for a while	
	Timely explicit positive feedback	Repeat parts of the partner's utterances, especially in rebuttal phase	
	Rude disagreements	Interruptive; loud; aggressive/angry face	

Turn Management	Avoid (non-collaborative) interruptions	Turn Grab acts: S wants to get the turn, which A currently has, before A assigns the turn to him or releases it.	
	Explicit turn accepting and turn assigning acts	Turn Accept: S agrees to take the turn, which A has given to him Turn Assign: S wants A to take the turn	
Discourse Structuring	Use meta-discoursive acts for structuring interaction, e.g. dialogue act announcements.	Dialogue act performed by the sender, S, in order to signal that he wants the addressee, A, to know that S is going to perform a certain dialogue act soon	
	Use Topic Shift announcement and Topic introduction acts if you want to shift a topic	S wants to change the topic S wants to introduce the topic mentioned in the semantic content.	
Own Communication Management: speech disfluencies	Too many and too long hesitations	2 or more hesitations per turn comprising 3 Task segments is 1 too many Long > 500ms	
	Many restarts and repairs	Every second Task segment with repair/restart inside is at least 1 too many	
Partner Communication Management	Provide collaborative completions	**S**peaker wants to help **A**ddressee to complete an utterance that A is struggling to complete	
	Provide corrections if you think partner made a speaking mistake, but do not interrupt (i.e. wait for pauses)	S wants to correct (part of) an utterance by A, believing that A made a speaking error	

Discourse (and argumentation) coherence, e.g. rhetorical relations and metadiscoursive acts, etc.	'mark' rhetorical relations	Use proper discourse marker. For example, Justify – because; Exemplify – for example, like;	
	Emphasize important message (or part of it) do no repeat it but paraphrase it	Use Restatement act	
Contact Management	Maintain contact with your partner during the whole conversation	Direct eye-contact Wait (pause) till the eye-contact of non-gazing participant is secured	
	Signal when you are ready to start or continue the interaction (or resume)	S wants A to know that S is ready to send messages to and receive messages from A	
	Check attention of you partner if you think he/she not paying attention or lost contact with you	S wants to establish whether A is ready to receive messages from and to send messages to S Verbally, Non-verbally or by changing you voice quality	