Instructional Designs for learning analytics and reflection support

The METALOGUE Consortium
October 2014

Version №: V1.0

Main Authors:
P. van Rosmalen, D. Börner (OUNL),
J. van Helvert (UESSEX) and V. Petukhova (UdS)

Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development
<table>
<thead>
<tr>
<th>Project ref. no.</th>
<th>ICT – 611073</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project title</td>
<td>METALOGUE – Multiperspective Multimodal Dialogue: dialogue system with metacognitive abilities</td>
</tr>
<tr>
<td>Document status</td>
<td>Final version</td>
</tr>
<tr>
<td>Contractual date of delivery</td>
<td>31 October 2014</td>
</tr>
<tr>
<td>Actual date of delivery</td>
<td>21 November 2014</td>
</tr>
<tr>
<td>Document number</td>
<td>D3.2</td>
</tr>
<tr>
<td>Deliverable title</td>
<td>Instructional designs for learning analytics and reflection support</td>
</tr>
<tr>
<td>Dissemination level</td>
<td>RE</td>
</tr>
<tr>
<td>Type</td>
<td>Report</td>
</tr>
<tr>
<td>Number of pages</td>
<td>44</td>
</tr>
<tr>
<td>WP contributing to the deliverable</td>
<td>WP 3</td>
</tr>
<tr>
<td>WP / Task responsible</td>
<td>OUNL / UESSEX</td>
</tr>
<tr>
<td>Contributing partners</td>
<td>OUNL, UESSEX, UdS</td>
</tr>
<tr>
<td>Reviewers</td>
<td>Sergey Sosnovsky (DFKI, UdS)</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Peter van Rosmalen, Dirk Börner, Joy van Helvert &amp; Olga Petukhova</td>
</tr>
<tr>
<td>EC Project Officer</td>
<td>Pierre-Paul Sondag</td>
</tr>
<tr>
<td>Keywords</td>
<td>Feedback, Sensors, Instructional Design, Multimodal Dialogue, Learner Analytics, Reflection Support</td>
</tr>
</tbody>
</table>
Table of Contents

1. EXECUTIVE SUMMARY ................................................................. 4
2. INTRODUCTION .................................................................................. 5
3. BACKGROUND .................................................................................... 7
   3.1 LEARNING ANALYTICS ................................................................. 7
   3.1.1 Generic Design Framework ......................................................... 7
   3.1.2 Quality Indicators .................................................................. 9
   3.1.3 Learning Dashboards ................................................................. 11
   3.2 REFLECTION SUPPORT .............................................................. 14
   3.2.1 Reflection-about-action ............................................................. 14
   3.2.2 Sense-Making .......................................................................... 15
4. TYPES OF DATA AND FEEDBACK .................................................... 18
   4.1 TYPES OF DATA ........................................................................... 18
   4.1.1 Speech signals .......................................................................... 18
   4.1.2 Visible movements .................................................................. 19
   4.1.3 Semantics of verbal contributions and pragmatics of multimodal input .............................................................................. 21
   4.2 TYPES OF FEEDBACK ................................................................. 24
5. APPLICATION EXAMPLE ................................................................. 26
   5.1 TUTORS AND LEARNERS ............................................................. 26
   5.2 SESSIONS AND ROUNDS .............................................................. 26
   5.3 ANALYSIS .................................................................................... 27
   5.4 REFLECTION DASHBOARD: UNDERLYING STRUCTURES .......... 27
   5.5 REFLECTION DASHBOARD: SCREEN MOCK-UPS ...................... 28
   5.5.1 Session Level ............................................................................ 28
   5.6 ROUND LEVEL ............................................................................ 29
6. INSTRUCTIONAL DESIGN BLUEPRINT ............................................. 34
   6.1 METALOGUE DEVELOPMENT ALIGNMENT ............................... 36
   6.2 INSTRUCTIONAL DESIGN ......................................................... 37
REFERENCES ..................................................................................... 43

© METALOGUE Consortium 2014  Page 3 of 44
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 learner analytics and reflection support of the METALOGUE system to support the training. The provided about-action feedback informs learners how they perform key skills and enables them to monitor their progress. The aim of the about-action feedback together with the in-action feedback (see D3.1) is to support the enhancement of the learners' metacognitive skills, such as self-monitoring, self-regulation and self-reflection. This deliverable discusses the role and scope of Learner Analytics in METALOGUE, what type of data is available and should be used, the use of Learning Dashboards and visualisations to enable the learner (and tutor) to access the outcomes of the learner analytics, a set of initial example visualisations to be used in METALOGUE, and, finally, it 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 introduces a framework for the instructional designs for learning analytics and reflection support and a set of indicators which should help to guide the evaluation of the tools to be prepared and their impact on the educational setting proposed. It continues with an overview of visualisations to use to report the outcomes based on a set of examples from similar contexts. Finally, it introduces so-called gaps, i.e. personal moments of struggle, angst or uncertainty, or success and how the approach of sense-making can support the learner. In chapter 4 the data available 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. 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), and voice quality, and finally movements (non-verbal behaviour).

Chapter 5 discusses the initial design of the about-action reflection tool. It starts with outlining the interactions in sessions and rounds and the role of learner and tutor and it concludes with a set of screen mock-ups demonstrating how the system will visualise the results of its analysis. 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 in-action and about-action feedback will be giving 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 will be explored: i.e. a business education context for training call centre employees to handle their customers successfully.

An important aspect therefore of the METALOGUE project is the development and implementation of the instructional design of the educational dialogue to enable the trainee to self-monitor, self-regulate and self-reflect. 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 further 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 also strategic feedback based on the cognitive modelling.

The aim of this deliverable D3.2 is to describe the use of learning analytics and reflection about action support. This implies that we will focus on the data to be used, the representation and visualisation and how the results are embedded in the instructional design. Other necessary parts to be able to use learner analytics such as collection, storing, cleaning, integration and analysis of the data (Kraan & Sherlock, 2013) will be covered in the technical WPs.

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 continuously to be aware of the effects of their arguments, of the use of their voice and of the use of their body language to their audience or opponents and therefore (metacognitive aspects) monitor, reflect and adapt when necessary. 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. D3.1 did discuss the Instructional Design approach chosen, i.e. 4C-ID (Van Merriënboer & Kirschner, 2013) and the theories relevant for the instructional design (chapter 3) and focussed on in-action feedback. In this deliverable we will build upon D3.1 and complement it with learning analytics and reflection support. For the sake of coherence and readability D3.1 did discuss the type of data and feedback for both in-action and about-action feedback (D3.1: chapter 4) and also the instructional design blueprint included both types of feedback (D3.1: chapter 6). For the same reason in this deliverable we also will include both chapters.

In chapter 3 we will start with a discussion of Learner Analytics (LA). We will review a general framework for LA and position METALOGUE on the dimensions on this framework. Next we will discuss a set of LA quality indicators and their relevance for METALOGUE. We conclude this chapter with an overview of Learning Dashboards and visualisation and show a number of examples upon which the METALOGUE visualisation will build. In chapter 4 we will give an overview of the data available for feedback and discuss their use both for in-action and about-action feedback (as mentioned partly already introducing D3.2). In chapter 5, we discuss a set of initial examples of LA visualisation to be used in METALOGUE. Finally, in the last chapter we will outline an instructional design blueprint.
3. Background

3.1 Learning Analytics

“Learning analytics (LA) is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs”\(^1\). In this section we will introduce the initial framework for the instructional designs for learning analytics and reflection support for METAQUEUE based on a framework proposed by Greller and Drachsler (2012) and a set of indicators which should help to guide the evaluation of the tools to be prepared and their impact on the educational setting proposed. We will conclude with an overview of visualisations to use to report the outcomes to our prospective learners based on a set of examples from similar contexts.

3.1.1 Generic Design Framework

Greller & Drachsler (2012) explored the key dimensions of learning analytics and proposed “a generic design framework that can act as a useful guide for setting up Learning Analytics services in support of educational practice and learner guidance, in quality assurance, curriculum development, and in improving teacher effectiveness and efficiency.”

The proposed model for the domain and application of LA in Figure 3.1 below considers six critical dimensions. Each of the dimensions can be subdivided into several instantiations falling into that dimension. For example, the generic “stakeholder” dimension can have instantiations (values) like “learners” and “teachers.” The list of instantiations in the diagram is not exhaustive and can be extended on a case-by-case basis. It is useful to note that through connecting various (and also multiple) different instantiations of each dimension, concrete use cases can be constructed. We call the dimensions “critical” in the sense that each of the six fields of attention is required to have at least one instantiation present in a fully formulated LA design. We realise, though, that some dimensions are vaguer than others in this respect. (Greller & Drachsler, 2012, p. 44-45)

---

\(^1\) 1st International Conference on Learning Analytics & Knowledge: https://tekri.athabascau.ca/analytics/
The authors also illustrated the purpose and possible usage of the framework with the help of an example use case. The example elaborates on a number of aspects out of the six dimensions. The use case can be used (1) as a checklist when designing a purposeful LA process; (2) as a sharable description framework to compare context parameters with other similar approaches in other contexts, or for replication of the scientific environment. Based on the example of Greller and Drachsler (2012) an initial use case has been elaborated for the METALOGUE context (Table 3.1). The main objective of the use case is set an initial framework for the instructional designs for learning analytics and reflection support.

**Table 3.1** Youth Parliament use case and values for dimensions (adapted from Greller & Drachsler, 2012)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Values</th>
</tr>
</thead>
</table>
| Stakeholders | *Data subjects:* Youth Parliament students.  
*Data clients:* Youth Parliament students. Tutor only as a role. The tutor in the final system will be represented by an ‘intelligent’ avatar. System Designers. |
| Objective | *Reflection:* Analyse student interactions in a debate to identify their strengths and weaknesses in order to help them master the constituent skills of debating including their self-monitoring, self-regulation and self-reflection skills. |
### Data
- **Protected dataset**: Student interactions in a debate both speech, voice and body language aspects.
- **Time scale**: a debating round i.e. between 3 - 10 minutes. Optionally, comparison between subsequent rounds.

### Instruments
- **Technology**: ASR, Voice Quality Analysis, Content Analysis, Prosody Analysis, 3D-Body and Face Analysis.
- **Presentation**: Scores, Diagrams, Timelines, Spider webs on single and integrated aspects.

### External limitations
- **Conventions**:
  1. **Privacy**: The students are informed about internal use and storage and publication (public open corpora) of data. Use, storage and publication of data will only be done with full consent.
  2. **Ethics**: The dangers of abuse/misguided use of the data is none. Participants data is gathered in a role-playing training setting.
- **Norms**: As far as known there are no legal data protection or IPR issues related to the use of the student data.
- **Time scale**: The student will directly benefit of the results of the analytics. Each debating round will be concluded with a reflection phase presenting the main outcomes.

### Internal limitations
- **Interpretation**: The students will reflect together with their tutor on the outcomes shown. This should assure that the students understand the feedback given and set the scope how to act.
- **Scoring**: The scoring of the categories is not directly available. It has to be derived by generalising of or comparing to expert observations.
- **Selection**: Various outcomes can be shown. Evaluation of the usefulness and effectiveness of the outcomes offered should enable selection as part of an incremental design.

### 3.1.2 Quality Indicators

Scheffel, Drachsler, Stoyanov and Specht (2014) recently carried out a Group Concept Mapping study (Kane & Trochim, 2007) to identify quality indicators for learning analytics which can contribute to a framework to regulate the evaluation of learning analytics tools and to provide a way to demonstrate the impact of learning. The study identified five topic areas that can be turned into evaluation criteria for a learning analytics framework. The topics selected, in principle, scored high both on feasibility and importance:
• The first topic Objectives is about awareness, reflection and behavioural change of students and teachers during the learning processes, i.e., it is about the educational aim.

• The second topic Learning Support relates to support for students and teachers during the learning process. It contains indicators for Perceived Usefulness, Recommendation, Activity Classification, and Detection of Students at Risk.

• The third topic Learning Measures and Output deals with the results at the end of the learning process. It is not primarily in relation to individual student performance, e.g., their grades, but refers to a learning analytics tool’s results and outcomes. It contains indicators for Comparability, Effectiveness, Efficiency, and Helpfulness.

• The fourth topic Data Aspects deals with data, algorithms, transparency and privacy.

• Finally, the last topic Organisational Aspects deals with organisational issues. It contains criteria related to Availability, Implementation, Training of Educational Stakeholders and Organisational Change.

Figure 3.2 shows an outline of the framework proposed with four indicators per topic area.

![Figure 3.2 Quality Indicators of Learning Analytics (Scheffel et al, 2014)](image)

In the context of this deliverable, i.e. the instructional design, in particular the topic areas of Objectives, Learning Support and Learning Measures are of importance. The data aspects are dealt with in WP1 where the METALOGUE system architecture is defined and the data formats and standards. Organisational aspects are dealt with in WP 7 ‘Deployment and Learner-Based Evaluation’ and WP 8 ‘Dissemination and exploitation’. From the quality
indicators proposed awareness and reflection (‘Objectives’), perceived usefulness (‘Learning Support’) and effectiveness (‘Learning Measures and Output’) are of importance for METALOGUE and will be discussed below. The perspective of discussion is mainly the learner, the stakeholder for the (in- and) about-action feedback.

**Awareness.** The feedback should enable the learner to become aware of their strong and weak points and their development. For the learner this would imply that they will learn and know which aspects are of relevance and, ultimately, would be able to recognise these aspects in their performance or the performance of others.

**Reflection.** Closely connected with awareness, reflection goes one step beyond. The feedback should enable the learner directly after doing a debate (or a call) to review, analyse, and evaluate the situation, to gain insight for improved practice in the future. Here, for the learner the ultimate goal is to train their self-monitoring, self-regulation and self-reflection. For the learner this would imply that as they practice through their tasks in a number of rounds that they stepwise seamlessly are able to adjust their performance with respect to their own utterances and behaviour and their opponent’s.

**Perceived usefulness.** The feedback given should be valued positively by the learner (and tutor). This means it should align with their practice and understanding on how they perceive or can be trained to perceive a good debater (or a good debate). Moreover, the coverage and accuracy of aspects given feedback upon should allow to act upon them and improve their skills. For the learner (and tutor) this implies that they will be asked to value both usability and usefulness aspects of the feedback offered.

**Effectiveness.** The feedback given should be effective. The objectives discussed above should be measurable. For the learner this implies that they will be assessed before (pre-assessment) and after being trained with the help of METALOGUE.

### 3.1.3 Learning Dashboards

A very important aspect of Learning Analytics is representation and visualisation. It should enable the users to inspect the results of the analysis and relate it to e.g. a pre-specified goal, directly if a recommendation is included or indirectly if it merely shows the results. There are several terms used for research dealing with the visualisation of data: scientific visualisation, information visualisation, data visualisation, visual analytics, infographics, etc. With regards to Learning Analytics, Verbert et al. (2013) presented a conceptual framework that helps to analyse learning analytics applications. The framework includes four process stages:

1. **Awareness.** This stage is concerned with just data, which can be visualized as activity streams, tabular overviews, or other visualizations.
2. **Reflection.** The reflection stage focuses on users’ asking questions and assessing how useful and relevant these are.
3. Sense-making. This stage is concerned with users’ answering the questions identified in the reflection process and the creation of new insights.

4. Impact. In the end, the goal is to induce new meaning or change behaviour if the user deems it useful to do so.

Just recently Verbert et al. (2014) also analysed learning dashboard applications that “capture data about learner activities and visualize these data to support awareness, reflection, sense-making, and impact, for instance by having an influence on behaviour change”. In a learning context they analysed relevant user actions, how data on these actions was captured, how the dashboard applications have been evaluated, and how their impact has been measured. The authors only present few systems that are comparable to the intended METALOGUE system, e.g. Yu et al. (2012) present a social interaction feedback system, measuring social signals from students in terms of politeness, stress, agreement, and disagreement. In the analysis the system is classified as using social interaction as main data source, tracking this data using microphones and cameras, and evaluating the dashboard in terms of effectiveness.

A wide variety of software exist offering tools to analyse and/or visualise existing data (see e.g. Kraan & Sherlock, 2013). Giving the aims of METALOGUE of particular interest is the use of visualisation tools as developed by the Flashmeeting project (http://flashmeeting.open.ac.uk; Scott, Tomadaki, & Quick, 2007). Although the tool has been developed to support online meetings, especially the analysis tools provide insights on
how a multimodal system analysis dashboard could look like. First of all the tool allows to replay the complete meeting, visualising the actual video replay as well as the broadcasting distribution over time (see Figure 3.3). A more detailed view is illustrated in Figure 3.4 showing the Meeting Analysis Overview. This view basically consists of the broadcasting distribution over time enriched with additional information, e.g. chat events, specific content annotations, as well as broadcasting events such as interruptions. The Broadcast Dominance view outlined in Figure 3.5 then also shows the broadcast ratio of the different participants, while the Keyword Cloud shown in Figure 3.6 even analyses the content and creates a tag cloud.

Figure 3.4 Flashmeeting Meeting Analysis > Overview

Figure 3.5 Flashmeeting Meeting Analysis > Broadcast Dominance

In the METALOGUE case the following visualizations will be of relevance:

- Occurrences of a single aspect (e.g. voice volume, confident posture) on a timeline (figure 3.3).
- Aggregation of a single aspect (e.g. time used) (figure 3.5)
- Occurrences of a number of aspects in relation to each other (Figure 3.4) in time.
- Integrated overview of various aspects e.g. AECAL (see section 4.1) as a score (compared to a norm such as occurrences of “bad”/“good” moves) or relative to prior
performance or to performance of peers (e.g. with the help of a spider web visualisation).

Finally, parts of the dashboard, in particular when an integrated aspect is shown, may be layered. If necessary the learner should be able to zoom-in into the underlying aspects for further clarification. In chapter 5 examples of METALOGUE feedback visualization to be developed will be discussed.

![Flashmeeting Keyword Cloud](image)

**Figure 3.6 Flashmeeting Keyword Cloud**

### 3.2 Reflection Support

Learning Analytics can support educational stakeholders in becoming aware of their actions and learning processes. Endsley (1995, 2000) has described being aware as a three level process consisting of the perception of elements in the current situation, the comprehension of the current situations and the projection of a future status (see for a more detailed discussion D3.1, section 3.2 ‘Reflection Processes’). Those three steps have to be seen as a prerequisite for making decisions and effectively performing tasks. Once people are aware of their situation, they can reflect on their actions, possibly adapt their behaviour and engage in a process of continuous learning (Schön, 1983).

#### 3.2.1 Reflection-about-action

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-about-action, deliverable D3.1 did focus on the reflection-in-action aspect.
3.2.2 Sense-Making

Dervin’s Sense-Making (Dervin, Foreman-Wernet, & Lauterbach, 2003) is a methodology drawn from the field of communications and grounded in the social paradigm of phenomenology. It provides a framework for investigation based on metacognitive moments when the making of ‘sense’ is interrupted, made or remade in communication. It theorises humans as being able to reflect on their own experiences in a structured way, therefore offering an approach that emphasises the voice of the human subject or learner alongside that of the researcher or tutor. In this frame, human experience is seen as being pervaded by ‘gaps’. These exist across time, between entities (human, system or otherwise) and between spaces, as Dervin (2003, pp. 270-271) explains:

*This discontinuity condition exists between reality and human sensors, between human sensors and the mind, between mind and tongue, between tongue and message created, between message created and channel [mode of communication], between human at time one and human at time two, between human one at time one and human two at time one, between human and culture, between human and institution, between institution and institution between nation and nation and so on. Discontinuity is an assumed constant of nature generally and the human condition specifically.*

As depicted in Figure 3.7 below, the human subject in Sense-Making is seen as moving through time-space, possessing an innate need to “bridge” any gaps that are encountered (here, “bridging” is the act of communicating, either through internal dialogue or external interaction). The subject is also seen as being ‘situated in cultural/historical moments in time-space, and that culture, history, and institutions define much of the world within which [they] live’ (Dervin et al. 2003: 139), at the same time the subject is assumed to construct their own personal sense of their relationship to such phenomena drawing on and interpreting their existing knowledge. Gaps, then are not rigidly or objectively defined, they are essentially personal moments of struggle, angst or uncertainty; moments where sense cannot immediately be made; moments of reaching out for clues from past experience, from current context or from future expectations, dreams or aspirations.
Sense-Making focusses on recurring patterns of gap defining and gap bridging in relation to particular circumstances; it asks what kinds of situations interrupt an individual’s journey, how do they conceptualise gaps, how do their past experiences relate to gaps and bridges, how do they construct their bridges, how they restart their journeys? (p68) Gathering accounts of gaps faced and gaps bridged in relation to the same phenomenon across numbers of individuals may provide insight, not into the phenomenon itself, but into the processes, patterns, themes and recursivities relating to the “experience” of interaction with the phenomena.

So, how might Dervin’s Sense-Making inform our approach to providing reflection tools for the Metalogue learner? With regard to playback and analysis of learners dialogues, perceived ‘gaps’ are generally unique to the individual and therefore may be difficult to systematically detect using gesture recognition, or dialogue analysis. Having said this, as gaps are often moments of angst or hesitation and there may at times be some physical and/or audible cues that could be registered. As individuals in Sense-Making are seen as the researchers of their own experience, a more reliable approach would be to afford learners some means of indicating their ‘gap’ moments as the real-time dialogue unfolds (this would require a small amount of conceptual training the first time they use the system). Signals could then be tagged allowing the learner to easily revisit the appropriate points in the video recording at a later date. The process of reflecting on and analysing each gap could be supported by a Sense-Making style template assisting learners through a framework of questions to recognise over time their personal patterns of successful and unsuccessful interaction. Some mapping to the game-theoretic analysis of the learner’s interaction may be necessary in order to fully interpret the learner’s activities.
Furthermore, the aggregation of data across numbers of individuals could be analysed for collective patterns and recursiveness in communicating behaviour. This could be used to build a knowledge base available both to the dialogue manager for use in real-time feedback, and to learners in their reflection process.
4. Types of Data and Feedback

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 duration 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. Schefflen (1964), Condon and Ogston (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. 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: Audibility, Engagement, Conviction, Authority and Likability (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:

\[
\begin{align*}
A &= \text{Argument (e.g. Marijuana should be legalized)} \\
R &= \text{Reason (e.g. It does not harm a human body)} \\
E &= \text{Evidence (e.g. According to recent research reported in Harm Reduction Journal, May 9 2006, frequent marijuana use is unlikely to be neurotoxic to the normal development of adolescent brain)}
\end{align*}
\]

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...’

---


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](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 D3.1 for 3 tables summarizing the indicators for dialogue acts, voice and body language).

As discussed in D3.1 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 D3.1 Annex 1 (Voice quality aspects) and D3.1 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. D3.1 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 (D3.1 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 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 Example

The aim of the About-action reflection tool, in the context of the proposed pilots (see D6.1), is to support the enhancement of the learners' metacognitive skills, such as self-monitoring, self-regulation and self-reflection, by allowing the tutor and learner to review and analyse the annotated video material generated during a METALOGUE performance. The preliminary designs have been developed with reference to the pedagogical framework and feedback categories described above, and reflect our discussions about what is possible, practical and valuable to put forward as a prototype for evaluation and development over the three pilot stages. Thus, the concepts and mock-ups presented below are design ideas rather than concrete proposals for the look, feel and structure of the dashboard. As we progress through each pilot and evaluation stage, the needs of the users (in this instance teachers and learners) will be better understood allowing us to refine and tailor the interface and the services provided.

While the technical development focus of the project is necessarily on the Youth Parliament scenario, given our current knowledge of the Call Centre domain, the data organisation and the interface frameworks are generic and potentially portable. We will continue to track this issue and ensure design portability as our knowledge of both the Youth Parliament and Call Centre domains develop over the life of the project.

The original design was inspired by the Flashmeeting project, described in section 3.1.3, with its video playback window, clickable timeline showing the utterances of each participant, and various analyses of the recorded interaction. These features map closely to the envisaged output for the reflection dashboard. Additional considerations include the tutor-learner relationship, the display of METALOGUE feedback events, performance analysis, and the session and round game structure (WP2) of the proceedings.

5.1 Tutors and Learners

The services provided by the dashboard are designed to support the dialogue between tutor and learner in the context of debate training. A tutor may be responsible for a number of learners who will train using the METALOGUE system, and, we have assumed, will need to create and manage learner data within the system. For example, this could mean the tutor creating the learner account and specifying elements of the learner experience to meet their needs in line with the instructional design framework. Towards the end of the prototype development the learner may be able to set their own parameters and types of analysis to view. It is therefore necessary for the system to hold user data and to identify and manage access for both the Tutor and the Learner.

5.2 Sessions and Rounds

Each learner interaction using the METALOGUE system will be structured by a game strategy which prescribes various sets of goals. The game is played out over a number of
Rounds, each round with new goals. A Round starts at the point the METALOGUE microphones and cameras commence recording, and ends when they stop, and it's envisaged that several of these Rounds will be played out sequentially (with only short breaks between i.e. less than an hour) making up a single Session. It will be the tutor’s task to create these Sessions in the system for the learner(s) she/he is responsible for, defining the number of rounds, the goals and eventually the analysis. The session is then initiated when the learner arrives and recording starts. The resulting data (annotated video) is stored for each learner at the round level.

5.3 Analysis

A selection of possible analyses relating to phenomenon such as voice quality, body movement, performance against goals etc. are ultimately envisaged for each round. Further analyses and comparisons across all rounds summarised at the session level are also proposed. Complexity is likely to range from simple voice volume levels over a performance, to complex concepts such as ‘likeability’ requiring the aggregation of multiple feedback events across multiple rounds. In the final stage of the project, tutors and learners will be able to tailor the reflection experience by pre-selecting the required analyses from a list.

The analysis algorithm then is another key concept; a Tutor may select multiple Analyses according to her/his learners’ needs. Each Analysis is defined by it’s unique algorithm, data parameters and function description. Like Learners and Tutors, Analysis is a dynamic entity allowing for the development and addition of new Analysis options. The detailed requirements for the algorithms have yet to be defined for each pilot, however all will be based on the list of feedback possibilities described in section 4.2 and developed in line with user needs gathered during evaluation.

Analysis at both the Round and the Session levels implies a two-tier user interface allowing the user to drill down into Round analysis or conversely aggregate up to Session level analysis. It is possible that a further level may be required at some point to accommodate a single learner interacting in multiple sessions.

5.4 Reflection Dashboard: Underlying Structures

So far we have seen that Tutor, Learner, Session, Round and, as the project progresses, Analysis and Analysis Selection are salient concepts for organising data and user services for the dashboard, and they are the same entities that underpin all user services within the METALOGUE system. The relationships between these real-world entities and the functionality available to the user are described, prior to technical design considerations, in Figure 5.1 below.
Figure 5.1 User Services Conceptual Class Diagram (simplified)

5.5 Reflection Dashboard: Screen Mock-ups

5.5.1 Session Level

At the end of a METALOGUE session, i.e. when all prescribed rounds have been completed, the tutor and learner will be able to access, review and analyse the learner’s performance using the About-Action Reflection Dashboard. Beyond logging-on to the system, the first thing the learner sees is a session level screen giving details of the learner, the tutor and the session reference, see Figure 5.2. Below this is a summary analysis window with tabs reflecting the key analysis categories defined in Section 4.2. The Overall tab is selected displaying a representation of the learner’s performance in comparison to other learners using the system. Other forms of analysis can be explored by selecting the different tabs. Currently, it is assumed there will be a single analysis option for each tab; however in later versions it is likely that multiple options will be available under each heading. This could include complex comparisons such as Spider Web or Radar charts that present, for example, the achievements of multiple learners over a range of factors, such as...
goals, content, delivery, emotion etc. (see example in Figure 5.3).

From the session level it is possible to navigate to the associated rounds, providing access to video of the learner’s performance and a range of more detailed round analysis.

**Metalogue Reflection Support**

![Session Level Screen Mock-up](image)

**Cassandra's Session Summary:**

![Spider Web Diagram](image)

**Round Details:**

- **Round 1:** Video still thumbnail
- **Round 2:** Video still thumbnail
- **Round 3:** Video still thumbnail
- **Round 4:** Video still thumbnail

![Example Spider Web Diagram](image)

**Figure 5.2 Session Level Screen Mock-up**

**Figure 5.3 Example Spider Web Diagram displaying an integrative 'delivery' perspective of the speaker (AECAL) and the overall Voice Quality and Body movement**

### 5.6 Round Level

By selecting a particular round, the learner and tutor gain access to a range of options for selecting and viewing the whole or segments of the learner’s performance during
the round, see Figure 5.4. Video material appears in the top left window with standard video controls immediately beneath. Below this is the timeline adapted from the Flashmeeting dashboard discussed in Section 3.1.3. This shows the utterances of both the learner (Cassandra) and her/his opponent, plus the METALOGUE feedback events against the timeline of the video. Clicking on an utterance block or a METALOGUE feedback symbol will display the corresponding video segment in the top left window. Similarly, clicking a particular point on the timeline will locate that point in the video.

The tabs along the top of the timeline window allow the user to view different types of feedback symbol; for example, Figure 5.4 shows that the voice tab has been selected. Accordingly, the symbols along the timeline represent all the METALOGUE feedback events relating to voice performance.

As the video plays, the top left window displays in detail any feedback events located on the timeline, this includes feedback given in-action, i.e. during performance, and other events detected by METALOGUE during performance but not displayed at the time to avoid overloading the learner. In Figure 5.4 the feedback is shown as a stopwatch symbol, providing positive feedback to the learner that her/his speech rate at this point was at an ideal level. It also provides clarification of what the symbol represents and provides links for further exploration. The use of symbols to provide feedback is not envisaged until at least pilot 2, the initial version of the system will use simple red/green flags.

The lower window is intended to provide various kinds of analysis depending on the tab selected on the timeline window above. For example, Figure 5.4 shows the voice tab is selected on the timeline window, therefore the available analysis options for voice are displayed along the top of the lower analysis window i.e. pause, emphasis, volume etc. The ‘Overall’ tab is shown as selected and the window displays an analysis of the learner’s voice performance for the round against the average performance of other learners training with the same game parameters.
A further example is given in Figure 5.5, which shows the Delivery tab selected above the timeline window. The positive and negative METALOGUE feedback events on the timeline now relate to the learner’s delivery during their performance. The top right window displays the feedback symbol with detailed advice, and the lower analysis window shows a range of delivery related analysis options with the Feedback tab selected. Again, to illustrate, a bar chart is displayed comparing the learner against her opponent in terms of the number of positive and negative delivery feedback events occurring during the round.
In the final example, Figure 5.6, the Round level screen shows the Gap tab selected and the learner’s Gap signals (hand icon) visible in the timeline window. As discussed in Section 3.2.2, gaps are personal moments of struggle, angst or uncertainty, moments where sense cannot be made; alternatively they may be moments of achievement. Essentially they are moments signalled by the learner in the course of the interaction that they wish to highlight and return to in their reflective discussions with their tutor. The mechanism allowing the learner to generate a signal has yet to be decided, preliminary thoughts include a hand held device or a hand signal that could be detected and recorded by the Kinect camera. However both options risk disrupting the body posture and gestures of the learner and would require careful experimentation.

In the lower analysis window a timeline for the whole video is shown. Above the timeline all the different METALOGUE feedback events appear as symbols at the point they occurred in the interaction; below the line the Gap signal symbols appear, again at the point they occurred in the interaction. This allows the learner to consider the system feedback in relation to her/his own metacognitive perceptions of performance. At some point it may also be possible to include a set of self-reflective discussion prompts that encourage the learner and tutor to consider patterns and recursivities in learner’s performance. Meanwhile, the top
right window displays the detail of the METALOGUE feedback events (i.e. all categories) as they occur. The issue of how to display multiple events occurring at the same point on the timeline would need to be resolved.

**Figure 5.6** Round level screen mock-up showing Gap feedback
6. Instructional Design Blueprint

Following the theoretical background discussed in chapter 3 (see also D3.1 chapter 3), the data and feedback categories considered in chapter 4 and the initial examples described in chapter 5, in this chapter we will outline the instructional design blueprint for METALOGUE following the 4C-ID model introduced in D3.1 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, 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’.

4 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.

<table>
<thead>
<tr>
<th>Task Complexity / level</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>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</td>
<td>Full topic. Limited number of arguments or argument exchanges</td>
<td>Full topic. Number of arguments or argument exchanges depend on the participants (or max 10 minutes)</td>
</tr>
<tr>
<td>Opposition</td>
<td>Agreeable opponent</td>
<td>Agreeable &amp; disagreeable opponent</td>
<td>Agreeable &amp; disagreeable opponent</td>
</tr>
<tr>
<td>Length</td>
<td>3-4 minutes</td>
<td>5 minutes</td>
<td>5-10 minutes</td>
</tr>
<tr>
<td>Set &amp; guard goals</td>
<td>Indicator: - overall dialogue performance (based on the available data)</td>
<td>Indicator: - overall dialogue performance (based on the available data)</td>
<td>Indicator: - overall dialogue performance - target achievement</td>
</tr>
<tr>
<td>Contents and organisation</td>
<td>-</td>
<td>Indicator/visualisation Argument use</td>
<td>Indicator/visualisation Argument – Reason – Evidence use</td>
</tr>
<tr>
<td>Delivery</td>
<td>Voice volume</td>
<td>Voice volume Speaking cadence</td>
<td>+ Overall Indicator/visualisation voice aspects</td>
</tr>
<tr>
<td>body language</td>
<td>Confident posture</td>
<td>Confident posture Hands &amp; arms usage</td>
<td>+ Overall Indicator/visualisation body language aspects</td>
</tr>
<tr>
<td>other</td>
<td>Relative speaking time; Relative turn time</td>
<td>Relative speaking time; Relative turn time DA: Communicative behaviour: Politeness</td>
<td>Indicator/visualisation AECAL Relative speaking time; Relative turn time</td>
</tr>
<tr>
<td>Emotion</td>
<td>-</td>
<td>Indicator/visualisation One Emotion – Response pair</td>
<td>Indicator/visualisation selected Emotions – Response pairs</td>
</tr>
</tbody>
</table>

### 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') (c.f. also D3.1); and at least one type of about-action feedback, 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\textsuperscript{5}) 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.

\textsuperscript{5} 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 videotaped 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 with and/or similar to a human qualitative assessment. For instance, translating a 'voice too low for 30 seconds' measurement to an summative judgement such as 'your use of voice volume is insufficient, sufficient or good' or alternatively to a formative judgement 'your use of voice volume is: not yet appropriate, sometimes appropriate, regularly appropriate, often appropriate or always appropriate'. As the system develops we will have to incrementally develop system output that provides meaningful formative or summative judgement by comparing and relating system measurements to human assessors (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”.

© METALOGUE Consortium 2014
Table 6.2 Simplified example of criteria (not exhaustive) to judge a single aspect.

<table>
<thead>
<tr>
<th></th>
<th>Not yet</th>
<th>Sometimes</th>
<th>Regularly</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>uses voice volume appropriately</td>
<td>Adequate Normal &lt; 60% or &gt; 90% (see Annex 1)</td>
<td>Adequate Normal &lt; 65% or Max Loud &gt; 4%</td>
<td>Adequate Normal between 65-75%</td>
<td>Max 4% Loud</td>
<td>Max 2% Loud</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max Soft &gt; 4%</td>
<td></td>
<td>Max 4% Soft</td>
<td>Max 2% Soft</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adequate Normal between 65-75%</td>
<td>Adequate Normal between 75-85%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


