An Experimental Environment for Analyzing Collaborative Learning Interaction

Yuki Hayashi, Yuji Ogawa, and Yukiko I. Nakano

Department of Computer and Information Science, Seikei University, Japan
hayashi@st.seikei.ac.jp

Abstract. In collaborative learning, participants progress their learning through multimodal information in a face-to-face environment. In addition to conversation, non-verbal information such as looking at other participants and note taking plays an important role in facilitating effective interaction. By exploiting such non-verbal information in the analysis of collaborative learning activities, this research proposes a collaborative learning environment in which the non-verbal information of participants is collected to analyze learning interaction. For this purpose, we introduce multimodal measurement devices and implement an integration tool for developing a multimodal interaction corpus of collaborative learning.

Keywords: Collaborative learning environment, multimodal interaction, gaze target, writing action.

1 Introduction

Collaborative learning is a learning style in which multiple participants study collaboratively to acquire knowledge about their subjects [1]. Since participants progress their learning through the exchange of ideas, many researchers have focused on analyzing and modeling the learning process in collaborative learning using dialogue data [2, 3]. In a face-to-face environment, participants generally interact with others by not only exchanging utterances but also using non-verbal behaviors such as looking at other participants and note taking [4]. By exploiting such multimodal information in the analysis of collaborative learning, collaborative learning support systems are able to intelligently mediate group interactions more effectively.

While several studies have analyzed group interaction based on non-verbal information such as gaze targets and speech intervals [5, 6] in the research field of computer-supported cooperative work, there is little research that deals with multimodal interaction during learning in order to facilitate a collaborative learning environment. In order to analyze the learning situation in detail, an interaction corpus that includes elaborative non-verbal information is of primary importance. However creating a multimodal corpus requires a large amount of labor. Thus, it would be very useful if the corpus could be generated automatically/semi-automatically.

S. Yamamoto (Ed.): HIMI/HCII 2013, Part III, LNCS 8018, pp. 43–52, 2013.
© Springer-Verlag Berlin Heidelberg 2013

As the first step for analyzing collaborative learning in terms of non-verbal information, the research objective is to propose a collaborative learning environment for collecting non-verbal information using multimodal measurement devices. The non-verbal information we extract consists of the gaze targets, speech intervals, and writing actions of participants. This exhaustive information allows us to analyze the interaction (e.g., mutual gaze). In order to integrate these primitive data, this research introduces an integration tool and attempts to gather the interaction corpus through collaborative learning in a face-to-face environment. We believe that the corpus can be used to detect learning situations such as when a participant is not actively engaged in the learning process or cannot effectively communicate with others.

2 Collecting Non-verbal Information in Collaborative Learning

In this research, we deal with collaborative learning among participants in small groups (three participants) who study/discuss in face-to-face situations. Through collaborative learning, they try to discuss and share their knowledge of the subject. Each participant has a piece of paper (note) for writing answers/ideas freely.

In collaborative learning, participants progress their learning not only by writing down the answers in their notes as individual learning, but also by looking around at others, listening to what someone is saying, and sometimes expressing his/her own ideas. In order to analyze these various types of interactions in learning, we focus on analyzing the non-verbal information that consists of *(i) gaze direction*, *(ii) speech*, and *(iii) writing action* as multimodal data of collaborative learning. Here, we do not target the verbal information such as utterance content. The following sections describe the methods of acquiring each type of non-verbal data.

2.1 Gaze Direction

The gaze directions of participants afford the clues needed to estimate *gaze targets*; i.e., which participant is gazing at another participant (or their note) at any particular time. The gaze direction of each participant is obtained by eye-tracking glasses[1]. The wearable glasses have a camera that can capture the scene (640*480 pixels, 30Hz) and record what the participant is looking at as coordinate data in the two-dimensional scene into an assistant recording device. In addition, the glasses recognize the identity numbers of infrared (IR) markers based on an IR-ray sensor, when such markers exist in the scene. These data are extracted using the eye-tracking software Tobii Studio[1].

In our learning environment, IR markers were put on each participant's neck and on his/her note. According to the aggregated eye-tracking data, the gaze targets were annotated by calculating the distance between tracked eye coordinates and IR coordinates based on the following equation.

[1] Tobii Glasses Eye tracker and Tobii Studio: Tobii Technology, `http://www.tobii.com/`

$$dist(i,t) = \sqrt{(x_e(t) - x_{ir}(i,t))^2 - (y_e(t) - y_{ir}(i,t))^2} \tag{1}$$

Here, *$(x_e(t)$, $y_e(t))$* and *$(x_{ir}(i,t)$, $y_{ir}(i,t))$* represent the coordinates of the eye-tracking position and IR marker *i*, respectively, for frame *t*. Fig. 1 shows an example of gaze target detection, representing the image captured by the camera of the eye-tracking glasses at time *t*. Red points indicate the eye coordinates of a participant, and devices enclosed by yellow circles are IR markers. In this case, two IR markers (IR *i* and IR *j*) exist in the view. When *dist(i, t)* is lower than certain threshold (in pixels), the gaze target is detected as the target marked by IR marker *i*. In this case, the gaze target is detected as participant *n* since IR marker *i* is proximate according to the results of the distance calculations. This detection process is conducted for every frame in which tracked eye coordination is normally recognized.

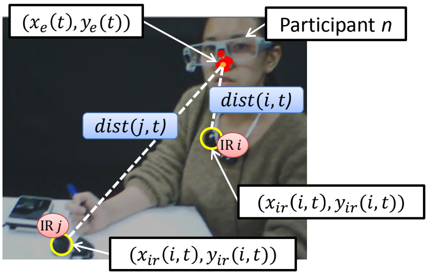


Fig. 1. Example of detection of gaze target

2.2 Speech

In order to detect the duration and identity of speakers at any particular time, the utterances of participants are gathered via a microphone. Audio data stream is transmitted via an audio interface device[2] and recorded independently as audio data (file format: wav, channel(s): 1, sampling rate: 16000 Hz). For speech input/interval detection, we use the Julius adintool software[3], which segments the audio data based on the amplitude level. That is, the start and end times of an utterance are detected when the audio level exceeds certain threshold and by silent intervals. The output is synchronized with the gaze target data. The utterance information gives the number of utterances for each participant and these intervals.

2.3 Writing Action

In order to extract the writing actions of participants, we introduced a digital pen device[4]. The pen performs in two types of modes. One is a mobile mode such that

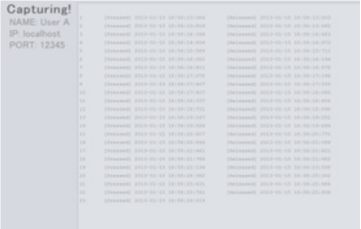
[2] UA-1000: Roland Co., `http://www.roland.com/products/en/UA-1000/`

[3] Adintool: Julius development team, `http://julius.sourceforge.jp/`

[4] airpenPocket: Pentel Inc., `http://www.airpen.jp/`

the written data are stored into a memory unit. The other is a mouse mode, where the pen is perceived as a computer mouse and the mouse move and click actions are controlled according to the writing motions by connecting the memory unit to a computer. The pen device can use a normal ballpoint-type refill, so that participants can write down his/her notes in the usual manner.

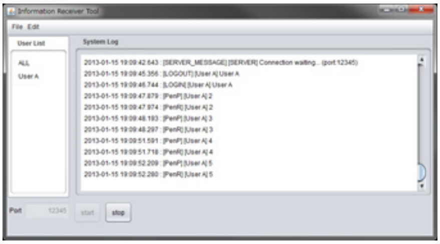
This research uses the mouse mode to judge whether participants are writing or not, expressed as mouse clicking actions based on the pressure-sensitive information collected by the device. In order to synchronize each participant's writing data, we construct a digital pen capturing tool, which sends pen pressed/released data to a writing data receiving tool. Fig. 2 shows an interface image of the pen capturing tool. The writing data stream of each participant is sent to a server computer through a socket connection configured for each participant in advance. Fig. 3 represents the receiving tool. When the data is entered, the tool stores the sequence of the pen pressed/released data with a unique timestamp.



Capturing!
NAME: User A
IP: localhost
PORT: 12345

[illegible]

Fig. 2. Digital pen capturing tool



```
2013-01-15 19:09:42.643 : [SERVER_MESSAGE] [SERVER] Connection waiting... (port:12345)
2013-01-15 19:09:45.356 : [LOGOUT] [User A] User A
2013-01-15 19:09:46.744 : [LOGIN] [User A] User A
2013-01-15 19:09:47.879 : [PenP] [User A] 2
2013-01-15 19:09:47.974 : [PenR] [User A] 2
2013-01-15 19:09:48.193 : [PenP] [User A] 3
2013-01-15 19:09:48.297 : [PenR] [User A] 3
2013-01-15 19:09:51.591 : [PenP] [User A] 4
2013-01-15 19:09:51.718 : [PenR] [User A] 4
2013-01-15 19:09:52.209 : [PenP] [User A] 5
2013-01-15 19:09:52.280 : [PenR] [User A] 5
```

Fig. 3. Writing data receiving tool

3 Experimental Environment

We have developed an experimental environment for monitoring the non-verbal behaviors of participants during collaborative learning. Fig. 4 shows the system architecture. In the environment, each participant wears eye-tracking glasses and a microphone (headset or tiepin type) and writes using the digital pen. The eye-tracking glasses record eye movement, IR data, and scene video files into secure digital cards in the assistant recording device. The microphones are connected to an audio interface device to create high-quality audio files. In order to gather the writing data, three computers in which digital pen capturing tools are running are used for the participants. In addition, a high-definition video camera is placed above the learning environment to record the overall learning scene.

Fig. 5 shows the layout of the participants and IR markers, and Fig. 6 shows a snapshot of the collaborative learning environment. Participants are arranged in a triangle formation around a square table (90 cm * 90 cm), and IR markers with unique IDs are placed on each participant and note to detect the gaze targets.

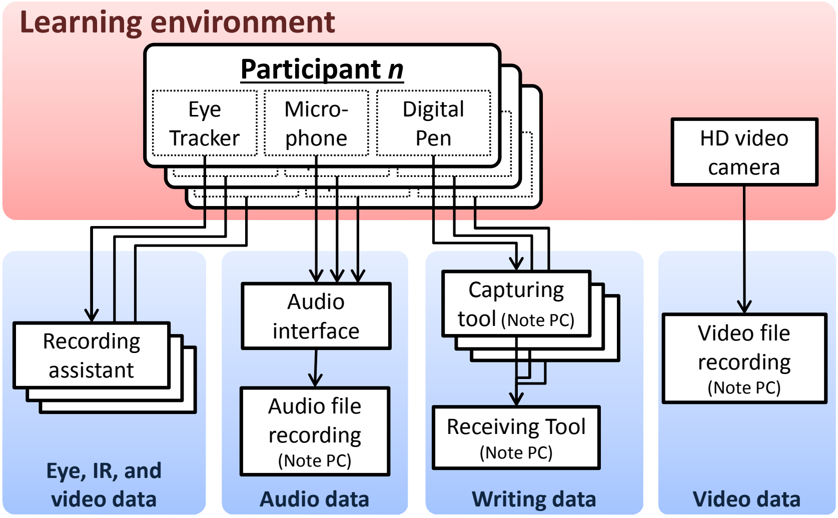
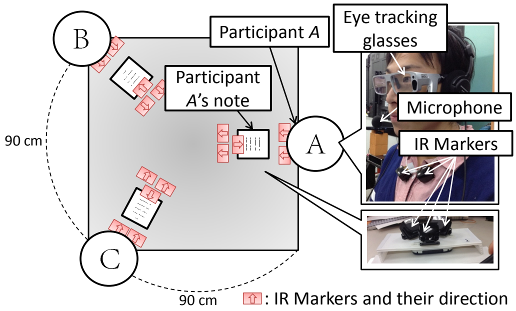


Fig. 4. System architecture of our collaborative learning environment



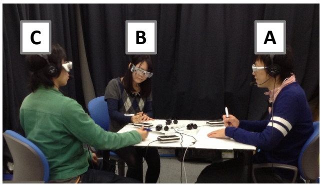


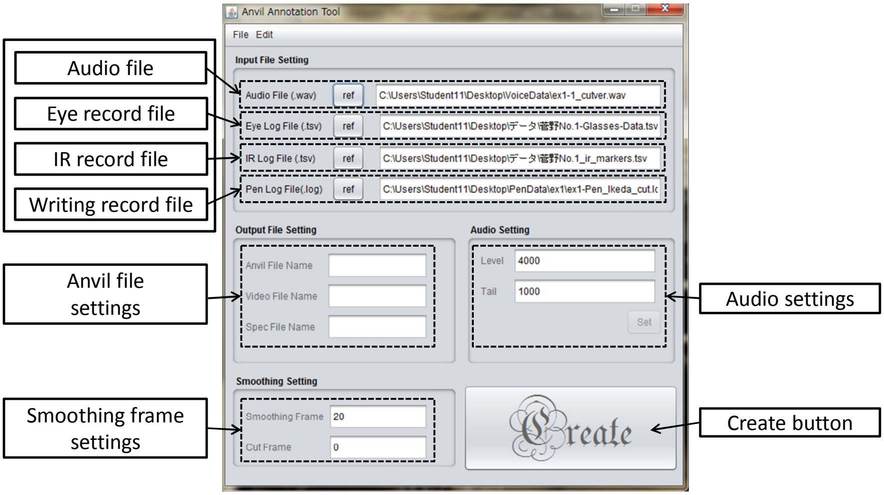
Fig. 5. Layout of participants and positions of IR markers

Fig. 6. Snapshot of collaborative learning

4 Integration of Multimodal Information

By analyzing the non-verbal information obtained from multimodal measurement devices, we can infer sophisticated meanings from the data array, such as "*who is gazing at the speaker*" and "*who is speaking to another participant who is taking notes*" in collaborative learning. In order to integrate these primitive data types, we developed a tool that integrates the multimodal data of each participant along with the timestamp information.

Fig. 7 represents the main interface of our integration tool. This tool requires a participant's audio file, eye record file, IR record file, and writing record file in the file



Input File Setting

Audio File (.wav) C:\Users\Student11\Desktop\VoiceData\ex1-1_cutver.wav

Eye Log File (.tsv) C:\Users\Student11\Desktop\データ\菅野No.1-Glasses-Data.tsv

IR Log File (.tsv) C:\Users\Student11\Desktop\データ\菅野No.1_ir_markers.tsv

Pen Log File(.log) C:\Users\Student11\Desktop\PenData\ex1\ex1-Pen_Ikeda_cut.lc

Output File Setting

Anvil File Name

Video File Name

Spec File Name

Audio Setting

Level 4000

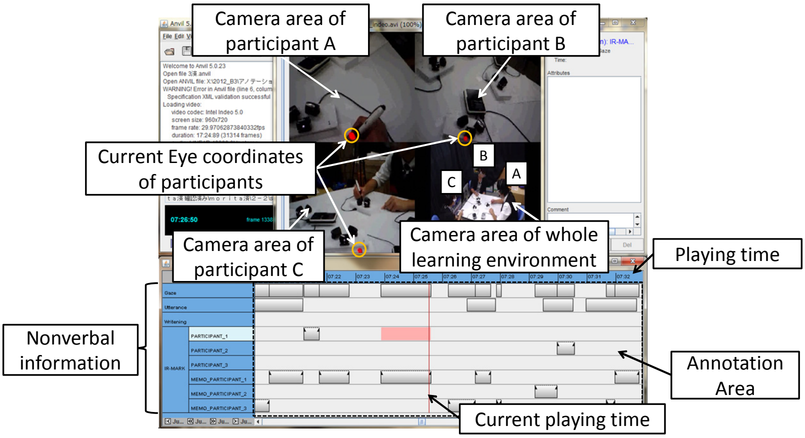
Tail 1000

Smoothing Setting

Smoothing Frame 20

Cut Frame 0

Fig. 7. Annotation tool for integrating non-verbal information



```
Welcome to Anvil 5.0.23
Open file 3済.anvil
Open ANVIL file: X:\2012_B3\アノテーショ
WARNING! Error in Anvil file (line 6, colum
  Specification XML validation successful
Loading video:
    video codec: Intel Indeo 5.0
    screen size: 960x720
    frame rate: 29.97062873840332fps
    duration: 17:24:89 (31314 frames)
```

07:26:50 frame 1338

Fig. 8. Example of annotated data visualized using Anvil interface

setting area. Our tool semi-automatically generates an annotated file that can be read by the generic annotation visualization tool, Anvil [7]. In the Anvil file settings, the parameters for the output Anvil file can be set. The audio settings allow the input amplitude level and silent interval to be set for segmenting the audio data as described in Section 2.2. For smoothing the gaze target data, the smoothing frame parameters can be set in the smoothing frame settings. This input is used to combine two gaze data segments if the adjacent same-target data are within the set frame. In addition, for display in the Anvil interface, we manually encode an integrated video file that includes the video stream of the three participants captured by the eye-tracking glasses shown in Fig. 1 and the video stream from the camera that looks at the learning environment.

Based on the created integrated data file and video file, we confirm the interaction in the collaborative learning using Anvil. Fig. 8 shows the integrated multimodal data in the interface. The horizontal axis represents time sequences and the vertical axis shows the ongoing multimodal interactions. The annotation area represents three types of non-verbal information: *gaze target*, *speech interval*, and *writing action* for each participant. Through the interface, we can not only check the annotated non-verbal data by comparing the segments with the video but also modify the data manually if they are incorrectly annotated.

5 Experiments of Data Acquisition

5.1 Experimental Setting

We conducted experiments for collecting multimodal information during collaborative learning. For the experiments, 30 participants (20 males and 10 females) participated in the collaborative learning experiment. Participants consisted of undergraduate and graduate students of our university. Each learning groups consisted of three participants, and 10 groups were created for the experiments.

Each group was asked to study with others for two sessions. In order to obtain various learning situations such as participants who frequently gave their knowledge to others or participants who learned more passively, we arranged the groups such that they contained participants who were both familiar and unfamiliar with the subjects to be learned. Each group consisted of two participants who majored in computer and information science and a participant whose major was in another field such as literature or economics. The groups were asked to work on exercises in the fields of computer and information science. We set two types of exercises: exercises in which participants took notes frequently to derive a unique answer (answer-derived type) and exercise in which participants mainly discussed and shared knowledge with others (open-ended type). Table 1 describes the learning exercises of the experiments. Exercise 1 consisted of radix conversion problems where an n-ary number needed to be converted to another m-ary number. Exercise 2 was a discussion on cloud computing by exchanging opinions for knowledge sharing. In order to avoid the situation where none of the members progressed in their learning, we told the participants who majored in computer and information science the kind of exercises that would be proposed. We asked them to review the exercise domain in advance.

Before the experiment, the participants calibrated their eye-tracking glasses. They were asked to study Exercise 1 and 2 by turns. We gave instructions to use a piece of paper for writing the solutions and answers in Exercise 1, and for writing useful comments in Exercise 2. The discussion times for all exercises lasted for more than 10 min. We observed the learning situation and stopped the discussion when the conversation quieted down.

Table 1. Learning exercises of the experiments

No.	Type	Contents
(1)	Answer-derived	**Radix conversion problems** - $(10101110)_2$ to octal, decimal, and hexadecimal - $(26584)_{10}$ to binary, octal, and hexadecimal - $(164701)_8$ to binary, decimal, and hexadecimal
(2)	Open-ended	**Discussion on cloud computing** - What is cloud computing? - What are the merits and demerits of cloud computing? - Are there any such services?

5.2 Experimental Results

The average times taken for Exercises 1 and 2 were 839 and 817 s, respectively. According to the non-verbal information extracted in the experiments, we created Anvil files of each session using the annotation tool described in Section 4 and the video files for Anvil using video-editing software. We confirmed the intervals of information and the correctness of the annotations using the Anvil interface.

Gaze Target. The average rates of eye-tracking data acquisition of Exercise 1 and 2 were 62.24% and 52.47%, respectively. One of the reasons for failed detection included eye blinking. According to the video observation from eye-tracking glasses of participants whose acquisition rate was low, some of them had certain characteristic eye-movements. For example, some participants looked down at their notes underneath the lenses of the eye-tracking glasses; there were others who only moved their eyes and peered over the glasses with their heads bent down when they looked at others (Fig. 9). The correlation coefficient indicated a positive correlation between the two exercises (0.59). That is, the participants whose acquisition rates were high or low in Exercise 1 tended to also have high or low acquisition rates in Exercise 2. Thus, the mannerisms of the eye movement of participants affected the data acquisition rates.

In addition, we observed many cases where IR markers were not identified even though eye coordination was detected. Our annotation tool does not annotate the gaze target when the IR markers are not identified, even if the acquisition rate of eye tracking is high. Because of these failed eye-tracking detections, often eye targets could not be annotated.

Speech Interval. We confirmed that the speech interval of each participant was almost exactly extracted according to the corresponding voice data. Using an audio interface contributed to the acquisition of high-quality audio data that includes few acoustic noises. Before the start of the experiments, we tuned the amplitude level of the microphones so that they were approximately equal. Depending on participants, however, we had to adjust the amplitude level in order to segment the audio data.

To extract the speech intervals completely automatically, a mechanism of adjusting the amplitude level (e.g., using average amplitude level information) is required.

Writing Action. Since both the eye-tracking glasses and the digital pen used in this research send signals using an IR-based mechanism, there was a risk of interference. In order to avoid IR interference in the experiments, we covered the IR receiver on the memory units for the digital pen (Fig. 10). However, according to the annotated intervals on the Anvil interface, many writing actions were not recorded. In order to improve the detection accuracy, we either need to develop the learning environment so that the interference of IR signals is avoided, or introduce an alternative writing detection device (e.g., tablet-type device with stylus pen).

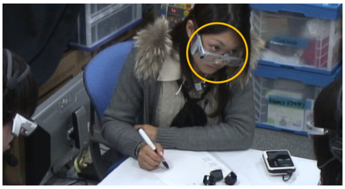


Fig. 9. Example of failed eye-tracking detection

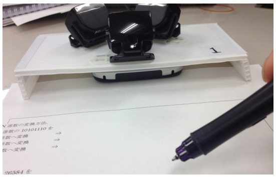


Fig. 10. Cover for digital pen unit

These results indicate that while the annotation of speech intervals was successfully extracted, the annotations of gaze targets and writing actions were not satisfactorily adequate. In order to extract these types of information accurately, we need to improve the data acquisition method. After the experiment, we annotated the correct eye targets and the writing action intervals based on the extracted eye-coordination points using Anvil to create the plenary annotated data. This collaborative learning corpus includes very exhaustive information regarding the interaction among participants that can be used for analyzing several learning situations.

6 Conclusion

In this research, we proposed an environment for monitoring the non-verbal information of participants to analyze collaborative learning interaction. We introduced various multimodal measurement devices into the environment to gather this non-verbal information. In addition, we proposed an integration tool that synchronized the multimodal data of participants. Experiments for the data acquisition showed that while speech intervals were detected with high accuracy, eye-targets and writing actions were not adequately annotated. This was caused by failed detection of eye-tracking, interference of the IR signal, and so on.

Through manual and automatic annotation of data, we now have an interaction corpus that includes plenary annotated non-verbal information in collaborative learning. Based on this corpus, we are currently implementing a visualization tool for briefly analyzing interaction sequences during learning. For future work, we intend to analyze the differences in the interaction that occurs in particular situations, such as when studying collaboratively or individually, or between the exercise types to facilitate group collaboration. In addition, we intend to consider additional methods for acquiring non-verbal data more accurately for gathering the collaborative learning data automatically.

References

1. Adelsberger, H.H., Collis, B., Pawlowski, J.M.: Handbook on Information Technologies for Education and Training. Springer (2002)
2. Soller, A., Lesgold, A.: Modeling the process of collaborative learning. In: Hoppe, U., Ogata, H., Soller, A. (eds.) The Role of Technology in CSCL, vol. 9, Part I, pp. 63–86. Springer (2007)
3. Inaba, A., Ohkubo, R., Ikeda, M., Mizoguchi, R.: Models and Vocabulary to Represent Learner-to-Learner Interaction Process in Collaborative Learning. In: Proc. of International Conference on Computers in Education, pp. 1088–1096 (2003)
4. Kreijns, K., Kirschner, P.A., Jochems, W.: Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. Computers in Human Behavior 19(3), 335–353 (2003)
5. Brennan, S.E., Chen, X., Dickinson, C.A., Neider, M.B., Zelinsky, G.J.: Coordinating cognition: the costs and benefits of shared gaze during collaborative search. Cognition 106(3), 1465–1477 (2008)
6. Kabashima, K., Nishida, M., Jokinen, K., Yamamoto, S.: Multimodal Corpus of Conversations in Mother Tongue and Second Language by Same Interlocutors. In: Proc. of 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction, Article No. 9 (2012)
7. Kipp, M.: Anvil – A Generic Annotation Tool for Multimodal Dialogue. In: Proc. of Eurospeech 2001, pp. 1367–1370 (2001)