Using Multimodal Annotation Tools in the Study of Multimodal Communication Involving Non-speaking Persons

Course paper in Multimodal Communication, Department of Linguistics, Göteborg University, 2003-03-27. Bitte Rydeman.

In my work as a speech- and language pathologist at a resource centre for communication aids, I meet many people who can't rely on speech to communicate. Some of them lack speech altogether because of anarthria, apraxia, severe aphasia or mental retardation. They have to rely on body language (gestures, body posture, facial expressions, eye gaze) and communication aids to communicate. Communication aids don't always fulfill the expectations, and professionals often ask themselves why non-speaking persons don't use them more. Many times they seem to prefer their natural means of communication, be it eye gaze, facial expressions, gestures, pointing, vocalisations or what little speech they may have to their disposal. Still, some non-speaking persons find communication aids very useful and it is very important to recognize the multimodal nature of non-speaking persons' communication and that different modes may serve different goals. In an attempt to begin to solve this puzzle regarding one non-speaking individual, I have looked at a 3 minute sequence where she communicates with her assistant. By using multimodal annotation tools I have attempted an in-depth study of some aspects of their interaction. The purpose of this paper has been twofold: to conduct a study using multimodal annotation tools and to evaluate the tools.

Method

My aim was to study a 3 minutes long video recording of the interaction between a non-speaking girl and her assistant. The video sequence had to be transcribed. Since only one of the participants in the conversation used speech, I had to decide on what else to annotate from the video and how to code it. This was also important for my choice of multimodal annotation tool. I wanted a tool that allowed me to decide what categories to use and what to put in them. Since I expected the use of gestures, facial expressions and eye gaze to differ from that of able-bodied persons, I didn't want to be limited to existing categories and coding schemes. However, I wanted to use existing schemes to find out if, or to what extent, they were applicable. So, my method consisted of:

- a review of available multimodal annotation tools, resulting in a choice of tools that could suit my needs;
- a review of suitable coding schemes, resulting in specially adapted codes;
- transcription and coding of the video sequence;
- tentative conclusions about the result.

Multimodal annotation tools

Several multimodal annotation tools have been developed by researchers in different parts of the world (Dybkjaer et al, 2001; Bernsen et al, 2002). Some of these tools only work on Macintosh or Unix/Linux systems and were therefore not possible for me to use for this assignment (CAVA, MediaTagger, Signstream and SyncWriter). Other tools are not yet available or only available to partners (ATLAS, EUDICO, SmartKom and TalkBank). Some are commercial (The Observer and SyncWriter), some use highly specialized coding schemes (CLAN) and yet others are not made for annotating gestures (MATE and CSLU Toolkit). Of the tools reviewed by Dybkjaer et al (2001) and Bernsen et al (2002) only two remained for me to try: **Anvil** (Kipp, 2001) and **MultiTool** (Grönqvist & Allwood, 1999) , both based on Java and possible to use on Windows systems. These tools are both meant for annotating video recorded conversations, including speech and gestures. They both seem to be very

flexible and share an appealing feature: a timeline based, colour coded overview of simultaneous annotations / codings. In Anvil this is called the Annotation Board, in MultiTool the Partiture view. I didn't, however, succeed in making Anvil work on any of my computers. I downloaded it from the Anvil website, together with the Java tools, and installed it on, first one, then another computer operating under Windows Me. Both times the installation went well, but the computer froze when I started to use Anvil. (There has since been released a new version that might have worked better.) MultiTool, however, worked well and thus became my main annotation tool.

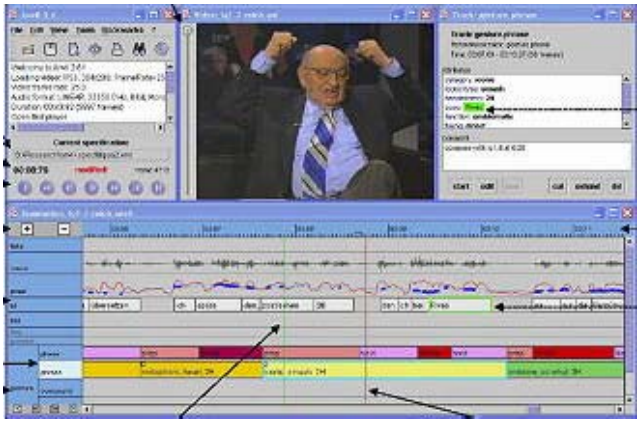

Figure 1: Anvil

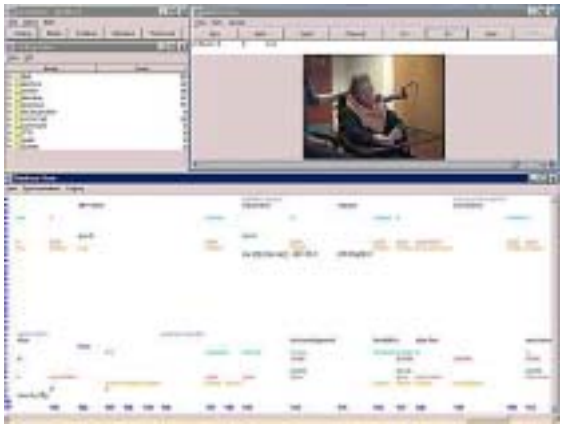

Figure 2. MultiTool

MultiTool suited my needs for a flexible annotation tool that would make it possible for me to construct my own coding schemes, code several persons' contributions simultaneously and look at both speech, communication aid use, gestures, facial expressions and eye gaze. For the transcription, however, it seemed that I would benefit from using a specific transcription tool other than MultiTool (Nivre et al, 1998; Gunnarsson, 2002) and then import the transcription into MultiTool for additional coding. The transcription was to be made using the Gothenburg Transcription Standard, **GTS** (former MSO6, Nivre, 1999), a standard supported by MultiTool. In Nivre et al (1998) there was a description of a tool, **TransTool**, that was supposed to make transcription with the GTS standard easier and less prone to errors. This tool was said to work in Unix, Macintosh and Windows environments and could be downloaded from the website of the Department of Linguistics at Göteborg University. When I found the TransTool the manual (Sofkova Hashemi, 1998) only specified its use on Unix and Macintosh systems, so I concluded I had to do the transcription manually. Another transcription tool, however, referred to in the article of Nivre et al (1998) turned out to be helpful. The tool **VoiceWalker** (du Bois et al, 1999) makes it possible to control the playback of recorded speech by setting the software to automatically loop back through short segments of the recording. It systematically steps through the recording (sound- or video file), repeating short segments for a specified number of repetitions, then moving on to the next segment.

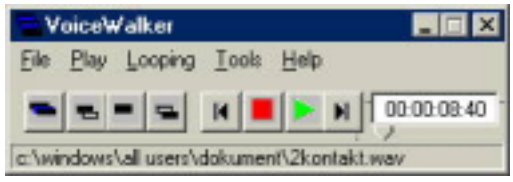

Figure 3: VoiceWalker

A more extensive tool that could be used for the same purpose is **Transana**. (1995-2003). Just like VoiceWalker, Transana can be used for transcription of video files and has features that facilitates the playback of the files during transcription. It supports Jeffersonian transcription, but the transcriber is free to use any transcription standard he or she likes. In addition to the playback features, Transana includes tools for identifying and organizing portions of videos,

attaching keywords to video clips and facilitating the organization and storage of large collections of digitized video.

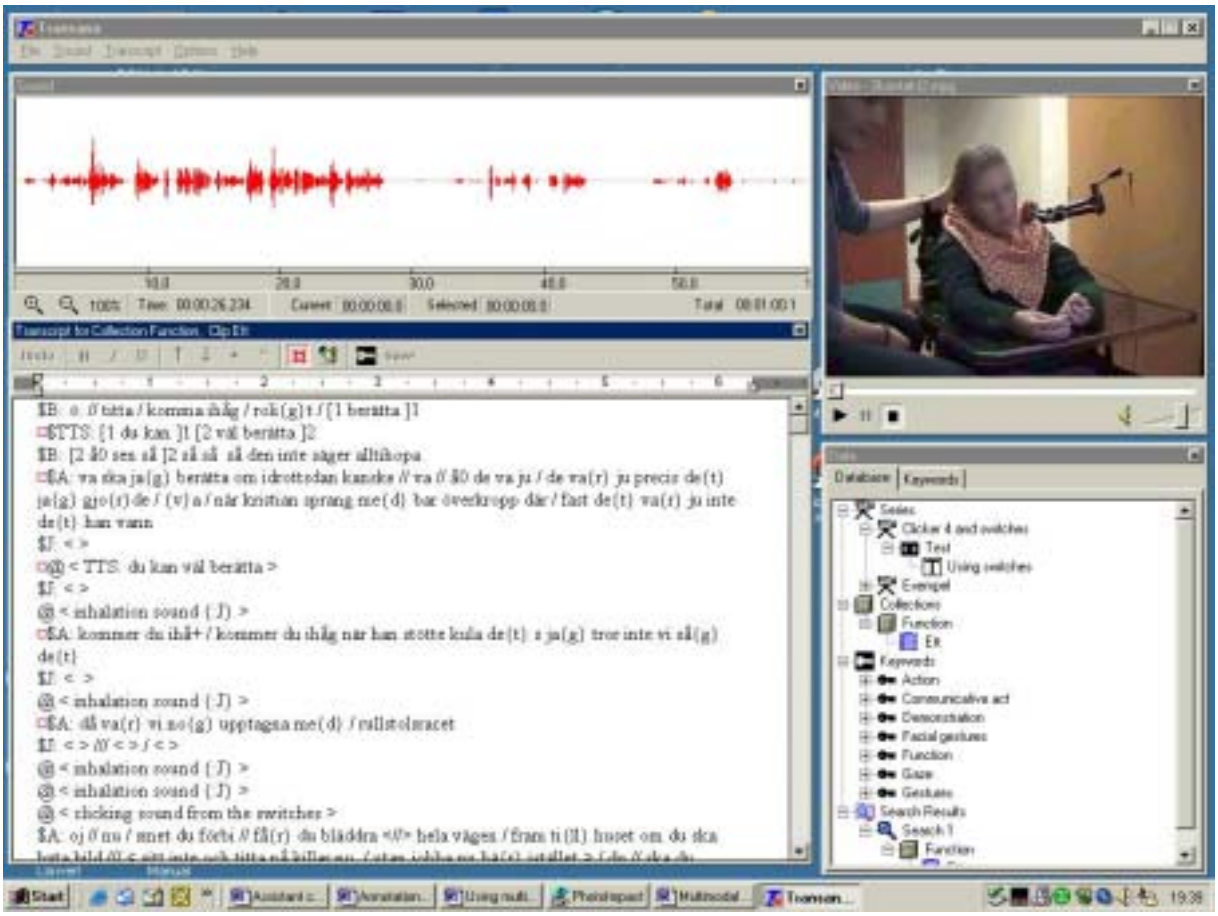

Figure 4: Transana. Including part of the transcription from the studied interaction.

Description of the participants and the studied activity.

In order to understand what is happening in a conversation, it is important to know certain things about the participants and about the activity in which the conversation takes place. This is especially important when the activity and/or the participants are non-typical, as in the conversation described here. One structured way to set the stage is to use social activity coding (Allwood, 2000; Allwood, 2002a). The social activity coding provides a summary of the purpose, function and procedures of the activity, the roles of the participants, the artefacts used and the social and physical environment. A social activity coding of the interaction studied in this paper is shown in table 1.

The main participants in the studied video clip are Jane and her assistant. Jane is 20 years old, has tetraplegic cerebral palsy, anarthria and has never been able to speak. Jane attends a senior high-school for pupils with intellectual disabilities. She is perceived as being very communicative and interested in people and in the activities that go on around her, despite very limited means of communication. Most of the time Jane communicates by eye pointing, by using facial expressions and by answering yes and no. Her gestures for yes and no are not the usual Swedish head movements. Instead, “yes” is indicated by moving the head backwards and looking up, “no” is indicated by moving the head forward and looking down.

Sometimes she points to symbols on a communication chart, using eye pointing or a light pointer attached to her head. The assistant has known Jane for several years. She knows what Jane likes to talk about and is good at interpreting Jane's signals.

Jane is learning to use a computer as a communication aid. To access the computer Jane uses two switches. With a switch located beside her left cheek, Jane scans through objects on the computer screen, with another switch, located behind her head, she activates her choices. The head movement she has to do to use the switch behind her head, resembles the way she expresses "yes".

In the video Jane and her assistant are sitting in front of a portable computer. Jane scans through photos in a program called Clicker 4. When she chooses a photo it is enlarged on the screen. At the same time Jane gets access to 4 spoken messages that she can chose to let the computer say: "Titta här!" (look here), "Kommer du ihåg?" (do you remember), "Så roligt det var!" (we had great fun) and "Du kan väl berätta." (please tell the story). Except for these four utterances, Jane has only access to her body movements (gestures, facial expressions and eye gaze) to communicate.

This is what the video taped situation looked like:

Figure 5: Drawing of Jane and her assistant in front of the computer.

What Jane saw on the computer screen was something like this:

Figure 6: Menu in Clicker 4. Each time Jane presses her switch the next square is highlighted

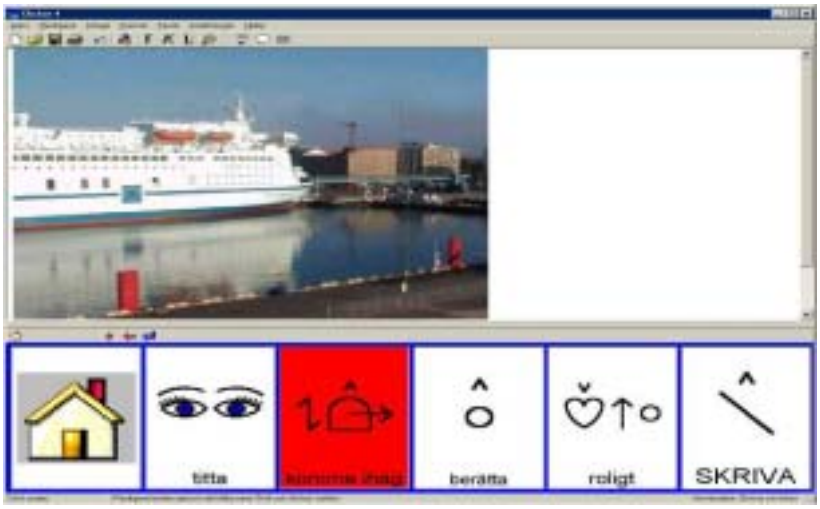

Figure 7: When Jane has chosen a picture in Clicker 4, it gets displayed on the screen. Jane then gets access to 4 spoken messages and 2 squares that lead to other pages in the program.

Table 1 , on the next page, shows a social activity coding of the video taped activity.

Table 1: Activity coding of a conversation about pictures: a non-speaking person and her assistant

PURPOSE Talking about pictures shown on the computer	**Activity structure** 1. J uses her switches to choose a certain picture to be displayed on the computer. 2. J uses the computer to comment on the picture and to ask A to talk about it. 3. A talks about the picture. 4. J gives feedback to A. 5. A tries to make J choose a new picture. 6. J chooses a new picture (->1) OR insists on staying with the same picture as before (->2).	**Sub goals** 1. Having a good time reminiscing on past events. 2. For J to learn to use the computer to communicate (switches, software, application). 3. Getting a good video recording of the interaction between J and A.	**Procedures** J is seated in her wheelchair by the computer. She uses a communication software to choose and display pictures and to make the computer voice say a few comments. To access the computer she uses one switch to scan the alternatives and another to choose. A sits beside J and holds one switch behind J's head. J and A are engaged in conversation about the events described in the pictures and about the task at hand - using switches to access the computer.

ROLES		**Competence**	**Rights**	**Obligation**
	Jane (J)	Non-speaking, severe physical impairment. Uses body movements and facial gestures to communicate + is learning to use the computer. Answers Yes by moving her head backwards.	Communicate by all available means. Use the computer to make A talk about the pictures. Pick and/or change topic. Decide when and if she wants to choose a new picture. Be a listener.	Allow A to be a speaker and a listener. Allow A to assist her.
	Assistant (A)	Knowledge about events and people J wants to talk about. Ability to motivate J and to assist her in using the computer.	Be a speaker and a listener. Teach J to use the computer.	Allow J to be a "speaker" and a listener. Assist J.
	B	Video recording the event. Assist with computer if malfunctioning.	Be a silent participant.	Not disturbing the interaction. Try to be invisible.
	G	Evaluate J's switch use.	Be a silent participant.	Not disturbing the interaction. Try to be invisible.

ARTI-FACTS	**Instruments**	**Media**
	Portable computer with communication software (Clicker 4). Wheelchair with one switch mounted beside the headrest. One switch held by A. Photos of interesting events (in the computer) .	Direct speech. Gestures (hands, head, trunk), facial gestures (smiles), eye gaze. Computerized voice output. (Video recorder)

ENVIRON-MENT	**Social–Cultural**	**Physical**
	A is J's assistant. They know each other well. B is a speech pathologist and G is an occupational therapist. All the participants have taken part in similar activities together a couple of times before.	J and A are sitting in front of a portable computer in a classroom. J is in a wheelchair. A is sitting beside her on a chair. B and G are sitting where J can't see them.

Coding schemas for multimodal interaction

Body movements are integrated with speech in normal communicative interaction (Allwood, 2002b). They can be divided into different groups, such as facial gestures, head movements, direction of eye gaze, lip movements, movements of arms and hands, postures, spatial orientation and touch. All these could be seen in the studied video recording.
In a gesture coding scheme distributed at this course (working paper by Cerrato & Allwood, 2002) movements believed to be non-verbal feedback expressions were listed. They consisted of nod, jerk, shake, waggle, side-way turn, move forward, move backward, hand, shrug, smile and laughter. On the same sheet of paper was a division of the expressions into head movements, facial expressions and gestures. In Allwood (2001) functions were attributed to the different gestures. Some of these functions were used in my coding sample.

The participants in the recorded activity had very different prerequisites for expressing gestures and other body movements. The assistant was able to speak and move freely. The girl, Jane, had a severe physical impairment that made it impossible for her to speak or use her hands. What she could use was her head, eyes and face. The situation was complicated by the fact that some of the same moments that made up Jane's gestures also were used to activate the switches.

In my coding of the bodily communication I tried to use the gesture coding mentioned above, but I also added other gestures that I saw on the video. I divided the body movements into gestures (including head, body and hands), facial gestures (lip movements and smiles) and gaze. I also coded direction as a separate feature, as I did with actions (involving things or people).

An important part of the recording and transcription consisted of text (= speech uttered by the assistant). The speech output from the computer was coded separately and called TTS (text-to-speech).

Two other features were coded from the video: functions (including feedback functions and other functions proposed by Allwood (2001) and communicative acts (Allwood, Ahlsén, Björnberg & Nivre, 2000).

Transcribing and coding the video recorded interaction

These were the steps I followed in my work with the recording:

1. Transcribing the 3 minute conversation, using VoiceWalker and the GTS (MSO6) standard. This transcription consisted mainly of text (speech) and some additional comments (see appendix 1).
2. Transcribing a small part of the video using Transana, confirming the two transcription tools' equal usefulness.
3. Importing the MSO6 transcription file into MultiTool.
4. Using MultiTool to code gestures, facial gestures, eye gaze, action, direction, function and communicative acts. (The partiture view of the full coding is shown in appendix 2)
5. Activity coding the conversation (table 1).
5. Extracting information from MultiTool to be analyzed.
6. Comparing the contributions from the two participants in the interaction.

Result and discussion

The transcription and annotation tools used proved to be very useful. The rewind / repeat features in VideoWalker and Transana facilitated the transcription of the spoken contributions to the conversation. The importation of the transcription file into MultiTool was swift, revealing a few transcription errors that were easily corrected. The partiture view in MultiTool was comfortable to use. It was easy to insert new codings and the opportunity to inset more coding points ensured a high degree of detail and (hopefully) accurateness in the coding. It was convenient to be able to look separately at the different coded features (gestures, gaze, facial gestures, functions etc.). Coding multimodal interaction is, however, very time consuming. Even as little as 3 minutes of conversation took more than a whole day's work to transcribe and code.

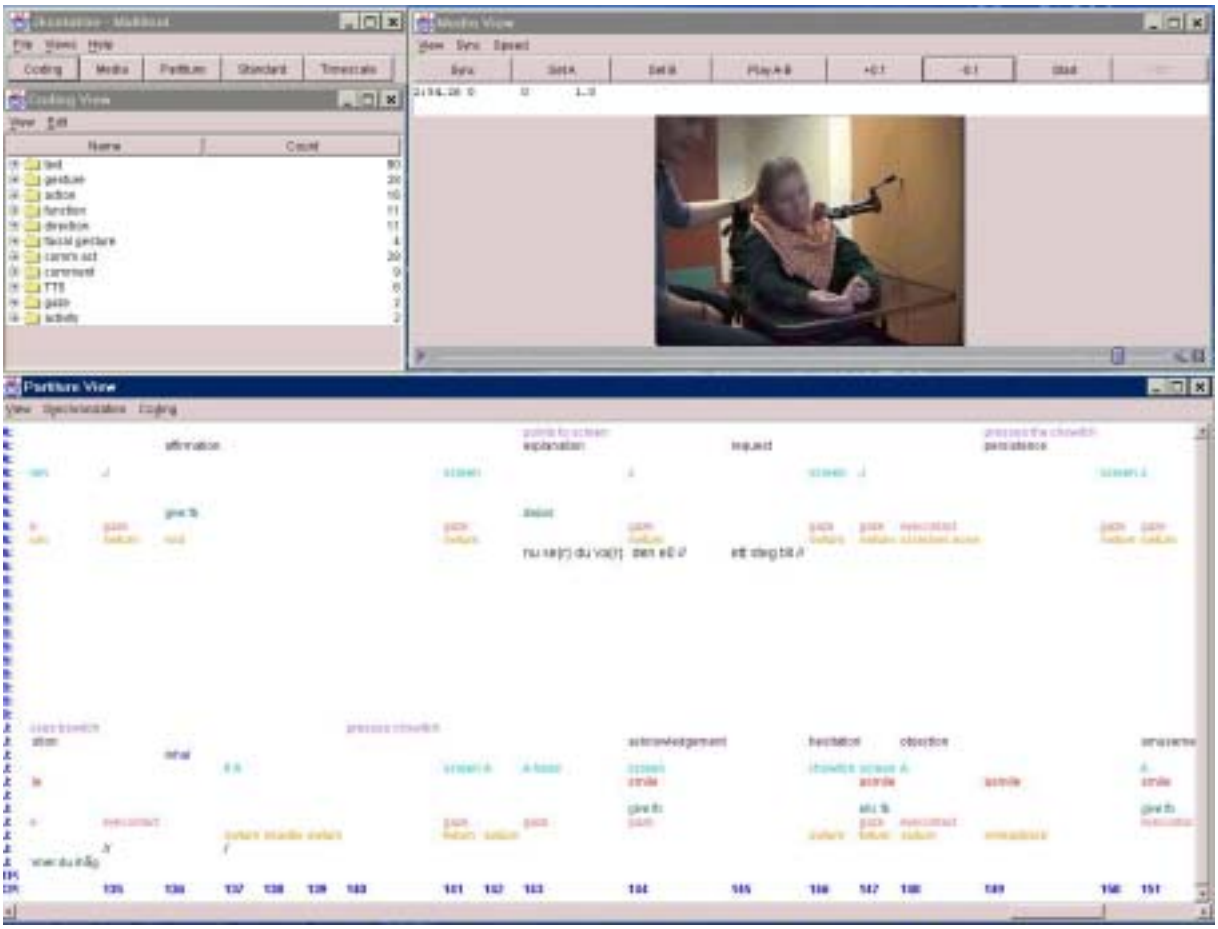

Figure 8: MultiTool. Partiture view of part of the finished coding of the conversation between Jane and her assistant. The full coding of the 3 minutes of conversation is shown in appendix 2.

You get a good overview and can follow the different turns and contributions in the partiture view in MultiTool. It is, however, not quite evident how to go about the analysis of all the assembled data. It seems like some "manual" labour is inevitable, in order to compare, find similarities and differences in the different contributions and codings, making use of the visuo-spatial feature of the partiture view. It would have been nice to have access to a database tool where you could select specific codings and cross reference them. There are however other ways. There is a function in MultiTool to export the transcription to MSO6. This did not work when tried for this transcription, but MultiTool automatically generates another file with the ending .mt. When opened in a word processor the content of the .mt-file looks like in figure 9.

54-55	A	text·/ du //
54-55	A	action·touches J
54-55	A	gesture·swturn
54-55	A	direction·screen
54-55	A	gaze·gaze
54-55	A	comm act·request attention
55-56	A	text·ska du bläddra fram ti{ll} huset
55-56	A	gesture·points to screen
55-56	J	gesture·headfw
55-56	J	gaze·gaze
55-56	J	direction·screen
55-56	A	gaze·gaze
55-56	A	direction·J
55-56	A	comm act·request
55-56	A	function·reinforce
56-57	J	text·
56-57	J	action·presses chswitch
56-57	A	gesture·swturn
56-57	A	gaze·gaze
56-57	A	direction·screen
56-57	J	comm act·acceptance

Figure 9: Part of the .mt-file from the MultiTool coding of the conversation between Jane and her assistant. From left to right you can see the interval (f ex 54-55), the person that displayed the coded behavior (A or J) and the coding.

By rearranging and counting the different contents of the file, the following summary of the codings was generated:

Table 2: Summary of the codings of the conversation between Jane and her assistant

	Assistant	**Jane**
Actions	8 actions involve the switches; she moves them, presses them etc. 1 action involves touching Jane	press the switches 16 touch the switches 2
Gestures	hand to chin 2 jerk 2 move back 1 move forward 1 nod 17 point 7 scratch nose 1 shake index finger 1 shake head 4 single nod (in direction of something) 3 turn head sideways (swturns) 40 waggle 1 = 40 turns of the head sideways + 40 other gestures	forward turn 2 head back 5 head back means Yes 3 head forward 11 head up 2 keep head back 1 move back 1 move forward 2 move head back 16 turn head sideways 23 = 23 turns of the head sideways + 43 other gestures"

	Assistant	**Jane**
Facial gestures	2 smiles annotated (other impossible to see from the video)	almost smiles 5 big smiles 10 mouth opening 1 smiles 17
Gaze	2 eye contacts annotated (other impossible to see from the video) gaze (looks at something) 47	eye contact 14 gaze (looks at smth) 16
Direction	forward 1 Jane 29 screen 26 switch 1 window 1 = 58 changes of direction Equally divided btw J and things in the environment	assistant 19 assistant's hand 2 down 1 forward + left 1 from A 3 screen 10 switch 1 window 1 = 38 changes of direction Equally divided btw A and things in the environment
TTS		3 "du kan väl berätta" 1 "kommer du ihåg"
Speech (text)	≈ 50 utterances	0
Function	deixis 3 elicit feedback 3 give feedback 11 indicating humour 1 reinforcement (using gesture) 14 support own neg. statement (u gest) 3	elicit feedback 3 give feedback 27 indicating humour 2
Communicative acts	acceptance 1 acknowledgement 3 (1 non-verbal) affirmation 3 (1 non-verbal) agreement 2 (1 non-verbal) clarification 1 comment 3 conclusion 1 description 1 elaboration 1 elicit agreement 3 explanation 6 joke 1 (non-verbal) objection 2 persistence 2 (1 non-verbal) question 6 request 12 specification 1 statement 1 supposition 1 = 54 (40 expressed by speech, 19 by speech + gesture, 5 non-verbal (= gesture, smile and/or action.)	acceptance 5 (3 activity) acknowledgement 9 affirmation 5 amusement 3 hesitation 1 objection 3 persistence 5 request 5 (4 TTS) seek agreement 1 =37 (7 expressed by smile, 5 by eye-gaze, 5 by head movement (=yes), 9 by smile + eye-gaze, 6 by smile + head movement (=yes) and 8 by activity (switches, TTS)

From the table summarizing the coding it is evident that there are many differences between the two participants in the conversation, but also similarities. They are both focused on the task at hand - using the computer to choose photos and then talk about the events depicted in these photos. Almost every **action** performed by Jane and the assistant, outside the talking and gesturing, is directed towards the switches. The assistant does all the talking (≈ 50 utterances), but Jane utters 4 spoken messages by means of the computer. As measured by the coding of direction, they both distribute their attention evenly between each other and things in the environment, most often the computer screen. The assistant, being more mobile and intent on monitoring Jane's actions and communicative signals, changes her direction much more often than Jane. She shows 58 changes of direction and Jane 38.

Eye gaze and **smiles are** very important features in Jane's communication. Three levels of smiling were coded in the conversation. The first type is rather a stretching of the lips than a smile, but in lack of a better expression I have called it "almost smile" (asmile). The other levels are "smile" and "big smile" (bsmile). The big smile is very strong and often accompanied by eye contact and sometimes close to laughing. If Jane had better lung function and coordination between breathing and vocal tract, presumably some of them would have been laughs. The relationship between smiling and eye-contact was very strong: 10 of the 14 coded eye contacts were accompanied by a smile or a big smile by Jane.
Unfortunately the video recording has a flaw when it comes to showing the assistant's face. She is placed to the side, so it is easy to see her gestures, head movements and direction changes, and it is evident which way her eye gaze goes, but when it comes to her making eye contact and smiling, this is very hard to see on the recording. It is therefore not possible to compare her with Jane in this respect, but per definition the number of eye contacts between the two has to be the same.

When I coded the **gestures** I included all the body movements, including the turning of the head from side to side. It is, however, questionable if these turns really could be called gestures, if you by gestures mean communicative body movements, since their main function is to change the direction and focus between the communication partner and something else, most often the computer screen. Most of the other gestures coded for the assistant were found in the existing coding schemes (Allwood, 2001). The most prominent gestures by the assistant were nods (17), pointing (7) and head shaking (4). She also displayed a type of nod that I didn't find in the coding schemas: a single directional nod, like pointing with the head. This occurred 3 times.

Almost none of the gestures from the existing coding schemes were used by Jane, except for moving her body back or forward. 38 of her 43 gestures (not counting the sideways turnings) involved the head. Sometimes it was evident that it really was a communicative gesture, for example to indicate yes; other times it was difficult to know if it was an intentional gesture or simply an attempt to reach the switches. Other head movements coded as gestures were clearly not gestures, but actions directed towards the switches or performed to change or maintain body posture or head control.

The coding of **function** gave an interesting result. I had coded almost the same number of functions for both participants: 35 for the assistant and 32 for Jane. This could be a function of my limited experience with this kind of coding, since I would have expected more from the assistant than from Jane. There were, however, important differences. The feedback function dominated the functions coded for Jane. There were 27 instances of giving feedback, 3 instances of eliciting feedback and the two remaining codings were about indicting humour.

The assistant also elicited feedback 3 times, but she gave feedback much less often, 11 times. Most of the other functions coded for the assistant regarded gestures. Some of the functions listed in Allwood (2001) were very prominent. They were reinforcement (by using gesture), that occurred 14 times, and support of own negative statement by using gesture, that occurred 3 times. The differences between the functions expressed by the two participants seemed to depend both on their respective roles in the specific activity and their different communicative skills (one non-speaking - the other doing a lot of talking).

The coding of **communicative acts**, although tentative, also showed interesting differences between the two participants. There were differences in the number of coded communicative acts (54 for the assistant, 37 for Jane), the number of different codings used (19 for the assistant, 9 for Jane) and the distribution among the coded acts. Most of the assistant's communicative acts were expressed by speech (40), 19 were expressed by speech + gesture and 5 were purely non-verbal (consisted of gesture, smile and/or action.). Most of her gestures were head movements, and there was no evident tendency to use the head movements in conjunction with specific communicative acts - they seemed to be evenly divided between the different types.
Of Jane's 37 communicative acts 5 were expressed by smiles, 5 by eye gaze, 5 by head movement, 10 by smile + eye gaze, 7 by smile + head movement and 8 by activity involving the switches. When the activity resulted in the computer speaking, it was coded as a request (based on the spoken message). The other times Jane pressed the switches it was coded as acceptance - she accepted the request given by the assistant that she should press the switches. Jane expresses yes by moving her head back, and this head movement (with or without accompanying smile) was used 4 of the 5 times an communicative act was coded as affirmation. Other times this head movement was used to express acknowledgement, acceptance, objection, persistence and request - these acts were however more often expressed by other means (smiles and/or eye gaze).

19 of the assistant's 54 acts were either a request, a question or an attempt to elicit agreement. Only 5 of Jane's 37 communicative acts were requests: 4 of these were expressed by computerized speech. 19 of Jane's communicative acts were either coded as acknowledgement, acceptance or affirmation. This is consistent with Jane's high frequency of giving feedback. In contrast to this, only 7 of the assistant's communicative acts consisted of acknowledgement, acceptance or affirmation. Instead, many of the remaining communicative acts coded for the assistant were acts like explanation, comment, description, clarification and elaboration, none of which were possible for Jane to express. Instead, her remaining communicative acts were coded as persistence, objection, amusement, hesitation and seeking of agreement.

The studied interaction between Jane and her assistant is not one of equal capacities and opportunities. In terms of power, the assistant controls the situation: she has staged the activity, started the computer, seated Jane in her chair and provided her with the switches, placed herself so that she can see Jane well and at the same time assist her with the switches and the computer. The assistant has specific goals for the activity: to teach Jane to use the computer and the switches and at the same time show the speech pathologist and occupational therapist the progress Jane has made. She has the power of speech and uses this to guide Jane through the activity through requests and questions. There is however an agreement between Jane and the assistant regarding the activity. Their mutual goal is to have a good time and despite her physical condition and total lack of speech, Jane also has a great deal of power in this situation. She can chose whether or not to participate and through the computer and her

body movements, smiles and eye gaze she can make the assistant do things. Through the computer she can choose the topic and she can make the assistant talk about the chosen topics. It is when she does this that the assistant performs the other communicative acts: she comments, describes, explains, clarifies, concludes, elaborates etc.

By studying the partiture view in MultiTool it is possible to find patterns that do not seem so evident when you just count the codings. It is possible to divide the conversation into coherent chunks and define specific parts of the activity structure. By doing so, it becomes more evident why the different communicative acts are performed by the participants.

1. In the beginning of the conversation Jane requests of the assistant that she talks about a certain event shown in a specific picture on the computer. The assistant agrees and starts talking (communicative acts: acknowledgment, elicit agreement, explanation, description, clarification). During this Jane gives feedback (affirmation, acknowledgement).

2. A new request from Jane, followed by more talking by the assistant (acceptance, question, explanation, elaboration). Feedback from Jane (acknowledgment, acceptance).

3. Jane responds to something the assistant has said by indicating acceptance, followed by seeking of agreement. The assistant agrees, but then Jane presses her switch so something not wanted by the assistant happens on the computer screen.

4. The assistant comments on what has happened and then expresses repeated requests that Jane corrects the error. Jane acknowledges and accepts and finally presses the switch. This is followed by more requests from the assistant to press the switches and Jane accepts.

5. Jane requests that the assistant talks again about the picture, which she does (question, conclusion, acknowledgment, comment), while Jane gives feedback (affirmation, acknowledgement).

6. Jane reaches for the switch in order to make a new request, but the assistant takes it away (joke) and suggest that Jane changes to a new picture (question, explanation). Jane objects and persists in trying to press the switch. The assistant gives in (affirmation, agreement) and Jane confirms (affirmation).

7. The assistant starts to talk about the picture again (agreement, statement), Jane gives feedback (indicates humour, acknowledgement, amusement).

8. Request from the assistant that Jane shows another picture on the computer screen, Jane persists in trying to request more information about the current picture. The assistant keeps trying to persuade her to show a new one (request, supposition, explanation, request). Jane shows amusement, accepts and starts pressing the correct switch.

9. The assistant requests that Jane confirms her choice of picture with the other switch. Jane accidentally(?) presses the wrong switch (objection?), the assistant objects and comments, Jane gives feedback (acknowledgement).

10. Question from Jane (“Do you remember?” TTS), the assistant confirms by nodding (affirmation), but then tries to make Jane scan to another picture (explanation, request, persistence, request). Jane gives and elicits feedback (acknowledgment, hesitation, objection, amusement).

Conclusions

The multimodal nature of human communication is important to take into account when studying the interaction between non-speaking persons and their communication partners. Computerized tools for transcribing, annotating and coding such interactions are very useful, and probably time- and cost effective. The tools used in this brief pilot study were found to function well for their intended tasks. The tools used were VoiceWalker and Transana, that both were helpful in the transcription process, and MultiTool, that proved to be a well functioning tool for annotating and coding video recorded interaction. The analysis used was that of simple counting of the respective codings for the two participants, combined with an overview of the activity structure of the conversation, facilitated by the partiture view in MultiTool . With more advanced tools for analysis, other interesting differences, similarities or connections would probably be revealed. Hopefully MultiTool will continue to develop or be completed with other tools for analysis and presentation of data.

References

Allwood, J, Nivre, J, & Ahlsén, E. 1992. "On the semantics and pragmatics of linguistic feedback), *Journal of Semantics*, vol. 9, no. 1, 1992.

Allwood, J. (2000). An Activity Based Approach to Pragmatics". In Bunt, H., & Black, B. (Eds.) *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*. Amsterdam, John Benjamins, pp. 47-80.

Allwood, J., Ahlsén, E., Björnberg, M. & Nivre, J. (2000). COMMUNICATIVE ACTS, Coding Manual. University of Göteborg, Dept of Linguistics.

Allwood, J. 2001. Cooperation and flexibility in multimodal communication. In *Cooperative Multimodal Communication* Harry Bunt and Robbert-Jan Beun, editors Lecture Notes in Computer Science 2155 Springer Verlag, Berlin/Heidelberg

Allwood, J. (2002a). Dialog Coding - Function and Grammar. Göteborg Coding Schemas, GPTL - Gothenburg Papers in Theoretical Linguistics.

Allwood, J. (2002b) Bodily Communication - Dimensions of expression and Content. In Björn Granström, David House & Inger Karlsson (Eds) *Multimodality in Language and Speech Systems* Kluwer Academic Publishers, Dordrecht, The Netherlands

Bernsen, N. O., Dybkjær, L. and Kolodnytsky, M.: THE NITE WORKBENCH - A Tool for Annotation of Natural Interactivity and Multimodal Data. Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'2002), Las Palmas, May 2002.

Cerrato, L. (2002) *Some characteristics of feedback expressions in Swedish*, Proceedings of Fonetik 2002. Speech, Music and Hearing Quarterly Progress and Status Report, vol 44, pp. 101-104. Stockholm, KTH.

Cerrato, L. & Allwood, J. (2002) Coding scheme for head movements and feedback (working paper, unpublished).

Dybkjaer, L., Berman, s., Kipp, M., Wegener Olsen, M., Pirrelli, V., Reithinger, N, & Soria, C. (2001). ISLE Natural Interactivity and Multimodality Working Group Deliverable D11.1. Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data, WP11.

Granström B, House, D. and Karlsson, I. (eds.) (2002). *Multimodality in language and speech systems.* Dordrecht: Kluwer Academic Publishers

Grönqvist, L. & Allwood, J. (1999) MultiTool - A tool for sunchronized work on multimodal dialog. In: Proceedings from the 3rd Swedish Symposium On Multimodal Communication, Linköping.

Gunnarsson, M. (2002). User manual for MultiTool. Department of Linguistics, Göteborg University. Information about the development of MultiTool and links to download the software are available at www.ling.gu.se/projekt/multitool .

Kipp, M. Anvil - A Generic Annotation Tool for Multimodal Dialogue. Proceedings of Eurospeech 2001, pp. 1367-1370, Aalborg, September 2001.

Nivre, J., Allwood, J., Holm, J., Lopez-Kästen, D., Tullgren, K., Ahlsén, E,. Grönkvist, L., & Sofkova, S, 1998. "Towards multimodal spoken language corpora: TransTool and SyncTool".*Proceedings of ACL-COLING 1998, June 1998.*

Nivre, J. (1999). Modifierad standardortografi (MSO6). Institutionern för lingvistik, Göteborgs universitet.

Nivre, J., Allwood, J., Ahlsén, E., Björnberg, M. & Weilenmann, A. (1999). FEEDBACK, Coding Manual. University of Göteborg, Dept of Linguistics.

Sofkova Hashemi, S. (1998). TransTool 2.0 User Manual, Department of Linguistics, Göteborg University.

Transana (1995-2003). A tool for the transcription and and analysis of audio / visual data. Originally created by Chris Fassnacht; further developed and maintained by David K. Woods at the Wisconsin Center for Education Research, University of Wisconsin-Madison. It is part of the Digital Insigth project. Freely available form www.transana.org.

Voice Walker 2.0, Digital Audio Transcription Utility (1999), by du Bois, J., Corston, S., Holton, G., Norris, R., & Field, K. University of California, Santa Barbara, available from www.linguistics.ucsb.edu/resources/computing/download/download.htm .

APPENDIX 1
Transcription of the conversation between J and her assistant, using MSO6.

@ Recorded activity ID: <not set>
@ Recorded activity date: <2001-11-12>
@ Recorded activity title: <Talking about pictures, using a computer with 2 switches>
@ Short name: <switchtalk>
@ Tape(s): <audiofile extracted from video>
@ Participant: J = Jane
@ Participant: A = Assistant
@ Participant: B = Visitor B
@ Participant: G = Visitor G
@ Transcription name: <not set>
@ Transcription System: MSO6
@ Duration: <00:03:10>
@ Transciber(s): <Bitte Rydeman>
@ Transcription date(s): <2003-01-19>
@ Transcibed segments: All
@ Checker(s): <not set>
@ Checking date(s): <not set>
@ Time coding: Yes
@ Section: Start
@ Section: End
$B: ö: // titta / komma ihåg / roli{g}t / [1 berätta]1
$TTS: [1 du kan]1 [2 väl berätta]2
$B: [2 å0 sen så]2 så så så den inte säger alltihopa
$A: va ska ja{g} berätta om idrottsdan kanske // va // å0 de va ju / de va{r} ju precis de{t} ja{g} gjo{r}de / {v}a / när kristian sprang me{d} bar överkropp där / fast de{t} va{r} ju inte de{t} han vann
$J: < >
@ < TTS: du kan väl berätta >
$J: < >
@ < inhalation sound {:J} >
$A: kommer du ihå+ / kommer du ihåg när han stötte kula de{t} s ja{g} tror inte vi så{g} de{t}
$J: < >
@ < inhalation sound {:J} >
$A: då va{r} vi no{g} upptagna me{d} / rullstolsracet
$J: < > /// < > / < >
@ < inhalation sound {:J} >
@ < inhalation sound {:J} >
@ < clicking sound from the switches >
$A: oj // nu / smet du förbi // få{r} du bläddra <//> hela vägen / fram ti{ll} huset om du ska byta bild /// < sitt inte och titta på killar nu / utan jobba nu hä{r} istället > / du // ska du bläddra fram ti{ll} huset
@ < sniffing sound >
@ < quiet >
$J: < >
@ < clicking sound from the switches >
$A: en ti{ll}

$J: < >
@ < gesture: >
$A: nä / en till <//> du // huset får du bläddra till
@ < clear throat {:unknown} >
$J: < >
@ < inhalation sound {:J} >
$A: nu är vi framt+ vill du ja{g} ska [berätta]
$J: [< >] < >
@ < TTS: du kan väl >
@ < TTS: berätta >
$A: ska vi / bl+
$J: < >
@ < TTS: du kan väl berätta >
$A: ska vi berätta igen
$J: < >
@ < inhalation and exhalation sound {:J} >
$A: du gillar de{t} // nä{r} kristian sprang med bar överkropp
$J: < >
@ < gesture: >
$A: ja o ja{g} förstå{r} de{t} / han va{r} o < > rikti{g}t tjusi{g} alltså / ja
@ < click {:A} >
$J: < >
@ < inhalation and exhalation sound {:J} >
$A: syn{d} inte han få{r} va{ra} me{d} på idrottsdan i år du
$J: < > < > < >
@ < clicking sound from the switches >
@ < inhalation and exhalation sound {:J} >
@ < inhalation and exhalation sound {:J} >
$A: ska vi byta bild
$J: < >
@ < gesture: >
$A: då få{r} du bläddra me{d} kinden / ti{ll}s du komme{r} ti{ll} huset
$J: < >
@ < gesture: >
$A: {j}a du ska titta på denna < a: se gjort >
@ <mumbling + other person mumbling >
$A: {j}a jo jo hen han e0 himla snygg / men de{t} e0 ju andra snygga bilder // ja{g} tycke{r} den hä{r} bilden när e du o kristian sitter när han har solglasögonen på sej / den tycke{r} ja{g} e0 snygg
$J: < >
@ < inhalation and exhalation sound {:J}
$A: kan vi inte få se den
$J: < >
@ < inhalation and exhalation sound {:J}
$A: du kan la visa den fö{r} bitte å0 gerd </> dom kanske ha{r} missat de{t} // få{r} du bläddra fram ti{ll} huset först //
@ < inhalation and exhalation sound {:J}
$J: < > < >
@ < inhalation and exhalation sound {:J}
@ < clicking sound from the switches >

$A: en till
$J: < >
@ < clicking sound from the switches >
$A: å0 så bekräfta{r} du me{d}/ nacken / hä{r} bak / ja de{t} e0 de{t} som e} så dumt // den få{r} liksom inte va{ra} fö{r} nära helle{r}
$J: < > // < > / < >
@ < TTS: kommer du ihåg >
@ < inhalation and exhalation sound {:J}
@ < clicking sound from the switches >
$A: nu se{r} du va{r} den e0 // ett steg till //
$J: < > < >
@ < inhalation and exhalation sound {:J}
@ < clicking sound from the switches >
$A. nu