# Multimodal Corpus for Psychotherapeutic Situations

**Masashi Inoue**[†], **Ryoko Hanada, Nobuhiro Furuyama, Toshio Irino**
**Takako Ichinomiya**, **Hiroyasu Massaki**

[†]Yamagata University/National Institute of Informatics
3-16, 4 Jyonan, Yonezawa-shi, Yamagata, Japan
mi@yz.yamagata-u.ac.jp

### Abstract

This paper presents a design principle for construction of an in-house multimodal corpus for computationally analysing and better understanding conversations during psychotherapy. We discuss some sharable information about research community data collection procedures such as recording devices, consent forms, and privacy considerations. We also explain multimodal coding schema and metadata that are needed in the domain. The corpus has three distinguishing properties: 1) it was constructed only for our research and not for public use; 2) the conversation and recording environment was in actual social situations and not controlled; 3) a multimodal coding schema that focuses on the co-construction nature of the conversation was used. Although the conversation contents are not sharable, the data collection procedure and the schema design for the psychotherapy corpus serve as an example of an initiative to construct a multimodal corpus.

## 1. Introduction

To better understand the nature of psychotherapy, we created a micro corpus of about 20 psychotherapeutic conversations whose situations are not suitable for public sharing. In this paper, we consider the possibility of exchanging information on corpus building while keeping the contents of the conversation private.

Recently, there has been growing interest in the situations of conversations. The situations are either physical, e.g., lighting conditions, noise level, temperature and room size, or social, e.g., such as conversation domain, the relationship between speakers, and the purpose of the conversation. We are interested in the social situations. In particular, we focus on domain specific characteristics of conversations. For this purpose, it was necessary to create our own corpora, even though they will be small compared to generic corpora. The content of our corpus will be kept private except for academic presentations due to the sensitive nature of the counselling; however, the procedures for constructing them can be made public for validation and information sharing purposes. In the following sections, we describe some requirements for the multimodal video corpora to better understand the interactions in a specialised setting.

## 2. Data collection procedure

### 2.1. Psychotherapeutic situations

In the field of psychotherapy, research is usually conducted as single-case analyses. In addition to those qualitative analyses, quantitative or mixed research approaches can be made possible by developing multimodal corpora with annotations. For this purpose, we collected psychotherapeutic interviews as data.

Among various psychotherapeutic situations, we focus on two: 1) psychotherapist training situations in which participants were psychotherapy graduate students at the Kyoto University of Education (**training**) and 2) clinical situations in which participants were chronic disease patients who also had some form of psychological problem (**clinical**).

The training situation consisted of 13 psychotherapeutic interviews. In the training situation, therapists had different skill levels. In this situation, the therapists formed a study group to help each other's practice sessions lasting around 20 to 50 minutes. Sessions were often split into several sub-sessions by inserting reflection periods conducted by the observers with a restriction that the counselling session must be completed within a single day, even if a reasonable solution was not found. Clients were students who talked about their problems, and the interviews were not role plays. In the clinical situation, qualified psychotherapists interviewed clients in parallel with physical disease treatment in the hospital. In this situation, therapy consisted of seven counselling sessions of various lengths, some of which are on-going.

### 2.2. Privacy and motivation

The most difficult issue in the construction of psychotherapeutic counselling conversation corpora is privacy. In counselling, participants talk about serious and sensitive topics. Thus, most clients do not want the contents of their sessions made public. In the **training** situation, because participants were psychotherapy graduate students, both therapists and clients were interested in the potential of using dialogue analysis to obtain insights into their psychotherapeutic interviews. Therefore, we were allowed to access all session data except for one session in which the participant refused to be video-recorded. In the **clinical** situation, participants were not particularly curious about the research, but were rather volunteers having good will.

### 2.3. Consent Form

To use therapeutic conversation for research, we needed to obtain participant agreement. We prepared a data usage consent form and asked participants to agree to allow us to disclose their counselling sessions for research purposes. From an ethical point of view, we have to maintain participant privacy. From a research point of view, there must be few restrictions on the use of collected data to extract

the maximum amount of knowledge from them. These two opposing points of view must be balanced in designing the consent form. We employed the usage log approach. In general, participating clients are often concerned with data accessibility, i.e., who will see the data. Therefore, the form contained a list of the people who were participating in the research project to clarify who will have access to the data and a list of the journals and conferences where the research results would be presented. However, research group members may change after the consent form was submitted by the participants and it is impossible to list all potential presentation places. Accordingly, these lists are regularly updated and can be accessed by the participants. Participants have the right to examine the list of members and presentations of the data including their change logs and to retract the use of their recorded data. Note that the consent form concerned data delivery and not the internal processing of data. Therefore, it is possible to add extra annotations to the data. A drawback of this approach is that participants cannot know the future use of the data when they sign the form, necessitating a good rapport with them. The utility of a consent form can be enhanced by not specifying the places of presentation and researchers who have access to the data. For example, the following expressions have been used (Clark, 2009): "The audio tape can be played at meetings of academics (e.g., conferences and presentations)" and "The written transcript can be used by other researchers". Although such option is attractive, we chose the restricted version that was more acceptable to many participants so as to increase the size of the corpus.

### 2.4. Non-invasive recording

Our data-recording environment was built to have minimal impact on counselling. The structure of psychotherapy in the study group (**training**) already incorporated the use of a video camera for reviewing purposes. Although not all participants used videos in their daily counselling activities, they were accustomed to the presence of a video camera. Microphones were placed some distance away so that they did not restrict the speakers' natural movement or influence their speech. Figure 1 shows how the counselling sessions were recorded. We also tested a moderately invasive recording setting shown in Figure 2, in which microphones were attached near the participant's mouth and an accelerometer was attached to the top of their head. The accelerometer was used to measure head movements as explained in 3.2.. We encountered only one client who felt this device was distracting.

We did not control the topic of counselling. The advantage of using conversations in the study group over using role-playing conversations (e.g., (Belvin et al., 2004)) was that the clients talked about their real problems. We were able to witness conversations that were from emotionally depressed or confused participants. Although some professionals can play the role of patients, client diversity cannot be achieved by such role play. Sometimes participants hesitated to openly describe their problems, and therefore, the therapists occasionally failed to grasp the client's problems, and the counselling sometimes failed or did not produce sufficient change in participants. These irregularities are of



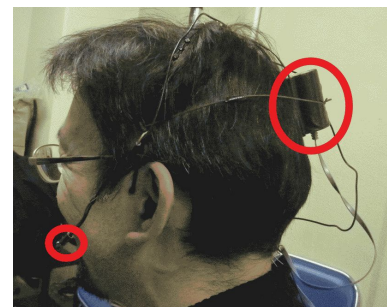Figure 1: Example of minimum invasive counselling conversation.



Figure 2: Example of counselling conversation with an accerelometer.

particular interest to us.

## 3. Multimodal coding schema

### 3.1. Gesture coding

One of our current focus is on gesture modality, particularly hand movements. Many researchers employ similar coding procedures for gestures based on McNeill's framework (McNeill, 1992). We also followed this framework, but with some operational modifications. Our gesture coding schema is summarised in Table 1. An important modification was the distinction between gestures and non-gestures. According to McNeill, a gesture is 'the movements of the hands and arms that we see when people talk'. It is natural to focus on the above-defined co-speech movements if the goal of the research is the psycholinguistic nature of speech and gesture production. For analysing psychotherapeutic conversations, however, the focus is more on the interaction that can lead to the solution of the problem. Therefore, in our schema, we included non-gestures such as self-touching hand movements called adaptors as pseudo gestures for counting occurrences. They may correspond to the mental status of the person; clearly, this is an important concept when dealing with psychotherapy. Another modification was the simplification of communicative gesture categories. Communicative gestures convey some meaning to receivers. McNeill has further divided communicative gestures phases (prep, stroke, hold, retract). Although these sub-categories may be informative, we have not incorporated them. In addition, in our gesture-coding schema, we omitted the spatial aspect of a gesture. That

Table 1: Summary of gesture coding schema

| gesture | communicative | **iconic, metaphoric, deictic** |
|---|---|---|
|  | non-communicative | **beat** |
| non-gesture |  | **adaptor** |

is, we did not look at hand position, direction or speed of movement. These factors may be included in our future studies. This hand gesture coding schema had been used for analysing the relationship between the frequency of gestures and semantic miscommunication (Inoue et al., 2011b).

### 3.2. Head gesture

Head movements are considered important to determine the characteristics of a conversation. This is because they convey the feelings of both speakers and listeners, often in the form of head nodding. We annotated head nodding based on vertical head movements. A series of up and down movements was considered as a single nod. Together with the accelerometer signals, the manual nodding annotations were used in analysing head nodding frequency and synchrony associated with the progress of the therapeutic conversation (Inoue et al., 2011a). Other head movements, such as tilts and shakes, play roles in conversations and should be studied in the future.

### 3.3. Emotion flow

We are currently investigating the assignment of emotional scores to video sequences. For this purpose, we developed a scoring interface named EMO (EMOtional MOvement Observation). EMO is designed for a continuous measurement of emotion in a conversation similar to the EMuJoy developed for music emotion measurement (Nagel et al., 2007). Videos of different views are displayed on a two-dimensional evaluation pane containing axes: pleasant/unpleasant, aroused/sleepy, dominant/submissive, credible/doubtful, interested/indifferent, and positive/negative. An evaluator moves a pointer on the evaluation pane using a computer mouse so that the emotional state of the conversation segment they are watching can be described by the score. The six axes used in our study had been used previously (Mori et al., 2011). We repeated the evaluation process three times, to span the six axes.

### 3.4. Therapeutic stages

We assumed that there were five stages in the therapeutic process: **introduction**, **elaboration**, **resistance**, **intervention** and **solution**. These five stages were given aliases based on stages in the flight of an airplane: **take off**, **cruising**, **turbulence**, **landing preparation** and **landing**. The meaning of the stages are summarised in Table 2. These codes were assigned by psychotherapy experts who have more than three years of experience after acquiring their counselling qualifications. Codes are mutually exclusive and every recorded session fell into one of them. The description of each label, listed in Table 2, represents a typical event during counselling stages. Other different but related activities could be given the same labels by experts. Note that not all counselling follows the same path of the

stages. For example, it may take too long to understand a client's problem and, as a result, the stage remains in **elaboration** for the entire session. Another example may be that the therapist finds a solution but the client is reluctant to accept it, and, thus, the **intervention** and **resistance** stages are repeated many times. The concept of therapeutic stages is similar to the stage definitions used in micro-counselling (Ivey and Ivey, 2002), which are **initiating the session**, **gathering data**, **setting a mutual goal**, **working** and **terminating the interview**. They represent the process of successful interviews and are useful for teaching because students can check to determine if their interviews follow the flow of stages. However, to describe varieties of interviews, we believe that there should be a stage definition of **turbulence**, since, in reality, many interviews do not go straight toward a solution.

## 4. Therapy-specific metadata

### 4.1. Therapeutic outcome

In addition to the process descriptions, the information on the entire session is of interest. We constructed metadata that were obtainable outside of the observed data. The outcome of a psychotherapeutic session is a controversial issue due to its importance and ambiguity. There are several measures for the outcome (Smock, 2011). We collected subjective evaluation for the entire session both from therapists and clients, specifically, whether therapists can listen to the client well and whether therapists can solve the client's problem. We have yet not used them in the research.

### 4.2. Participants' background

There are two types of background information: background of the therapists' and of the clients'. We kept the information as metadata for each session. Therapist background information includes their sex and age, and we were most interested in their expertise. Because the skill levels of therapists are difficult to measure, we used an objective scale: the length of service after acquiring psychotherapist certification. This measure can be used as a rough approximation of expertise, and we believe it reasonable to use as a means of distinguishing novice therapists from experts. Alternatively, we can use the number of cases the therapist worked. However, not all therapists keep a record of their cases, and the community does not currently collect this information. Another important, yet difficult to obtain, information is the school of psychotherapy followed by the therapist. The techniques used and the goal of the sessions depend on the school the therapist follows. However, some therapists do not follow particular school or mix knowledge of different schools. Therefore, in our corpus, school information serves as reference material.

Clients' background information includes sex, age, education and family structures, which are important to understanding the context of the therapy sessions. However, due

Table 2: List of therapeutic stage codes

| Stage name | Alias | Description |
|---|---|---|
| **introduction** | **take off** | Initiating session, constructing rapport. The client introduces oneself with the problem. |
| **elaboration** | **cruising** | Exploring the situation and searching for the solution. The therapist tries to find some clues for the solution. |
| **resistance** | **turbulence** | Struggle due to miscommunication or resistance by clients. The client feels uncomfortable or resists the therapy |
| **intervention** | **landing prep** | Determining the candidate action toward solution. A route to a solution has entered the therapist's mind. |
| **solution** | **landing** | Concluding interview. The clients could rethink their problem in a better way. |

to the sensitive nature of the these types of information, we do not record it in any files and exchange it orally. Since the number of cases was limited and at least one of our project members was involved in each session, either as the participant or as the supervisor, we could recall the background information when necessary. When publishing or presenting the research results, we sometimes alter a clients' background information for the purpose of anonymity, following psychotherapy research conventions.

## 5. Conclusion

In this paper, we explained our procedure for constructing a multimodal video corpus to better understand conversations that occur during psychotherapy. An empirical understanding of psychotherapeutic conversations is needed, and the corpus we built can be an initial step toward a corpus-based study of psychotherapy. In this paper, we illustrated the importance of paying particular attention to the nature of psychotherapy regarding three aspects: the sensitive nature of a conversation and privacy and motivation issues, the special data collection environment necessary for reducing disturbance in the conversation, and the need for a particularly tailored coding scheme and metadata.

Although our strategy can be improved, we believe the information provided here is a useful reference for researchers constructing multimodal corpora in similar social situations, e.g., medical situations, to gain a better understanding of conversations. Also, we can consider sharing anonymous data from the corpus, such as the frequency of gestures or head noddings, as well as sensor signals that do not reveal the contents of the conversation. Privacy-preserving techniques can also be applied as well for the comparison of corpus statistics (Wong and Fu, 2010; Aggarwal and Yu, 2008).

## 6. Acknowledgements

## 7. References

Charu C. Aggarwal and Philip S. Yu, editors. 2008. *Privacy-Preserving Data Mining: Models and Algorithms*. Springer.

Robert Belvin, Win May, Shrikanth S. Narayanan, Panayiotis G. Georgiou, and Shadi Ganjavi. 2004. Creation of a doctor-patient dialogue corpus using standardized patients. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, May.

Shannon Jay Clark. 2009. *Getting personal: Talking the psychotherapy session into being*. Ph.D. thesis, The Australian National University, July.

Masashi Inoue, Toshio Irino, Nobuhiro Furuyama, Ryoko Hanada, Takako Ichinomiya, and Hiroyasu Massaki. 2011a. Manual and accelerometer analysis of head nodding patterns in goal-oriented dialogues. In *Human-Computer Interaction, Part II, HCII 2011*, volume LNCS 6762, pages 259–267.

Masashi Inoue, Mitsunori Ogihara, Ryoko Hanada, and Nobuhiro Furuyama. 2011b. Gestural cue analysis in automated semantic miscommunication annotation. *Multimedia Tools and Applications*, pages 1–14, January.

Allen E. Ivey and Mary Bradford Ivey. 2002. *Intentional Interviewing and Counseling*. Brooks/Cole, 5th edition.

David McNeill. 1992. *Hand and mind*. The University of Chicago Press.

Hiroki Mori, Tomoyuki Satake, Makoto Nakamura, and Hideki Kasuya. 2011. Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics. *Speech Commun.*, 53(1):36–50, January.

Frederik Nagel, Reinhard Kopiez, Oliver Grewe, and Eckart Altenmuller. 2007. EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, 39(2):283–290.

Sara A. Smock. 2011. A review of solution-focused, standardized outcome measures and other strengths-oriented outcome measures. In *Solution-focused brief therapy: A handbook of evidence-based practice*. Oxford University Press.

Raymond Chi-Wing Wong and Ada Wai-Chee Fu. 2010. *Privacy-Preserving Data Publishing: An Overview*. Morgan and Claypool Publishers.