2016 Conference of The Oriental Chapter of International Committee
for Coordination and Standardization of Speech Databases and Assessment Technique (O-COCOSDA)
26-28 October 2016, Bali, Indonesia

Multimodal Interaction Data between Clinical Psychologists and Students for Attentive Listening Modeling

Hiroki Tanaka, Koichiro Yoshino, Kyoshiro Sugiyama
Satoshi Nakamura
Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan 630-0101
Email: {hiroki-tan,koichiro,sugiyama.kyoshiro.sc7,
s-nakamura}@is.naist.jp

Makoto Kondo
Health Care Center
Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan 630-0101
Email: makoto-k@hcc.naist.jp

Abstract—This paper represents new multimodal interaction data recorded during conversation between clinical psychologists and students. For our recording, students talked small talk and topics they are working hard recently, and clinical psychologists attentively listened. Four clinical psychologists and 20 students were participated in this multimodal interaction, and we transcribed and annotated the data. Before and after recording, the participants answered questionnaires. Our preliminary analysis revealed that there were no strong correlations between clinical psychologists and students in most parts of subjective ratings. However, the student's rating of "I was able to talk topics I prepared" was correlated to the clinical psychologist's rating. We plan to create attentive dialogue systems, model good attentive listeners by using this data, and distribute this data to researchers of related fields.

I. INTRODUCTION

The skill of attentive listening is a significant part of communication competence [1], and social interaction [2] as well as treatment outcome and counseling rapport [3]. Modeling of attentive listening can be applied to develop a spoken dialogue system with a function of listening. Attentive listening system may be used to monitor isolated elderly people [4] because loneliness and social isolation is risk factors for mortality [5]. Also, attentive listening modeling can be applied to automated social communication training systems such as [7]. Thus, collecting attentive listening data and modeling it are important and challenging theme.

Previous works have attempted to collect attentive listening data between counselors and participants. An example of actual counseling data was collected by [8], in which real dialogue between professional clinical psychologists and drug addiction patients towards modeling therapist empathy. Sim-Sensei Kiosk was a virtual human interviewer with empathic speaker as well as an attentive listener [9]. Kawahara et al., [10] have tried to model backchannels towards creating attentive listening agents from interview data between counselors and university students.

In this paper, we attempt to collect human-human interaction data as a first step to model attentive listening. We recorded dialogue between professional clinical psychologists and students. Our data includes rich of movie, speech, and language information by using several sensors as well as transcription and annotation. Before and after interaction, we asked to answer questions about subjective impressions for both clinical psychologists and students.

We preliminary analyzed the questionnaires collected from both of clinical psychologists and students that conjugate each other. In real dialogue, it often happens that dialogue attendees have some gaps on their feeling. For example, one dialogue attendee feels that the partner comfortably speaks even if he or she does not feel so. We can observe such gaps of feeling of dialogue attendees from the mismatch of collected questionnaires.

II. PRELIMINARY RECORDING

We recorded interaction between clinical psychologists and students. Before collecting a large number of data, three Japanese students (2 males and 1 female, ID: 1 to 3) and one professional clinical psychologist (male) who usually work as a school counselor more than three years participated in our preliminary recording. This data was not actual counseling, in which the participants were asked to come to the room for dialogue data collection, not for counseling. However, they were asked to talk about their real personal topic.

This preliminary analysis was performed in order to examine variations of recording settings for further collecting. We fixed the clinical psychologist and changed types of interaction of the students. The students separated into three types of condition as shown in Table I. First, we changed whether small talk and main topic are explicitly separated or not. Second, we changed whether the clinical psychologist observe prepared students' personal information, before and during the recording. These two parameters were selected because we

2016 Conference of The Oriental Chapter of International Committee
for Coordination and Standardization of Speech Databases and Assessment Technique (O-COCOSDA)
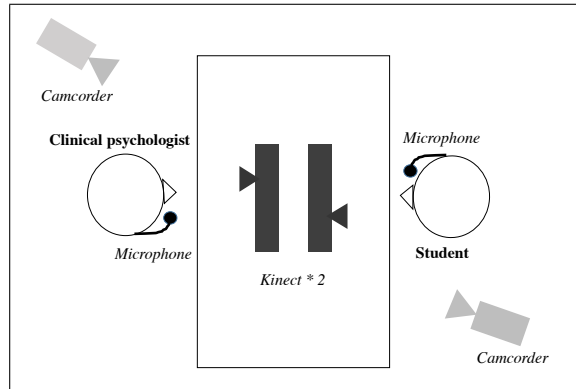26-28 October 2016, Bali, Indonesia

Fig. 1. Data recording setting.

Fig. 2. Data recording without real participants. iPad was placed on a table in order to monitor a progress.

would like to confirm how the clinical psychologist obtain users' knowledges in attentive listening.

TABLE I
THREE SETTINGS OF OUR PRELIMINARY RECORDING.

ID	Gender	Separation	Observing students' information
1	male	Yes	Yes
2	male	No	Yes
3	female	Yes	No

We recorded all three dialogue for a day. As shown in Figure 1 and 2, camcorder, microphone, and iPad were placed. Two Microsoft Kinect sensors were also placed in front of two participants. Three assistants were helped to record. During recording, there was no one in the room without two participants, and the assistants monitored the recording progress at a different room by using Skype on iPad. After the dialogue, we asked all participants about impressions of interaction, and analyzed recorded data.

As a result, the clinical psychologist satisfied the ID 3's setting because it was easy to change topics and did not interrupt the dialogue procedure. Also, this setting is closer to a real dialogue system, which doesn't have prior knowledge. We use the case of ID 3 in the following larger scale recording.

III. DATA RECORDING

A. Participants

Nara Institute of Science and Technology approved this research and written informed consent was obtained from all clinical psychologists and students.

We recruited 20 Japanese students from Nara Institute of Science and Technology. For clinical psychologists, we recruited a total of four people (2 males and 2 females) who have a license of clinical psychologist and currently work on school or health care center as counselor or therapist. Here, one clinical psychologist is a participant of our preliminary recording (in the previous section). The clinical psychologists also helped and suggested about our recording room and environment to make our recording close to actual counseling.

Each clinical psychologist interacted with five students for a day.

B. Procedure

We prepared two types of questionnaires set of before and after the interaction. Creating questions was conducted based on discussion of a professional clinical psychologist and previous works [8], [10].

1) Pre questionnaire: Before the interaction, the students were directed to fill out the pre-questions, print it out, and bring it to the recording room. The items of pre questions were made as follows.

preQ1 Regarding name, age, affiliation, and school year.
preQ2 Regarding research theme, research topic, hobby, favorite food/movie/activity, recent memorable story, and most memorable story in a life.
preQ3 What I am working hard recently.

We told the students that the clinical psychologists do not know the contents of your answer except preQ1, but know the question items.

2) Dialogue: The dialogue session consists of a part of small talk (about preQ2) and a part of main topic (about preQ3), and the participants were directed to naturally change the two parts. The dialogue session lasted around 40 minutes. The clinical psychologists were permitted to take a note about students' important information, and observe general information such as name, age, affiliation, and school year written in preQ1.

3) Post questionnaire: After interaction, two assistants asked the students and the clinical psychologists to answer the post-questions at different rooms. Answering post questions were conducted by watching recorded camcorder videos of a dialogue partner (e.g. a clinical psychologist watches a student's talk). This lasted around 10-15 minutes.

The following is a question set asked to the students. A question set asked to the clinical psychologists is similar type

2016 Conference of The Oriental Chapter of International Committee
for Coordination and Standardization of Speech Databases and Assessment Technique (O-COCOSDA)
26-28 October 2016, Bali, Indonesia

of questions. All participants also described a timing and a reason of each answer by watching the video.

Q1   Did you feel that the conversation was smooth?
Q2   Were you able to trust clinical psychologists?
Q3   Did you feel rapport?
Q4   Did you feel empathy?
Q5   Were you able to talk topics you prepared?
Q6   Did you think that the clinical psychologist was interested in your talk?
Q7   Did you feel happy after interaction?
Q8   Did you feel that an emotional distance of the clinical psychologist was appropriate?
Q9   Did you feel that you could have enough small talk?
Q10 Did you feel rapport in small talk?
Q11 Were you able to trust the clinical psychologist in small talk?
Q12 Did you feel that an emotional distance of the clinical psychologist was appropriate in small talk?
Q13 Did you feel that the conversation was easy? [Yes/No]
Q14 Did you have any topics that you could not talk? [Yes/No]

The rating of Q1 to Q12 was Likert scale, indicating 1 is "I don't think so" and 7 is "I think so". The rating of Q13 and Q14 was a binary, which was answered by yes or no.

C. Transcription and annotation

The transcription and annotation of the recorded data were done by experts according to a guideline. In this procedure, we consider not only creating transcription but also dialogue acts which frequently appear in attentive listening (see Table II). We asked experts to transcribe and annotate with audacity software[1].

TABLE II
ANNOTATION LABEL.

Reactive token	(R): Reaction / backchannel without content words.
Laughter	(L): Laughs with appropriate text (e.g. hahaha).
Silence	(P): Silence / breath more than 500ms in a utterance.

IV. QUESTION ANALYSIS

We analyzed questionnaires obtained from each participant. We examined an inter-question correlation and an agreement between the clinical psychologists and the students by calculating Pearson's correlation coefficient.

A. Inter-question correlation

Figure 3 shows a correlation matrix of Q1 to Q12 in the clinical psychologists. The correlation coefficients of each question were relatively high (e.g. 0.78 ($p < .05$) between Q1 and Q2), indicating the clinical psychologists strongly biased by their impression from the students through the dialogues. In contrast, there were no strong positive correlations between
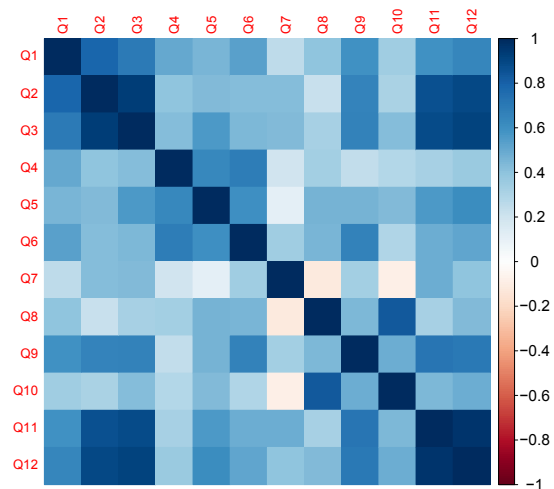
[1]http://www.audacityteam.org/

Fig. 3. Correlation matrix of 12 questions answered by the clinical psychologists. The blue color (value: 0 to +1) shows positive correlation and the red color (value: 0 to −1) shows negative correlation.
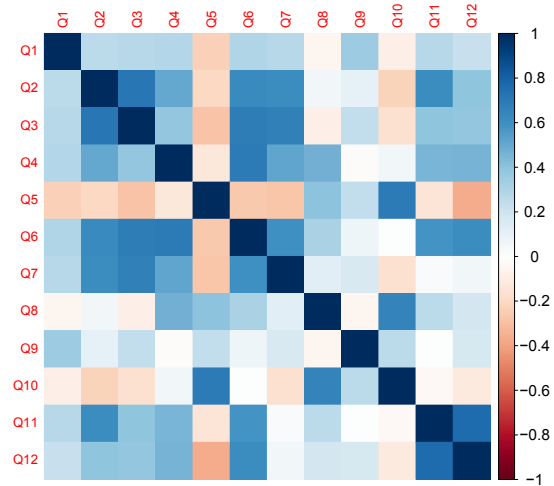

Fig. 4. Correlation matrix of 12 questions answered by the students. The blue color (value: 0 to +1) shows positive correlation and the red color (value: 0 to −1) shows negative correlation.

questions in the students (e.g. 0.26 ($p > .05$) between Q1 and Q2) as shown in Figure 4. This indicates that the students evaluated independently for each question.

B. Correlation between clinical psychologists and students

As shown in Table III, most of the ratings between the clinical psychologists and the students were not correlated (around 0.1 to 0.2). For example, there was no correlation regarding Q1 (0.14, $p > .05$) and especially Q8 (0.07, $p > .05$). This shows some gaps between dialogue attendees in their feelings because the students may not be familiar with this type of conversation and questionnaire compared to the clinical

2016 Conference of The Oriental Chapter of International Committee
for Coordination and Standardization of Speech Databases and Assessment Technique (O-COCOSDA)
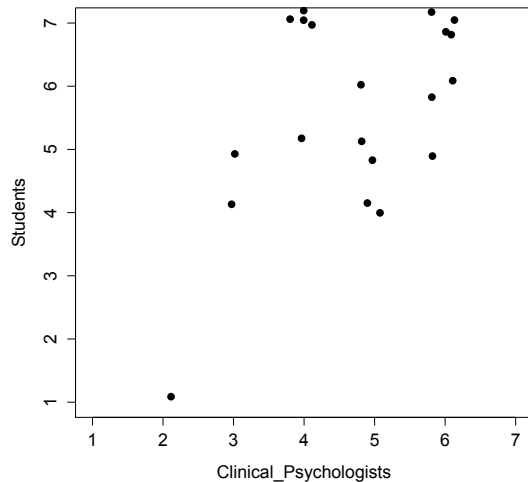26-28 October 2016, Bali, Indonesia

Fig. 5. Plots of Q5 in the clinical psychologists and the students. A small amounts of noise was added to separate same value points.

psychologists as referred to the previous subsection. However, we found that Q5 (I was able to talk topics I prepared) was significant correlated as shown in Figure 5. This indicates that the clinical psychologists obtain information that the partner has something he or she want to talk.

Here, the question Q13 and Q14 were a binary type of question. For Q13, we found that nine out of 20 was answered as yes in the clinical psychologists, and six out of 20 was answered as yes in the students. An agreement rate of Q13 was 0.65. For Q14, we found that four out of 20 was answered as yes in the clinical psychologists, and four out of 20 was answered as yes in the students. An agreement rate of Q14 was 0.60. These agreements were a moderate. This also shows some gaps between dialogue attendees in their feelings.

TABLE III
CORRELATION COEFFICIENTS BETWEEN THE CLINICAL PSYCHOLOGISTS
AND THE STUDENTS (*: P < .05).

Question	Correlation coefficient
Q1	0.14
Q2	0.19
Q3	0.15
Q4	0.13
Q5	0.51*
Q6	0.15
Q7	0.14
Q8	0.07
Q9	0.18
Q10	0.13
Q11	0.21
Q12	0.23

V. CONCLUSION

This paper represents new data collection of multimodal information between clinical psychologists and students to model attentive listening. We obtained data of a total of 23 dialogues, and transcribed the data with dialogue acts annotation. We initially analyzed the pre and post questions of the recorded interaction. We investigated gaps of feeling of dialogue attendees and good attentive listening from the collected questionnaires by calculating correlations of inter-question and between clinical psychologists and students. The results showed that the clinical psychologists answered consistently. In contrast, the students answered with larger variations for each question. We also confirmed that most questions were not related between the clinical psychologists and the students. However, there was a correlation regarding I/he/she was able to talk topics I/he/she prepared.

Limitations of this work are as follows: 1) this data is not actual counseling as we mentioned in the previous section, 2) we did not control clinical psychologists' skills (e.g. years of experience and types of current work). We plan to control these parameters and recruit unexperienced populations such as university students as attentive listeners. This data can be used for creating dialogue system incorporating with attentive listening and to model good attentive listeners. To do so, we need to model multimodal verbal and non-verbal cues of good attentive listeners [8], [10].

REFERENCES

[1] A. Bachelor, "Clients' perception of the therapeutic alliance: A qualitative analysis." *Journal of Counseling Psychology*, vol. 42, no. 3, p. 323, 1995.

[2] A. S. Bellack, *Social skills training for schizophrenia: A step-by-step guide.* Guilford Press, 2004.

[3] G. W. Joe, D. D. Simpson, D. F. Dansereau, and G. A. Rowan-Szal, "Relationships between counseling rapport and drug abuse treatment outcomes," *Psychiatric services*, 2001.

[4] S. Ohta, H. Nakamoto, Y. Shinagawa, and T. Tanikawa, "A health monitoring system for elderly people living alone," *Journal of telemedicine and telecare*, vol. 8, no. 3, pp. 151–156, 2002.

[5] J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson, "Loneliness and social isolation as risk factors for mortality a meta-analytic review," *Perspectives on Psychological Science*, vol. 10, no. 2, pp. 227–237, 2015.

[6] M. I. Tanveer, E. Lin, and M. E. Hoque, "Rhema: A real-time in-situ intelligent interface to help people with public speaking," in *Proceedings of the 20th International Conference on Intelligent User Interfaces.* ACM, 2015, pp. 286–295.

[7] H. Tanaka, S. Sakti, G. Neubig, T. Toda, H. Negoro, H. Iwasaka, and S. Nakamura, "Automated social skills trainer," in *Proceedings of the 20th International Conference on Intelligent User Interfaces.* ACM, 2015, pp. 17–27.

[8] B. Xiao, P. G. Georgiou, Z. Imel, D. Atkins, and S. S. Narayanan, "Modeling therapist empathy and vocal entrainment in drug addiction counseling," in *International Speech and Communication Association*, 2013.

[9] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems.* International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068.

[10] T. Kawahara, M. Uesato, K. Yoshino, and K. Takanashi, "Toward adaptive generation of backchannels for attentive listening agents," in *International Workshop Serien on Spoken Dialogue Systems Technology*, 2015, pp. 1–10.