Manual and Accelerometer Analysis of Head Nodding Patterns in Goal-oriented Dialogues

Masashi Inoue[1,2], Toshio Irino[3], Nobuhiro Furuyama[4,5], Ryoko Hanada[6], Takako Ichinomiya[7], and Hiroyasu Massaki[8]

[1] Graduate School of Science and Engineering, Yamagata University, Yonezawa, Japan
[2] Collaborative Research Unit, National Institute of Informatics, Tokyo, Japan
[3] Faculty of Systems Engineering,Wakayama University, Wakayama, Japan
[4] Information and Society Research Division,
National Institute of Informatics, Tokyo, Japan
[5] Department of Computational Intelligence and Systems Science,
Tokyo Institute of Technology, Tokyo, Japan
[6] Graduate School of Clinical Psychology/Center for Clinical Psychology and Education,
Kyoto University of Education, Kyoto, Japan
[7] Graduate School of Clinical Psychology, Kyoto University of Education, Kyoto, Japan
[8] Department of Informatics, Sokendai, Tokyo, Japan
mi@yz.yamagata-u.ac.jp, irino@sys.wakayama-u.ac.jp,
furuyama@nii.ac.jp, {hanada,din95002}@kyokyo-u.ac.jp,
ssaki@goo.jp

Abstract. We studied communication patterns in face-to-face dialogues between people for the purpose of identifying conversation features that can be exploited to improve human-computer interactions. We chose to study the psychological counseling setting as it provides good examples of task-oriented dialogues. The dialogues between two participants, therapist and client, were video recorded. The participants' head movements were measured by using head-mounted accelerometers. The relationship between the dialogue process and head nodding frequency was analyzed on the basis of manual annotations. The segments where nods of the two participants correlated were identified on the basis of the accelerometer data. Our analysis suggests that there are characteristic nodding patterns in different dialogue stages.

Keywords: Dialogue analysis, Face-to-face, Head nodding, Accelerometer.

1 Introduction

We shall describe our method of analyzing face-to-face interactions to find structural patterns in dialogues that may be useful in designing human-computer interactions. Note that various modalities have been investigated, such as the relationship between hand gestures and verbal miscommunications [1] and the relationship between body synchronicities and empathy in dialogues [2, 3].

Our focus is on head gestures, especially nodding. We quantitatively analyzed the usage of head nods in dialogues. Nodding is important because it is strongly tied with

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2011, LNCS 6762, pp. 259–267, 2011.

the empathy that is the basis of the bond between participants of the interaction and has received much attention among various head movements [4]. The relationship between nods and the mood of the interaction can be better understood by observing task-oriented dialogues, especially those of psychological counseling. As Rogers pointed out [5], it is critically important for a therapist to listen to the client skillfully for the counseling to be successful. We expected that gestures, especially the time course of the nodding pattern, would play a certain role in giving an impression that the client is being attended by the therapist.

The two participants in counseling play quite different roles: therapists and clients. We measured the head nodding frequencies and their timings in relation to the changes in the dialogue's mood. Manual annotations made by a psychotherapist and a mechanical recording of head movements by using an accelerometer were used to analyze the relationship. The results showed that the changes in the nodding patterns are related to dialogue stages.

The present research project is concerned with the following two questions: 1) Does the frequency of nods of the therapist and the client change as therapeutic stages (initial stage, exploration stage, struggling stage, and closing stage) progress? 2) Does the phase-lag of nods between the therapist and the client change as the therapeutic stages progress?

1.1 Related Work

Head nodding in dialogues has been studied as a nonverbal behavior in the field of psychology [6, 7, 8]. However, these studies were either experimental ones that did not correspond to real dialogues or were summarizations of entire dialogues. We studied the time course of nodding patterns during actual task-oriented dialogues by using detailed manual coding and objective sensor data. The concept of dialogue stages is especially important when we want to apply our knowledge on effective interaction to computer systems that manage complex dialogue tasks rather than single-shot exchanges of messages. Another important aspect is that we contrasted dialogues conducted by specialists of different skill levels. Even psychotherapists can fail in counseling because the task is so complex. This aspect has not been well explored in the previous studies.

2 Data Collection

We recorded two counseling interviews. The therapists, clients, and topics varied. Each dataset was recorded on different dates and completed in a single interview session. The length of the interview was not pre-determined and could be controlled by the participants during the dialogue. The problems discussed were actual problems and not ones of role-plays that used specific roles and scenarios. The properties of the datasets are listed in Table 1. All participants used Japanese as their language of communication. The participants consented to being video recorded for research purposes. The first counseling was carried out by a novice therapist who is a graduate student studying psychotherapy. The second counseling was done by an expert therapist who supervises students majoring in psychotherapy. Both sessions completed

Table 1. Overview of datasets

Dataset	Duration (min.sec)	Therapist (level of experience)	Client
1	47.36	Female (beginner)	Male
2	29.49	Female (expert)	Female

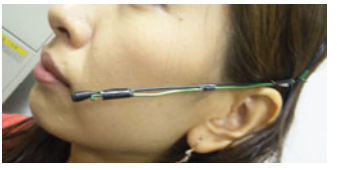

Fig. 1. Microphone placed near the mouth

Fig. 2. Accelerometers mounted in the occipital region

successfully but in different ways. In the first counseling, the client felt that he could organize his ideas and was relieved by talking. In the second counseling, the therapist proposed an intervention and the client agreed to try it with satisfaction.

All of the interviews were video-recorded. The participants faced each other, and the video camera was situated by their side and it captured images of their whole bodies. We used microphones (DPA 4060-BM) to clearly record their voices and triaxial head-mounted accelerometers (Kionix KXM52-1050) to record their head movements, as shown in Figure 1 and 2. Speech and acceleration were recorded at a 50-kHz (first data) and 20-kHz (second data) sampling rate. Although the accelerometers recorded triaxially, we used only the acceleration along the vertical axis; that is, we measured only the up and down movements that corresponded to nodding movements. The sensor itself was 1 square centimeter in size and could not be seen by the other person when mounted in the occipital region. Thanks to the non-invasive characteristics of our sensor mounting systems, the participants reported that they could maintain a natural dialogue.

3 Method

We conducted two analyses. The first was based on the manual coding of the differences in head nod frequencies at different stages of the dialogue. In the first analysis, a non-participating psychotherapist used ELAN annotation software to

manually annotate nods in both recordings. Each nod was identified as a segment from the beginning of the head movement to its end. The second analysis dealt with the degree of nodding synchronization between participants as revealed by the sensor time series data of the accelerometer. Therapists sometimes moved their head simultaneously with clients; while in other situations, they delayed in nodding. These differences may correspond to the task that the therapists have to conduct during the stages. The acceleration signals of head in the upward and downward directions were used as is regardless if they were manually annotated as nods or not. Cross correlation was calculated by using a 0.5 second window (sliding in 0.1 second increments) on the data from the two accelerometers. This analysis was meant to determine if therapists coordinate their nods with the clients' and if the degree of synchronization is associated with any events in the counseling sessions. The second analysis is explained in detail in the next section.

The usage of nods by therapists varies from stage-to-stage because the different stages correspond to the different roles taken to achieve therapeutic goals. In a past study, we found that there are different therapeutic stages in interview dialogues and they can be characterized by the occurrence patterns of speech types [9]. In this study, we tried to determine whether the nodding pattern changes according to the goals that therapists try to achieve in each stage.

3.1 Calculation of Cross Correlations

Acceleration data. The acceleration data were used as measures for the nodding movements. The accelerometer (KXM52-1050) outputs analog waveforms corresponding to the acceleration values along three orthogonal axes (X, Y, and Z). The acceleration waveforms and the speech of the therapist and the client were simultaneously digitized by using an eight channel data recorder (EZ7510) with a sampling rate of 50kHz (first data) and 20 kHz (second data). For the measurement setup, the nodding movements roughly corresponded to the Y values. The cross correlation was calculated for the Y data of the therapist and the client.

Cross correlation. The acceleration data was resampled at the sampling rate of 1 kHz and filtered by a notch filter of 60 Hz and a lowpass filter to reduce hum and high frequency noise. Denoting the denoised signal as $\mathbf{x} = (x_1, x_2, x_3, \cdots, x_N)$, the normalized signal, $\mathbf{x}_A$, is

$$\mathbf{x}_A = \frac{\mathbf{x} - \bar{\mathbf{x}}}{\| \mathbf{x} - \bar{\mathbf{x}} \|}$$

where $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} x_i$ (average) and $\| \mathbf{x} \| = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$ (rms value). A window function, $\mathbf{w}$, was applied to the normalized signal, $\mathbf{x}_A$, to extract a short frame (or segment) of the data:

$$\mathbf{x}_w = \mathbf{x}_A \cdot \mathbf{w}.$$

The duration of the window function was 500 ms, and the frame shift was 100 ms. The window function had a flat part of 400 ms and tapers conforming a raised cosine

function (50 ms × 2). The cross correlation $R_w(n_\tau)$ for one frame between the windowed signals for the client, $\mathbf{x}_{w_1}$, and the therapist, $\mathbf{x}_{w_2}$, is

$$R_w(n_\tau) = \sum_{n=-\infty}^{\infty} x_{w_1}(n+n_\tau) \cdot x_{w_2}(n),$$

where n_τ is the lag time between the client's nod and the therapist's nod in the sample. The correlation value, r, is the peak value (either positive or negative) of the cross correlation $R_w(n_\tau)$:

$$r = \max(\text{abs}(R_w(n_\tau)). \quad (1)$$

The correlation lag, $n_{\tau p}$, is defined as the lag at the peak. The correlation value was derived for every frame, and the sequence was derived for the whole therapy session. The maximum absolute value of r of all frames between the first and last three minutes of the session was used to normalize r. Thus, the peak value of r was unity in the session, excluding first and last three minutes.

4 Results and Discussion

4.1 Nodding Frequency

The nodding frequencies of the therapist (Th) and client (Cl) are shown in Figures 3 and 4, where the height of each bar represents the frequencies in 50 second segments and the horizontal axes represent time from the beginning of the session. To understand the overall trends rather than individual values in the time segments, we smoothed the frequencies by averaging the five preceding and five subsequent bins around the target bin and plotted these averages as lines.

The first analysis revealed that there were M-shaped changes in the therapists' nodding frequencies in both counseling dialogues, as shown in the upper graphs of Figure 3 and Figure 4. From the beginning of dialogue, the frequencies increased and then dropped for a while. Then it increased again, before decreasing toward the end of the session. The initial increase and final decrease can be understood as natural changes that occur at the starts and ends of dialogues. However, the reason for the drops in the middle of the dialogues is not clear. It is also interesting that the timing of the drops differed between the therapist and client. Accordingly, we closely looked at what was happening during these video segments. The lowest frequency of the first (beginner) therapist occurred when the client uttered "I'm not characterized as a parent" referring to the result of personality test he had taken. The drop might have happened because the client humbled himself by talking about his immaturity whereas the therapist wanted to mediate him and could not respond with an affirmative nod. The lowest frequency of the first client occurred when the therapist complimented the client by saying, "You can see yourself very well", but the client responded humbly by saying "No no, only this time". The drop seemingly occurred when the therapist was working on the client but the client could not accept her advice. These two drops were in different sequences but have a commonality that they happened when the

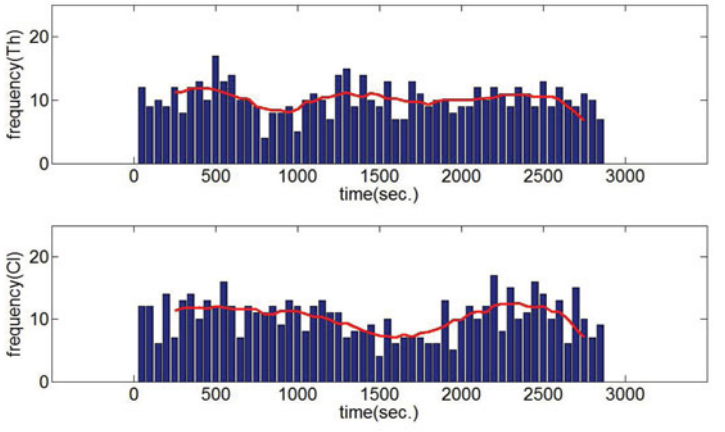

Fig. 3. Nodding frequency for the first dialogue

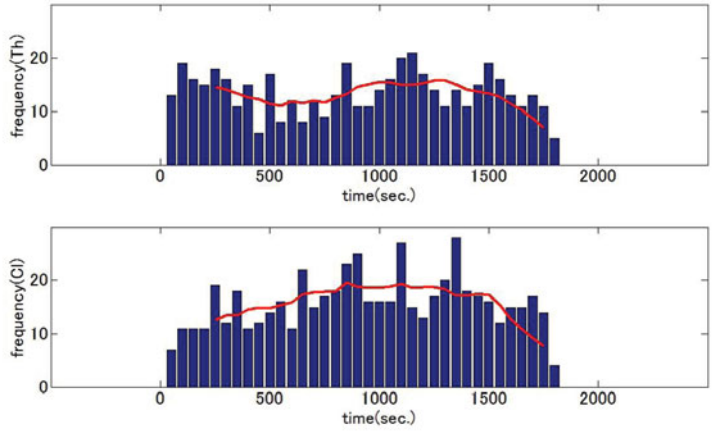

Fig. 4. Nodding frequency for the second dialogue

counseling went into the deepening stage of dealing with the client's problem. In that stage, utterances occurred to which the listeners would not or could not agree. That is, the nodding pattern changed when the counseling transited from the information gathering stage involving simple questions and answers to the stage where introspection played a central role and involved the client's monologue-like utterances.

The lower graph of Figure 3 shows that the nod frequency of the client interviewed by the novice therapist also had an M-shaped change. On the other hand, the client interviewed by the expert therapist showed an inverted U-shaped nodding frequency pattern (the lower graph of Figure 4). The lowest frequency of the first client occurred when the therapist was about to take the initiative in counseling. Until then, the client

had spontaneously detailed the problem and tried to figure out how to solve it, but she had failed in her mind to produce any clue to a solution. From that point, the therapist had to lead the session. This shift of dialogue stages, from the therapist listening to the client to her asking the client questions, might be reflected in the nodding frequency pattern of the therapist. On the other hand, there was no drop in frequency on the client's side, because the client did not go into the introspection but continued interaction with the therapist as their dialogue deepened. From the viewpoint of expertise, the first dialogue was led by the client and the inexperienced therapist simply reacted. In contrast, the second dialogue was coordinated by the experienced therapist, who was oriented toward solving the problem rather than letting the client analyze herself.

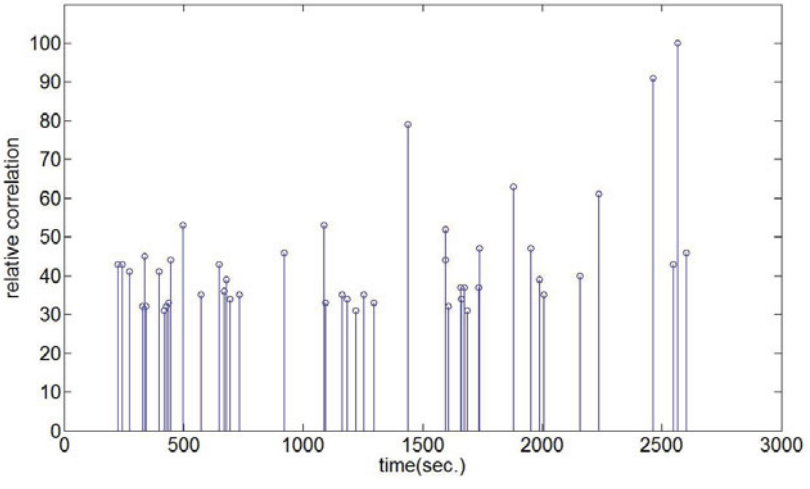

Fig. 5. Cross-correlation value of head accelerations for the first dialogue

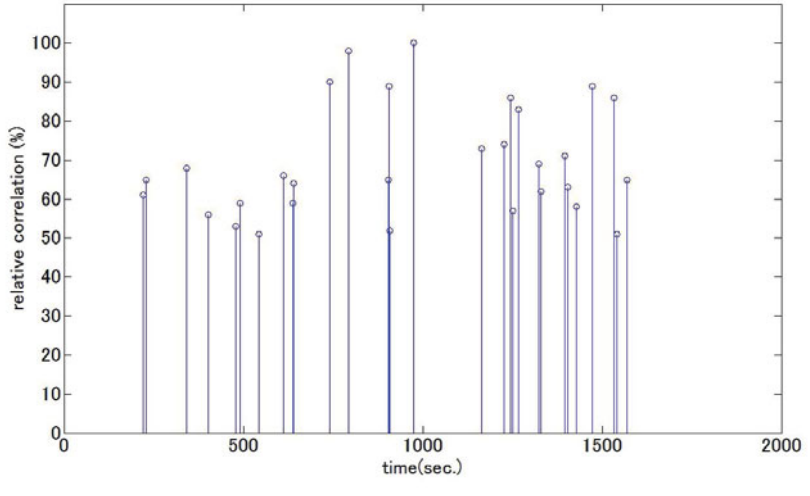

Fig. 6. Cross-correlation value of head accelerations for the second dialogue

4.2 Timing of Nodding

Figures 5 and 6 show the synchronies of head movements between therapist (Th) and client (Cl). The height of each bar represents the value of the cross correlation calculated by Equation 1, and the horizontal axes represent the time from the beginning of the session. We used the biggest acceleration value and speculated that the timing of the value may often correspond to when the head movement starts. For the ease of visual interpretation, we set threshold values as 30% and 50% of the maximum value in each figure and displayed bars higher than the thresholds.

In the first dialogue, the strongest synchrony occurred at the end of the dialogue when the client mentioned how nice his parents were. The therapist was trying to confirm the client's positive perception by using strongly correlated nodding. In the second dialogue, strong synchronies occurred in two places. In the first interaction, the client who had been replying negatively to the therapist eventually replied positively. At that point, the therapist seized upon the chance and tried to reach an agreement by nodding her head. In the second interaction, the therapist checked the facts surrounding the problem and asked for confirmation by the client in each utterance. The confirmation was accompanied by nodding. After this interaction, the client started talking lively.

The second analysis revealed that the degree of nodding synchronization was not segmental as seen in the frequency patterns; rather, it corresponded to particular mood shifts. By analyzing the content, we identified these interesting points where nodding synchronized differently between the two therapists. In the counseling of the novice therapist, nods synchronized as the client talked freely, while in the counseling of the experienced therapist, nods became strongly synchronized when the therapist started to intervene in the problem of the client. This difference can be interpreted as due to a similar difference of activity and passivity as in the first analysis.

5 Conclusion

To obtain insights for developing better human-computer interactions, we analyzed head nodding patterns in task-oriented dialogues of psychotherapy. By annotating nodding in videos of client-therapist interactions, we found that there are patterns of nodding frequency changes as the counseling proceeds. The degree of synchronization between two participants was measured by using accelerometers. We found that distinctive communication events occurred when head movements strongly correlated. In addition, by comparing therapists having different degrees of expertise, we identified the possibility that these head nodding patterns vary according to the skill level of the therapist.

In the future, we will examine the relationship between hand-annotated head nodding segments and acceleration signals and increase the amount of data to be analyzed in order to generalize the head nodding patterns found in this research. Other modalities such as speech and hand gestures will be also considered, especially as regards their interactions with nods.

Acknowledgments. This research was partially supported by Grants-in-Aid for Scientific Research 19530620 and 21500266, and the research grant from Kayamori Foundation of Informational Science Advancement.

References

1. Inoue, M., Ogihara, M., Hanada, R., Furuyama, N.: Gestural cue analysis in automated semantic miscommunication annotation. Multimedia Tools and Applications (to appear)
2. Maurer, R.E., Tindall, J.F.: Effect of postural congruence on client's perception of counselor empathy. Journal of Counseling Psychology 30, 158–163 (1983)
3. Nagaoka, C., Komori, M.: Body movement synchrony in psychotherapeutic counseling: A study using the video-based quantification method. IEICE Trans. Info. Sys. E91-D(6), 1634–1640 (2008)
4. Harrigan, J.A.: Proxemics, kinesics, and gaze. In: The Handbook of Methods in Nonverbal Behavior Research. Oxford University Press, New York (2008)
5. Rogers, C.: Client-Centered Therapy: Its Current Practice, Implications, and Theory. Houghton Mifflin, Boston (1951)
6. Matarazzo, J.D., Saslow, G., Wiens, A.N., Weitman, M., Allen, B.V.: Interviewer head nodding and interviewee speech durations. Psychotherapy: Theory, Research & Practice 1(2), 54–63 (1964)
7. McClave, E.Z.: Linguistic functions of head movements in the context of speech. Journal of Pragmatics 32, 855–878 (2000)
8. Maynard, S.: Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation. Journal of Pragmatics 11, 589–606 (1987)
9. Inoue, M., Hanada, R., Furuyama, N.: Assessment of counseling stages to better understand the process toward solutions. In: 2008 Conference on Solution-Focused Practice, Austin, Texas, November 12-16 (2008)