Electronic Preprint for Journal of Information Processing Vol.25

Technical Note

Analyzing Listeners' Empathy by Their Nonverbal Behaviors in Bibliobattle

Kohei Matsumura[1,a)] Yasuyuki Sumi[2,b)] Mitsuki Sugiya[2]

Received: August 5, 2016, Accepted: December 1, 2016

Abstract: Nonverbal information plays an important role to convey feelings and/or interests of the people in conversations. Since Bibliobattle, a book-review game, has pleasant features to investigate non-verbal information on conversation settings, we conduct a series of experiments on Bibliobattle settings. In Bibliobattle, each speaker presents his/her own recommended book to listeners as a bibliobattler in 5 minutes. At the end of all presentations, everyone votes for the champion book. We analyzed a series of Bibliobattle experiments by video investigation. In the analysis, we focused on the listeners' non-verbal information, in particular, nods, laughs and change postures. Our results showed that there are co-occurrence of nonverbal action among the audience in Bibliobattles. The frequency of co-occurrence of positive non-verbal information were assumed to be excitement of the presentation. However, interestingly, the results showed that the frequency does not affect the result of voting for the champion book in Bibliobattle. We discuss the cause of the results in the paper.

Keywords: nonverbal information, Bibliobattle, co-occurrene

1. Introduction

When people do a presentation, they may express their expression not only by their spoken language, but also by gestures. However, it has not well studied the role of the gestures in a presentation. We conducted a series of experiments on Bibliobattle settings because Bibliobattle has pleasant features to investigate nonverbal information on presentation settings. Bibliobattle is proposed by Taniguchi et al. [1]. Bibliobattle is regulated in time, i.e., individual presenters have five minutes presentation on books and two to three minutes follow for question and answering, so we can systematically observe semi-structured speech as well as interactive conversation. Also, we can have voting results by the participants in Bibliobattle so that we can evaluate our method to evaluate each listener's level of interest to each presentation.

We conducted four Bibliobattle games and recorded book presentations by video cameras. We also used an optical motion capture system and recorded head movements of the participants. In order to investigate the relationship between nonverbal actions of the audience and the result of votes, we classified the audience's actions into three types,"nod","laugh" and "change posture," and analyzed co-occurrence of these actions among the audience. Our results showed that there are co-occurrence of nonverbal actions among the audience in Bibliobattles, however, it is hard to find that a relationship between nonverbal actions of the audience and the result of votes. We discuss the cause of the results in the paper.

The contributions of the study are as follows:

- Proposed to use Bibliobattle games to investigate nonverbal information on presentation settings.
- Investigated co-occurrence pattern of body gestures using motion capture system.
- Found roles of empathy in the presentation game.

2. Related Works

Our proposed approach is to infer listeners' empathy and agreement to a speech by the listeners' nonverbal responses. This is related to the general domain of conversation analysis, a study of social interaction in situations of formal meetings and everyday casual conversations. Traditional linguistic researches have focused on verbal information such as semantic analysis and its contextual understanding of contents of speech. But recently, there have been many works focusing on nonverbal information such as hand-gesture, head-gesture, gaze, turn-taking, posture, etc. produced by speakers and listeners during the conversations [2]. The paper is in the context of the approach by focusing on listeners' head-nods, laughter, and posture changing as responses to a speech.

There have been works for detecting participants' influence levels in meeting, predicting next speakers [3], and indexing meetings [4] in order to realize smart systems that are socially embedded in our daily social setting and timely response to our intention. Rienks et al. addressed the problem of automatically detecting participant's influence levels in meetings by easily sensed features such as the number of turns and successful interruptions [5]. Otsuka et al. propose a probabilistic framework to predict next speaker in casual multi-party conversation by gaze patterns, head directions and utterances [6]. Kawahara et al. have attempted to detect listeners' understanding and level of interest on poster presentations by multimodal information (pointing, gaze,

[1] Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan
[2] Future University Hakodate, Hakodate, Hokkaido 041–8655, Japan
[a)] matsumur@acm.org
[b)] sumi@acm.org

nodding, backchannel and laughter) sensed by multiple types of sensors [7] and applied its findings to a smart digital signage system [8]. Katagiri et al. proposed a statistical method to estimate proposals in multi-party consensus-building conversations with head direction/movement and speech power [9].

We address the problem of inferring listeners' empathy and agreement to a short speech from nonverbal responses such as nodding and laughter. Back-channel signals have a powerful effect on speakers [10]. Listeners' head-nods and smiles give a speaker positive feedback, for example agreement and approval. The absence of back-channel signals is taken as a negative reaction. We assume listeners' empathy to speech is roughly evaluated by the number of noddings and laughters. Stivers discussed the role of nodding [11]. Our study also contribute the realm of stance-taking on conversation analysis field.

3. Experiment

3.1 Experimental Field: Bibliobattle

We conducted four Bibliobattles in a series of experiments. Twelve unique subjects aged 21 to 45 (11 males, 1 female) participated in the series of experiments. Three presenters and audience of three are participated in a Bibliobattle. That is, a cumulative total of 24 subjects participated in the series of experiments. Each Bibliobattle consists of three sessions and voting. The details of each session and voting are as follows:

Session Each session consists of two periods, five minutes for a presentation period and three minutes for a Q&A period. In the presentation, a presenter introduces his/her recommended book for 5 minutes without any materials excluding the book. The presenter also acts as audience while another presenter is giving his/her presentation. That is, in the presentation, there will be six participants, a presenter and audience of five. In addition, the order of presentation is decided by Scissor-Paper-Rock.

Voting After all three presentations, participants vote by a show of hands on the most interesting presentation introduced by others. The vote-winner will be the champion book of the Bibliobattle. This voting can be utilized as one of an evaluation of a series of presentation. We investigate the relationship between nonverbal information and this evaluation (i.e., result of the voting) in Bibliobattle.

Figure 1 shows our experimental setup in Bibliobattle; the audience sat on a box in a semicircle with about 0.5 m spacing, and listened to presentations. The seating order was random. That is, each participant sat down in an unoccupied seat (i.e., box).

3.2 Equipment

In order to measure participants' head motions, we employed an optical motion capture system, OptiTrack V120. The system consists of 14 cameras and its measurement area is about 8×8 m. We attached an optical marker-set onto the top of a cap (shown in the bottom left of Fig. 1) and asked each participant to wear the cap. Using the motion capture system, it will be possible to detect participants' gestures and joint attention automatically. In addition to the motion capturing, we recorded each Bibliobattle using four video cameras. As shown in Fig. 1, four cameras surrounded

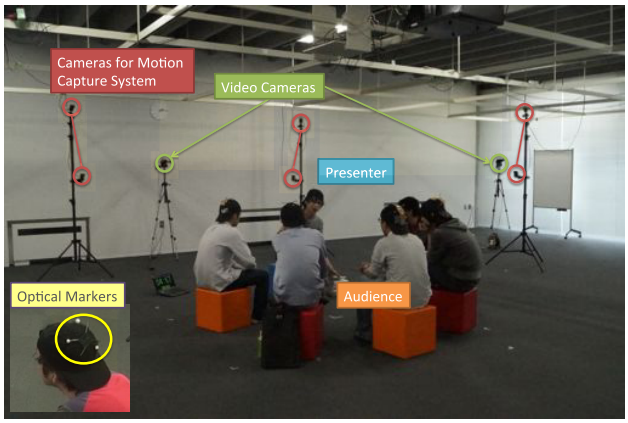

Fig. 1 The experimental setup of a Bibliobattle experiment. We employed a motion capture system and four video cameras to record participants' behavior in Bibliobattles.

participants with 3 to 5 m distances. The recorded videos are used for transcription of each presentation and investigating audience behavior.

3.3 Analysis: Co-occurrence of Actions

In order to investigate the relationship between nonverbal information expressed by the audience and the results of the vote in Bibliobattle, we analyzed recorded videos in terms of co-occurrence of actions within the audience in particular. In the analysis, we firstly classified audience's actions into three types, "nod", "laugh" and "change posture." The details of each type are as follows:

Nod When the audience repeatedly moves his/her head up and down, we classify the action as "nod." This action expresses one's assent or understanding in general. The duration of the action is varied from less than a second to a couple of seconds.

Laugh When the audience laughs in response to the presentation, we classify the action as "laugh." We distinguish the action from wry smile because the expressions of these two actions seem to differ from each other.

Change Posture When the audience changes his/her posture, we classify the action as "change posture." This type of action includes re-cross his/her legs action, reseat action and lean forward/backward action. This type of action indicates changes of the audience's interest level or increase/decrease in concentration.

We then manually make annotations into each audience's movement by watching the series of recorded videos using a video annotation software, iCorpusStudio [12]. Since the software can play multiple videos and annotated timeline simultaneously (**Fig. 2**), it is suitable for the purpose.

Figure 3 shows that an annotated timeline of a session in the Bibliobattle. We manually enumerated co-occurrence actions among the audience. As shown in Fig. 3, we enumerate strictly overlapped annotations among subjects as co-occurrence points. We not only enumerated the co-occurrence points, but also counted the number of people involved in each co-occurred action. We enumerated strictly overlapped annotations among subjects as co-occurrence points. That is, if there is a gap be-

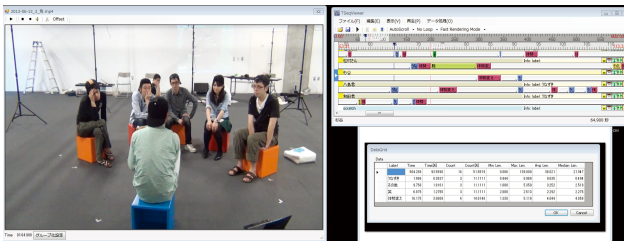

Fig. 2 A screenshot of iCorpusStudio. The software plays multiple video files and annotated timeline data simultaneously.

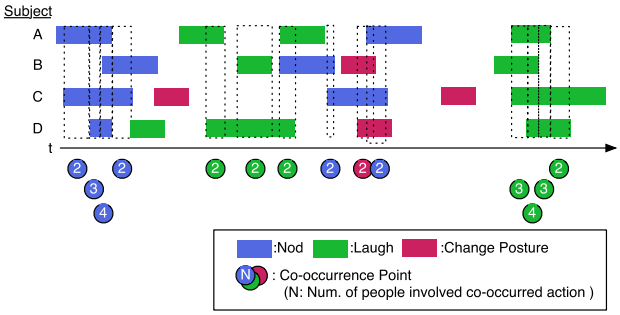

Fig. 3 An annotated timeline of a session in the Bibliobattle. We manually enumerate co-occurrence points (overlapped annotations among subjects).

Table 1 The results of votes and the numbers of the audience's actions in Bibliobattle. Where **N** means 'Nod', **L** means 'Laugh' and **CP** means 'Change Posture'.

Bibliobattle 1

Subject	Num. of Votes	Number of Actions N	L	CP
A	**3**	46	29	22
B	1	11	8	51
C	2	43	17	47

Bibliobattle 2

Subject	Num. of Votes	Number of Actions N	L	CP
D	-	29	11	27
C	1	142	32	39
B	**5**	90	48	44

Bibliobattle 3

Subject	Num. of Votes	Number of Actions N	L	CP
E	2	58	43	27
F	-	31	29	31
A	**4**	60	28	37

Bibliobattle 4

Subject	Num. of Votes	Number of Actions N	L	CP
G	-	24	18	24
B	**4**	83	8	44
C	2	144	22	40

tween two actions that occured almost the same time, we did not consider them as co-occurred actions no matter how small gap.

4. Result

As the result of the series of experiments, we collected motion data as well as videos for four Bibliobattles, 12 presentations in total. **Table 1** shows the results of each Bibliobattle. The number of votes indicated in bold face means the champion book of each Bibliobattle. The subject A, B, A and B wins Bibliobattle 1, 2, 3 and 4 respectively. This table also shows the total numbers

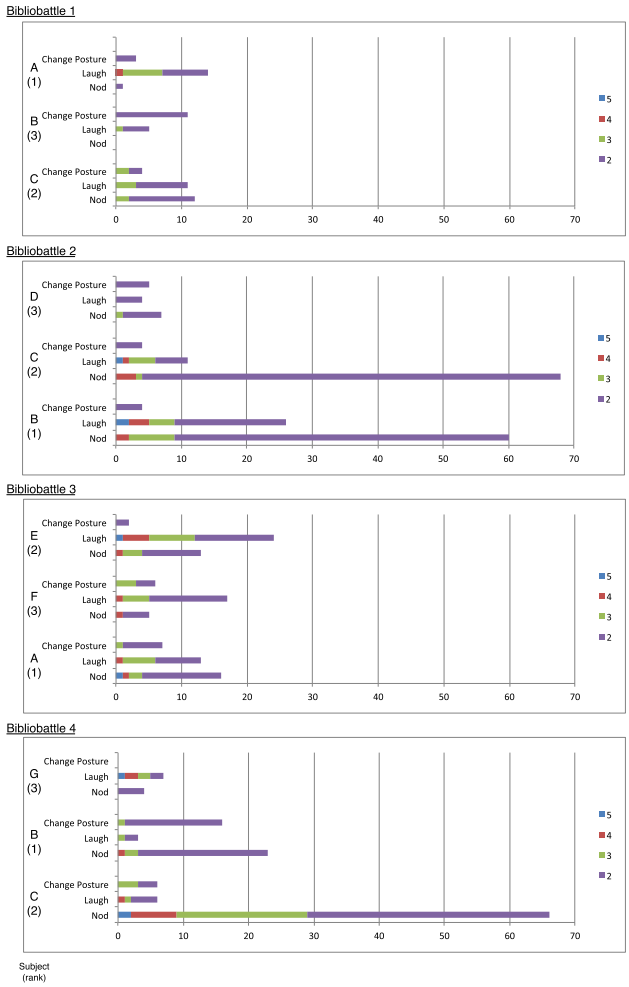

Fig. 4 The numerical result of co-occurrence actions. The color means the number of subjects involved in a co-occurrence action; blue, red, green and purple means 5, 4, 3 and 2 respectively. The characters placed at left side of each graph means subject ID and rank for the Bibliobattle. For example, "A(1)" means subject ID is A and he/she get the 1st place in the Bibliobattle.

of three actions, "nod", "laugh" and "change posture," of the audience in each presentation. We could not find any implications from the table excluding one fact that the total number of actions does not affect the result of the vote.

4.1 Co-occurrence

We analyzed co-occurrence of actions among the audience. The action of the audience often co-occured with one acted by the others. In order to investigate a relationship between nonverbal information and the results of the vote in Bibliobattle, analyzing co-occurrence of the actions seems to be promising. **Figure 4** shows the numerical result of co-occurred actions.

The frequency of co-occurrences of positive non-verbal information, "Laugh" in this analysis, seems to be excitement of the presentation, also, "Nod" seems to be agreement of the presenter's argument. The results showed that the frequency does not directly affect the result of voting for the champion book in Bibliobattle; for example, in Bibliobattle 3, subject A wins the battle where the frequency of co-occurrence of "Laugh" action is much lower compares to the others. However, it showed that it is necessary to have an excitement (i.e., "Laugh") and an agreement (i.e.,

"Nod") to win the Bibliobattle.

The result leaves much room for discussion. We will further investigate the implications of the results in the future. The following points of view are considerable for the future work:

Classifying audience's actions We classified the audience's action into three types, "nod", "laugh" and "change posture." However, the nod action may include both positive and negative emotions. When we listen to a poor presentation, the nod action means encouraging his/her presentation. We should distinguish the nod action into positive and negative. Also the criteria for classifying actions are considerable.

Criteria for determining co-occurrence point We enumerated strictly overlapped annotations among subjects as co-occurrence points. That is, if there is a gap between two actions that occurred almost the same time, we did not consider them as co-occurred actions no matter how small gap. However, the durations of nonverbal actions are varied; for example, the duration of the nod action will be less than a second to a couple of seconds. It seems to be worthy to make gap allowance to determine co-occurrence point.

Relationship between presenter and the audience In this study, we focused on the audience's nonverbal actions. However, the interactions between presenter's action and/or utterance and the audience's action are worthy to investigate. Also, the conversational analysis (i.e., the content of utterances of the presenter) is required to understand co-occurrence actions.

5. Concluding Remarks

In this paper, we analyzed the audience's empathy by their nonverbal behaviors in Bibliobattle. Biblobattle is a game where the audience decides the most popular book by voting on books. Since this voting can be utilized as evaluations of the presentation, it is suitable for investigation.

We conducted a series of experiments on Bibliobattle settings. Four Bibliobattles are held in the series of experiments. We employed an optical motion capture system and four video cameras to record the presentations and participants' motions in Bibliobattle.

We analyzed recorded videos in terms of co-occurrence of actions within the audience. Our result showed that there are co-occurrence of nonverbal actions among the audience in Bibliobattles, however, it is hard to find relationships between nonverbal actions of the audience and the result of votes.

The result leaves much room for discussion and raises our future work. In this study, we investigated the relationship between the audiences' nonverbal actions occurred in presentations and the results of the vote, however, there are no strong correlation between them. Since the audience's empathy may increase when the audience involved with the others interactively, we will widen the scope of investigation to the whole of the session.

This study focused on the audience's nonverbal actions; however, the interactions between presenter's action and/or utterance and the audience's action are worth to investigate. We plan to make transcripts of presentations and to investigate interactions between presenter's action and the audience's (re)action. It will contribute to clarify the role of the audience's gestures in a presentation.

References

[1] Taniguchi, T., Kawakami, H. and Katai, O.: Bibliobattle: Informal community scheme based on book review sessions, *Proc. 8th Int. Workshop on Social Intelligence Design* (*SID 2009*), pp.92–98 (2009).

[2] Kendon, A.: *Gesture: Visible Action as Utterance*, Cambridge University Press (2004).

[3] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Analysis of Respiration for Prediction of Who Will Be Next Speaker and When? in Multi-Party Meetings, *Proc. 16th International Conference on Multimodal Interaction*, pp.18–25, ACM (2014).

[4] Stiefelhagen, R., Yang, J. and Waibel, A.: Modeling focus of attention for meeting indexing based on multiple cues, *IEEE Trans. Neural Networks*, Vol.13, No.4, pp.928–938 (online), DOI: 10.1109/TNN.2002.1021893 (2002).

[5] Rienks, R. and Heylen, D.: Dominance detection in meetings using easily obtainable features, *2nd International Workshop on Machine Learning for Multimodal Interaction* (*MLMI 2005*), Lecture Notes in Computer Science, Vol.3869, pp.76–86, Springer (2006).

[6] Otsuka, K., Sawada, H. and Yamato, J.: Automatic inference of cross-modal nonverbal interactions in multiparty conversations: "who responds to whom, when, and how?" from gaze, head gestures, and utterances, *Proc. 9th International Conference on Multimodal Interfaces* (*ICMI 2007*), pp.255–262, ACM (2007).

[7] Kawahara, T., Setoguchi, H., Takanashi, K., Ishizuka, K. and Araki, S.: Multi-modal recording, analysis and indexing of poster sessions, *INTERSPEECH-2008*, pp.1622–1625 (2008).

[8] Tung, T., Gomez, R., Kawahara, T. and Matsuyama, T.: Group dynamics and multimodal interaction modeling using a smart digital signage, *ECCV 2012* (*Springer LNCS7581*), pp.362–371 (2012).

[9] Katagiri, Y., Matsusaka, Y., Den, Y., Enomoto, M., Ishizaki, M. and Takanashi, K.: Implicit proposal filtering in multi-party consensus-building conversations, *Proc. 9th SIGdial Workshop on Discourse and Dialogue*, pp.100–103 (2008).

[10] Argyle, M.: *Bodily Communication 2nd Ed.*, International Universities Press (1988).

[11] Stivers, T.: Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation, *Research on Language and Social Interaction*, Vol.41, No.1, pp.31–57 (online), DOI: 10.1080/08351810701691123 (2008).

[12] Sumi, Y., Yano, M. and Nishida, T.: Analysis environment of conversational structure with nonverbal multimodal data, *12th International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction* (*ICMI-MLMI 2010*), ACM (2010).

Kohei Matsumura is an Assistant Professor at Ritsumeikan University. He received his Ph.D. in Knowledge Science from Japan Advanced Institute of Science and Technology in 2010. His broad research interests include human-computer interaction, embodied cognitive science, urban informatics and novel HCI techniques. He is a member of IPSJ and ACM.

Yasuyuki Sumi is a Professor at Future University Hakodate since 2011. Before joining the university, he has been a researcher at ATR and an Associate Professor at Kyoto University. He received his B.Eng. from Waseda University in 1990, and his M.Eng. and D.Eng. degrees in Information Engineering from the University of Tokyo in 1992 and 1995. His research interests include knowledge-based systems, creativity supporting systems, interface/social agents, ubiquitous/wearable computing, Web intelligence, multimedia processing, and their applications for facilitating human interaction and collaboration.

Mitsuki Sugiya was born in 1991. She graduated from Future University Hakodate in 2014. In the university, she worked on the realm of human-computer interaction especially on nonverbal communication analysis. She is currently working as a web application developer.