Synthetic Evidential Study for Deepening Inside Their Heart

Takashi Ookaki, Masakazu Abe, Masahiro Yoshino, Yoshimasa Ohmoto, and Toyoaki Nishida(✉)

Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto, Japan
{ookaki,abe,yoshino}@ii.ist.i.kyoto-u.ac.jp,
{ohmoto,nishida}@i.kyoto-u.ac.jp

Abstract. Synthetic evidential study (SES) is a novel technology-enhanced methodology that combines dramatic role play and group discussion to help people learn by spinning stories comprised of partial thoughts and evidence. The SES Support System combines a game engine and an augmented conversation technology to help with the facilitation of SES workshops. We performed a feasibility study with a partial implementation of SES and simplified SES sessions and obtained numerous supportive findings.

Keywords: Inside understanding · Group discussion and learning · Intelligent virtual agents · Theatrical role play · Narrative technology

1 Introduction

Inside understanding is deemed mandatory in establishing empathic relationships in a hybrid society that contains not only people but also social agents as first citizens. The long-term research goal underlying this paper is to establish a methodology that not only helps people achieve an inside understanding of each other but also endows artificial agents with the same ability, enabling both people and artificial agents to live and work in harmony with each other. Our approach relies on the sharing hypothesis [1], which implies that sharing the common ground is a key to empathy. Although group discussions and learning are powerful means of sharing knowledge about inside understanding, the success of such techniques depends largely on the cognitive ability of the participants, as the discussions are mostly based on indirect experiences brought about by secondary presentation materials. Moreover, the methodology is not directly applicable to help artificial agents feel empathy. We believe a solution to these problems is dramatic role playing, which not only helps people simulate subjective experience to delve deep inside a given role but also provides a useful way for artificial agents to obtain data about how people deepen their inside understanding. In fact, dramatic role playing has been used in various contexts of education [2]. Meanwhile, Cavazza et al [3] demonstrated that immersive interactive story telling is useful in gaining the first-person understanding of the story in the area of intelligent virtual agents. In order for such an approach to scale up, a significant work is necessary for effective content production.

M. Ali et al. (Eds.): IEA/AIE 2015, LNAI 9101, pp. 161–170, 2015.
DOI: 10.1007/978-3-319-19066-2_16

Synthetic evidential study (SES) combines dramatic role play and group discussion to help people spin stories by bringing together partial thoughts and evidence [4, 5]. The SES framework consists of SES sessions and interpretation archives. In each SES session, participants repeat a cycle consisting of dramatic role play, its projection into an annotated agent play, and a group discussion. In an SES workshop, one or more successive SES sessions are executed until participants come to an initial agreement.

In the dramatic role play phase, participants play respective roles to demonstrate their first-person interpretation in a virtual space. This allows them to interpret the given subject from the viewpoint of an assigned role. In the projection phase, an annotated agent play is produced on a game engine from the dramatic role play in the previous phase by applying oral edit commands (if any) to dramatic actions by the actors elicited from the behaviors of all the actors. We use this play for refinement and extension in the later SES sessions or even adapt it for reuse in a new context. In the group discussion phase, the participants or audience members share a third-person interpretation of the play performed by the actors for criticism. The actors then revise the virtual play until they are satisfied. The understanding of the given theme will be progressively deepened by repeatedly looking at embodied interpretation from the first- and third-person points of view.

The SES Support System provides computational support for the entire SES process. It combines a game engine and a conversation augmentation technology we developed for research on conversational informatics [6]. This system allows us to capture a group dramatic performance and create an agent play on Unity 3D so that participants can switch between objective and subjective views for each scene in an immersive environment consisting of large surrounding displays that provide a 360-degree view at any point in a virtual shared space.

We conducted a feasibility study with a partially implemented SES Support System. There were numerous encouraging results regarding the typical behavioral patterns of participants in role playing and how the inside understanding was deepened through the role play and discussion in the SES framework.

In this paper, we explain how SES is applied to deepen inside understanding in a group discussion. First we describe the conceptual framework of SES and discuss the technologies used to support SES with an emphasis on capturing group performance. We then present the insights obtained from the feasibility study. An interpretation archive is used to provide logistical support for the SES sessions. The annotated agent plays and stories resulting from SES workshops are decomposed into components for later use so that participants in subsequent SES workshops can adapt previous annotated agent plays and stories to use as a part of the present annotated agent play.

2 SES Support System

The aim of the SES Support System is to help facilitate the smooth management of an SES session cycle, as shown in Fig. 1. It supports the capture of human performances, which the system then converts into virtual agent behavior in the dramatic role play phase. Specifically, it provides agent role plays in an immersive virtual space, edits agent motions in the virtual space in the projection phase, and accumulates and structures potential knowledge from the discussion for evidential study.

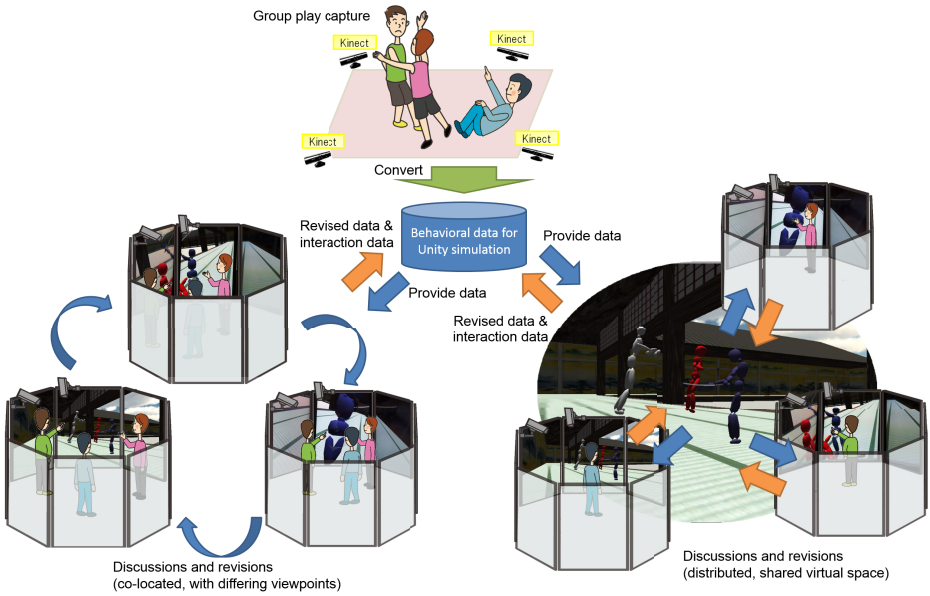


Fig. 1. SES Support System

2.1 The Core

The core of the SES Support System consists of an immersive collaborative interaction environment (ICIE) and a distributed elemental application linker (DEAL). ICIE is an immersive interaction environment made available by a 360-degree display and surround speakers and audio-visual sensors for measuring user behaviors. It consists of an immersive audio-visual display and plug-in sensors that capture user behaviors therein. The user receives an immersive audio-visual information display in a space surrounded by eight large displays about 2.5 m in diameter and 2.5 m in height. We have prototyped motion capture consisting of multiple depth sensors that can sense the user's body motion in this narrow space. This platform constitutes a "cell" that will allow users to interact with each other in a highly immersive interface. Cells can be connected with each other or with other (e.g., robot) interfaces so that multiple users can participate in interactions in a shared space.

DEAL is a software platform that permits individual software components to cooperate in providing various composite services for the interoperating ICIE cells. Each server has one or more clients that read/write information on a shared blackboard on the server. Servers can be interconnected on the Internet and the blackboard data can be shared. One can extend a client with plug-ins using DLL. Alternatively, a client can join the DEAL platform by using DLL as a normal application that bundles network protocols for connecting a server.

Leveraging the immersive display presentation system operated on the DEAL platform and the motion capture system for a narrow space allows the behaviors of human users to be projected onto those of an animated character who habits are shared in a virtual space. The ICIE+DEAL platform is coupled with the Unity (http://unity3d.com/) platform so that participants can work together in a distributed environment. This permits the participants to take the form of animated Unity characters in the virtual space.

2.2 Dramatic Group Play Capture Subsystem

The Dramatic Group Play Capture (DGPC) system can capture the 3D surface model data and skeleton model data of multiple persons without markers and can convert the data for using the Unity environment in a short time.

The participant-motion-estimation subsystem uses the skeleton data from multiple Kinect v2 sensors to estimate the motion of a participant. First, personal IDs are given to each piece of skeleton data of each Kinect v2 sensor. Second, each piece of skeleton data is projected onto the integrated coordination and the personal IDs are integrated based on the overlap of each skeleton's coordination data. Third, each joint coordinate of each piece of skeleton data is integrated. Each joint coordinate is weighted based on various heuristic conditions, such as how far the joint is from the Kinect v2 sensor, whether the joint is occluded or not, how many sensors capture the joint, and whether the captured person is facing the Kinect v2 sensor. In order to avoid any misunderstanding between right and left joints, a subsystem checks the time series data to confirm the consistency of the joint recognition.

The conversation-scene-reconstruction subsystem builds a 3D model for the conversation scene by integrating the 3D surface model and skeletons of participants in the conversation. Currently, a simple method is employed to integrate the multiple Kinect v2 sensor coordination. In order to decrease the overhead for calibrating the multiple Kinect v2 sensor coordination, the subsystem integrates them by using skeleton data and a depth map. First, a person is captured by multiple Kinect v2 sensors. Second, the skeleton data is used to calculate the relative positions of the Kinect v2 sensors. Third, measured 3D points near the matched skeleton points in the depth map are used for fine adjustment by using a least-square method. This method can recalculate the sensor coordination after the measuring by using the skeleton data from the measured scenes if the data has been correctly captured.

The system can capture multiple user interactions including up to four persons without markers at the same time. Each piece of skeleton data captured by the participant-motion-estimation subsystem can be output in real-time. In multiple user interactions, however, it is hard to identify adequate skeleton data because of occlusion and contact between participants. Therefore, in the conversation-scene-reconstruction subsystem, depth maps of each sensor are used for modifying the surface and skeleton model data. The processing takes 100 second to 300 second. Furthermore, the system cannot always correctly modify the data when the skeleton data is captured by less than two sensors. These problems should be improved in the future.

The projection subsystem can convert the captured skeleton data into character motion data in the Unity virtual environment and can play the motion data using the character. After the data has been converted, users can easily change the projection. For example, the same data can be shown on both a big flat display and a 360-degree immersive display. In addition, users can change the position of virtual cameras for showing the data on a screen. They can select an objective position for the overview an entire role play scene using a flat display in discussion and one of the role player's subjective position for experiencing the player's situation using an immersive display. Figure 2 shows multiple views of one shot of the role play data in an actual SES workshop. In this example, the blue character attacked the red character and the white character was a spectator. Figure 2 (a) shows a snapshot from a recorded video and

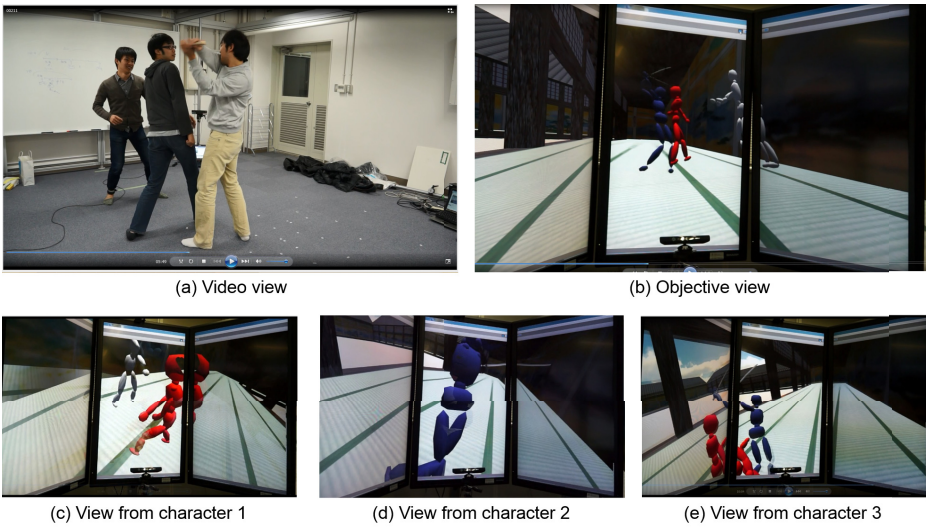


Fig. 2. Group play capture by SES Support System

Fig. 2 (b) shows the same view reconstructed in the Unity virtual environment. Figure 2 (c) (d) (e) shows the subjective views of each character. In the actual SES session, the scene was played as a movie on a life-size screen. As shown in Fig. 2 (d), the perception from the person who was attacked is very different from the objective view shown in Fig. 2 (b). The perception from the person who was a spectator is also different from the objective view. These indicate that different perspectives provide useful clues to investigate principles of human behavior in SES sessions.

3 A Feasibility Study

We conducted a feasibility study using one simplified SES workshop. The simplified SES session consisted of the group discussion phase based on document-based discussions (about 30 minutes) followed by the role play phase (about 20 minutes) and the projection phase (about 10 minutes) in which participants discussed with each other while switching between the objective and subjective views of one of the three roles in the scenario. A partial implementation of the SES Support System was used to support the feasibility study. Figure 3 shows an overview.

3.1 Experiment Setting

We used the "Matsu no Roka" ("Hallway of Pine Trees") passage of Chushingura as a sample scenario. This fictionalized account based on a historical incident roughly reads like this:

> *At the beginning of the 18th century, a feudal lord named Asano Takumi-no-kami Naganori was in charge of a reception for envoys from the Imperial Court in Kyoto. Another feudal lord, Kira Kozuke-no-suke Yoshinaka, was appointed to instruct Asano in the ceremonies. On the day of the reception, while Kira was talking with Yoriteru*

Kajikawa, a lesser official, at "Matsu no Roka" ("Hallway of Pine Trees") in Edo Castle, Asano came up to them screaming "This is for revenge!!" and slashed Kira twice with a short sword. Soon after the incident, Kajikawa restrained Asano, who was then imprisoned. The reason for the attack was not known, though it was widely believed that Kira had somehow humiliated Asano. Ultimately Asano was sentenced to commit seppuku, a ritual suicide, but Kira went without punishment.[1]

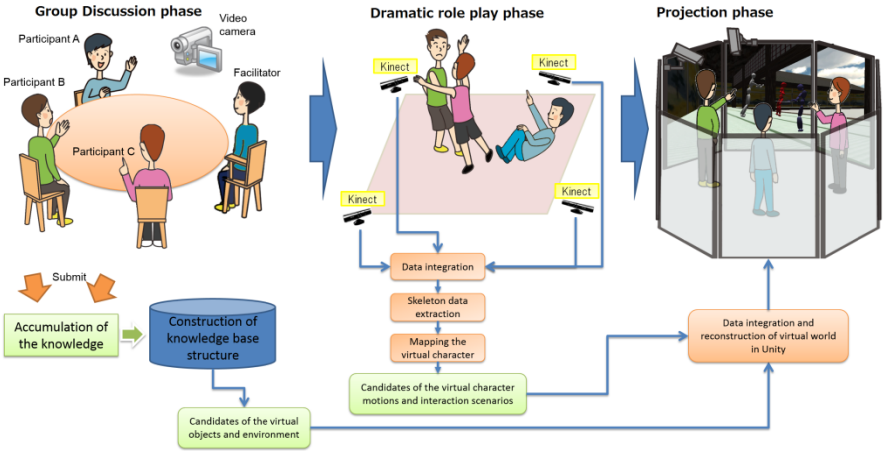


Fig. 3. Procedure for Feasibility Study

We hired three participants (males) who knew each other and asked them to serve as hypothetical inspectors of historical affairs and determine a sentence for Asano as a result of group discussions. The participants went through a simplified SES session consisting of three phases: group discussion, role play, and projection. The participants were asked over the course of these three phases to investigate the incident, propose a legal judgment for it, and explain their reason for reaching their decision. The conclusion had to be unanimous within the group. At the end of each phase, we asked the participants to write a brief explanation of how each of them interpreted the "Matsu no Roka" affair. We videotaped the discussions for later analysis. This simplified version of an SES session took about one hour.

1) Group Discussion Phase

We provided the participants with historical documents about Asano, Kira, and Kajikawa and had them conduct a group discussion about the affair for about 30 minutes. The documents were based on Wikipedia.

2) Role Play Phase

We allocated the roles of Asano, Kira, and Kajikawa to the three participants and had them reproduce the "Matsu no Roka" affair on the basis of their interpretation. We captured the dramatic role play by the DGPC subsystem to produce an agent play on Unity 3D. It took about 30 minutes.

[1] Based on multiple sources, e.g., `http://en.wikipedia.org/wiki/Ch%C5%ABshingura`.

3) Projection Phase
We used ICIE to show the participants the agent play reproduced from their role play. Objective and subjective views from each player were displayed repeatedly until the participants were satisfied. We did not allow the participants to revise any parts of this agent play. It took about 10 minutes.

3.2 Findings

We obtained five major findings from this feasibility study.

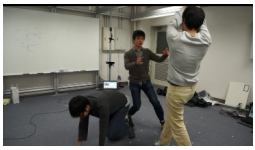
1) Participants' Behavior in Role Playing
The behavior of the participants in the role play phase can be classified into four patterns: (a) acting behavior, (b) commenting behavior, (c) oral editing behavior, and (d) idling behavior. Acting behavior is what the participants do when they are actually acting. Commenting behavior is a critique of the incidents and the acting, including reasoning, discussion and thinking aloud. The editing behavior is suggested revisions to the acting, e.g., "This action could have been better." It can be accompanied by gestures and examples of preferred acting behavior. Idling behavior includes all actions that are not classified above—in other words, irrelevant movement, such as stretching or making the motion of swinging a bat to loosen up the body.

Further analysis of the role play phase revealed that it could be roughly classified into two parts: rehearsal acting scenes and production acting scenes. A rehearsal acting scene is a scene in which acting and discussions are mixed. Acting a scene while checking the position and behavior of each character in the discussion and thinking aloud (essentially, a continuation of the discussion in the group discussion phase) was observed. In a rehearsal acting scene, special acting behavior such as repeatedly mimicking the acting of different roles was observed. It seems that mimicking the acting of others was an unconscious attempt to experience the viewpoint of others. A production acting scene is a scene in which the acting is performed from the beginning to the end, and discussion and thinking aloud are not observed. Switches between the rehearsal acting scene and production acting scene can be clearly identified by eye. For example, just before the production acting scene, explicit signaling behavior such as giving-a-cue was observed.

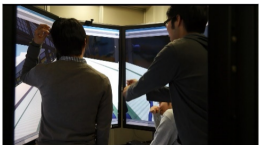
In a rehearsal acting scene, acting behavior, oral editing behavior, speaking his/her role, commenting behavior, thinking aloud, and idling behavior were observed. On the other hand, in a production acting scene, acting and speaking his/her role were observed while oral editing, commenting, thinking aloud, and idling were not.

Twelve detailed behaviors were observed in the role play phase: (1) acting, (2) commenting, (3) oral editing, (4) idling, (5) speaking his/her role, (6) acting + thinking aloud, (7) acting + commenting, (8) acting + oral editing, (9) acting + speaking his/her role, (10) idling + commenting, (11) idling + oral editing, and (12) idling + speaking his/her role.

2) Deepening by Playing
In the role play phase, progress of the discussion occurred. We confirmed details that are not usually described in historical materials by remarks from the participants such



(a) Acting is modified: "Kajikawa jumped over the Kira that fall, and go straight toward Asano (rather than to take a roundabout path)."



(b) Discussions with gestures and body movements

Fig. 4. Snapshots from the feasibility study

as, "He could probably help earlier", "Probably, he thrust with the sword rather than swinging the sword down", and "Perhaps, the distance between the two should have been closer", as shown in Fig. 4(a). However, the content of the discussion in the role play phase was biased toward content that had already been discussed in the group discussion phase and revealed through the acting of each person.

3) Deepening Based on Contrasting Objective and Subjective Views

In the projection phase, we observed contrasting objective and subjective views. After experiencing all perspectives on the action of Kira falling prone, one participant remarked that "It seems strange in the objective viewpoint, but when I experience the subjective viewpoint, it looks like a natural movement", and everyone agreed. They considered it and remarked "I don't know. Which viewpoint is correct?" This event is evidence that the presentation of a subjective viewpoint provided the participants with a more intense experience than could be obtained from the objective viewpoint alone.

4) Deepening Based on Subjective View Transfer

In the projection phase, when the participants had experienced Kajikawa's subjective viewpoint, the participant who had played the role of Kira in the acting phase (the Kira player) remarked, "(Kajikawa was too slow to) try and hold (Asano) down after having been slashed" while gesturing (Fig. 4(b)). In contrast, the participants who played the roles of Kajikawa and Asano had no such feeling. This probably means that the Kira player thought that Kira had hoped to be helped earlier. We suspect that the memory of acting affects the subsequent thinking and discussion. In addition, soon after this, the Kira player remarked, "When (Asano) swung his sword up the first time, Kajikawa might have been farther away from Asano (and therefore, Kajikawa couldn't quickly hold down Asano)". This remark suggests that, through repeatedly experiencing Kajikawa's viewpoint, the Kira player could imagine the feeling of Kajikawa and thus change his opinion.

5) Implicit Opinion Expression by Acting

In the acting phase, we observed an interesting case. The participant playing Kira continued to repeat the action of falling prone in the production acting scene, in spite of the opinion that "It is unnatural for Kira to fall prone", which was shared with everyone in the rehearsal acting scene. The Kira player himself remarked that falling prone is unnatural, and he actively tried to collapse on the spot instead. Furthermore, the other two participants demonstrated the action of staggering behind, but still he

never changed. We feel this phenomenon is an example of implicit opinion expression via acting. In the acting phase, the change of a participants' inner heart, which cannot be observed in the form of speech, may be expressed by acting.

According to the notes after the acting phase, the participants described Kira as having collapsed on the spot. However, in the memo after the projection phase, the Kira player wrote, "If the assumption that Kira fell prone is correct, I think that our acting is correct." This means that an opinion that was not observed previously had been exposed by reviewing their acting from the subjective/objective viewpoint. We feel that this implicit opinion expression has been overlooked in traditional workshops with theater. By modifying acting as editing behavior in the projection phase, further progress on the inner understanding of participants can be expected.

4 Discussions

The results of the feasibility study are very promising in that we have obtained plenty of evidence to support our initial expectation that SES is an effective method for deepening inside understanding. One concern during the initial discussions was the trade-off between physical role play and agent play. Physical role play appeared more realistic and depended too much on the skill of the actor. In this feasibility study, non-experienced participants were the actors, and their interpretation was not always properly expressed as far as the video analysis is concerned. For example, they did not wear historically accurate clothing, and their unintentional shyness appeared to hinder realistic acting from time to time. In contrast, although agent play was less expressive, it hid excessive details and allowed for soliciting space for interpretation in the discussion phase. Overall, the employment of agent play powered by Unity 3D was effective. In addition to our initial expectation, it seems that participants were also able to deepen their understanding of how other participants interpret the role he or she played.

There are several issues remaining for future research. We believe that the following are three of the most challenging. The first is intelligent role play support that can watch the role play like a stagehand and offer assistance if needed, e.g., setting-up/updating the background, virtually creating props or historical clothing, and virtually editing the film as the participants like, to name just a few. The second is story archiving that allows for community-wide sharing of agent plays so that interested members can refine or extend their interpretations. The third is discussion facilitation in which an intelligent agent mediator facilitates group discussions.

5 Conclusion

In this paper, we presented initial findings obtained in a feasibility study of SES. The results were quite encouraging. Major findings include participants' behavior in role playing, deepened understanding by playing, deepened understanding based on contrasting objective and subjective views, deepened understanding based on subjective view transfer, and implicit opinion expression by acting.

Acknowledgment. This study has been carried out with financial support from the Center of Innovation Program from JST, JSPS KAKENHI Grant Number 24240023, and AFOSR/AOARD Grant No. FA2386-14-1-0005

References

1. Nishida, T.: Towards Mutual Dependency between Empathy and Technology, 25th anniversary volume. AI & Society **28**(3), 277–287 (2013)
2. Hawkins, S.: Dramatic Problem Solving: Drama-Based Group Exercises for Conflict Transformation, Jessica Kingsley Publishers (2012)
3. Cavazza, M., Lugrin, J.L., Pizzi, D., Charles, F.: Madame Bovary on the holodeck: immersive interactive storytelling. In: MULTIMEDIA 2007: Proc. of the 15th International Conference on Multimedia, ACM, pp. 651-660 (2007)
4. Nishida, T., Abe, M., Ookaki, T., Lala, D., Thovuttikul, S., Song, H., Mohammad, Y., Nitschke, C., Ohmoto, Y., Nakazawa, A., Shochi, T., Rouas, J.-L., Bugeau, A., Lotte, F., Zuheng, M., Letournel, G., Guerry, M., Fourer, D.: Synthetic evidential study as augmented collective thought process – preliminary report. In: Nguyen, N.T., Trawiński, B., Kosala, R. (eds.) ACIIDS 2015. LNCS, vol. 9011, pp. 13–22. Springer, Heidelberg (2015)
5. Nishida, T., Nakazawa, A., Ohmoto, Y., Nitschke, C., Mohammad, Y., Thovuttikul, S., Lala, D., Abe, M., Ookaki, T.: Synthetic Evidential Study as Primordial Soup of Conversation. In: Chu, W., Kikuchi, S., Bhalla, S. (eds.) DNIS 2015. LNCS, vol. 8999, pp. 74–83. Springer, Heidelberg (2015)
6. Nishida, T., Nakazawa, A., Ohmoto, Y., Mohammad, Y.: Conversational Informatics–A Data-Intensive Approach with Emphasis on Nonverbal Communication, Springer (2014)