Synthetic Evidential Study as Primordial Soup of Conversation

Toyoaki Nishida[1,*], Atsushi Nakazawa[1], Yoshimasa Ohmoto[1], Christian Nitschke[1], Yasser Mohammad[2], Sutasinee Thovuttikul[1], Divesh Lala[1], Masakazu Abe[1], and Takashi Ookaki[1]

[1] Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto, Japan
{nishida,nakazawa.atsushi,
ohmoto,christian.nitschke}@i.kyoto-u.ac.jp,
{thovutti,lala,abe,ookaki}@ii.ist.i.kyoto-u.ac.jp
[2] Assiut University, Assiut, Egypt
yasserfarouk@gmail.com

Abstract. Synthetic evidential study (SES for short) is a novel technology-enhanced methodology for combining theatrical role play and group discussion to help people spin stories by bringing together partial thoughts and evidences. SES not only serves as a methodology for authoring stories and games but also exploits the framework of game framework to help people sustain in-depth learning. In this paper, we present the conceptual framework of SES, a computational platform that supports the SES workshops, and advanced technologies for increasing the utility of SES. The SES is currently under development. We discuss conceptual issues and technical details to delineate how much we can implement the idea with our technology and how much challenges are left for the future work.

Keywords: Inside understanding, group discussion and learning, intelligent virtual agents, theatrical role play, narrative technology.

1 Introduction

Our world is filled with mysteries, ranging from fictions to science and history. Mysteries bring about plenty of curiosities that motivate creative discussions, raising interesting questions, such as "how did Romeo die in Romeo and Juliet?", "why was Julius Caesar assassinated?", and "why did Dr. Shuji Nakamura, 2014 Nobel laureate, leave Japan to become a US citizen?" to name just a few.

It is quite natural that mysteries motivate people to bring together the partial arguments, ranging from thoughts to evidences or theories, to derive a consistent and coherent interpretation of the given "mystery". Let us call this discussion style an *evidential study* if it places a certain degree of emphasis on logical consistency. Discussions need not be deductive so long as they are based on a discipline such as an abduction. On the one hand, rigorous objective discussions were considered mandatory in an academic

* Corresponding author.

W. Chu et al. (Eds.): DNIS 2015, LNCS 8999, pp. 74–83, 2015.

disciplines such as archeology and evidential methods broadly observed in empirical sciences. On the other hand, the constraints are much weaker in the folk activities such as those workshops in schools. Even though they may sometimes run into flaw, their utility in education, e.g., motivating students, is even higher due to its playful aspects.

Unfortunately, "naive" evidential study has its own limitations as well even with enhancement by conventional IT. Although normal groupware greatly helps large scale online discussions and accumulate their results for later use, it does not permit the participants to examine the discussion from many perspectives, which is critical in evidential study. In general, conventional software falls short for allowing participants to visualize their thoughts instantly on the spot for sharing better understanding. In particular, evidential study cannot be deepened without going deeply into the mental process of the characters in question by simulating how they perceive the world from their first person view, it is critical to uncover the mental process of a given character by deeply into partial interpretation in particular.

Synthetic evidential study (SES for short) is a novel technology-enhanced methodology for combining theatrical role play and group discussion to help people spin stories by bringing together partial thoughts and evidences [1]. SES practically implements the idea of primordial soup of conversation in conversational informatics [2], in which common ground and communicative intelligence may co-evolve through the accumulation of conversations. SES leverages the powerful game technologies [3], and is applicable to not just mysteries but also a wide range of application in science and technology, e.g., [4-6].

2 The Conceptual Framework of SES

SES combines theatrical role play and group discussion to help people spin stories by bringing together partial thoughts and evidences. The conceptual framework of SES is shown in Fig. 1.

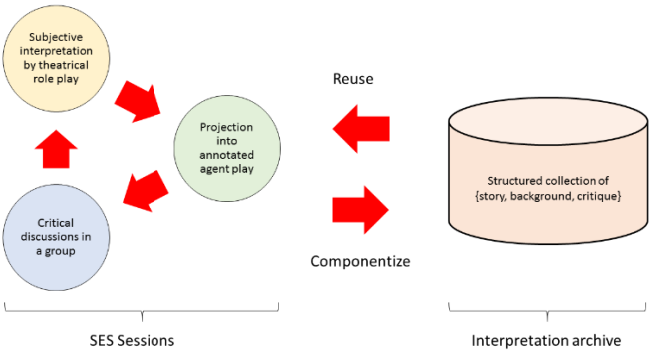

Fig. 1. The conceptual framework of SES

At the top level, the SES framework consists of the SES sessions and the interpretation archives. In each SES session, participants repeat a cycle of a theatrical role play, its projection into an annotated agent play, and a group discussion. One or more

successive execution of SES sessions until participants come to a (temporary) satisfaction is called a SES workshop.

In the theatrical role play phase, participants play respective roles to demonstrate their first-person interpretation in a virtual space. It allows them to interpret the given subject from the viewpoint of an assigned role.

In the projection phase, an annotated agent play on a game engine is produced from the theatrical role play in the previous phase by applying the oral edit commands (if any) to theatrical actions by actors elicited from the all behaviors of actors. We employ annotated agent play for reuse, refinement, and extension in the later SES sessions.

In the critical discussion phase, the participants or other audience share the third-person interpretation played by the actors for criticism. The actors 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 views.

The interpretation archive logistically supports the SES sessions. The annotated agent plays and stories resulting from SES workshops may be decomposed into components for later reuse 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.

Let us use a hypothetical example to illustrate on more details. Consider the following story is given as a subject for a SES workshops:

> The Ushiwaka-Benkei story: When Ushiwaka, a young successor to a noble Samurai family who once was influential but killed by the opponents, walked out of a temple in a mountain in the suburbs of Kyoto where he was confined, to wander around the city as daily practice, he met Benkei, a strong priest Samurai on Gojo Bridge. Although Benkei tried to punish him as a result of having been provoked by a small kid Ushiwaka, he couldn't as Ushiwaka was so smart to avoid Benkei's attack. After a while, Benkei decided to become a life-long guard for Ushiwaka.

One interest here is to figure out exactly what happened during the encounter of Ushiwaka and Benkei on Gojo Bridge. Let us assume that three people participate in the SES session, to discuss the behaviors of Ushiwaka, Benkei, and a witness during the encounter. It is assumed that before a SES workshop starts, a preparatory meeting is held in which the role assignment is made so each participant is asked to think about the role she or he is to play.

In the theatrical role play phase of the SES session, the participants may do a role play for the given subject, which may be similar to, but slightly different from the one as shown in Fig. 2(a). It is similar, in the sense that each participant tries to bodily express her/his interpretation. It is different, as participants are expected play in a shared virtual space in the SES theatrical role play phase and we do not assume that participants need to physically play together, need carry props such as a sword prop, or physically jump from a physical stage, though production of a virtual image and motion is a challenge.

Fig. 2. (a) Left: Theatrical role play by the SES participants is similar to, but slightly different from the role play like this. (b) Right: Agent play to be reproduced from theatrical role play on the left. One or more parts of the agent play may be annotated with the player's comments.

A think aloud method is employed so each actor can not only show her/his interpretation but also show the rationale for the interpretation, describe their intention of each action using an oral editing command, or even criticize role play by other actors. The actor's utterances will be recorded and used as a resource for annotations associated with the agent play. It is an excellent opportunity for each participant to gain a pseudo-experience of the situation through the angle of the given role's viewpoint. Each actor's behaviors are recorded using audio-visual means.

The goal of the projection phase is to reproduce an annotated agent play such as shown in Fig. 2(b) from the theatrical role play.

Ideally, the projection should be automatically executed on-line while participants are acting in the theatrical role play phase. In order to do so, separation of the behaviors of each actor into the genuine theatrical role play, the editing behavior, and the meta-level actions such as commentating.

In the critical discussion phase, the participants watch the resulting annotated agent play and discuss the resulting interpretation from various angles (Fig. 3).

Fig. 3. Critical discussion based on the annotated agent play will help participants to deepen objective understanding. In actual SES sessions, the participants may well meet virtually to criticize the annotated agent play.

Again, we do not necessarily assume that the participants need to hold a physical meeting. Most importantly, the participants are provided with an opportunity to discuss the theatrical role play from the objective third person view, in addition to the subjective view gained in the theatrical role play phase. A 3D game engine such as Unity allows the group to switch between the objective and subjective views of a given role to deepen the discussion. Furthermore, it might be quite useful if the participants can modify the annotated agent play on the fly during the discussions, while we

assume that it will be more convenient to make major revisions in the theatrical role play phase. That is why we have assumed that the SES sessions will be repeated.

3 Computational Platform for SES

We are building the SES computational platform by combining a game engine (Unity 3D) and the ICIE+DEAL technology we developed for conversational informatics research [2]. ICIE is an immersive interaction environment made available by a 360-degree display and surround speakers and audio-visual sensors for measuring the user's behaviors. ICIE 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 8 large displays about 2.5 m diameter and 2.5 m in height (Fig. 4). We have implemented a 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 for the user to interact with each other in a highly-immersive interface. Cells can be connected with each other or other kinds of interfaces such as a robot so that the users can participate in interactions in a shared space.

(a) the device (b) immersive interaction in ICIE

Fig. 4. The hardware configuration of ICIE

DEAL is a software platform that permits individual software components to cooperate to provide various composite services for the interoperating ICIE cells. Each server has one or more clients that read/writes 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.

By leveraging the immersive display presentation system operated on the DEAL platform and the motion capture system for a narrow space it allows for projecting the behaviors of a human to those of an animated character who habits in a shared virtual space. The ICIE+DEAL platform is coupled with the Unity (http://unity3d.com/) platform so the participants can work together in a distributed environment. It permits the participants to animated Unity characters in the virtual space.

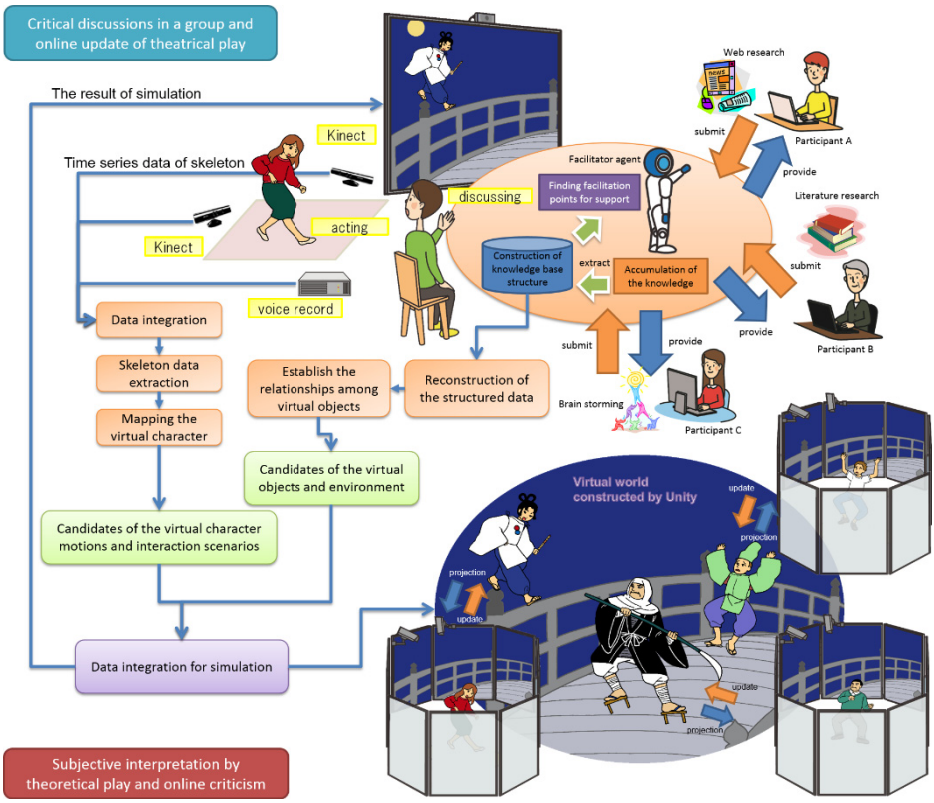

Fig. 5. The architecture of the SES computational platform

Fig. 5 shows how ICIE and DEAL can be put together to support theatrical role play and group discussion in SES. The upper half of the diagram depicts a subsystem for supporting group discussions. It consists of a group discussion support for conventional evidential study based on literature and review of theatrical role play. A facilitation agent is employed to help participants bring together partial knowledge and criticisms to formulate a consistent flow of theatrical role plays as a synthetic interpretation of the subject. The major role of the facilitation agent is to estimate and track tacit preferences of participants underlying discussions and navigate the discussion by conducting appropriate facilitation behaviors such as encouraging remark by a participant with a conflicting opinion at a proper timing. In addition, the subsystem permits the participants to revise the details of the behavior of the theatrical role play on the fly by using a mini-studio adjacent to the discussion table.

The lower half of Fig. 5 depicts a subsystem for supporting a theatrical role play in a distributed environment. ICIEs can be interconnected on the net to allow participants from remote sites to share a virtual space for interaction to collaboratively coordinate a theatrical role play to build up their subjective interpretation of the subject. Think aloud method is used to permit the participants to make critical or even

editorial comments during their play. The behaviors of participants are not only recorded for review but also projected onto synthetic characters by incorporating editorial comments and isolating critical comments into a separate track.

4 Technologies for SES

Technologies we have developed so far can be adapted to plug in to the SES computational platform.

4.1 Capturing the Realworld Environment

The background for a theatrical role play is an integral part of the evidential study. It is often the case that we can import a useful background data from the Internet. Flexible Communication World (FCWorld) [7] allows us to project an external source on the net such as Google Street View to the immersive display. An alternative approach is to build the data for the shared virtual space by measuring the realworld. A wide-range area visualizer [8] can automatically construct a 3D panoramic model for an outdoor scene from a collection of ordinary 2D digital photo images for the place by combining structure from motion, multi view stereo, and other standard techniques. Capturing first person view and gaze is critical in estimating the mental process of actor. Our corneal imaging method allows for going beyond the current human view understandings that uses only the 'point' of the gaze information in a static planar environment to capture the whole peripheral visual field in an arbitrary dynamic 3D environment without the need for geometric calibration [9].

4.2 Capturing Realworld Interaction

Capturing the motion of an actor in an immersive interaction cell becomes ready with existing technologies, by mapping the skeleton data from a Kinect to animate a Unity character. However, we need to overcome several limitations.

A theatrical role play by multiple actors can be captured by our 3DCCbyMK technology [2]. It exploits Kinect technology to measure and critique physical display of interpretation played together one or more local participants. The system allows us to simultaneously capture the behaviors of up to four participants, by integrating data captured by multiple Kinects, thereby to project the interaction by multiple participants to the shared virtual space as it is. It assumes that the scene consists of the static background and moving people in the foreground.

Two subsystems were developed for capturing the static background and the dynamic foreground. OpenNI is used to capture an RGB image, depth map, skeleton data, user masks and tracking IDs. To reconstruct a static 3D model for the surrounding background, a single Kinect is carried around in the environment to gather data from continuous viewpoints. SLAM (simultaneous localization and mapping) is used to build a local 3D model from the data from each point. Image features calculated by the SURF method [10] are used to calculate the similarity to integrate the local 3D

model into the global 3D model. The LMedS method [11] is used to reduce errors in image feature matching.

The motion estimation subsystem of the 3DCCbyMK technology estimates the motion of each participant using the skeleton data from multiple Kinect sensors. Time series data are checked to reduce the confusion between the left and right joints. The output of the two subsystems are integrated to produce an interaction scene. A problem remains regarding incorporating the results into the 3D coordinate system of a game engine.

4.3 Estimating Attitudes and Emotion

Mind reading or estimating the internal mental state of the human from external cues or social signals is necessary to build an advanced service for the SES participants, such as discussion support as discussed later in this section or even producing a better annotated agent play in the theatrical role play.

DEEP [12] is a method for estimating a dynamically changing emphasizing point by integrating verbal behaviors, body movements, and physiological signals of the user. DEEP is applied to situations where many known/unknown factors must be considered in choosing a satisfactory option. DEEP repeats the cycle consisting of explanation, demand-seeking, and completion-check until the user is satisfied with the proposed option. User's emphasizing point is obtained at the same time. gDEEP is a method for estimating emphasizing point for a group using DEEP as a component. gDEEP repeats the cycle similar to DEEP.

4.4 Discussion Support

Our discussion support method centers on a chairperson agent that can support discussions by estimating distribution of opinions, engagement, and emphasizing points. We addressed interactive decision-making during which people dynamically and interactively change the emphasizing points and have built several prototypes of a group decision-making support agent [12-14].

A recent prototype [14] can guide the divergence and convergence processes [15] in the facilitation by producing appropriate social signals as a result of grasping the status of the decision-making process by the group of participants. The system uses the gDEEP method to estimate the emphasizing point of the group from verbal presentation of demands and nonverbal and physiological reactions to the information presentation by the agent. If it has turned out that the group has not yet well formulated an emphasizing point, the system will present information obtained from a broad search in the problem space with reference to the emphasizing point so that it can stimulate the group's interest to encourage divergent thought. When it has turned out that the group has formulated an emphasizing point, the system will focus on the details to help the group carry out convergent thought for making decision.

Our technologies allow for capturing not only explicit social signals that clearly manifest on the surface but also tacit and ambiguous cues by integrating audio-visual and physiological sensing.

4.5 Fluid Imitation Learning

Learning by mimicking is a computational framework for producing the interactive behaviors of conversational agents from a corpus obtained from the WOZ experiment. In this framework, the learning robot initially "watches" how people interact with each other, estimates communication principles underlying the way the target actor communicates with other partners, and applies the estimated communication patterns to produce the communicative behaviors of the conversation agent.

Currently, we focus on nonverbal behaviors and approximate the communicative behaviors as a collection of continuous time series. A suite of unsupervised learning algorithms [16,17] are used to realize the idea.

By having this algorithm in combination with action segmentation and motif discovery algorithms that we developed, we have a complete fluid imitation engine that allows the robot to decide for itself what to imitate and actually carry on the imitation [18].

We designed, implemented and evaluated of a closed loop pose copying system [19]. This system allows the robot to copy a single pose without any knowledge of velocity/acceleration information and using only closed loop mathematical formulae that are general enough to be applicable to most available humanoid robots. This system makes it possible to reliably teach a humanoid by demonstration without the need of difficult to perform kinesthetic teaching.

5 Concluding Remark

Synthetic evidential study (SES) combines theatrical role play and group discussion to help people spin stories. The SES framework consists of (a) the SES sessions of theatrical role playing, projection into the annotated agent plays and a critical group discussions, and (b) a supporting interpretation archive containing annotated agent plays and stories. We have described the conceptual framework of SES, a computational platform that supports the SES workshops, and advanced technologies for increasing the utility of SES. The SES is currently under development. So far, we have implemented basic components. The system integration and evaluation are left for future work. Future challenges include, among others, automatic production of actor's intended play, self-organization of the interpretation archive, and automatic generation of the virtual audience from critical discussions.

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., et al.: Synthetic Evidential Study as Augmented Collective Thought Process – – Preliminary Report. ACIIDS 2015, Bali, Indonesia (to be presented, 2015)
2. Nishida, T., Nakazawa, A., Ohmoto, Y., Mohammad, Y.: Conversational Informatics: A Data-Intensive Approach with Emphasis on Nonverbal Communication. Springer (2014)

3. Harrigan, P., Wardrip-Fruin, N.: Second Person: Role-Playing and Story in Games and Playable Media, MIT Press (2007)
4. Thovuttikul, S., Lala, D., van Kleef, N., Ohmoto, Y., Nishida, T.: Comparing People's Preference on Culture-Dependent Queuing Behaviors in a Simulated Crowd, In: Proc. ICCI*CC 2012, pp. 153–162 (2012)
5. Lala, D., Mohammad, Y., Nishida, T.: A joint activity theory analysis of body interactions in multiplayer virtual basketball. Presented at 28th British Human Computer Interaction Conference, September 9-12, Southport, UK (2014)
6. Tatsumi, S., Mohammad, Y., Ohmoto, Y., Nishida, T.: Detection of Hidden Laughter for Human-agen Interaction. Procedia Computer Science (35), 1053–1062 (2014)
7. Lala, D., Nitschke, C., Nishida, T.: Enhancing Communication through Distributed Mixed Reality. AMT 2014, 501–512 (2014)
8. Mori, S., Ohmoto, Y., Nishida, T.: Constructing immersive virtual space for HAI with photos. In: Proceedings of the IEEE International Conference on Granular Computing, pp. 479–484 (2011)
9. Nitschke, C., Nakazawa, A., Nishida, T.: I See What You See: Point of Gaze Estimation from Corneal Images. In: Proc. 2nd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 298–304 (2013)
10. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
11. Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. Image Vis. Comput. 15(1), 59–76 (1997)
12. Ohmoto, Y., Miyake, T., Nishida, T.: Dynamic estimation of emphasizing points for user satisfaction evaluations. In: Proc. the 34th Annual Conference of the Cognitive Science Society, pp. 2115–2120 (2012)
13. Ohmoto, Y., Kataoka, M., Nishida, T.: Extended methods to dynamically estimate emphasizing points for group decision-making and their evaluation. Procedia-Social and Behavioral Sciences 97, 147–155 (2013)
14. Ohmoto, Y., Kataoka, M., Nishida, T.: The effect of convergent interaction using subjective opinions in the decision-making process. In: Proc. the 36th Annual Conference of the Cognitive Science Society, pp. 2711–2716 (2014)
15. Kaner, S.: Facilitator's guide to participatory decision-making. Wiley (2007)
16. Mohammad, Y., Nishida, T., Okada, S.: Unsupervised simultaneous learning of gestures, actions and their associations for Human-Robot Interaction. In: IROS 2009, pp. 2537–2544 (2009)
17. Mohammad, Y., Nishida, T.: Learning interaction protocols using Augmented Bayesian Networks applied to guided navigation. In: IROS 2010, pp. 4119–4126 (2010)
18. Mohammad, Y., Nishida, T.: Robust Learning from Demonstrations using Multidimensional SAX. Presented at ICCAS 2014, October 22-25, Korea (2014)
19. Mohammad, Y., Nishida, T.: Tackling the Correspondence Problem - Closed-Form Solution for Gesture Imitation by a Humanoid's Upper Body. In: Yoshida, T., Kou, G., Skowron, A., Cao, J., Hacid, H., Zhong, N. (eds.) AMT 2013. LNCS, vol. 8210, pp. 84–95. Springer, Heidelberg (2013)