Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 7 – 16

Procedia Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Conversational informatics: Toward cultivating wisdom from conversational interaction

Toyoaki Nishida[a,*]

[a]*Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*

Abstract

In this paper, I present a data-intensive approach to conversational informatics. It not only brings about quantitative understanding, permitting us to turn a great accumulation of keen observations into a pile of computational models, but also helps to build conversational agents by virtue of recent progress in machine learning and data mining. The topics include a smart conversation space for allowing people to engage in conversation in a cyber-physical space, conversational interaction capture for conversation modeling and content production, learning by imitation for producing conversational interactions, cognitive design for understanding mental processes of conversational actors, and synthetic evidential study for collaborative study on understanding social processes.

Peer-review under responsibility of KES International

Keywords: Conversational informatics; Synthetic evidential study

1. Introduction

People converse with each other for many reasons: to exchange information, to discuss an issue, to resolve a conflict, to increase mutual understanding, to compose a joint story, or just for fun. Conversation will remain as a vital means for people to communicate with other people and autonomous agents in the emerging human-agent symbiotic society. People are sufficiently proficient and adaptive in expressing and interpreting thoughts and feelings by exploiting a sophisticated structure and dynamism of conversational interaction.

* Corresponding author. Tel.: +81-75-753-5371.
E-mail address: nishida@i.kyoto-u.ac.jp

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2015.08.099

Conversational informatics[1,2] is a field of research that focuses on conversational interaction. Rather than conceptualizing conversation as typical talking activities around a round table or space, it attempts to shed light on a much broader class of social interactions in the wild involving a great deal of interactions between the participants and environment. On the scientific side, it attempts to unveil how mental processes interact with each other to share thoughts and feelings using social signals. On the engineering side, it aims at designing and implementing cognitive artifacts that can fluently interact with people and possibly with other cognitive artifacts in a conversational fashion. Typical applications include conversational agents and robots that can interact with the user in a friendly fashion.

Synthetic evidential study (SES for short)[3,4] is a unique showcase application area of conversational informatics. SES leverages powerful game technologies[5] to engage participants in a collaborative study on unveiling mysteries and other kinds of social processes by combining dramatic role play, agent play and group discussions to spin a story as a joint interpretation. Work is in progress to build a SES support system by integrating techniques that have been cultivated in conversational informatics.

In this paper, I present a data-intensive approach to conversational informatics. It not only brings about quantitative understanding, permitting us to turn a great accumulation of keen observations into a pile of computational models, but also helps to build conversational agents by virtue of recent progress in machine learning and data mining. The topics include a smart conversation space for allowing people to engage in conversation in a cyber-physical space, conversational interaction capture for conversation modeling and content production, learning by imitation for producing conversational interactions, cognitive design for understanding mental processes of conversational actors, and synthetic evidential study for collaborative study on understanding social processes.

2. A brief historical overview of inquiries into conversations

Conversation has been a vital subject of philosophical study since early days of mankind. Ancient Greek philosophers[6] sublimated conversation into dialectic, an interactive approach to truth. Saussure[7] contrasted parole, or use of language, from langue, the system of language. Wittgenstein's view of conversation as a language game[8] that encompasses the use and the meaning of language resulted in the establishment of social sciences, social constructionism in particular. Erving Goffman[9,10,11,12] introduced numerous ideas to frame conversation as social activities. Kendon[13,14] and McNeill[15] conducted a substantial work on nonverbal communication behaviors, followed by conversational analysis by Sacks,[16] Schegloff,[17] Goodwin,[18] and others. Austin[19] and Searle[20] proposed speech act theory to investigate pragmatics of language in conversational interactions. Clark[21] provided with a comprehensive theory of language use, known as cognitive linguistics.

Apart from those standard perspectives, I believe two more viewpoints are important: conversation as narrative and cognitive processes underlying conversations. The former allows us to view conversation at a coarse approximation level in which a conversation is characterized as a process of exchanging small talks or chitchats that are regarded as components of larger stories. It sheds light on the content-oriented aspects of conversation, particularly how pieces of conversation contribute to the structure and organization of the content to be shared among participants or an individual participant's personal memory organization. In the meanwhile, the cognitive process viewpoint enables us to look inside the mental space of each participant by focusing on the mental processes for interpreting or producing verbal and nonverbal behaviors in conversation.

On the engineering side, a synthetic approach has been taken to build conversational systems since early days of artificial intelligence research in the 1960s. After the initial success of natural language question answering systems, numerous AI researchers became interested in extending them as interactional systems that could pursue goal-oriented dialogue,[22,23] spoken dialogue,[24] or multi-modal dialogue.[25] Eventually, those endeavors converged as embodied conversational agents or intelligent virtual agents, after the groundbreaking Knowledge Navigator movie based on a book by Sculley and Byrne[26] released by Apple, Inc. in 1987. This movie eloquently illustrated how an artificial intelligence system employed as an embodied conversational agent could help people. Inspired researchers started to build embodied conversational agents and intelligent virtual agents, e.g., Peedy,[27] which bore key features of the agent illustrated in The Knowledge Navigator movie, such as, anthropomorphism and verbal-nonverbal interactions.

From methodological points of view, approaches to build conversational systems can be classified into three groups. The earliest approaches[28] placed an emphasis on the principles and the system architecture. Discussions were

made on finding the architecture on bringing about modularity and flexibility. Researchers made a certain amount of trial-and-error to realize a mostly partial set of functions necessary for building conversational systems. The most classic approach is a hierarchical method in which the processing proceeds from the perception of the environment, followed by a semantic processing, from which the output is generated. The blackboard architecture is employed to achieve flexibility in control. The most successful case was HEARSAY-II speech dialogue system.[24]

In the second approaches,[29] scripts and mark-up languages were developed as a handy means for developing conversational systems. Scripts and markup languages are used to specify the behavior of a conversational agent. They allow authoring conversational agent scenarios that can interact with the user or other agents without committing to implementation details. A script language is more like a high-level programming language. An interpreter takes an expression in a given script language to produce animation for a given situation. In contrast, a markup language is less procedural, allowing the target animation to be specified without complete procedural information, which introduces some flexibility. An action planner and an action realizer are required for procedural interpretation of markup language expressions to produce animation.

In the third approaches,[1] corpora are used to base target behaviors of conversational agents on varieties observed in existing conversations. In corpus-based generation of conversation agent behaviors, one first creates an interaction corpus or a database containing instances of actual conversation behaviors, from which specification of interactions is generated in a markup language. By examining the behaviors of humans quantitatively, one might be able to determine the key features that differentiate appealing gestures from unappealing ones.

Evaluation is indispensable to establish a solid understanding technology. Generally, evaluation criteria can be classified as usability and user perception.[30] Usability is related to task performance and can be investigated in terms of learnability, efficiency, and error. On the other hand, user perception is concerned with the way a user perceives the conversational system. User perception may be judged with respect to satisfaction, engagement, helpfulness, trust, believability, likeability, and entertainment.

The previous research in building conversational systems is limited due to the poverty of common ground for human agent interaction. Common ground is a shared knowledge and belief among participants. According to Clark,[21] common ground consists of communal and personal common ground. The former comprises human nature, communal lexicons, cultural facts, norms, and procedures, while the latter involves perceptual bases, actional bases, personal diaries including discourse, friends and strangers, and personal lexicons. Just as common ground is necessary for human-human conversations to proceed efficiently and reliably, it is also indispensable for proficient and trustable human-agent interactions. I proposed a sharing hypothesis,[31] arguing that common ground is mandatory for empathic relations. Even after an endeavor of research over fifty years till today, we do not still have an effective method of making common ground ready for use in conversational systems. It also leads to another limitation of the previous research on conversational system that little efforts have been made on transactional, or narrative-oriented, aspects of conversation, while a majority of previous research has been focusing on an interactional aspects. As a result, the conversational systems still remain as an alien to humans and premature for sharing and cultivating wisdom in a community. Due to the large volume and tacitness of communal common ground, it is not easy to establish a dependable common ground for ordinary conversation. It is not feasible to hand-craft the communal common ground due to its volume. It is still beyond the reach of machine learning due to its tacitness. Although it is likely that a good delineation of communal common ground could be obtained through conversation, good conversation cannot be obtained without common ground. It is indeed a chicken-and-egg problem.

3. A data-intensive approach to conversational informatics

A data-intensive approach to conversational informatics[2] is concerned with acquisition and utilization of data regarding how participants interact with each other, what information to be shared, and which aspects of the environment are relevant. Its primary aim is to build a computational framework for sharing and cultivating wisdom through enhancing conversational interactions and facilitating conversational content in a community. We are keen to provoke and support empathic conversation in which participants are engaged in a game-like activity to make tacit thoughts explicit and organize them into a larger discourse in a very effective trial-and-error fashion. Common ground building is a key issue to make it happen. In order to break through the chicken and egg deadlock, we attempt

to realize a *primordial soup of conversation*, an initial core beyond a critical mass so that it can embrace nutrition rich enough to attract participants' engagement in productive conversations from which conversational systems can learn to improve common ground. Toward this end, we equally place an emphasis on interactional and transactional aspects of conversation. Thus, conversational environment and content is an integral part of the topics covered in conversational informatics. Conversational environment provides the participants with a place for interaction and an inventory of potential referents which might be referred to in conversation. Conversation content represents what is manifested during conversation. Conversation content may be produced offline or online and consumed during a conversation. Conversation content might be transformed and accumulated by a person to produce larger units of a story.

The primary theoretical backbone is conversation quantization that integrates the interactional and transactional aspects of conversation. The conversation quantization is a theory that characterizes conversation as a series of conversational quanta, each of which packages information about relevant participants, references to the objects and events discussed in the discourse, a series of verbal and nonverbal utterances exchanged by the participants, commitments to previous discourse (themes), and new propositions in the discourse (rhemes).[2] Although we assume that each conversation quantum can be seen as a dictation of conversation segment from a real or hypothetical observer, we also assume that each participant can normally interpret the conversation segment in a similar fashion to internally produce conversation quanta for later use to interpret conversational scenes or produce conversational behaviors in similar situations. The episodic memory of each participants might be regarded as a structured collection of conversation quanta while her/his semantic memory may consist of more abstract or prototypical entities. Long-term memory processes will intervene to generalize episodic memory into semantic memory to serve as a prototype for coping with broader discourses in the future.

On the computational side, it consists of smart conversation space, conversation capture, conversation production, cognitive approach, and an integrated approach. In the following sections, I will describe them in more detail.

4. Smart conversation space

Smart conversation space is a shared interaction environment, created by mixed reality ranging from virtual reality to augmented reality, which encompasses participants and referents of conversation. It is not only used to entertain the users by providing with an intelligent environment for conversation, but also to scientifically investigate conversation under various designed settings. We have been working with two types of smart conversation space: an open smart space where actors can move around to dynamically interact with each other and the environment, and an immersive interaction space where actors are individually embedded and interact with each other through interconnected ambient audio-visual cells. In this paper, we focus on the latter.

4.1. Immersive Collaborative Interaction Environment (ICIE)

Our Immersive Collaborative Interaction Environment (ICIE)[2] is a facility that can embrace the user in a space cylindrically surrounded by 8 large displays about 2.5 m diameter and 2.5 m in height and multiple speakers together with plug-in audio-visual sensors for capturing user behaviors therein. It features rich ambient information for the user to enjoy with low cognitive load and the ability to capture human motion with few physical constraints. It allows the user to control an avatar or tele-operated robot in a situated fashion in a human-agent interaction (HAI) environment from a first-person perspective. More than one ICIE can be connected with each other to virtually compose a shared space for conversation. Alternatively, an ICIE can be connected to a mobile robot with an omnidirectional camera to allow the user in an ICIE to interact with people using a mobile robot as her/his physical surrogate in a remote environment. ICIE has been used for numerous projects including virtual meeting, video-game, tele-presence, to name just a few. From the viewpoint of conversational informatics, ICIE provides with a useful environment for collecting behavioral data on conversation in various settings designed by researchers. In addition, we have recently introduced a dome display that can produce much stronger immersion.

An alternative approach to a shared virtual space is to employ head mount displays (HMDs), which provide the user with a portable environment to enter the virtual space from almost everywhere. Although ICIE has no such

portability, it can provide the user with a stable means and wide view angle for sharing the virtual environment with a usually small number of physically co-located people without wearable apparatus such as HMD.

4.2. Background capture

The role of background capture is to build a 3D model for a portion of the physical world to be projected on the immersive display from given viewpoints. Different techniques are used to capture the outdoor and indoor scenes. The former requires only a digital camera to capture the outdoor scene. It combines multiple computer-vision techniques including structure from motion, multi-view stereo, and depth map.[32] Given an outdoor scene, our system will first reconstruct a rough 3D geometry from real world photos using the stereo method. After that, it will build panorama images. The system will also create a depth-map for each panorama image from the 3D geometry. It will interpolate between panorama image pairs if the user moves to a position for which image data does not exist. As the user walks in the virtual space, a panoramic image will be updated accordingly almost in real time. In a preliminary evaluation, we have found that the current version can automatically reconstruct from approximately 5,000 digital photos a 3D scene for a 50m x 50m space in one day.

To capture the indoor scenes, one system[31] reconstructs camera poses and a polygon mesh from digital photos. It uses random sampling like RANSAC to estimate camera pose to cope with white balance difference between images. The system extracts the ground level from a polygon mesh and smoothens it. Another system uses the SLAM (simultaneous localization and mapping) method based on the scanned data of the environment from a single Kinect sensor. It reconstructs a 3D model of the surrounding environment using image features and the depth map. It initially estimates the relative 3D location of the scanning Kinect sensor using image features. After the estimation of the Kinect sensor position of each scanned frame, it will integrate the scanned depth map data and reconstruct the scene. We can also connect the outdoer and indoor capture to produce a unified virtual space.

5. Capturing conversational interaction

Our repertoire of conversational behavior capture consists of capturing group interaction in a narrow space, capturing first-person view by corneal imaging, and estimating the physiological indices of the user in conversational interaction.

5.1. Capturing group interaction in a narrow space

Three-dimensional conversation capture by multiple Kinects (3DCCbyMK) [2,34,35,36] measures conversational interactions in a space with a 3-m diameter to produce a 3-D movie that allows one to observe the activities virtually from any viewpoints together with the background. The target may include up to four participants and objects such as a table and a wall. The participants may move around the space and talk with each other with nonverbal behaviors. The background is assumed to be static. The current version of 3DCCbyMK captures the participants' behavior and the background separately and merge the resulting 3D models into one. The background is captured using the indoor background capture method described in the previous section. To capture the interactions of participants, skeleton data from multiple Kinect sensors are used to estimate the motion of the participant. Firstly, personal IDs are allocated to each set of skeleton data for each Kinect sensor. Secondly, each set of skeleton data is projected into an integrated coordination where the personal IDs are integrated based on the overlap of the skeleton coordination data. Finally, each joint coordinate of the skeleton data is integrated with weights determined by various heuristics, such as how many sensors have captured the joint. Time series data are also checked for the consistency of joint recognition. To reduce the overhead of calibrating multiple Kinect sensor coordination, the 3DCCbyMK integrates them using the skeleton data and depth map.

5.2. First-person view by corneal imaging

Capturing the first person view of the world provides a valuable means for estimating the mental status of a human. In fact, we have found that first-person view may bring about quite different emotional state from the third-

person view in human-robot interaction.[37] Our corneal imaging technology determines the point of gaze (PoG) and estimating the visual field from reflecting light at the corneal surface using a closed-form solution. Our method allows for equipment and calibration-free PoG and peripheral vision estimation.[38,39]

5.3. Physiological indices

The user's emotion and stress need to be monitored for a conversational system to sustain pleasant interaction with the user. Agent's advices may not be well-accepted by the user unless she or he is open to advices, i.e., not occupied by some thoughts or committed to some thought. Physiological indices provide us with a useful means for estimating the user's mental state and process. For example, the mental stress may be estimated by measuring the low frequency/high frequency heart rate variability (LF/HF). We have found that the combination of LH/HF and the skin conductance response (SCR) allows us to estimate the degree of concentration during VR exercise games.[40] Furthermore, we have found that the user accepts advices from a conversational agent when the mental status of the user is taken into account.

6. Producing conversational behaviors

Learning by imitation allows conversational agents to estimate communication principles from demonstrated sample of interactions. Our early work of learning by imitation focused on nonverbal communications. Our method[41,42,43,44] consists of four stages. At the discovery stage, the basic actions and commands will be discovered. At the association stage, a probabilistic model will be generated which specifies the likelihood of the occurrence of observed actions as a result of observed commands. Granger causality is used to discover natural delay. At the controller generation stage, the behavioral model will be converted into an actual controller to allow the robotic agent to act in similar situations. Finally, the gestures and actions learned from multiple sessions will be combined into a single model at the accumulation stage. Although this framework can demonstrate the basic ability of learning by imitation, it is limited in that the building blocks are rather primitive (the low-level primitive problem) and it cannot decide for itself when it needs to imitate the instructor and improve its skills by watching other people doing the skills it already learned (the context problem).

To address the low-level primitive problem, we designed, implemented and evaluated of a closed loop pose copying system.[45] 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.

The context problem contains three challenges: the data-sparsity challenge (the number of available demonstrations is usually limited), the distortions challenge (demonstrations may not be perfect in the sense that some of the demonstrations may be failed-attempts, or some of them may be contaminated by noise or bursts of nonlinear signal distortions), and the confusion challenge (demonstrations are not always correctly segmented).

To address the second and third problems, we have developed a new algorithm called SAXImitate, which utilizes a novel learning from demonstration system called MSAX[46] that in turn builds on the Symbolic Aggregate approXimation (SAX[47]) time series symbolization algorithm to handle multi-dimensional time series. SAX uses a symbolic representation of time series that was first introduced by Lin *et al*. SAX takes as input a single dimensional time series of length T and produces its representation in the form of a string of N characters ($N \leq T$) where each character is taken from a M valued alphabet (e.g. the numbers from 1 to M). We introduced three methods to extend SAX to handle multi-dimensional data. SAXImitate takes as input a set of K multidimensional time series X_i of lengths T_i (possibly different) and dimensionality D that represent different instances of the demonstrated action to be learned, and produces Z-Matrix that assigns the relative importance/confidence to every input demonstration for every T/N steps (now represented as individual symbols) of the original task.

The main strength of SAXImitate is in being able to handle distortions and confusions in the demonstrations that are expected when the robot is collecting its own demonstrations and automatically segmenting them. SAXImitate provides a simple approach to alleviate the problem of utilizing the information in multiple demonstrations by using

the model learned through SAXImitate (which encodes useful information from all demonstrations) as an input to the dynamic motor primitive (DMP) learner.

On the practical side, the learning by imitation method needs to sustain the motivation of demonstrators. To address this issue, we have conducted three studies to explore the effects of imitating a robot (back imitation) on human's perception of this robot in terms of imitative skill, interaction quality, humanness of the robot and intention of future interaction. The results of these studies taken together show that subjects preferred the robot that they previously imitated in terms of imitation skill, naturalness, and motion human-likeness compared with the robot that they did not imitate.[37] There was no difference between the simple back imitation and the more complex mutual imitation conditions in the main study. This result supports the use of a back imitation familiarization session before learning form demonstration sessions. This result emphasizes the importance of taking the interaction context (and history) into consideration when attributing differences in people's responses to robots.

7. Cognitive design

In order for wisdom to emerge and develop in conversation, we need to design a cognitive model for conversational agents so that they can sense social signals, estimate the tacit intentions underlying them, manifest enough presence as a communication partner, and even lead discussions.

7.1. Communication competence

One of the challenging goals in human-agent interaction is to design dependable autonomous agents that can be deemed as our partner to collaborate with us to accomplish a task in a complex domain. Towards this end, we designed a virtual basketball game[48] to investigate how people behave and how much communicability we can endow artificial agents in a multi-player, real-time, situated joint activities where collaboration and competence are deemed critical factors. We modeled the basketball game using Clark's joint activity theory. We characterized the top level of the basketball game as a collection of joint projects, including those for getting the ball into the opposition's hoop and stopping the other team from scoring. We modeled lower levels in terms of joint projects for passing and catching the ball, or running certain plays. By observing how people use their avatars to play virtual basketball, we identified various signals players use intentionally to communicate with each other, such as a gaze towards a team-mate to indicate a pass. We identified key modalities players were using explicitly or implicitly to send signals. For example, we found that players used rich body expressions. We prototyped as a research platform a virtual basketball game that will be played in a cyberspace by an ensemble of avatars and agents.

We compared the players' perception of an agent team mate with a higher basketball ability against one with higher communication ability. We found that people were able to distinguish between the two agents, and preferred the one with a higher communication ability even though no difference was perceived in the intelligence of the agents.[49] This would suggest that users prefer the communication ability to the task ability in this environmental, though this could depend on the nature of the game.

7.2. Inducing intentional stance

Intentional stance[50] is a mental abstraction that one may adopt toward an entity when it is best regarded as a human-like being that has intention. We hypothesized that people could take the intentional stance toward an agent if the agent performs goal-oriented actions in human-agent interaction. In one study,[51] we compared a trial-and-error agent that performs goal-oriented actions using multimodal behavior against a text display agent that displays its behavioral intention via text. As a result, we have found that participants continuously tried to communicate with the trial-and-error agent that keeps some reticence while performing the task. We found that the participants felt that the agent using multimodal nonverbal behavior was more goal-oriented, more intelligent and understood their intentions more than the text-display agent.

In another study,[52] we attempted to endow an agent with an ability of dynamically changing its strategy based on user's estimated state to encourage the human partner to take the intentional stance towards the agent. The hypothesis was that if the agent constantly estimates the mental status of the user and motivate her/him only when it

detects a significant drop of her/his motivation, the participants will adopt the intentional stance towards the agent. It was empirically supported by an experiment.

7.3. Discussion support by dynamically estimating the preference structure

People often reorganize the preference structure in dynamic and interactive fashion. The conversational agent should be able to dynamically estimate the partner's preference structure to control its own communicative behaviors accordingly. We represent the structure of potential decisions as a three-dimensions consisting of factor, aspect, and choice. A factor is a concrete feature specifying a property that can be judged in a rather objective fashion. The aspect is an abstract conceptual feature that participants share in characterizing each decision. Unlike factors, we cannot assume that participants always share the definition of aspect. Choice is a potential decision that participants could make as a conclusion of the discussion.[2]

DEEP[53,54] is a method for estimating dynamic preference structure in human-agent dialogues. It integrates nonverbal expressions and physiological indices to repeat cycles of explanation, demand seeking, and decision to jointly formulate a preference structure shared by the human and the agent.

DEEP is extended to gDEEP[55] for estimating the status resulting from participants' preference structure in a group discussion. Our method is used to estimate the divergence and convergence processes in group discussions and produce appropriate social signals as a result. If it has turned out that the group has not yet well formulated an emphasizing point, our 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.

8. Synthetic evidential study as a showcase of conversational informatics

Synthetic evidential study (SES) is a novel method of collaborative study on addressing social processes, mysteries in particular that range from fictions to science and history. 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 dramatic 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 dramatical 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 dramatic role play in the previous phase by applying the oral edit commands (if any) to dramatic 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 first- and 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 repeating the above cycle. 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 as a part of the present annotated agent play.

As a result of preliminary experiment with an SES workshop, we have found that participants successfully elaborate the interpretation by engaging in a play as a first-person and contrasting the first- and third- person views in an immersive virtual environment enabled by our technology of conversational informatics.[56]

The SES scheme may be employed in a broad range of joint activities, such as academic research, profiling, planning and strategy formation, and training and education. SES will contribute to the progressive formation of common ground which is knowledge that participants believe to share. According to Clark,[21] common ground is either communal, consisting of human nature, communal lexicons, and cultural facts/norms/ procedures or personal consisting of perceptual bases gestural indications, partner's activities, salient perceptual events, actional bases, and personal diaries.

The SES support system is indispensable to make the SES methodology for everybody. Work is in progress to build the SES support system by combining technologies we developed for conversational informatics research.[2] It consists of an immersive interaction & collaboration environment for the shared virtual world, dramatic group play capture, creating agent play, discussion support, and conversation quantization.

9. Conclusion

Conversational informatics is a field of research that focuses on conversational interaction. A data-intensive approach to conversational informatics is concerned with acquisition and utilization of data regarding how participants interact with each other, what information to be shared, and which aspects of the environment are relevant. I have presented a smart conversation space, conversational interaction capture, learning by imitation, and cognitive design, as major topics in conversational informatics. Synthetic evidential study serves as a showcase for conversational informatics that allows for collaborative study on understanding social processes.

Acknowledgements

This study has been partially supported by JSPS KAKENHI Grant Number 24240023 and 15K12098, the Center of Innovation Program from JST, and AFOSR/AOARD Grant No. FA2386-14-1-0005.

References

1. Nishida T, editor. *Conversational informatics: an engineering approach*. London: John Wiley & Sons Ltd; 2007.
2. Nishida T, Nakazawa A, Ohmoto Y, Mohammad Y. *Conversational informatics–a data-Intensive approach with emphasis on nonverbal communication*. Tokyo: Springer; 2014.
3. Nishida T *et al*. Synthetic evidential study as augmented collective thought process–preliminary report. In: Nguyen NT et al, editors. *ACIIDS 2015, Part I*, LNAI 9011 2015; p. 13–22.
4. Nishida T *et al*. Synthetic evidential study as primordial soup of conversation. In: Chu W *et al*. editors. *DNIS 2015*, LNCS 8999 2015; p. 74–83.
5. Harrigan P, Wardrip-Fruin N, editors. *Second person: role-playing and story in games and playable media*. Cambridge: MIT Press; 2007.
6. Brunet, J. *Early Greek philosophy*, Fifth edition, Cambridge: Makiaea Press; 2010.
7. Saussure, Ferdinand de. *Course in general linguistics*. Edited by Roy Harris, Bloomsbury Academic; 2013.
8. Wittgenstein L, Anscombe GE.M (English translation), Hacker PMS, Schulte J. (editor). *Philosophical Investigations*. Revised 4th edition, Wiley-Blackwell; 2009 (Originally Blackwell Publishing, 1953).
9. Goffman E. *Behavior in public places*. New York: The Free Press; 1963.
10. Goffman E. *Interaction ritual: essays face-to-face behavior*. Chicago: Aldine; 1967.
11. Goffman E. *Frame analysis: an essay on the organization of experience*. New York: Harper & Row; 1974.
12. Goffman E. *Forms of talk*. Pennsylvania: University of Pennsylvania Press; 1981.
13. Kendon A. *Conducting interaction: patterns of behavior in focused encounters*. Cambridge: Cambridge University Press; 1990.
14. Kendon A. *Gesture*. New York: Cambridge University Press; 2004.
15. McNeill D. *Gesture and thought*. Chicago: The University of Chicago Press; 2005.
16. Sacks H, Schegloff EA, Jefferson GA. A simplest systematics for the organization of turn-taking in conversation. *Language* 50: p. 996–735; 1974.
17. Schegloff EA, Jefferson G, Sacks H. The preference for self-correction in the organization of repair in conversation. *Language* 53(2): p. 361–382; 1977.
18. Goodwin C. *Conversational organization: interaction between speakers and hearers*. New York: Academic Press; 1981.
19. Austin J. *How to do things with words*. Cambridge: Harvard University Press; 1962.
20. Searle J. *Speech acts*. Cambridge: Cambridge University Press; 1969.
21. Clark H. *Using language*. Cambridge: Cambridge University Press; 1986.
22. Weizenbaum J. ELIZA–a computer program for the study of natural language communication between man and machine. *Commun ACM* 9(1):36–45; 1966.
23. Winograd T. *Understanding natural language*. Orlando: Academic Press Inc; 1972.
24. Erman LD, Hayes-Roth F, Lesser VR, Reddy DR. The Hearsay-II speech-understanding system: integrating knowledge to resolve uncertainty. *ACM Comput Surv* 12(2) 1980; p. 213–253.
25. Bolt RA. "Put-that-there": voice and gesture at the graphics interface. SIGGRAPH Comput Graph 14(3) 1980; p. 262–270.
26. Sculley J, Byrne JA. *Odyssey: pepsi to apple: a journey of adventure, ideas, and the future*. New York: HarperCollins; 1987.
27. Ball G et al: Lifelike computer characters: the persona project at Microsoft research. In: Bradshaw JM (ed) *Software agents*. Menlo Park: AAAI/MIT Press; 1997.

28. Cassell J, Sullivan J, Prevost S, Churchill E, editors. *Embodied conversational agents*. Cambridge: The MIT Press; 2000.
29. Prendinger H, Ishizuka M, editors. *Life-like characters: tools, affective functions and applications*. Heidelberg: Springer; 2004.
30. Ruttkay Z, Pelachaud C. *From brows to trust: evaluating embodied conversational agents*. Berlin: Springer; 2004.
31. Nishida T. Towards mutual dependency between empathy and technology, 25th anniversary volume, *AI & Society* 28(3) 2013; p. 277–287.
32. Mori S, Ohmoto Y, Nishida T. Constructing immersive virtual space for HAI with photos. *GRC 2011*; p. 479–484.
33. Mori S. *Constructing realistic virtual spaces where agents move smoothly*. Unpublished master thesis, Graduate School of Informatics, Kyoto University, 2013 (in Japanese).
34. Yano M. *Construction of 3-Dimensional Recording Environments for Multi-party Conversation with RGB-Depth Sensors*. Unpublished master thesis, Graduate School of Informatics, Kyoto University, 2012 (in Japanese).
35. Izukura T. *Indices extracting system for ballroom dance through tracking important body parts referred to in teaching*. Unpublished master thesis, Graduate School of Informatics, Kyoto University, 2013 (in Japanese).
36. Yoshino M. Constructing knowledge structure of dance from teacher's instruction behavior and process of student proficiency. Unpublished master thesis, Graduate School of Informatics, Kyoto University, 2015 (in Japanese).
37. Mohammad Y, Nishida T. Why should we imitate robots? effect of back imitation on judgment of imitative skill. *Int J of Soc Robotics*; Published online, 2015.
38. 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*) 2013; p. 298–304.
39. Nakazawa A, Nitschke C, Nishida T. Non-calibrated and real-time human view estimation using a mobile corneal imaging camera. to be presented, *WEaAX* 2015.
40. Takeda S, Nishida T, Ohmoto Y. Method of Estimating Concentration in Exercise Game by Combining Multiple Physiological Indices. *JSAI Annual Convention* 2015 (in Japanese).
41. Mohammad Y, Nishida T, Okada S. Unsupervised simultaneous learning of gestures, actions and their associations for human-robot interaction, In: Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems (IROS); p. 2537–2544.
42. Mohammad Y, Nishida T. Learning interaction protocols using Augmented Bayesian Networks applied to guided navigation. In: Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 4119–4126.
43. Mohammad Y, Nishida T. Learning interaction protocols by mimicking: Understanding and reproducing human interactive behavior, *Pattern Recognition Letters*; Available online, 11 December 2014.
44. Mohammad Y, Nishida T. Shift density estimation based approximately recurring motif discovery, *International Journal of Applied Intelligence* 2015; 42(1): 112–134.
45. Mohammad Y, Nishida T. Human-like motion of a humanoid in a shadowing task. In: *Proc. International Conference on Collaboration Technologies and Systems* (*CTS*) 2014; p. 123–130.
46. Mohammad Y, Nishida T. Robust learning from demonstrations using multidimensional SAX. *ICCAS 2014*; 64–71.
47. Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. In: *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery* 2003; p. 2–11.
48. Lala D, Nitschke C, Nishida T. User perceptions of communicative and task-competent agents in a virtual basketball game. *ICAART* 2015.
49. Lala D, Mohammad Y, Nishida T. A joint activity theory analysis of body interactions in multiplayer virtual basketball, *28th British Human Computer Interaction Conference* 2014.
50. Dennett D. *The intentional stance*. Cambridge: The MIT Press; 1987.
51. Ohmoto Y, Furutani J, Nishida T. Induction of intentional stance in HAI by presenting goal-oriented behavior using multimodal information. *CogSci*; 2015.
52. Suyama T, Ohmoto Y, Nishida T. Improving engagement of users by changing agent's strategy action dynamically based on the observed user's state. *JSAI Annual Convention* 2015 (in Japanese).
53. Ohmoto Y, Miyake T, Nishida T. Dynamic estimation of emphasizing points for user satisfaction evaluations. *Proceedings of the 34th Annual Conference of the Cognitive Science Society* 2012; p. 2115–2120.
54. Ohmoto Y, Kataoka M, Nishida T. Extended methods to dynamically estimate emphasizing points for group decision-making and their evaluation. *The 9th International Conference on Cognitive Science* 2013.
55. Ohmoto Y, Kataoka M, Nishida T. The effect of convergent interaction using subjective opinions in the decision-making process. *The 36th Annual Conference of the Cognitive Science Society* 2014.
56. Ookaki T *et al*. Synthetic evidential study for deepening inside their heart. In: *IEA/AIE 2015*; p. 161–170.