1

Introduction

Toyoaki Nishida

1.1 Conversation: the Most Natural Means of Communication

Conversation is the most natural and popular means for people to communicate with each other. Conversation is everywhere around us, and even though you may feel that making conversation is relatively effortless, a closer look shows that a tremendous amount of sophisticated interaction is involved in initiating and sustaining conversation.

Figure 1.1 illustrates a normal conversation scene in our daily research activities. People use a rich set of nonverbal communication means, such as eye contact, facial expressions, gestures, postures and so on, to coordinate their behaviors and/or give additional meaning to their utterances, as shown in Figures 1.1(a) and (b), where participants are passing the initiative in the discussion from one to the other, or keeping it, by quickly exchanging nonverbal signs such as eye contact and facial expression as well as voice. People are skilled in producing and interpreting these subtle signs in ordinary conversations, enabling them to control the flow of a conversation and express their intentions to achieve their goals. Occasionally, when they get deeply involved in a discussion, they may synchronize their behavior in an almost unconscious fashion, exhibiting empathy with each other. For example, in Figures 1.1(c)–(f), a couple of participants are talking about controlling robots using gestures; they start waving their hands in synchrony to refer to the behavioral modality of a particular gesture. After a while, they become convinced that they have established a common understanding. In addition, their behavior enables the third and fourth participants to comprehend the situation and note it as a critical scene during the meeting.

In general, nonverbal means of communication play an important role in forming and maintaining collaborative behaviors in real time, contributing to enhancing engagement in conversation. The process normally takes a short period of time and is carried out almost unconsciously in daily situations.

Nonverbal interaction makes up a significant proportion of human–human interactions (Kendon 2004). Some authors consider that nonverbal interaction reflects intentions at a deeper level and precedes verbal communication in forming and maintaining intentions at the verbal level. McNeill suggests that verbal and nonverbal expressions occur in parallel for some psychological entities called growth points (McNeill 2005).

Conversations consisting of verbal and nonverbal interactions not only provide the most natural means of communication but also facilitate knowledge creation through such mechanisms as heuristic production

Conversational Informatics: An Engineering Approach Edited by Toyoaki Nishida

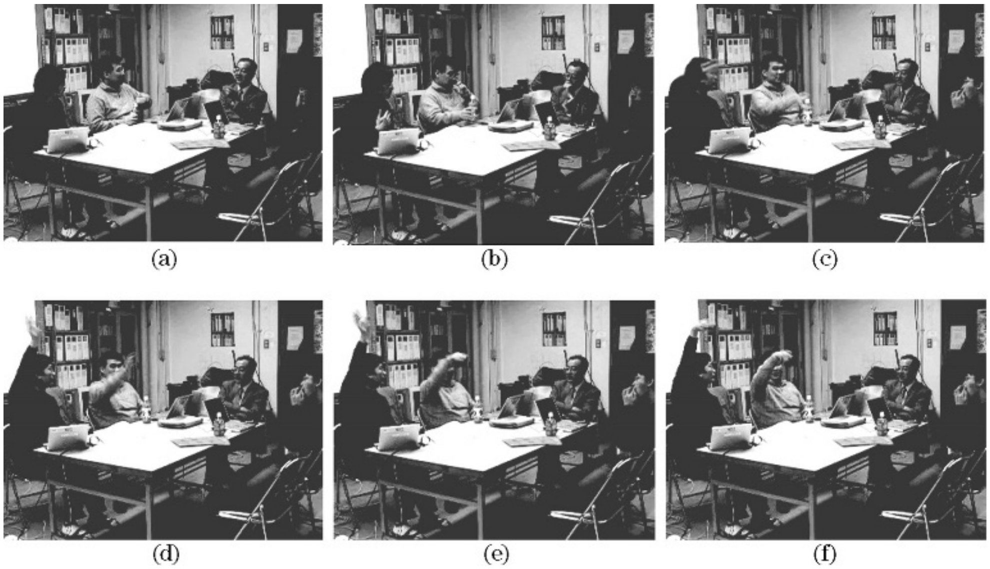


Figure 1.1 Conversation as a knowledge process

of stories from different points of view, tacit–explicit knowledge conversion, and entrainment[1] to the subject.

- *Heuristic production of stories from different points of view.* Conversation can be seen as an improvisational social process that allows each participant to bring together small fragments of stories into larger pieces in a trial-and-error fashion. In a business discussion, for example, participants attempt to find a useful shared story that could benefit themselves and their colleagues. Conversation provides an effective method of negotiating by taking into account the reactions of participants on the fly. The entire process is formulated as a mixture of verbal and nonverbal communication governed by social conventions that reflect a shared cultural background. This shared background constrains choices concerning a common discussion theme, setting up an agenda for the discussion, identifying critical points, raising proposals, arguing for and against proposals, negotiating, establishing a consensus, voting, deriving conclusions, and so on. Effective use of nonverbal signals makes the process of negotiation much more efficient than it would be otherwise.
- *Tacit–explicit knowledge conversion.* Knowledge is explicit or tacit depending on how clearly it is represented by words or numbers. Nonaka pointed out that conversion between tacit and explicit knowledge plays a key role in knowledge creation (Nonaka and Takeuchi 1995). Conversations provide an opportunity and motivation to externalize tacit knowledge. During a discussion, each participant tries to figure out how to express an idea to achieve the intended effect; for example, participants may propose, counter propose, support, challenge, negate, etc., during conversational discourse. In a debate, participants try to be the most expressive speaker in their search for new points that will win the contest. In contrast, conversations often encourage participants to look at knowledge from different perspectives. As a result, they may discover the incompleteness of existing explicit knowledge, which

[1] The tendency for two or more objects to behave in a similar rhythm.

may in turn lead to the formation of new knowledge. This might be tacit initially until an appropriate conceptualization is found.

- *Entrainment to the subject.* To gain an in-depth understanding of a given subject, individuals need to possess a sense of reality about the situation and to capture the problem as their own. Conversations help people react subjectively as they get involved in a given situation as a role player. Thus, conversations may remind people of past experiences similar to the given situation that can serve as a source of knowledge creation triggered by new analogies or hypotheses. Subjective understanding is indispensable to formulating mutual understanding in a community.

However, conversations also have shortcomings as a communication medium. Arguments with a logically complicated structure or those loaded with references to multimedia information are not well communicated by spoken language alone. Furthermore, conversations are volatile and not easily extensible beyond space and time. Both the content and context of a conversation may be lost quite easily. Even if conversations are recorded or transcribed, it is difficult to capture the subjective nature of a conversation after it has ended.

Advanced information and communication technologies offer huge potential for extending conversation by compensating for its limitations. For example, there are emerging technologies for capturing a conversation together with the surrounding situation. These technologies enable previous conversations to be recovered in a realistic fashion beyond temporal or spatial constraints. They also offer the possibility of building synthetic characters or intelligent robots that can talk autonomously with people in an interactive fashion on behalf of their owner.

1.2 An Engineering Approach to Conversation

Conversation has been studied from various perspectives. Philosophers have considered the relationship between thought and conversation. Linguists have attempted to model the linguistic structure underlying conversation. Anthropologists have worked on describing and interpreting the conversational behaviors that people exhibit in various situations. Communication scientists have investigated how components of conversation are integrated to make sense in a social context.

We take an engineering approach, using engineering techniques to measure and analyze conversation as a phenomenon (Nishida 2004a,b). Our aim is to exploit state-of-the-art technology to capture the details of conversations from various viewpoints and to automatically index the content. Annotation tools are widely used to analyze quantitative aspects of conversation.

We are also attempting to build artifacts that can participate in conversations. A typical example is the construction of embodied conversational agents (ECAs), which are autonomous synthetic characters that can talk with people (Cassell *et al.* 2000; Prendinger and Ishizuka 2004). This is a challenge because conversation is a sophisticated intellectual process where meaning is associated with complex and dynamic interactions based on collaboration between the speaker and listener.

An even more challenging goal is to build communicative robots capable of becoming involved in the conversational process and producing fluent interactions at an appropriate knowledge level (Figure 1.2). Robots that can merely exchange sentences with humans cannot participate meaningfully in human conversations in the real world because people make extensive use of nonverbal means of communication such as eye contact, facial expressions, gestures, postures, etc., to coordinate their conversational behaviors.

Figure 1.3 suggests how an interaction with a robot could go wrong. Suppose a robot is asked to describe an object. Scenario A shows a normal sequence in which the robot attempts to attract joint attention to the object by pointing and looking at the object. In contrast, in Scenario B, the agent points to the object while making eye contact with the human, which could cause an inappropriate emotional reaction on the part of the human. Thus, even a tiny flaw in nonverbal behavior could result in a significant difference in the outcome.

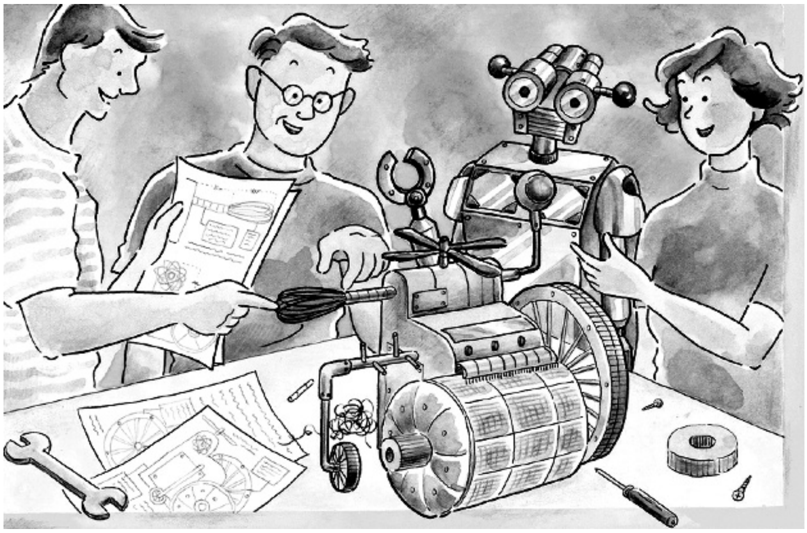


Figure 1.2 Communicative robots that can be involved in the conversation process

Conversation is a process that is initiated, sustained and ended as the result of collaboration between the speaker and listener. For example, when the speaker looks at an object and starts talking about it, the listener should also look at it to demonstrate that he/she is paying attention to the explanation. If the listener loses the trail of the discourse during the speaker's explanation, he/she may look at the speaker's face and probably murmur to signal the fact (Nakano *et al.* 2003). The speaker should recognize the communication flaw and take appropriate action such as suspending the flow of explanation and supplementing it with more information (Figure 1.4).

Robots need to be able to detect the subtle signs that participants produce, capture the meaning associated with interactions, and coordinate their behavior during the discourse. In other words, robots need to be able to play the role of active and sensible participants in a conversation, rather than simply standing still while listening to the speaker, or continuing to speak without regard for the listener.

In contrast to this view of *conversation-as-interaction*, it is important to think about how meaning emerges from interaction; i.e., the *conversation-as-content* view. Consider a conversation scene where the speaker is telling the listener how to take apart a device by saying "Turn the lever this way to detach the component" (Figure 1.5). Nonverbal behaviors, such as eye contact and gestures made by the speaker, not only control the flow of conversation (e.g., the speaker/listener may look away from the partner to gain time to briefly deliberate) but also become a part of the meaning by illustrating the meaning of a phrase (e.g., a gesture is used to associate the phrase "Turn the lever this way" with more detailed information about the way to turn the lever). It also helps the listener find a referent in the real world by drawing attention to its significant features.

1.3 Towards a Breakthrough

Conversation quantization is a conceptual framework for integrating the *conversation-as-interaction* and *conversation-as-content* views to build artifacts that can create or interpret meaning in conversation by interacting appropriately in real time (Nishida *et al.* 2006). The key idea is to introduce conversation quanta, each of which is a package of interaction and content arising in a quantized segment of conversation.

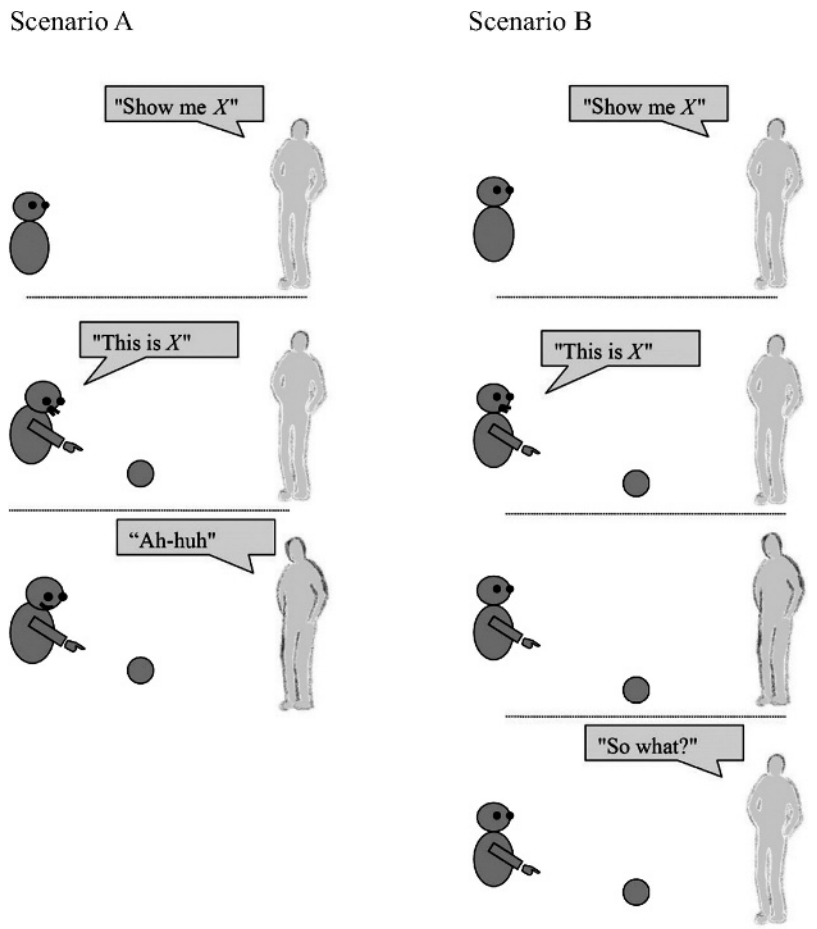


Figure 1.3 Robot is requested to explain object to person

Figure 1.6 illustrates how each conversation quantum represents a fragment of conversation from the situation in Figure 1.5. It contains a description of a visual scene in which the speaker gives an answer to the listener in response to a question that he/she has raised. It also contains a description of an interaction where the speaker is giving an explanation by pointing out a component to the listener, who is listening while paying attention to the object.

Conversation quanta may be acquired and consumed in an augmented environment equipped with sensors and actuators that can sense and affect conversations between people and artifacts. Conversation quanta have a twofold role. Firstly, they may be used to carry information from one conversation situation to another. Secondly, they may serve as a knowledge source that enables conversational artifacts to engage in conversations in which they exchange information with other participants.

Conversation quantization enables conversation quanta to be acquired, accumulated, and reused (Figure 1.7). In addition to the basic cycle of acquiring conversation quanta from a conversational situation and reusing them in a presentation through embodied conversational agents, conversation quanta may be aggregated and visually presented to the user, or may be manipulated, for example by summarization, to

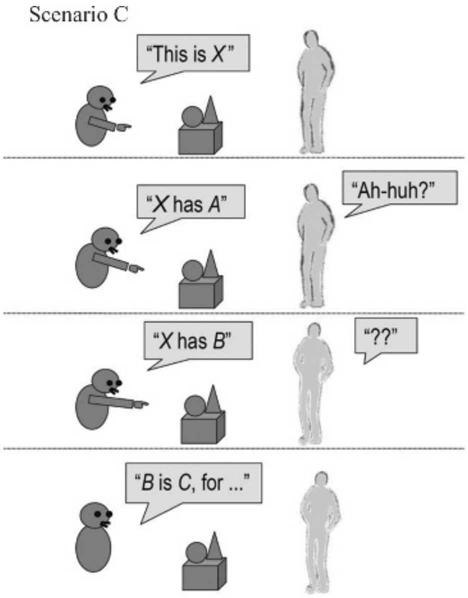


Figure 1.4 Repairing a grounding error

transform one or more conversation quanta into another. They can also be converted to and from various kinds of information archives.

We take a data-intensive approach based on sophisticated measurement and analysis of conversation in order to create a coupling between interaction and content. It will work even if we cannot implement fully intelligent or communicative artifacts from the beginning. In some situations, robots may be regarded as useful even if they can only record and replay conversations. The more intelligent algorithms are introduced, the more autonomous and proficient they will become, which will enable them to provide

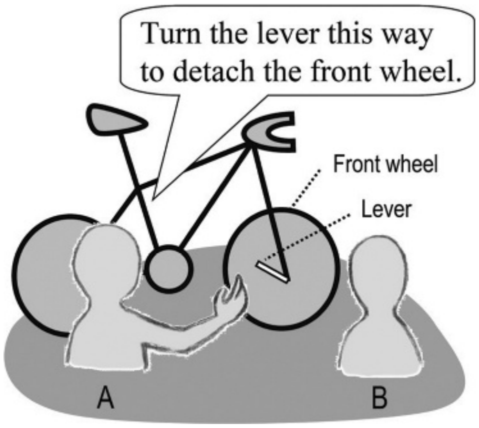


Figure 1.5 Example of a conversation scene

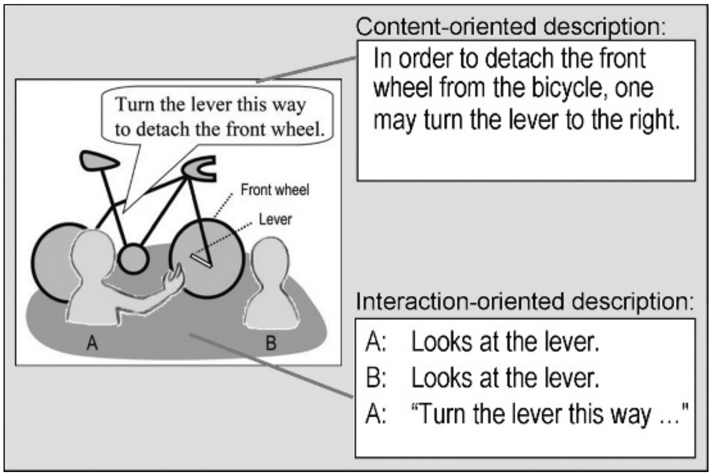

Content-oriented description:
In order to detach the front wheel from the bicycle, one may turn the lever to the right.
Turn the lever this way to detach the front wheel.
Front wheel
Lever
A
B
Interaction-oriented description:
A: Looks at the lever.
B: Looks at the lever.
A: "Turn the lever this way ..."

Figure 1.6 Example of a conversation quantum

services in a more interactive fashion. For example, artifacts may be able to follow nonverbal interaction even without an in-depth understanding, if they can mimic conversational behavior on the surface. In fact, a robot will be able to establish joint attention by creating eye contact with an object when the partner is recognized as paying attention to that object. The media equation theory (Reeves and Nass 1996) suggests that superficial similarities might encourage people to behave as if they were real.

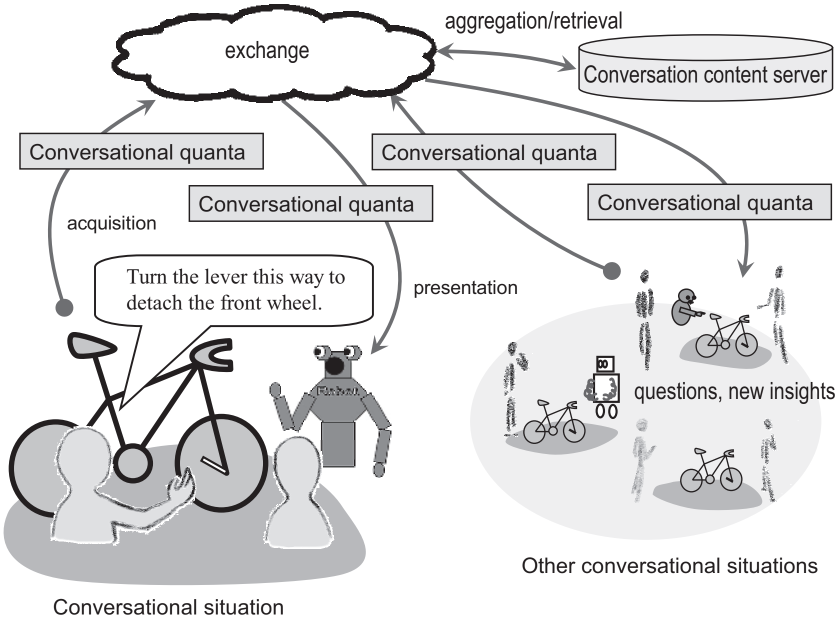


Figure 1.7 Conversation quantization as a framework for circulating conversational content

1.4 Approaches Used in Conversational Informatics

Until recently, various aspects of conversation have been investigated in multiple, disparate fields of research because it is such a complex topic and the theory and technology was so premature that researchers did not consider it feasible to place the entire phenomenon of conversation within a single scope. However, advances in technology have completely changed this situation, enabling us to take an entirely new approach to conversation. Computer scientists have succeeded in developing realistic, embodied conversational agents that can participate in conversation in a realistic setting. Progress in intelligent robotics has enabled us to build various kinds of conversational robots that can share the physical environment with people and undertake a collaborative task. Acoustic stream technology allows us to locate, single out, and track an acoustic stream at a cocktail party, while computer vision and sensor fusion technology is an inexpensive means of recognizing and tracking conversations in the physical environment in real time.

By exploiting these newly developed intelligent information media and processing technologies, conversational informatics brings together fields of research related to the scientific or engineering aspects of conversation. This has given rise to a new research area aimed at investigating human conversational behaviors as well as designing conversational artifacts such as synthetic characters that show up on the computer screen or intelligent housekeeping robots that are expected to interact with people in a conversational fashion.

Conversational informatics covers both the investigation of human behaviors and the design of artifacts that can interact with people in a conversational fashion. It is attempting to establish a new technology consisting of environmental media, embodied conversational agents, and management of conversational content, based on a foundation provided by artificial intelligence, pattern recognition, and cognitive science. The main applications of conversational informatics involve knowledge management and e-learning. Although conversational informatics covers a broad field of research encompassing linguistics, psychology and human–computer interaction, and interdisciplinary approaches are highly important, the emphasis is on engineering aspects, which have been more prominent in recent novel technical developments such as conversational content acquisition, conversation environment design, and quantitative conversational modeling. Current technical developments in conversational informatics center around four subjects (Figure 1.8).

The first of these is conversational artifacts. The role of conversational artifacts is to mediate the flow of conversational content among people. To succeed in this role, conversational artifacts need to be fluent in nonverbal interactions. We address how to build artifacts, such as synthetic characters on a computer screen or intelligent robots that can help the user by making conversation not only using natural language but also using eye contact, facial expressions, gestures, or other nonverbal means of communication.

The second subject is conversational content. Conversational content encapsulates information and knowledge arising in a conversational situation and reuses it depending on a given conversational situation. We examine methods of capturing, accumulating, transforming, and applying conversational content. We address building a suite of techniques for acquiring, editing, distributing, and utilizing content that can be produced and applied in conversation.

The third subject is conversational environment design. The role of a conversation environment is to sense and help actors and artifacts make conversation. We address the design of an intelligent environment that can sense conversational behaviors to either help participants become involved in a collaboration even though they may be in a separate location, or record conversation accompanying the atmosphere of specific conversational behavior for later use or review.

The last subject is conversation measurement, analysis, and modeling. Motivated by scientific interest, we take a data-driven quantitative approach to understanding conversational behaviors by measuring these behaviors using advanced technologies and building detailed quantitative models of various aspects of conversation.

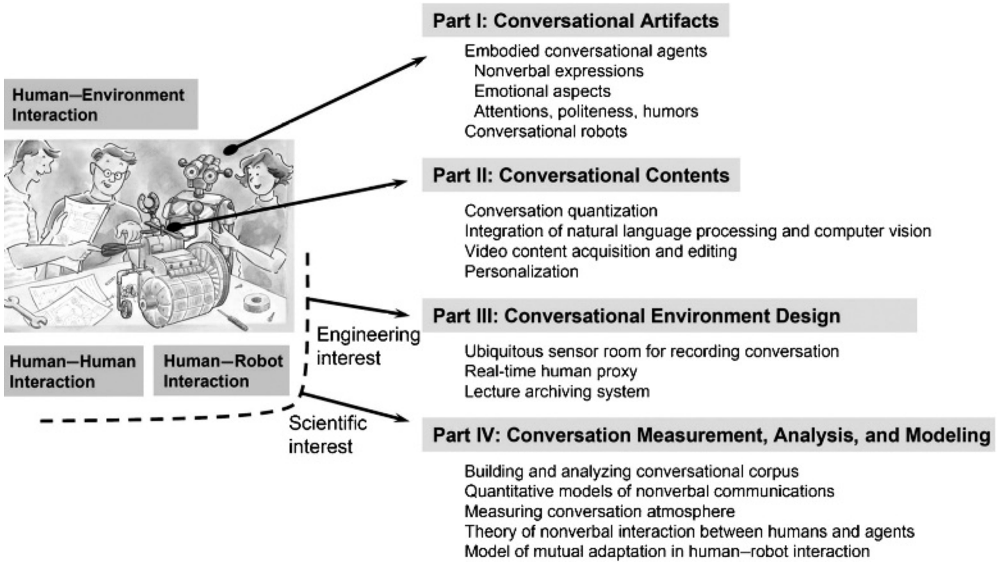

Part I: Conversational Artifacts
Embodied conversational agents
Nonverbal expressions
Emotional aspects
Attentions, politeness, humors
Conversational robots
Human–Environment Interaction
Part II: Conversational Contents
Conversation quantization
Integration of natural language processing and computer vision
Video content acquisition and editing
Personalization
Part III: Conversational Environment Design
Engineering interest
Ubiquitous sensor room for recording conversation
Real-time human proxy
Lecture archiving system
Human–Human Interaction
Human–Robot Interaction
Scientific interest
Part IV: Conversation Measurement, Analysis, and Modeling
Building and analyzing conversational corpus
Quantitative models of nonverbal communications
Measuring conversation atmosphere
Theory of nonverbal interaction between humans and agents
Model of mutual adaptation in human–robot interaction

Figure 1.8 Conversational informatics. This figure originally appeared in Toyoaki Nishida: Prospective view of conversational informatics, *Journal of JSAI* (Japanese Society for Artificial Intelligence), Vol. 21, No. 2, pp. 144–149, 2006. (Reproduced by permission of Japanese Society of Artificial Intelligence)

1.5 Conversational Artifacts

Conversational artifacts fall into several categories. The simplest ones are autonomous, text-based dialogue systems, such as ELIZA (Weizenbaum 1996). Even though they do not have physical body, it is known that people sometimes feel as if they really exist due to media equation effects (Reeves and Nash 1996). More complex systems include embodied conversational agents, which are synthetic characters capable of talking with people. Conversational robots are at the high-end. They are physically embodied and share physical spaces with people.

1.5.1 Embodied Conversational Agents

Conversational informatics is mainly concerned with communicating a rich inventory of conversational content with appropriate use of nonverbal means of communication (Figure 1.9).

Examples of major technological contributions include knowledgeable embodied conversation agents that can automatically produce emotional and socially appropriate communication behaviors in human–computer interaction. IPOC (Immersive Public Opinion Channel; see Chapter 5) enables conversation quanta to be expanded in a virtual immersive environment. Users can interact with conversational agents in a story-space in which a panoramic picture background and stories are embedded. The embedded stories are presented on user demand or spontaneously according to the discourse. The stories are used to represent a discourse structure consisting of more than one conversation quantum.

To generate a more complex set of nonverbal behaviors in an immersive environment, theories of nonverbal communication have been extensively studied and incorporated in systems for generating agent behavior. Conversation management includes the ability to take turns, interpret the intentions of participants in the conversation, and update the state of the conversation. Collaboration behavior

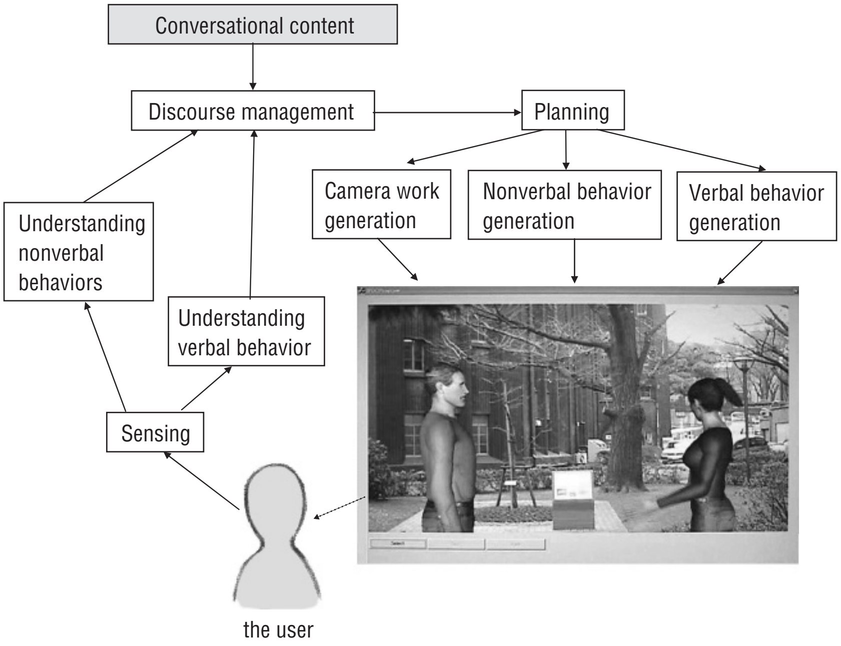

Conversational content
Discourse management
Planning
Camera work generation
Nonverbal behavior generation
Verbal behavior generation
Understanding nonverbal behaviors
Understanding verbal behavior
Sensing
the user

Figure 1.9 Architecture of embodied conversational agents

determines the agent's next action in order to accomplish the goal of the conversation and collaboration with the user. Engagement behaviors consist of initiating a collaborative interaction, maintaining the interaction, and disengaging from the interaction.

The Interaction Control Component (ICC) of the IPOC system interprets inputs from a speech recognizer and a sensor system and generates verbal and nonverbal behaviors performed by conversational agents. The ICC includes a conversation manager (CM), which maintains the history and current state of the conversation, a collaboration behavior generation module (CBG), which selects the next utterance and determines the agents' behaviors in telling a story, and engagement behavior generation (EBG), which determines appropriate engagement behaviors according to the state of the conversation.

In addition to basic communication abilities, it would be much nicer if embodied conversational agents could exhibit humorous behaviors. Nijholt *et al.* (see Chapter 2) consider that humorous acts are the product of an appraisal of the conversational situation. They attempted to have ECAs generate humorous behaviors by making deliberate misunderstandings in a conversation.

To make ECAs believable and sociable, it seems critical to incorporate emotion and personality components. Becker and his colleagues analyzed the notions underlying "emotion" and "personality" and described an emotion simulation system called Max (see Chapter 3). They argue that Max's emotional system increases his acceptance as a coequal conversational partner, and describe an empirical study that yielded evidence that the same emotional system supports the believability and lifelike quality of an agent in a gaming scenario.

Sidner *et al*. (2003) proposed that conversation management, collaboration behavior, and engagement behaviors were communicative capabilities required for collaborative robots. Morency *et al* (see Chapter 7) proposed to combine information from an ECA's dialogue manager and the prediction of head nodding and shaking to predict contextual information for an ongoing dialogue. They point out that a subset of lexical, punctuation, and timing features available in ECA architectures can be used to learn how to predict user feedback.

To make ECAs act politely, Rehm and André (see Chapter 4) applied an empirical study on the impact of computer-based politeness strategies on the user's perception of an interaction. They used a corpus to analyze the co-occurrence of gestures and verbal politeness strategies in the face of threatening situations.

Facial gestures such as various nodding and head movements, blinking, eyebrow movements, and gaze play an important role in conversation. Facial displays are so important as a communication channel that humans use them naturally, often subconsciously, and are therefore very sensitive to them. Zoric *et al.* (see Chapter 9) attempted to provide a systematic and complete survey of facial gestures that could be useful as guideline for implementing such gestures in an ECA.

1.5.2 Robots as Knowledge Media

Another approach is the use of robots as knowledge media, where the role of the robot is to acquire or present information and knowledge contained within the discourse by engaging in appropriate behavior as a result of recognizing the other participants' conversational behaviors.

Nishida *et al*. (2006) focus on establishing robust nonverbal communication that can serve as a basis for associating content with interaction. The proposed communication schema allows two or more participants to repeat observations and reactions at varying speeds to form and maintain joint intentions to coordinate behavior called a "coordination search loop". The proposed architecture consists of layers to deal with interactions at different speeds to achieve this coordination search loop (Figure 1.10).

The lowest layer is responsible for fast interaction. This level is based on affordance (Gibson 1979), which refers to the bundle of cues that the environment gives the actor. People can utilize various kinds of affordances, even though these are subtle. The layer at this level is designed so that a robot can suggest its capabilities to the human, coordinate its behavior with her/him, establish a joint intention, and provide the required service.

The intermediate layer is responsible for interactions at medium speed. An entrainment-based alignment mechanism is introduced so that the robots can coordinate their behaviors with the interaction partner by varying the rhythm of their nonverbal behaviors.

The upper layer is responsible for slow and deliberate interactions, such as those based on social conventions and knowledge, to communicate more complex ideas based on the shared background. Defeasible interaction patterns are employed to describe typical sequences of behaviors that actors are expected to show in conversational situations. A probabilistic description is used to cope with the vagueness of the communication protocol used in human society.

The three-layer model serves as a basis for building listener and presenter robots with the aim of prototyping the idea of robots as embodied knowledge media. The pair of robots serves as a means of

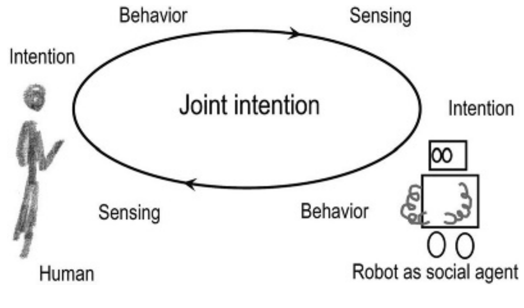
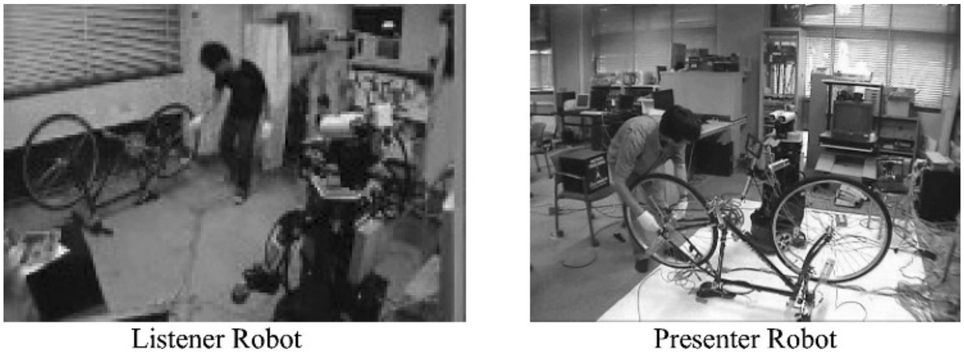


Figure 1.10 Architecture consisting of layers to deal with interactions at different speeds



Listener Robot

Presenter Robot

Figure 1.11 Listener and presenter robots in action

communicating embodied knowledge (Figure 1.11). The listener robot first interacts with the human with knowledge to acquire conversational contents. The presenter robot, equipped with a small display, then interacts with a human to show the appropriate content in situations where this knowledge is considered to be needed. Research is in progress to use conversation quanta as a means for representing knowledge transferred from the listener robot to the presenter robot.

Sidner *et al.* proposed the use of visual processing to extract nonverbal features, in particular those relating to engagement, which would enable robots to talk with humans in a collaborative fashion. As an application, they are working on a robot that collaborates with a human on an equipment demonstration (see Chapter 6).

1.6 Conversational Content

The success of Conversational Informatics depends on the amount of conversational contents circulated in a community. The major technical concern related to conversational content is cost reduction in content production. One approach to this goal is to increase the reusability of content by inventing a method of managing a large amount of conversational content to enable the user to easily create new content by combining and incrementally improving existing content. The other approach is to introduce media processing and artificial intelligence techniques to reduce the cost of content acquisition (Figure 1.12).

- An example of the former approach is the sustainable knowledge globe (SKG), which helps people manage conversational content by using geographical arrangement, topological connection, contextual relation, and a zooming interface. Using SKG, a user can construct content in a virtual landscape and then explore the landscape within a conversational context (see Chapter 10).
- An example of the latter approach is the integration of natural language processing and computer vision techniques. Kurohashi *et al.* (see Chapter 11) succeeded in automatically constructing a case frame dictionary from a large-scale corpus. The case frame dictionary can be used to analyze discourse for content production, including automated production of conversational content from annotated videos or natural language documents.

Nakamura (see Chapter 12) combines an environmental camera module and content-capturing camera modules to capture conversation scenes. In addition, he uses optimization and constraint–satisfaction methods to choose parameters that can be adjusted depending on the purpose of the videos. Using this

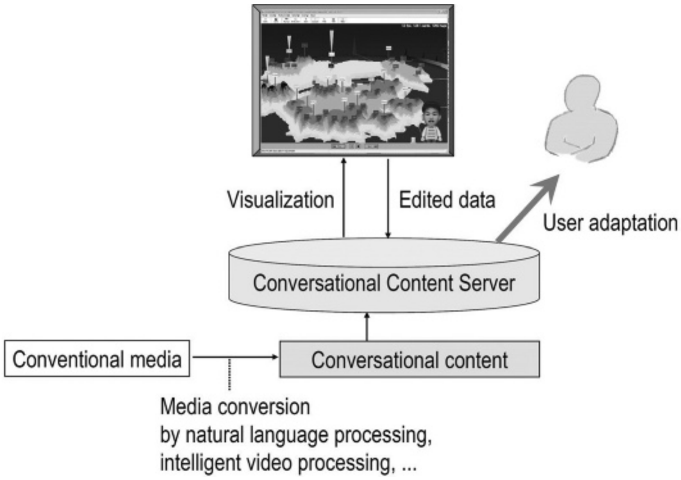


Figure 1.12 Technical issues underlying conversational content

technical platform, he has built a system that can answer and interact with users in a conversational environment.

Content production may also become more effective with the introduction of user adaptation. Babaguchi (see Chapter 13) introduced a technique for acquiring the preferences of the user by observing the user's behavior to implement a personalized video service with tailored video content. The method is applied to personalized summaries of broadcast sports video.

1.7 Conversational Environment Design

One approach to making conversational communication more effective is to embed sensors and actuators into the environment so that conversations can be recorded or measured by sensors and accurate feedback can be given to the participants (Figure 1.13).

Sumi *et al.* (see Chapter 14) built a ubiquitous sensor room for capturing conversations situated in a real space using environment sensors (such as video cameras, trackers, and microphones set up ubiquitously around the room) and wearable sensors (such as video cameras, trackers, microphones, and physiological sensors). To supplement the limited capability of the sensors, LED tags (ID tags with an infrared LED) and IR trackers (infrared signal-tracking devices) are used to annotate the audio/video data with positional information. Significant intervals or moments of activities are defined as interaction primitives called “events”. Currently, five event types are recognized: “stay”, “coexist”, “gaze”, “attention”, and “facing”. Events are captured by the behavior of the IR trackers and LED tags in the room. For example, a temporal interval will be identified as a joint attention event when an LED tag attached to an object is simultaneously captured by IR trackers worn by two users, and the object in focus will be marked as a socially important object during the interval. The accumulated conversation records can be presented to the user in various way, such as using automated video summarization of individual users' interactions by chronologically synthesizing multiple-viewpoint videos, a spatio-temporal collage of videos to build a 3D virtual space for re-experiencing shared events, or an ambient sound display for increasing the level of awareness of the real space by synthesizing spatially embedded conversations.

Nishiguchi *et al.* (see Chapter 16) proposed ambient intelligence based on advanced sensing technology that can record human communication activities and produce feedback without hindering communicative

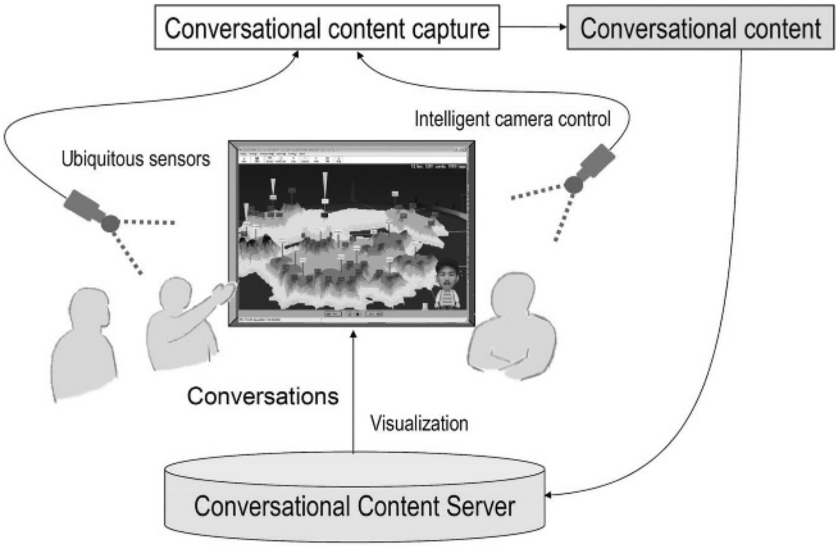

Conversational content capture
Conversational content
Intelligent camera control
Ubiquitous sensors
Conversations
Visualization
Conversational Content Server

Figure 1.13 Technical issues underlying conversational environment design

activities. The concept, which is called “environmental media”, is implemented as an augmented classroom that can automatically highlight and record significant communicative behaviors in a classroom.

Taniguchi and Arita (see Chapter 15) proposed a virtual classroom system using a technique called “real-time human proxy” in which the communicative behaviors of participants at distant sites are encoded using computer vision; a virtual classroom scene in which all the participants are represented by avatars is synthesized in real time.

1.8 Conversation Measurement, Analysis, and Modeling

By exploiting advanced sensor technologies, we are now able to measure various aspects of conversational behaviors, particularly nonverbal behaviors, in greater detail than ever. The insights gained can be applied to the design of conversational artifacts and conversational environments (Figure 1.14).

For example, Xu *et al.* (2006) designed a human–human WOZ (Wizard of Oz) experimental setting to observe mutual adaptation behaviors. The setting enables the researcher to set up an experiment with an appropriate information overlay that may be hidden from subjects. Experiments conducted using the environment, coupled with a powerful annotation and analysis tool such as ANVIL, have yielded interesting findings about mutual adaptation, such as the emergence of pace keeping and timing of matching gestures, symbol-emergent learning, and environmental learning.

Den and Enomoto (see Chapter 17) focused on the responsive actions of listeners including backchannel responses, laughing, head nodding, hand movements, and gazing. They built a multimodal corpus containing records of three-person Japanese conversations annotated with speech transcriptions and the nonverbal communicative behaviors of participants. When they analyzed the distribution of the listeners' responsive actions, they gained some interesting insights; e.g., “there is not only a tendency for the next speaker to gaze at the current speaker toward the end of his turn, but also a tendency for the other listener, who keeps silent during the next turn, to gaze at the next speaker around the end of the current turn and before the next speaker starts speaking”.

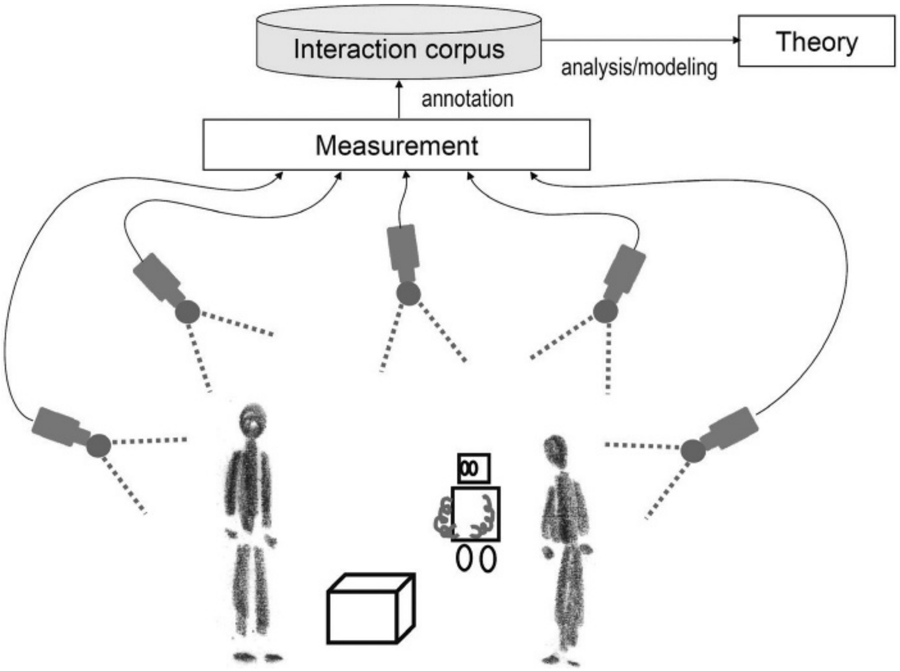


Figure 1.14 Conversation measurement, analysis, and modeling

Kopp *et al.* (see Chapter 8) studied the use of gestures in giving directions. They proposed a framework for analyzing how such gestural images are associated with semantic units (image description features) and morphological features (hand shape, trajectory, etc.).

Nagaoka *et al.* (see Chapter 18) studied the synchrony tendency, a typical phenomenon observed in conversation in which participants' nonverbal behaviors, such as body movement and speech interval, tend to synchronize and become mutually similar. Based on an empirical study in a therapist–client conversation situation, they discuss how various behavioral measures can be used to evaluate the degree of synchrony tendency.

Rutkowski and Mandic (see Chapter 19) proposed three characteristics of a communication atmos phere – based on social features, the mental characteristics of the communicators, and physical features – and discussed how to evaluate the atmosphere by tracking audio-visual signals. Potential applications include semi-automated indexing of multimedia archives and a virtual chat room.

Matsumura (see Chapter 20) proposed a method called an influence diffusion model to identify the roles of individuals in communication (i.e., leader, coordinator, maven, and follower) by analyzing the log of threaded messages. The method is expected to be applied to analyzing conversation records as well as to identifying the influence of various participants.

Interactive agents such as pet robots or adaptive speech interface systems that require forming a mutual adaptation process with users require two competences. One of these is recognizing reward information from the users' expressed paralanguage information, and the other is informing the learning system about the users by means of that reward information. The key issue here is to clarify the specific contents of reward information and the actual mechanism of the learning system by observing how two people could create smooth nonverbal communication, similar to that between owners and their pets.

Ueda and Komatsu (see Chapter 21) conducted a communication experiment to observe how human participants created smooth communication through acquiring meaning from utterances in languages they did not understand. The meaning acquisition model serves as a theory of mutual adaptation that will enable both humans and artifacts to adapt to each other.

1.9 Underlying Methodology

What is common to the above studies? A close look at the details of research in conversational informatics shows that the following methodological features underlie the work described in this book.

1. *Observation, focusing, and speculation.* Conversation is a highly empirical subject. The first thing to do in starting research in this area is to closely observe the conversation process as a phenomenon and focus on a few interesting aspects. Although looking at the entire process of conversation is important, a focus is necessary to derive a useful conclusion. In-depth speculation based on the existing literature is an inevitable part of setting up an innovative research program into highly conceptual issues such as humor or politeness. This approach is effectively used in Chapter 2.
2. *Building and analyzing a corpus.* Intuitions and hypotheses should be verified by data. Building and analyzing an accumulated corpus is effective in both engineering and scientific approaches. In fact, the authors of Chapters 4, 5, and 8 succeeded in deriving useful insights in building embodied conversational agents by analyzing corpora. The authors of Chapter 17 built a novel corpus containing records on multi-party conversations, from which they successfully built a detailed model of eye gaze behaviors in multi-party conversations. The authors of Chapter 14 present a powerful method of using a ubiquitous sensor room to build a conversation corpus.
3. *Sensing and capture.* By exploiting the significant progress made in capturing events taking place in a real-world environment, using a large number of sensing devices embedded in the environment and actors, we can now shed light on conversations from angles that have not been possible before. Conversational environment design, reported in Chapters 14, 15, and 16, has opened up a new research methodology for understanding and utilizing conversations.
4. *Building a computational model and prototype.* A computational model helps us establish a clear understanding of the phenomenon and build a powerful conversation engine. In this book, computational models are used effectively to reproduce the emotional behaviors of embodied conversational agents (in Chapter 3), facial gesture (in Chapter 9), or conversational robots (in Chapters 6 and 7).
5. *Content acquisition and accumulation.* Content is a mandatory constituent of conversation in the sense that we cannot make conversation without content. In addition, conversation can produce high-quality content. Conversational informatics is greatly concerned with capturing and accumulating content that is produced or consumed in a conversation. Chapters 10 and 12 describe new efforts in this direction based on integrating state-of-the-art technologies. Chapter 13 reports the use of personalization to increase the value of content to individual users. These studies are supported by technologies for sensing and capturing.
6. *Media conversion.* Media conversion aims at converting conversational content from/to conventional archives such as documents. Technologies in natural language processing, acoustic stream processing, and image processing are used to achieve these goals, as described in Chapters 11 and 12.
7. *Measurement and quantitative analysis.* Quantitative analysis based on various kinds of measurements provides an effective means of understanding conversation in-depth. Chapter 19 introduces a new measure of evaluating communication atmosphere and Chapter 20 proposes a way of quantifying how much an actor influences other participants in e-mail communications, which may well be applicable to the analysis of face-to-face conversation.

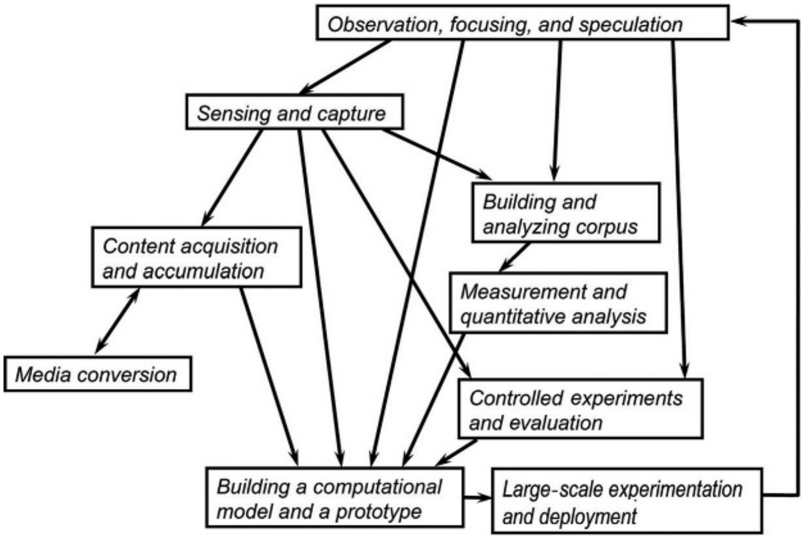

Observation, focusing, and speculation
Sensing and capture
Building and analyzing corpus
Content acquisition and accumulation
Measurement and quantitative analysis
Media conversion
Controlled experiments and evaluation
Building a computational model and a prototype
Large-scale experimentation and deployment

Figure 1.15 Relationship between methodological components: each directed edge represents that the insights or results obtained in the source are used in the destination

8. *Controlled experiments and evaluation.* Designing controlled experiments to verify a working model is a common methodology used in experimental psychology. It can also be applied effectively to experimentally prove novel features of conversation, as shown in Chapters 18 and 21, where models for embodied synchrony and mutual adaptation are validated in psychological experiments.

Figure 1.15 shows the relationship between methodological components. Each arrow shows that the insights or results obtained in the source are used in the destination. One can proceed along the edges as research proceeds. Alternatively, one can focus on one methodological component, assuming the existence of the upstream and evaluating the result in the context of the downstream.

References

Cassell J., Sullivan J., Prevost S. and Churchill E. (eds) (2000) *Embodied Conversational Agents.* MIT Press.

Gibson J.J. (1979) *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

Kendon A. (2004) *Gesture: Visible action as utterance*. Cambridge University Press.

McNeill D. (2005) *Gesture and Thought*. University of Chicago Press.

Nakano Y.I., Reinstein G., Stocky T. and Cassell J. (2003) Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (ACL 03), pp. 553–561.

Nishida T. (2004a) Towards intelligent media technology for communicative intelligence (Keynote speech). In *Proceedings of International Workshop on Intelligent Media Technology for Communicative Intelligence* (IMTCI), pp. 1–7.

Nishida T. (2004b) Conversational knowledge process for social intelligence design (Invited talk). In A. Aagesen *et al.* (eds): *Intellcomm 2004*, LNCS 3283, Springer, pp. 28–42.

Nishida T., Terada K., Tajima T., Hatakeyama M., Ogasawara Y., Sumi Y., Xu Y., Mohammad Y.F.O., Tarasenko K., Ohya T. and Hiramatsu T. (2006). Towards robots as an embodied knowledge medium (Invited paper). Special Section on Human Communication II, *IEICE Transactions on Information and Systems*, vol. E89-D, no. 6, pp. 1768–1780.

Nonaka I. and Takeuchi H. (1995) *The Knowledge-creating Company: How Japanese companies create the dynamics of innovation*. Oxford University Press.

Prendinger H. and Ishizuka M. (eds) (2004) *Life-Like Characters: Tools, Affective Functions, and Applications*. Springer.

Reeves B. and Nass C. (1996) *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Sidner C.L., Lee C. and Lesh N. (2003) Engagement rules for human–robot collaborative interactions. In *Proc. IEEE International Conference on Systems, Man & Cybernetics* (CSMC), vol. 4, pp. 3957–3962.

Weizenbaum J. (1996) “ELIZA”: a computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery* **9**, 36–45.

Xu Y., Ueda K., Komatsu K., Okadome T., Hattori T., Sumi Y. and Nishida T. (2006) WOZ experiments for understanding mutual adaptation. Presented at Social Intelligence Design, 2006.