Editorial Introduction to the Special Articles in the Winter Issue

Turn-Taking and Coordination in Human-Machine Interaction

Sean Andrist, Dan Bohus, Bilge Mutlu, David Schlangen

■ *This issue of* AI Magazine *brings together a collection of articles on challenges, mechanisms, and research progress in turn-taking and coordination between humans and machines. The contributing authors work in interrelated fields of spoken dialog systems, intelligent virtual agents, human-computer interaction, human-robot interaction, and semiautonomous collaborative systems and explore core concepts in coordinating speech and actions with virtual agents, robots, and other autonomous systems. Several of the contributors participated in the AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction, held in March 2015, and several articles in this issue are extensions of work presented at that symposium. The articles in the collection address key modeling, methodological, and computational challenges in achieving effective coordination with machines, propose solutions that overcome these challenges under sensory, cognitive, and resource restrictions, and illustrate how such solutions can facilitate coordination across diverse and challenging domains. The contributions highlight turn-taking and coordination in human-machine interaction as an emerging and evolving research area with important implications for future applications of AI.*

A long-standing pursuit in the field of artificial intelligence has been to enable humans and computers to collaborate seamlessly on shared tasks through natural interaction. Autonomous or semiautonomous systems with collaborative competency enable a myriad of current and future applications from virtual personal assistants embedded in smartphones to robots assisting surgeons in the operating room.

At the core of any collaborative process lies the ability to *coordinate* individual contributions to the process. Humans involved in a collaborative activity continuously coordinate with each other to establish and maintain mutual understanding, called *grounding,* and to coconstruct solutions for their shared tasks. Such coordination happens along multiple dimensions. Due in part to the inherent structure of tasks and in part to cognitive and resource limitations of the interaction channel, the actions produced by collaborators are temporally coordinated. The relative order and timing with which collaborators perform actions is critical for the success of the collaboration. Beyond the temporal domain, participants must coordinate with each other to establish and maintain a shared cognitive basis for future actions and future understanding.

A quintessential example of fine-grained temporal coordination in human-human interactions is the process by which we take turns to speak in a group conversation. This conversational turn-taking process is largely unconscious, as we rarely think of it, allowing us to know intuitively when it is our time to talk, seamlessly finish each other's sentences, and have the ability to inject a quip in a group conversation at exactly the right moment. Research in sociolinguistics, psycholinguistics, and conversational analysis has revealed that turn-taking is a mixed-initiative, locally coordinated process, in which a variety of verbal and nonverbal cues such as eye gaze, body pose, head movements, hand gestures, intonation, hesitations, and filled pauses play a very important role. We continuously produce and monitor each other for these signals and can coordinate seamlessly at the scale of hundreds of milliseconds across these different channels with multiple actors. Can we achieve such precise but powerful coordination with machines?

Beyond turn-taking, coordination on concepts is commonly exhibited in interactions between experts and novices, where the novices quickly learn to adapt to the terminology used by the experts, for example when naming tools. Can we achieve this type of learning and implicit teaching with machines?

Five of the articles in this issue of *AI Magazine* focus on issues of turn-taking, coordination, and collaboration in human-machine interaction. The contributing authors have been working in interrelated fields of spoken dialog systems, intelligent virtual agents, human-computer interaction, human-robot interaction, and semiautonomous collaborative systems, exploring core mechanisms that facilitate coordinating speech and actions with virtual agents, robots, and other autonomous systems. Several of the contributors participated in the AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction, held in March 2015, and several articles in this issue are extensions of work presented at the symposium.

Nigel G. Ward and David DeVault open this issue with an article outlining a core set of 10 challenges for highly interactive dialog systems. These challenges span aspects of modeling, architectures, component technologies, and methods, synthesized from ongoing research in spoken dialog systems, intelligent virtual agents, and human-robot interaction. Next, Gabriel Skantze contributes an article that delves deeper into the technical aspects of temporal coordination, investigating turn-taking in face-to-face interactions. The article explores how the detection and production of both verbal and nonverbal signals from the face and voice can enable fine-grained coordination between a person and robot engaged in a joint activity. In the next article, Joyce Y. Chai and colleagues focus on the cognitive and perceptual aspects of human-robot interaction, exploring techniques for mapping human language to a form understandable by a robot with very different perceptual capabilities from its human counterpart. The article also discusses the degree of collaborative effort that robots might make in order to improve coordination.

The final two articles illustrate the diversity and complexity of challenges involved in human-computer coordination. In the first, Jeremy D. Frank, Kerry McGuire, Haifa R. Moses, and Jerri Stephenson discuss coordination at NASA between an astronaut crew, mission control, and autonomous software in missions planned beyond the moon. Such missions will have to overcome a fundamental obstacle to tight coordination: long communication delays. In the second article, Jill Fain Lehman and Iolanda Leite present two case studies of multiparty interactions between groups of children and an autonomous virtual character. Challenges span from accurately inferring the addressee of utterances (children could be talking to each other or to the character) to maintaining a fun level of engagement without letting the interaction collapse into "chaos" beyond the autonomous capabilities of the virtual character.

Overall, the diversity of problems and techniques in this collection of papers (from coordinating in time to coordinating a shared perceptual basis and from the challenges of interacting with children to the challenges of long temporal delays in space exploration) reflects the natural heterogeneity of an emerging and evolving research area that cuts across traditional boundaries in AI. A dictionary definition of *to coordinate* is "to combine in harmonious relation or action." Unfortunately, interactions between humans and machines are often less than harmonious, marked by awkward speech turns, limited contextual awareness, or errors in action coordination. Progress in this space continues apace and, in the long run, will constitute a core component in enabling seamless, zero-impedance, fully harmonious collaborations between humans and machines.

Sean Andrist is a PhD candidate and research assistant in the Department of Computer Sciences at the University of Wisconsin-Madison.

Dan Bohus is a senior researcher at Microsoft Research.

Bilge Mutlu is an associate professor of computer science, psychology, and industrial engineering at the University of Wisconsin-Madison.

David Schlangen is a professor of applied computational linguistics at Bielefeld University.