Articulatory Adaptation in Multimodal Communicative Action

Melanie Baljko
melanie@cs.utoronto.ca
Department of Computer Science
University of Toronto
Toronto, Canada
M5S 3G4

in Proceedings of the NAACL Workshop on Adaptation in Dialogue Systems,
pages 73-74, PA, June 2001.

Abstract

This paper describes *articulatory adaptation* in human-to-human communication, a computational model of such adaption by human communicators, and the application of this work — intended primarily for the design of communication aids, but with potential application to multimodal communicative agents.

Background

Communicators must adapt their strategies not only in response to the communicative acts performed by *other* interlocutors, but also in consideration of their *own* means for the perception and articulation of acts of communication. The impact of the latter — the focus of our current research — is especially relevant to individuals who are affected by expressive communication disorders. Past research has addressed adaptation by a communicative agent to uncertainty with respect to its own "percepts" (Paek and Horvitz, 2000); in contrast, our work concerns adaptation by a communicator to uncertainty with respect to his or her own embodiment (and the affordances of that embodiment for performing acts of communication). And while past research concerns the design of communicative agents for use in dialogue systems, our work concerns the development of communicative agents for use in computational simulations.

Individuals with expressive communication disorders must adapt their strategies in response to constraints arising from physical disorder, which can impair the individual's ability to perform certain acts of communication and can have intermittent or chronic effects (e.g., neuromuscular dysfunction, resulting in spasticity, atonicity, or excessive fatigue). For instance, the intelligibility of an individual's speech or gestures might vary over the course of an interaction, from somewhat intelligible to not at all. When functional communication is not possible, clinical interventions — possibly in the form of a voice-output communication aid, a type of augmentative and alternative communication (AAC) device — can provide interlocutors with an additional, *aided* mode of communication in the form of synthesized speech. And while these devices can be useful, they are slow and extremely costly to use in terms of physical exertion. In addition, the interfaces of these devices often compete with or subvert the communicator's adaptive strategies. For instance, an interlocutor, anticipating her next conversational turn, must look down at the display of the device in order to compose a spoken utterance precisely at the point where eye gaze is important function in regulating turn-taking.

Such communicators must also adapt to their interlocutors, whose level of familiarity with these devices can signal which communicative strategies are likely to be fruitful. A familiar partner is likely to understand a strategy of using abbreviations or gestures (which is effort-saving, but requires a pre-existing, shared understanding), while an unfamiliar partner is not. An poorly-chosen strategy can result in great expenditure of effort to signal and to repair any resulting misunderstanding — often more effort than would have been exerted had a more costly strategy been chosen (Baljko, 2000b).

So in addition to perceptual uncertainty, aided communicators also face articulatory uncertainty — that is, uncertainty about which modes of communication (aided or unaided) are most appropriate. The strategy for mode use must be chosen with care; a conservative choice can lead to unnecessary exertion, while a risky choice can lead to the breakdown of communication. Compounding the difficulty, the choice of strategy depends on communicator's uncertain perceptions of his own, possibly-shifting expressive capabilities and of the common ground accumulating with his communication partner.

Current Research

The focus of our current research is the production of *multimodal referring expressions* — that is, communicative acts that indicate reference, possibly using multiple modes. We have implemented computational simulations of dyads of agents who perform a task of communicating a reference by multimodal means. Human communicators performing this task under varying conditions demonstrate articulatory adaptation, so there exists empirically-derived baselines with which to evaluate the simulations (Baljko, 2001). In our simulations, we manipulate the independent variables (the parameters of the agents), while the agents repeatedly perform this task. Evidence of articulatory adaptation is then derived by a longitudinal analysis of the agent's behaviour, which serves as the experimental dependent variable. The longterm goals of this research is to derive accurate models of the way in which human interlocutors adapt the use of their modes of communication and to use such models to investigate the impact of alternative designs of communication aids.

In this task, one agent, which represents an individual with a communication disorder, selects one of a set of possible referents, and produces a multimodal expression to convey it to the other agent, which represents a non-disabled communicator. Both agents are instantiations of a single underlying agent model; they differ only in the values assigned to their parameters. These parameters represent the agent's familiarity with AAC devices and its capacity for expressive communicative behaviour — the degree to which eye gaze, the speech-sound articulators, and gesture are subjected to disorder, whether the communicator has the mode of synthesized speech available, and the conflicts among the defined modes of communication, such as the way the gestural mode and the aided mode of synthesized speech can both simultaneously demand the use of the neuromuscular articulators of the hand and arm.

The process of generating multimodal acts of communication has been implemented as solving a constraint-satisfaction problem (CSP) (Baljko, 2000a). In this formulation, each solution to the CSP is an event-based representation of a multimodal communicative act that serves to indicate the identity of a given referent (which is an input to the CSP). These acts vary with respect to their effectiveness, their mode usage, and their physical expense to the communicator. The constraints assert that modes of communication may not be used in a conflicting manner, that the partial ordering of the sub-constituents of the semantic representation of the intended referent may not be violated, and that the Gricean maxim of quantity must not be violated. Since this problem is under-constrained, we use a set of soft constraints in addition to these hard constraints. The soft constraints express the agent's preference for certain types of multimodal acts, given the characteristics of the communicative scenario (level of ambient noise), the status of the communicator's own articulators for communication, and the common ground shared with the communication partner. These factors were derived from the AAC research literature and our analyses of videotaped dyads (Baljko, 2001). Our preliminary evaluations of selected conditions show that the agent behaviour does emulate the adaptive strategies of aided communicators (Baljko, 2000a;2000b).

Future Work

In our current model, the agent's basis for communicative action is preference with respect to soft constraints. We are currently working on a decision-theoretic framework in which the expected utility of each potential communicative action is evaluated; the agent then selects the communicative action with this highest expected utility. We are also investigating mechanisms to incorporate the agent's uncertainty about its own perceptions (which, in turn, affect the utility scores) into the simulations.

The application of models of articulatory adaption to dialogue systems is theoretical at this stage. For instance, dialogue systems (or even intelligent multimedia presentation systems) operating over a network could be subject to constraints analogous to the expressive communication disorders that affect human communicators; adaptive strategies might be required in the face of such constraints. Also, in order to parse and to interpret the multimodal communicative acts of their human users — and even human communicators can become functionally disabled by fatigue or environmental factors (Newell et al., 1995), dialogue systems may need to maintain increasingly-sophisticated user models that capture the articulatory adaptation of their human interlocutors in response to the system's own actions.

Acknowledgements

This work is supported by the Natural Sciences and Engineering Research Council of Canada. For discussions that influenced the course of this work, I am grateful to Graeme Hirst and to Fraser Shein of the Bloorview MacMillan Rehabilitation Centre.

References

Melanie Baljko. 2000a. The computational simulation of multimodal, face-to-face communication constrained by physical disabilities. In *Proceedings of ESSLLI 2000 Workshop on Integrating Information from Different Channels in Multi-Media-Contexts*, pages 1–10, Birmingham, UK. European Association for Logic, Language, and Information.

Melanie Baljko. 2000b. Incorporating multimodality in the design of interventions for communication disorders. In Patric Dahlqvist, editor, *Proceedings of the 4th Swedish Symposium on Multimodal Communication (SSoMC)*. Stockholm University.

Melanie Baljko. 2001. The evaluation of microplanning and surface realization in the generation of multimodal acts of communication. In *Proceedings of the Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Fifth International Conference on Autonomous Agents*. Montreal, Canada, May.

Alan F. Newell, John L. Arnott, Alistair Y. Cairns, Ian W. Ricketts, and Peter Gregor. 1995. Intelligent systems for speech and language impaired people: A portfolio of research. In Alistair D. N. Edwards, editor, *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilities*, pages 83–101. Cambridge University Press.

Tim Paek and Eric Horvitz. 2000. Conversation as action under uncertainty. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, pages 455–464, Stanford, CA, June.