in Proceedings of the Workshop on Multimodal Communication and Context in Embodied Agents, pages 89-94, Montreal Canada, May 2001.

The Evaluation of Microplanning and Surface Realization in the Generation of Multimodal Acts of Communication

Melanie Baljko
Department of Computer Science, University of Toronto
10 King's College Road
Toronto, Canada, M5S 3G4
melanie@cs.utoronto.ca

Abstract

In this paper, we describe an application domain which requires the computational simulation of human-human communication in which one of the interlocutors has an expressive communication disorder. The importance and evaluation of a process, called here *microplanning and surface realization*, for such communicative agents is discussed and a related exploratory study is described.

1 Introduction

For a spoken act of communication, Speech Act Theory tells us that for each physical utterance, three acts are actually performed [1] — a locutionary act (the act of uttering a sequence of words, such as shouting or whispering), an illocutionary act (the act performed *in* saying, such as requesting, asking, telling, suggesting, or greeting), and a perlocutionary act (the act that actually results of the utterance, such as impressing, persuading, or embarrassing). Locutionary acts and illocutionary acts stand in a many-to-many relationship with one another; multiple possible illocutionary acts can correspond to a particular locutionary act, and a particular illocutionary act can be accomplished by multiple possible locutionary acts. If acts of communication performed using multiple modes of communication are considered,[1] then there are even more possible locutionary acts that could accomplish a particular illocutionary act. For example, to refer to an entity, a speaker might use, in isolation or in combination, the modes of speech (e.g., through the use of various deictic linguistic expressions), gesture (e.g., through the use of various types of gestures, deictic or otherwise), facial expression, gaze, torso and head movement and so on.

Such a distinction can be made for embodied conversational agents (ECAs), too. The illocutionary act that is to be performed is represented in the agent's architecture as a planned communicative action, in an intermediate, functionally-specified form. For example:
$\mathcal{S}$ INFORM $\mathcal{L}$ THAT $\mathcal{X}$ [22], or
SPEECH ACT TEMPLATE: Describe(object $\mathcal{Y}$, aspect $\mathcal{Z}$) [8]
But such acts might be accomplished by multiple possible locutionary acts. Alternatively, communicative plans may have multiple possible surface realizations (depending on the communicative articulators afforded by the agent's embodiment). Functionally-specified plans describe *what* the agent should communicate, but not necessarily *how* the act should be conveyed (the illocutionary component of the planned communicative action is given, but not which of the several possible locutionary acts should be performed).

This task of deriving a surface realization for a particular communicative plan can be called *microplanning and surface realization* (MP&SR). MP&SR in the generation of multimodal utterances for a communicative agent differs from MP&SR for natural language generation (NLG). The output of a MP&SR module in an agent architecture may yet be preliminary (e.g., it may be subsequently passed to some other module, such as for graphics rendering), while the output of MP&SR in NLG is readable text. Also, not all agent behaviours are the output of the agent's MP&SR module — if the architecture includes a reactive layer, then the production of some communicative behaviours bypasses the two-step process of plan generation and MP&SR (this strategy may avoid unnecessary computational processing). For NLG, however, MP&SR is the last step for all output. But for both types of MP&SR, *content determination* is an important preliminary task; for communicative agents, this task is subsumed by the agent architecture.

When humans perform multimodal acts of communication, the use of the spoken and non-spoken modes of communication is coordinated — the relative timing of the subconstituent, mode-specific actions of the overall multimodal communicative act can affect the act's overall interpretation. The MP&SR module of an ECA must too carefully coordinate the way in which the agent's modes are used — not necessarily in the interest of emulating human communicative behaviour accurately (in fact, this is not required nor desired for some application domains), but rather in the interest of avoiding the unintended or undesirable consequences of a poorly coordinated multimodal communicative act (or, conversely, in the interest of capitalizing on the benefits of a well-coordinated one). In order to use modes *synergistically*, an MP&SR module must be able to identify and capitalize on inter-mode interactions. Additionally, it must ensure consistency across the modes, as users may be particularly sensitive to this [20].

We would argue that the process of MP&SR can be found, at least implicitly, in those agent architectures that, for at least

[1] In this generalization of Speech Act Theory, acts of communication may be performed not only with the mode of speech, but also with other modes of communication. However, this generalization has some theoretical consequences [3]. In order for an illocutionary act to be performed successfully, the listener must recognize the *illocutionary intention* of the speaker. But for multimodal acts of communication, intent may not be so easily identified, even in principle. If, as some argue, certain components of multimodal communication are epiphenomenal, then it doesn't make sense to attribute intentionality to them. The result is a slippery slope in which the distinction between a communicative process and an information process becomes blurred.

some communicative behaviour, make use of an intermediate, functionally-specified communicative plan and that have sufficiently articulated embodiments. And although such architectures could potentially weigh and choose from among multiple possible surface realizations for a given communication plan, this has not been explored in detail [8, 10, 27]. In the subsequent section, we describe our particular application domain in which such a selection mechanism, in an explicitly-defined MP&SR module, is important.

2 Current Research

A goal of our current research is to model human-human communication which involves an interlocutor who has an expressive communication disorder (arising from, say, cerebral palsy or amyotrophic lateral sclerosis), has little or no functional speech or writing, and uses an *AAC device* — a clinical intervention that can provide an individual with the mode of *synthesized speech*. We focus on communication disorders that arise from physical disorder of an individual's *communicative articulators*, which include the speech-sound articulators (which underlie the modes of speech and vocalization) and other neuromuscular articulators (which underlie the modes of gesture, facial expression, gaze and so on). Communicative articulators are the physical means with which acts of communication can be performed and stand in many-to-many relationships with the modes of communication; the status of the communicative articulators determines which modes of communication are available and to what degree. (The term "communication mode" has several different usages in the literature; we use it here to refer to the physical manifestation of the communicative action, but also, to some degree, the abstract properties of the shared system of meaning in which that act is interpreted.)

The multimodal utterances produced by individuals whose articulators are constrained are not merely simplified, paler versions of those produced by individuals who do not have physical disabilities [4, 5]. Aided communicators are adaptive in performing multimodal utterances [6]. The aided mode of synthesized speech is used selectively. It is often the only linguistic mode available. Also, a communicator has a higher likelihood of being understood properly if this mode is used. Unfortunately, the aided mode is fatiguing to use and slow (possibly to the point that a breakdown in communication occurs). 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 in regulating turn-taking. So the use of the aided mode may preclude the use of other unaided modes to communicate. This interference arises because the use of the aided and unaided modes both depend upon a common set of communicative articulators (in the example above, the oculomotor articulators). Thus, the aided mode of synthesized speech is not a replacement for the existing unaided modes, but rather a sometimes-intrusive and sometimes-valuable addition to the individual's repertoire.

Current AAC devices still only partially mediate communication disorders and are incremental improvements over their precursors. Increased multimodality is seen as a possible improvement to AAC design [26], but this requires a better understanding of the interrelationships between the aided and unaided modes. The goals of our research are, first, to formalize these interrelationships and to model accurately the communication that is mediated by existing devices and then subsequently to model the effects of alternative, prototypical AAC devices, thereby providing a development and testing environment for the design of AAC systems. These simulations involve the use of communicative agents as models of human communicators, but these agents differ from those in other work. For example, the agent BALDI is used as an instructional tool for deaf children so that they may learn how to use their speech sound articulators in the absence of auditory feedback [19]. Performative Faces is a simulation tool in which the joint production of facial expressions and speech is investigated [22]. But in our application, the articulators that are afforded by the communicator's embodiment themselves are parameters of the MP&SR process. These parameters determine what types of surface realizations the MP&SR module can produce. By varying these parameter values, we simulate a wide variety of communicators — both those with and without physical disabilities. To constrain scope of possible behaviours, we focus on *multimodal referential communication*. Although the production of referring expressions is a more specific domain of communicative action than that of other agents, we felt is was an appropriate starting place, as referring expressions are both are fundamental in everyday conversation (and, by extension, to aided communicators) and are often highly multimodal [2].

3 Evaluation Methodologies for MP&SR

3.1 Motivation

The evaluation of MP&SR is important in agent evaluation, yet is still poorly defined. As described earlier, an MP&SR module is, in effect, responsible for the selection of one surface realization over another. Since the illocutionary act to be performed by the agent has been derived prior to the invocation of this module, only the act's surface realization is within the module's scope of influence. Thus, we could say that the module's effect is local. On the other hand, the selection of a surface realization is tantamount to the specification of agent behaviour and can affect the appearance of its communicative strategy. (This determination should, in principle, take into account a variety of global considerations, many of which would not be — and, given the principles of modularity, should not be — visible or available to an MP&SR module.) Thus, the module's effect is global. So the quality of an MP&SR module in an agent architecture can have far-reaching consequences — the locutionary component of the multimodal utterances performed by an agent are one of the primary sources of information upon which human users base their subjective judgments and assessments of the agent's personality, competence, and abilities. In addition, they are also one of the primary means upon which the human user relies in order to perform any required tasks.

The surface realizations produced by the MP&SR module of a NLG systems can be evaluated with respect to the standards of written language (e.g., grammaticality, lexical choice, etc), but no analogous standard exists for the surface realizations produced by MP&SR module of an agent architecture. Rather, the quality of the output depends upon a standard that is context-dependent. The surface realizations required by, say, an agent serving as a collaborator on a particular task (e.g., a real estate agent working with a buyer [8]) may differ from those required by a pedagogical agent [17], by an agent that serve as an instructor [19], or by an agent that acts as a research model [22]. Rather than posit some *a priori* standard, we can instead make use of the following four dimensions in determining which evaluation criteria are appropriate.

Ability to handle a variety of plans An MP&SR module should handle all of the communicative plans within its input domain, although the scope of the domain depends upon the application. For example, an MP&SR module within an architecture for a generalized ECA must be able to derive surface realizations for a wide variety of illocutionary acts, while for a modeling application, the module might specialize in derivations of surface realizations for a particular type of communicative function.

Resemblance to human behaviour An MP&SR module should model human behaviour, but only with respect to the properties or dimension that are relevant to the communicative context. For example, a pedagogical agent in some instances might caricature human behaviour but not in others (e.g., for savvy users, the perception of competence might be particularly important), but it still should follow the rules of turn-taking. Another way of describing this evaluation criterion is *model adequacy* — a *model* of human expressive communicative behaviour can be construed from the agent's architecture, in the sense that its implementation describes the human process by simulation, invoking similarities between a computational process and the behaviour under investigation. The *adequacy* of this model is determined by its ability to **account for** the range of behaviours within its scope. "Lacuna-based" evaluations [8] relate to this criterion of model adequacy. Note that model adequacy should be distinguished from theoretical adequacy — for model adequacy, a range of human communicative behaviours must be accounted for, but not necessarily by the same mechanisms that humans employ (and, thus, model adequacy can be thought of as I/O equivalence). Model and theoretical adequacy have been alternatively described as "phenomenological" and "process" [20].

Consistency across modes and consideration of inter-mode interactions An MP&SR module should take into account inter-mode interactions (to avoid the problems arising from poor coordination, or to capitalize from the benefits from subtle coordination), but the potential for such interactions depends upon the embodiment of the agent and the types of possible communicative action.

Adaptation An MP&SR module might need to adapt to changing conditions. An agent's ability to adapt its global communication strategy is a factor in the agent's usefulness [21], but adaptation with respect to the characteristics of the surface realizations that the agent's MP&SR module derives are also important [6]. Certain surface realizations are more beneficial than others (where benefit can be expressed by measures such as likelihood of being interpreted correctly, or positive impact on the communicator's perceptions of the agent), so an agent might need to adapt its surface realizations to suit the changing conditions of the interaction.

3.2 Related Research

For our application domain, it is particularly important that the surface realizations generated by our MP&SR module emulate those that would be produced by human communicators in analogous circumstances. Whether *model adequacy* served as the basis for the evaluation of other MP&SR modules was relevant to our research. This question, however, presupposes a certain conceptual framework for MP&SR. For instance, in our agent architecture, a MP&SR module is defined explicitly, whereas in other architectures, the functionality exists implicitly.

Two basic classes of ECA-related evaluations have been described: *inter-interface* evaluations — which assess the degree to which or the manner in which an interface that makes use of an embodied agent offers an advantage over other types of interfaces (such as menu- or phone-based interfaces, for example [10]); and *inter-ECA* evaluations — which assess the degree to which or the manner in which one ECA offers an advantage over another ECA (for example [9]). Whether an evaluation is carried out using a Wizard of Oz simulation or an actual implementation of the system is considered to be an orthogonal issue.

Inter-interface evaluations Interfaces based on ECAs differ from other types of interfaces with respect to many factors, only one of them being manner or quality of surface realization. The level at which these factors can be controlled for is fairly coarsely-grained, so it is problematic to attribute some particular evaluation outcome with the specific factor that relates to the performance of an ECA's MP&SR module. Our intuition tells us that a correlation should exist — MP&SR determines at least a subset of the agent's outward, expressive behaviour, which, in turn, can affect the measures that relate to the user's subjective experience arising from the use of one type of interface vs. another (e.g., such as *preference* [10] and assessments of *autonomy* [8]). Additionally, measures that relate to the user's behaviour or performance when using one interface vs. another would be influenced by the ECA's MP&SR module (such as *rate of error* [23], or *attention and retention* [20]).

Inter-ECA evaluations Inter-ECA comparisons vary with respect to fewer factors than inter-interface comparisons, so it is possible to manipulate a specific factor while holding fixed the remainder. A comparison could be made of, say, two versions of an ECA in which the underlying architectures are identical in all respects except the MP&SR modules. We would hypothesize that the effect of different MP&SR modules could be demonstrated in both subjective and behavioural measures (e.g., effects in the user's assessment of the ECA's personality, attractiveness, and competence could be elicited by manipulating the manner and style of the multimodal utterances that the MP&SR module produces). Such effects, however, might not be solely attributable to the differences in the MP&SR module. In repeated, uncontrolled interactions, a user might elicit different ECA behaviours. This could result in variations in the global communicative strategy undertaken by the agent. Thus, the

Multimodal Referential Communication Task	
1.	Two subjects, C (for "chooser") and L (for "listener") face one another, with a set of objects positioned on a table between them.
2.	C chooses an item and communicates the identity of the selected item to the other subject L. Any desired mode of communication, alone or in combination, can be used.
3.	L is asked by the experimenter which entity C chose. This should be done in a way to avoid confounding effects which might arise from L anticipation of C's reaction. The purpose of this step is to determine whether the task was performed successfully.

Table 1: An overview of the multimodal referential communication task.

types of communication plans that are generated as input to the MP&SR module would not be controlled for. Additionally, this experimental condition would not control the degree to which the reactive layer contributed to the agent's behaviour. This is another source of potential confounding, as users are likely sensitive to the degree of consistency between the utterances generated by the MP&SR module and the behaviour that is generated by the agent's reactive layer [20]. Therefore, in order to avoid the potential for confounding between these factors, the types of communicative plans generated as input to the MP&SR module must be also be controlled for.

3.3 Background to Exploratory Study

In previous work [4, 5], we isolated the MP&SR module of our communicative agent in the following way. First, we constructed a task for the agents to perform, a specific type of referential communication task, as shown in table 3.2. (We focus on the "Visible Situation Use" [13], in which the intended referent is visible to both of the interlocutors.)

This task was designed to be performed by both human subjects in face-to-face communication and computational agents in simulated face-to-face communication. Utterances that convey definite reference have been elicited from human subjects using similar tasks. In pioneering work on referential communication [14, 15, 16], a task was used in which one subject must get the other to arrange ten hard-to-describe figures in a particular order. In subsequent work [11, 12], a similar type of collaborative task was employed, although Tangram figures were instead used. While these tasks were useful for exploring the role of accumulating common ground in the production of utterances that convey definite reference, they also required that the subjects not be visible to one another; *only* the mode of speech was studied. Thus, the referential communication task was modified to elicit *multimodal* utterances. In our simulations, we used communicative agents to simulate the behaviour of the human subjects. The agent architecture was designed so that an agent in the role of C would know to select randomly an intended referent (say R) and to perform utterances that convey the semantic information required to identify R. An individual with a expressive communication disorder was simulated by agent C (parameters values representing properties of the agent's embodiment are used to simulate a variety of disorders). Any of the available modes of communication may be used, but not in a conflicting manner; as well, neither the partial ordering over the sub-constituents of the semantic representation of the intended referent, nor the Gricean maxim of quantity may be violated. This task is under-constrained, so the MP&SR module selects non-deterministically any valid utterance. This condition guarantees that the plan that is provided to the agent's MP&SR module is fixed. We implemented "behaviour observers" in the computational simulations in order to record the multimodal utterances generated by the MP&SR module of agent C.

We devised two conditions, A and B, in which agent C repeatedly performs the multimodal referential communication task for L. For both conditions, the parameters of the MP&SR module were considered the independent variables and the output from the MP&SR module was considered the dependent variable of the simulation, although in condition B, we manipulate the values of the independent variables. We are interested in two types of analysis — *intra-invocation analyses* are based on data derived from condition A. The values of the dependent variable are compared to an empirically-derived baseline, which reveals whether the surface realizations produced by the MP&SR module resemble the utterances produced by human subjects (because the agent behaviour is non-deterministic, a large sample of utterances is needed). For example, the agent should employ the aided mode if and only if the conditions correspond to the conditions in which a human communicator would. Also, the agent should exhibit the same degree of multimodality as human communicators under analogous conditions. *Inter-invocation analyses* are based on data derived from condition B. The longitudinal data are then compared to empirically-derived baselines. For example, the surface realizations produced by the MP&SR module should demonstrate the same types of adaptations as human communicators (e.g., aided communicators switch modes depending on the type of communicator partner, and exhibit varying degrees of multimodality). The empirical baselines used in these preliminary evaluations we derived from the empirical studies described in the AAC research literature (such as [7, 18]), but these are coarsely-grained. A more finely-grained specification is needed. In the subsequent sections, our current research to derive such baselines is described.

3.4 Exploratory Study

We have conducted an exploratory study in order to investigate the following two issues. First, the MP&SR module was implemented so that the format of its output (the multimodal utterances generated by the agent) adhered to a particular event-based representation in which the actions that are performed with respect to each of the modes of communication are represented, as well as the temporal interrelationships [4, 5]. Can this representation formalism serve as a coding scheme for real-world empirical data? Second, the perception and interpretation of multimodal utterances is subjective to some degree. Does this subjectivity have an adverse effect on inter-judge reliability? To what degree can the details of an utterance's surface realization be decomposed and distinguished from the utterance's overall interpretation or "gestalt"?

3.4.1 Materials

A set of videotaped dyads was used as the data to be coded. All dyads involved two subjects in a job interview scenario. The interviewer I and the questions she asked were the same across all dyads. Four different subjects, S_1, S_2, S_3, and S_4, participated in the role of applicant. Each performed the interview under two conditions — C_1, in which a communication board served as the AAC device, and C_2, in which a voice-output communication aid served as the AAC device.

The coding scheme was based on the output specification of our MP&SR module. Our focus was on coding the actions of the S_i's. A coding variable was defined for each of the subject's modes of communication. Shifts of eye gaze, gestures (deictic and other), and vocalizations were coded. The spoken utterances of I were coded (using [25]). We used the exploratory sequential data analysis (ESDA) software package MACSHAPA [24] to analyze and to code the video. It provided a facility in which either digital or analog video can be coded on a frame-by-frame basis and in which the coding variables could be appropriately customized (e.g., as opposed to ATLAS.ti© The Knowledge Workbench). It was also more affordable (c.f. the more-expensive package The Observer© Noldus Information Technology). In order to prepare the data, the videotapes were digitized. We experimented with various video compression algorithms until we found one (Sorensen) that balanced the competing requirements of relatively low disk space and adequate frame resolution (400×600) and rate (30 frames/sec). Unfortunately, some of the data for condition C_2 could not be digitized.

3.4.2 Judges

Two judges participated, the author of this paper and a research assistant. The judges first coded half of the materials and then met to discuss the problems and issues that arose while applying the coding scheme to the data. The coding scheme was refined, and the data were re-coded. After multiple iterations of coding the data and refining the scheme, the remaining data were coded.

3.4.3 Discussion

Coding initially took approximately two hours per dyad (which ranged from 4:30 to 5:31 in length), but required approximately half as much time with increased experience. For each of the dyads, there were eight conversational turns. The judges agreed with respect to the number and type of each multimodal event, but disagreed with respect to the start and finish of each action. A summary of the inter-judge disagreements is given in table 2. Onset and offset Δ values for each mode-specific action are given in milliseconds (msec) and were calculated by averaging the absolute difference in the timestamps assigned between judges J_1 and J_2. Much of the disagreement arose from the effects of muscle spasticity in the subjects. It was often difficult to determine the exact start of the preparation phase and the finish of the retraction phases for deictic gestures. Coding onset was particularly problematic for subject S_3 (inter-judge disagreement, 1040msec), and retraction for subjects S_2 and S_4 (inter-judge disagreement, 1952msec and 1004msec, respectively). We also coded shifts in eye gaze, which varied considerably by subject (counts varied from 12 to 37). Average inter-judge disagreement was 262msec and 306msec for onset and offset, respectively. In these preliminary materials, the subject's eye gaze was periodically obstructed by their eyeglass frame or the brow ridge, but this can be avoided in the future through an improved camera angle (for the level of detail required by our coding scheme, the use of an eye tracking device is probably not warranted.) None of the subjects used the mode of speech. There was acceptable agreement on the coded duration of I's speech in each of her conversational turns (average difference, 240msec and 189msec for onset and offset, respectively).

	Onset Δ (msec)				Offset Δ (msec)			
	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
				Eye Gaze				
C_1	467	118	93	368	189	381	285	675
C_2	-	341	-	-	-	329	-	-
				Gesture				
C_1	81	419	1040	196	726	1952	467	1004
C_2	-	292	-	-	-	558	-	-
				Speech				
C_1	106	117	243	496	131	90	83	454
C_2	-	111	-	-	-	37	-	-

Table 2: Average inter-judge differences in the coding of mode-specific events (by variable, subject, and condition).

This study (albeit exploratory, as the task wasn't controlled and some data were re-coded) has provided answers to our preliminary, yet crucial, questions of subjectivity and inter-judge agreement with respect to the coding of multimodal utterances made by aided communicators. Several modifications to the coding scheme were made on the basis of the coders' discussions — we found that in order to characterize the surface realization of a multimodal utterance, it is important to avoid descriptions that appeal to an interpretation of communicative intent. However, this can lead to a slippery slope — if the coder is not permitted to take intent into consideration, then it is difficult to exclude any behaviour from coding. Even with changes to the coding scheme, some effect remained from the factor of subjectivity on the perception and interpretation of certain multimodal action, although the impact of this factor seems to be correlated with the degree to which the intentionality of a communicator's behaviour is ambiguous. We are currently in the process of designing a subsequent data-gathering study so that an empirically-derived baseline that is more relevant to the production of multimodal referring expressions (the communicative behaviour produced by the participants S_i included non-verbal referring expressions, but none in the "Visible Situation Use").

4 Conclusion

In this paper, we described the application domain of simulating communicators with physical disabilities and discussed the particular importance of the process we have called microplanning and surface realization (MP&SR). We described our approach to the evaluation of MP&SR and have advocated controlling for confounding factors through the use of a specific eliciting task for communicative action. Through iterations of evaluation and model refinement, we hope to achieve a communicative agent that coordinates the use of its modes of communication in a way that is appropriate to the context, to the environment in which it is communicating, the interlocutor, and, especially, to the status of the communicative articulators afforded to it by its embodiment.

Acknowledgments This work is supported by the Natural Sciences and Engineering Research Council of Canada. For her careful work as a research assistant, I am thankful to Agnes Lisowska. 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. I am also grateful to Joe Laszlo and Petros Faloutsos for their assistance with the preparation of the digitized video and the use of the Digital Graphics Project laboratory equipment. The empirical data were kindly loaned to me by Professor Bernard O'Keefe of the Department of Speech-Language Pathology of the University of Toronto.

References

[1] J. L. Austin. *How to Do Things with Words*. Harvard University Press, 1962.

[2] M. Baljko. Object reference in augmented and assisted communication. In submission to *Human-Computer Interaction, Special Issue on Talking About Things: Mediated Conversations about Objects*.

[3] M. Baljko. What is a multimodal speech act? Manuscript in preparation.

[4] M. Baljko. The computational simulation of multimodal, face-to-face communication constrained by physical disabilities. In *Proceedings of ESSLLI 2000 Workshop Integrating Information from Different Channels in Multi-Media-Contexts*, pages 1–10, Birmingham, UK, August 6–10 2000. European Association for Logic, Language, and Information.

[5] M. Baljko. Incorporating multimodality in the design of interventions for communication disorders. In P. Dahlqvist, editor, *Proceedings of 4th SSoMC, the Fourth Swedish Symposium on Multimodal Communication*, pages 13–14. Stockholm University/KTH, October 26-27 2000.

[6] M. Baljko. Articulatory adaptation in multimodal communicative action. In C. Thompson, T. Paek, and E. Horvitz, editors, *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) Workshop on Adaptation in Dialogue Systems*. Pittsburgh, PA, June 2001.

[7] D. M. Blischak and L. L. Lloyd. Multimodal augmentative and alternative communication: Case study. *Augmentative and Alternative Communication*, 12(1):37–46, March 1996.

[8] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan. Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 29–63. MIT Press, Cambridge, MA, 2000.

[9] J. Cassell and K. Thórisson. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Journal of Applied Artificial Intelligence*, 12(3):519–538, 1999.

[10] J. Cassell and H. Vilhjálmsson. Fully embodied conversational avatars: Making communicative behaviours autonomous. *Autonomous Agents and Multi-Agent Systems*, 2:45–64, 1999.

[11] H. H. Clark. *Arenas of Language Use*. The University of Chicago Press and the Center for the Study of Language and Information, 1992.

[12] H. H. Clark. *Using Language*. Cambridge University Press, 1996.

[13] H. H. Clark and C. R. Marshall. Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, and I. Sag, editors, *Linguistic Structure and Discourse Setting*, pages 10–63. Cambridge University Press, Cambridge, 1981.

[14] R. M. Krauss and S. Weinheimer. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1:113–114, 1964.

[15] R. M. Krauss and S. Weinheimer. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 14:343–346, 1966.

[16] R. M. Krauss and S. Weinheimer. Effect of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior*, 6:359–363, 1967.

[17] J. C. Lester, S. G. Towns, C. B. Callaway, J. L. Voerman, and P. J. FitzGerald. Deictic and emotive communication in animated pedagogical agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 123–154. MIT Press, Cambridge, MA, 2000.

[18] J. Light, B. Collier, and P. Parnes. Communicative interaction between young nonspeaking physically disabled children and their primary caregivers: Part III – Modes of communication. *Augmentative and Alternative Communication*, 1(4):125–133, 1985.

[19] D. W. Massaro, M. M. Cohen, J. Beskow, and R. A. Cole. Developing and evaluating conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 287–318. MIT Press, Cambridge, MA, 2000.

[20] C. Nass, K. Isbister, and E.-J. Lee. Truth is beauty: Researching embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 374–402. MIT Press, Cambridge, MA, 2000.

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

[22] I. Poggi and C. Pelachaud. Performative faces. *Speech Communication*, 26(1–2):5–21, October 1998.

[23] G. A. Sanders and J. Scholtz. Measurement and evaluations of embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 346–373. MIT Press, Cambridge, MA, 2000.

[24] P. M. Sanderson, J. J. P. Scott, T. Johnston, J. Mainzer, L. M. Watanabe, and J. M. James. MacSHAPA and the enterprise of Exploratory Sequential Data Analysis (ESDA). *International Journal of Human-Computer Studies*, 41:633–668, 1994.

[25] D. Schiffrin. *Approaches to Discourse*. Blackwell, Oxford, UK, 1994.

[26] F. Shein, N. Brownlow, J. Treviranus, and P. Parnes. Climbing out of the rut: The future of interface technology. In B. Mineo, editor, *Proceedings of the Visions Conference: Augmentative and Alternative Communication in the Next Decade*, University of Delaware/Alfred I. duPont Institute, Wilmington, DE, 1990.

[27] K. R. Thórisson. *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*. PhD thesis, Massachusetts Institute of Technology, 1996.