



Jahresübersicht 2004

Veröffentlichungen des Jahres 2004 inklusive aller verfügbaren Abstracts

Becker, C., Kopp, S., Wachsmuth, I. Simulating the emotion dynamics of a multimodal conversational agent

In *Proceedings of Affective Dialogue Systems: Tutorial and research workshop (ADS 2004)*, Kloster Irsee, Germany (revised papers, LNCS 3068, pp. 154-165). Berlin Heidelberg: Springer, 2004.

Abstract:

We describe an implemented system for the simulation and visualisation of the emotional state of a multimodal conversational agent called Max. The focus of the presented work lies on modeling a coherent course of emotions over time. The basic idea of the underlying emotion system is the linkage of two interrelated psychological concepts: an emotion axis - representing short-time system states - and an orthogonal mood axis that stands for an undirected, longer lasting system state. A third axis was added to realize a dimension of boredom. To enhance the believability and lifelikeness of Max, the emotion system has been integrated in the agents' architecture. In result, Max's facial expression, gesture, speech, and secondary behaviors as well as his cognitive functions are modulated by the emotional system that, in turn, is affected by information arising at various levels within the agents' architecture.

Biermann, P. & Jung, B. Variant Design in Immersive Virtual Reality: A Markup Language for Scalable CSG Parts

In *Proceedings of Articulated Motion and Deformable Objects (AMDO-2004)*, Palma de Mallorca, Spain. LNCS 3179, Berlin Heidelberg: Springer, 2004, pp. 123-133.

Abstract:

In many product areas, a growing trend can be observed towards variant design, i.e. the development of customized designs based on variations of mature product models. We have developed a Virtual-Reality (VR) system for variant design that supports the real-time scaling and subsequent simulated assembly of hierarchical, CSG-like parts. An XML-based format, VPML, serves as description for the scalable CSG parts. VPML part descriptions determine how the scaling behavior of the whole part affects the scaling of its subparts, constrain the translation and rotation of subparts w.r.t. their parent parts, and define the scalable parts' dynamic mating properties. The part descriptions are utilized by several submodules of the overall VR system including: a) algorithms for real-time CSG visualization, b), the updating of part geometry using the ACIS CAD kernel, and c), the assembly simulation engine. The VR system runs in a CAVE-like large screen installation and enables interactive variant design using gesture and speech interactions.

Biermann, P. & Wachsmuth, I. Non-physical simulation of gears and modifiable connections in virtual reality

In *Proceedings Sixth Virtual Reality International Conference (VRIC 2004)*, Laval, France, 2004, pp. 159-164.

Abstract:

In this paper we present the functional (non-physical) modeling of gear couplings and adjustable building parts in a system for Virtual Assembly. In this system the user can multimodally interact in a CAVE-like setup using gesture and speech to instantiate, connect and modify building parts. The building parts, which are modeled in an XML description language, can have parametrically modifiable subparts, and ports as assembly points. The parameters of these parts can be linked in order to simulate hinges and the transmission ratio of gears. Special nodes in the scene graph so called Constraint Mediators are established to watch the port connections for propagation and adjustment of the motion of the connected parts. After the virtual assembly of such parts the user can interactively explore the functional effects of the simulation, e.g., the propagation of movements.

Keywords: Virtual Assembly, Virtual Prototyping, Mechanical Simulation, Multimodal Interaction

Kopp, S., Sowa, T. & Wachsmuth, I. Imitation Games with an Artificial Agent: From Mimicking to Understanding Shape-Related Iconic Gestures.

In Camurri, A., & Volpe, G. (eds.): "Gesture-Based Communication in Human-Computer Interaction",
International Gesture Workshop 2003, Genua, Italy.
Revised Papers, LNAI 2915, Springer, 2004, pp. 436-447.

Abstract:

We describe an anthropomorphic agent that is engaged in an imitation game with the human user. In imitating natural gestures demonstrated by the user, the agent brings together gesture recognition and synthesis on two levels of representation. On the mimicking level, the essential form features of the meaning-bearing gesture phase (stroke) are extracted and reproduced by the agent. Meaning-based imitation requires extracting the semantic content of such gestures and re-expressing it with possibly alternative gestural forms. Based on a compositional semantics for shape-related iconic gestures, we present first steps towards this higher-level gesture imitation in a restricted domain.

Kopp, S., Tepper, P., & Cassell, J. Towards integrated microplanning of language and iconic gesture for multimodal output

In *Proceedings of the International Conference on Multimodal Interfaces (ICMI'04)*,
Penn State University, PA (pp. 97-104). ACM Press, 2004.

Abstract:

When talking about spatial domains, humans frequently accompany their explanations with iconic gestures to depict what they are referring to. For example, when giving directions, it is common to see people making gestures that indicate the shape of buildings, or outline a route to be taken by the listener, and these gestures are essential to the understanding of the directions. Based on results from an ongoing study on language and gesture in direction-giving, we propose a framework to analyze such gestural images into semantic units (image description features), and to link these units to morphological features (hand shape, trajectory, etc.). This feature-based framework allows us to generate novel iconic gestures for embodied conversational agents, without drawing on a lexicon of canned gestures. We present an integrated microplanner that derives the form of both coordinated natural language and iconic gesture directly from given communicative goals, and serves as input to the speech and gesture realization engine in our NUMACK project.

Kopp, S., & Wachsmuth, I. Synthesizing multimodal utterances for conversational agents

In *Computer Animation and Virtual Worlds*, 15(1), 39-52

Abstract:

Conversational agents are supposed to combine speech with non-verbal modalities for intelligible multimodal utterances. In this paper, we focus on the generation of gesture and speech from XML-based descriptions of their overt form. An incremental production model is presented that combines the synthesis of synchronized gestural, verbal, and facial behaviors with mechanisms for linking them in fluent utterances with natural co-articulation and transition effects. In particular, an efficient kinematic approach for animating hand gestures from shape specifications is presented, which provides fine adaptation to temporal constraints that are imposed by cross-modal synchrony.

Kranstedt, A., Kühnlein, P., & Wachsmuth, I. Deixis in Multimodal Human Computer Interaction: An Interdisciplinary Approach

In Camurri, A., & Volpe, G. (eds.): "Gesture-Based Communication in Human-Computer Interaction",
International Gesture Workshop 2003, Genua, Italy.
Revised Papers, LNAI 2915, Springer, 2004, pp. 112-123.

Abstract:

This paper presents interdisciplinary work on the use of co-verbal gesture focusing on deixis in human computer interaction. Empirical investigations, theoretical modeling, and computational simulations with an anthropomorphic agent are based upon comparable settings and common representations. Findings pertain to the coordination of verbal and gestural constituents in deictic utterances. We discovered high variability in the temporal synchronization of such constituents in task-oriented dialogue, and a solution for the theoretical treatment thereof is presented. With respect to simulation it is exemplarily shown how the influence of situational characteristics on the choice of verbal and nonverbal constituents can be accounted for. In particular, this depends on spatio-temporal relations between speaker and the objects referred to.

Kranstedt, A., & Wachsmuth, I. Situated Generation of Multimodal Deixis in Task-Oriented Dialogue

In Belz, A., Evans, R., & Piwek, P.: *INLG04 Posters: Extended Abstracts of Posters Presented at the Third International Conference on Natural Language Generation*, Technical Report No. ITRI-04-01, University of Brighton, 2004.

Abstract:

This poster describes ongoing work concerning the generation of multimodal utterances, animated and visualized with the anthropomorphic agent Max. Max is a conversational agent that collaborates in cooperative construction tasks taking place in immersive virtual reality, realized in a three-side CAVElike installation. Max is able to produce synchronized output involving synthetic speech, facial display, and gesture from descriptions of their surface form [Kopp and Wachsmuth, 2004]. Focusing on deixis here it is shown how the influence of situational characteristics in face-to-face conversation can be accounted for in the automatic generation of such descriptions in multimodal dialogue.

Leßmann, N., Kranstedt, A., & Wachsmuth, I. Towards a Cognitively Motivated Processing of Turn-Taking Signals for the Embodied Conversational Agent Max

In *Proceedings of the Workshop Embodied Conversational Agents: Balanced Perception and Action* (pp. 57-64). Conducted at AAMAS '04, New York, July 2004.

Abstract:

Max is a human-size conversational agent that employs synthetic speech, gesture, gaze, and facial display to act in cooperative construction tasks taking place in immersive virtual reality. In the mixed-initiative dialogs involved in our research scenario, turn-taking abilities and dialog competences play a crucial role for Max to appear as a convincing multimodal communication partner. The way how they rely on Max's perception of the user and, in special, how turn-taking signals are handled in the agent's cognitive architecture is the focus of this paper.

Pfeiffer, T., & Latoschik, M.E. Resolving Object References in Multimodal Dialogues for Immersive Virtual Environments

In *Proceedings of the IEEE VR2004* Chicago, USA, March 2004.

Abstract:

This paper describes the underlying concepts and the technical implementation of a system for resolving multimodal references in Virtual Reality (VR). It has been developed in the context of speech and gesture driven communication for Virtual Environments where all sorts of temporal and semantic relations between referential utterances and the items in question have to be taken into account during the analysis of a user's multimodal input. The system is based on findings of human cognition research and handles the resolving task unifyingly as a constraint satisfaction problem, where the propositional value of each referential unit during a multimodal dialogue updates the active set of constraints to be satisfied. The system's implementation takes VR related real-time and immersive conditions into account and adapts its architecture in terms of the established system access, interface and integration into VR-based applications to well known scene-graph based design patterns by introducing a so-called reference resolution engine. Regarding the conceptual work as well as regarding the implementation, special care has been taken to allow further refinements and modifications to the underlying resolving processes on a high level basis.

Tepper, P., Kopp, S., & Cassell, J. Content in context: Generating language and iconic gesture without a gestionary

In *Proceedings of the Workshop Embodied Conversational Agents: Balanced Perception and Action*, (pp. 79-86). Conducted at AAMAS '04, New York, July 2004.

Abstract:

When expressing information about spatial domains, humans frequently accompany their speech with iconic gestures that depict spatial, imagistic features. For example, when giving directions, it is common to see people indicating the shape of buildings, and their spatial relationship to one another, as well as the outline of the route to be taken by the listener, and these gestures can be essential to understanding the directions. Based on results from an ongoing study on gesture and language during direction-giving, we propose a method for the generation of coordinated language and novel iconic gestures based on a common representation of context and domain knowledge. This method exploits a framework for linking imagistic semantic features to discrete morphological features (handshapes, trajectories, etc.) in gesture. The model we present is preliminary and currently under development. This paper summarizes our approach and poses new questions in light of this work.

A. Kranstedt, 2.02.2005