# Graduiertenkolleg Cross-modal Interaction in Natural and Artificial Cognitive Systems (CINACS)

**Vogt-Kölln-Str. 30, 22527 Hamburg; Tel.: (040) 428 83-2430, Fax: (040) 428 83-2397, http://www.cinacs.org/1. Zusammenfassende Darstellung**

**Angehörige des IGrK (der Universität Hamburg):**

*ProfessorInnen:*

Dr. Jianwei Zhang (Koordinator des Kollegs, TAMS); Dr. Christian Büchel (UKE); Dr. Andreas Engel (UKE), Dr. Carola Eschenbach (WSV); Dr. Christopher Habel (WSV); Dr.-Ing. Wolfgang Menzel (NATS); Dr. Brigitte Röder (FB Psychologie)

*KollegiatInnen:*

MS Cengiz Acartürk (WSV); Dipl.-Psych. Patrick Bruns (FB Psychologie); MS Tian Gan (NATS); Dipl.-Inform. Sascha Jockel (TAMS); MS Tobias Kringe (UKE), Dipl.-Psych. Mario Maiworm (FB Psychologie); Dipl.-Psych. Andreas Marschner (UKE); MLitt Patrick McCrae (NATS); Dipl.-Psych Klaus Nuißl (UKE); MS Inga Schepers (UKE); Dipl.-Inform. Martin Weser (TAMS)

**Laufzeit des Projektes**

Seit 4/2006; zweite Förderungsphase: 10/2010 bis 3/2015

**Allgemeiner Überblick**

The International Graduate College CINACS is carried out jointly by the departments of Informatics, the University Medical Center Hamburg-Eppendorf and the department of Psychology of the University of Hamburg, as well as by the University of Tsinghua in Beijing, China. The following groups of the department of Informatics are involved: the TAMS group (as coordinators) as well as the groups NATS and WSV.

The following, except in the general introduction to CINACS, refers to the CINACS-relevant activities of the members of the department of Informatics.

Within this international graduate college the principles of cross-modal interactions in natural and cognitive systems are investigated to implement them in artificial systems. Research primarily considers three sensory systems (vision, hearing and haptics) and their interactions. Multisensory interaction in natural systems is studied using behavioural, electrophysiological and neuroimaging techniques. Different paradigms including cross-modal association learning, sensorimotor control, cross-modal illusions and multisensory language perception is used to uncover the principles of multisensory processes and multimodal representation.

The research program aims at understanding the biological mechanisms of cross-modal processing and its role in perception and behavioural control. Furthermore, our goal is to design models, implement algorithms and architectures for more robust artificial multimodal systems which can function like natural systems.

The members of the department of Informatics are carrying out research in the subject areas „Development of grounded multimodal memory in robots“, „High-level information fusion for speech and language“ and „Interaction of representational modalities in communication“.

**Forschungsschwerpunkte**

Natural cognitive systems profit from combining the input of the different sensory systems not only because each modality provides information about different aspects of the world but also because the different senses can jointly encode particular aspects of events, e.g. the location or meaning of an event. However, the gains of cross-modal integration come at a cost: since each modality uses very specific representations, information needs to be transferred into a code that allows the different senses to interact. Corresponding problems arise in human communication when information about one topic is expressed using combinations of different formats such as written or spoken language and graphics.

The research program will aim at understanding the biological mechanisms of cross-modal processing and its role in perception and behavioural control. Furthermore, our goal is to design models, implement algorithms

and architectures for more robust artificial multimodal systems which can function like natural systems, i.e. even if their input is imprecise, ambiguous, incomplete, or incoherent. For this purpose, a system needs to form supplementary cross-connections between the sensory receptor level of a given type and the higher stages of processing specific to another sensory modality.

CINACS will combine the relevant methods, in particular behavioural techniques, EEG, fMRI, TMS, multi-electrode recordings, simulation, artefact construction, computer and robot experiments. This combination of approaches is only possible because CINACS comprises the disciplines of neuroscience, psychology, linguistics, computer science, robotics and bio-engineering. From this synergy we expect major advances in the fields of multisensory learning, attention, memory and sensorimotor control. Although cross-modal issues have attracted more and more interest in each individual field, interdisciplinary studies comprising informatics, neuroscience and psychology are still rare. Therefore, we expect a profound impact on future and emerging technologies, especially in human-computer interaction, human-robot communication, sensory substitution for rehabilitation, hybrid technology to restore sensory loss, hybrid bionic systems, a better understanding of information processing and functions in the human brain.

**Projekt Development of grounded multimodal memory in robots**

Weser, Martin, Dipl.-Inform; Jockel,Sascha, Dipl.-Inform; Zhang, Jianwei, Prof. Dr.

*Laufzeit:*

4/2006 bis 9/2010

*Projektbeschreibung:*

Multimodal representations play an important role for mobile robots and service robots. In this subproject we will use some common fusion methods, but will focus on building grounded memories of robot actions instead of solely a sensorimotor controller. The subproject will share modelling methods but will use real-world visual, audio and tactile data collected by the robot. Jointly utilizing the data of different modalities will enrich the robotic memory and increase the robustness of both representation and retrieval process.

According to empirical investigations, episodic memory represents one of the most important components of human intelligence. The act of remembering as well as mental simulation and planning use episodic memory as their basis. The diverse multisensory high-bandwidth data of our robot such as vision data, joint angles, positions, force profiles etc., can obviously not be saved in their raw format for an arbitrarily long period of time. Therefore, coding approaches based on appearances and features are suggested for summarizing and generalizing experiences from successfully performed operations. The multisensory trajectories and the motor signals need to be “grounded” in the learned operation sequences. It is extremely important to develop a memory mechanism for robots which allow using incomplete and distorted single sensor data but fuse them in a unified framework. We expect such a representation will increase the robustness of memory formation, the easiness of memory retrieval and the possibility of symbol-grounding.

*Schlagwörter:*

Robot intelligence, multimodal representation, sensori-motor skills, information retrieval

**Subprojekt 1: Multisensory Memory Representations of Robot Actions**

Weser, Martin, Dipl.-Inform.

*Laufzeit des Projektes:*

5/2006 bis 4/2008

*Projektbeschreibung:*

Service robots are still in a state of development and are only at the threshold to commercial use in the primary target group of private households. In order to successfully cross this threshold, the robots’ flexibility as well as safety in the natural surroundings of people has to be ensured. The safety of actions carried out by robots can only be ensured by massive interaction with their surroundings in the form of sensorial supervision and active examination. As different sensors cover different work areas and feature different characteristics, the use of various sensor modalities is a basic prerequisite for sensor-based actions.

In the last year the service robot control architecture was extended to a selective attention module. If a certain object is required, all available sensors are focussed on this target. Active movements of the robot are used to shift the sensors’ different fields of view to the goal object. This form of attentive examination includes the

physical approach to and the haptic palpation of objects. Currently a small set of predefined objects are available. Arrangements to include scalable learning algorithms are made by adopting a modular approach.

*Schlagwörter:*

Multimodal; sensor fusion; robot action; artificial cognition; action-oriented perception

**Subprojekt 2: Real-World Planning Based on the Retrieval of Episodic Memory**

Jockel, Sascha, Dipl.-Inform.

*Laufzeit des Projektes:*

7/2006 bis 6/2008

*Projektbeschreibung:*

Information-specific memory systems are differentiated according to their memory content into item-specific and relational information. Episodic memory is the memory of a certain episode fixed in time and space and relates to that episode's repercussions (e.g. the memory of your last holiday). Past episodes influence the current behaviour of people. The notion was already coined in the beginning of the 70ies, but to date the encoding in the learning phase has been receiving the main attention, while the recalling processes received practically none.

Episodic memory offers mechanisms to access past events in time and space – so-called experiences – and take these into account in the current planning (of actions). In spite of long-term research into the biological model of man by psychologists and neuropsychologists, the development of predictable models of these principles has been largely ignored to date.

The aim of this project is the research and development of an episodic memory to improve action planning and action prediction in the context of service robotics. A system design for our robot will make it possible to compare current actions with past episodes – e.g. successful and comparable prior operations – and to adapt and apply them to the current situation.

A Framework called EPIROME has been developed to collect high-level episodic memories which should be used to model and compare episodic memories of high-level events for technical systems. Now, research will focus on the process of episodic memory retrieval and prediction capabilities of event-sequences in detail.

*Schlagwörter:*

Artificial intelligence; memory; episodic memory; action planning; memory architecture

**Projekt High-level information fusion for speech and language**

McCrae, Patrick, MLitt; Gan, Tian, MS; Menzel, Wolfgang, Prof. Dr.-Ing.

*Laufzeit des Projektes:*

4/2006 bis 9/2010

*Projektbeschreibung:*

This project sets out to investigate cross-modal facilitating effects in technical systems for speech and language processing. Such a kind of system-integration differs from other problems of information fusion by the need to deal with highly abstract representations for which no straightforward mapping to the corresponding sensory input exists. Speech input, for instance, needs to be interpreted in terms of word or phrase meaning before it can be related to the visual experience of the hearer. Despite the enormous gap between the representations involved, the human cognitive system is able to make the complementary contributions from other modalities almost instantaneously available to the speech and language processing faculty, establishing the prerequisites for a synergy which contributes considerably to the rapid and robust behaviour of human multi-modal communication and sets it apart from any current technical solutions.

To gain deeper insights into the fundamental requirements for high-level information fusion, selected problem areas are studied in two PhD-projects, namely audio-visual speech recognition and the (dynamic) influence of context on sentence processing.

*Schlagwörter:*

Information fusion, multi-modal communication, speech recognition, sentence processing

**Subprojekt 1: Audio-visual speech recognition**

Gan, Tian, MS

*Laufzeit des Projektes:*

7/2006 bis 6/2008

*Projektbeschreibung:*

Audio-visual speech recognition makes use of extra video data, in particular lip-reading information, to improve the performance of a traditional acoustic-only speech recognizer. Based on state-of-the-art speech recognition technology, information fusion between acoustic and visual cues is attempted on the level of phones, where the visual stimulus is described by means of an underspecified phone representation. These two representations are then to be combined by a third component for word recognition. Moreover, compared to the conventional feature processing techniques, Articulatory Features can be used as an intermediate representation, capturing relevant characteristics of the speech production information. In a two-level AVSR system, we applied Hidden Markov Models (HMM) for modelling abstract articulatory classes, and then designed an N-best decision schema to decide the best articulatory feature tuples, in order to achieve a better recognition performance.

*Schlagwörter:*

Bimodal speech recognition, automated speech recognition, image processing, Hidden Markov Models (HMM), Articulatory Features

**Subprojekt 2: Cross-modal compensation in language processing**

McCrae, Patrick, MLitt

*Laufzeit des Projektes:*

8/2006 bis 7/2008

*Projektbeschreibung:*

During sentence processing cross-modal semantic influences often result in an (early) disambiguation, whenever there are several competing structural interpretations for an utterance, which are equally plausible with respect to lexical preferences and static world knowledge. Contextual information, as obtained from sensory input, e. g. via cross-modal perception, can then provide additional cues. While a wide range of studies have systematically investigated the impact of context and world knowledge upon structural disambiguation in human sentence processing, surprisingly few attempts set out to model the integration of dynamic cross-modal context in natural language processing applications. The projects will study approaches for integrating extra-sentential context modelled to stem from sensory modalities into the decision process on the most plausible structural interpretation. Context integration will be performed by using a parsing application based on weighted structural constraints in combination with a custom-developed context integration component.
Expected benefits of the successful integration of context knowledge into syntactic parsing include an increase in parsing accuracy and robustness, as well as an efficiency increase in the parsing process as such. Insights gained from this research can provide useful input to the improvement of applications in areas such as machine translation, natural language understanding, and automated information retrieval.

*Schlagwörter:*

Natural language processing, dependency parsing, context modelling, information fusion

**Projekt: Interaction of representational modalities in communication**

Habel, Christopher, Prof. Dr.; Eschenbach, Carola, Dr., Acartürk, Cengiz, MS; assoziiert: Mohammed Elmogy, M.Sc.Eng.

*Laufzeit des Projektes:*

4/2006 bis 9/2010

*Projektbeschreibung:*

People make use of distributed internal and external representations to perform higher level tasks in their daily life, e.g. they use paper and pencil to solve numerical problems, they construct and exploit graphs and diagrams to analyze the behaviour of physical, economical or social systems, or they make use of maps for route planning. Additionally, in solving problems cooperatively, people communicate, i.e. they use external

representations as language or pictorial means of communication. The subject of their communication is internal representations, i.e. their ideas and plans, as well as entities of the environment including external representations, like graphs, diagrams, tables or maps. To use external representations successfully it is necessary to integrate the information provided by different representational modalities, like language, diagrams, pictures, etc. In this project we focus on the use of diagrams and maps in communication and problem solving: For people seeing pictorial representations, language can be used to give additional information or to focus on specific aspects or parts of maps and diagrams; currently we investigate this aspect primarily in the interaction of language and information graphics. In contrast, for visually impaired people the auditory and haptic sensation can be a way to experience maps and diagrams by using only the visual and the haptic channel of sensation.

*Schlagwörter:*

Representation, multimodal; representation, graphical; haptic perception; information graphics

**Subprojekt: Multimodal comprehension of text and graphics**

Acartürk, Cengiz, MS;

*Laufzeit des Projektes:*

7/2006 bis 6/2008

*Projektbeschreibung:*

Information graphics, in the most frequently used sense of the term, are visual representations for the display of quantitative information. Contrary to numerical representations of data, for example tables, graphs display visual characteristics of data such as patterns, trends, comparisons, interactions. Nevertheless, visualizations are usually not self-explaining. Even simple pictorial illustrations need symbolic components or textual information to make the communicated message explicit. For this reason, graphs are generally presented together with accompanying linguistic constituents, thus constituting multimodal documents. Specifically, our focus in this project is on multimodal documents which include graphs presented on-screen (i.e. in the visual modality) either together with accompanying text on the same screen (i.e. in the visual modality) or together with narration (i.e. in the speech modality). Our purpose is to experimentally investigate the interaction between representational modalities, i.e. language and information graphics via the investigation of co-referenced relations between graph and text. This investigation is accompanied by several research questions: Concerning the analysis of co-referenced relations, what types of semantic patterns are used in verbal constituents either during the verbal description of graphs or during communication using graphs? How do the semantic patterns change when linguistic input is given under two different sensory modalities, namely the visual modality like text on-screen and the auditory modality of speech? How could available frameworks in linguistics be extended to include the comprehension of multimodal documents? The results of this study will be applicable in the development of NLG models for comprehension and generation of multimodal interfaces in Human Computer Interaction and Artificial Intelligence as well as design and implementation of computer-based tools to assist learners to interpret, create and manipulate graphs. In addition, the analysis of multimodal graph and text comprehension from the cognitive perspective has the potential to contribute to the development of haptic graph visualization, i.e. the haptic equivalent of visual graphs.

*Schlagwörter:*

Multimodal comprehension; graph comprehension; text comprehension; representational modalities.

**Wissenschaftliche Zusammenarbeit**

Tsinghua University, China

**Finanzmittel**

| | |
|---|---|
| Geldgeber: | DFG |
| Laufzeit der Förderung: | 4/2006 bis 9/2010 |
| Sachmittel: | € 726.671 |
| Personalmittel: | € 1.229.580 |

## 2. Publikationen und weitere Leistungen

**Wissenschaftliche Publikationen im Berichtszeitraum**

Siehe Einträge der Kollegsangehörigen in den Berichten der AB NATS, TAMS und WSV.

**Wichtige Publikationen aus den vergangenen Jahren 2003-2005**

Siehe Einträge der Kollegsangehörigen in den Berichten der AB NATS, TAMS und WSV.

**Wissenschaftliche Vorträge**

Siehe Einträge der Kollegsangehörigen in den Berichten der AB NATS, TAMS und WSV.

3. Wichtige weitere Aktivitäten

CINACS Summer School September 2007 in Beijing