

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

Vogt-Kölln-Str. 30, Haus F, 22527 Hamburg; Tel.: +49 40 428 83-2430, Fax: +49 40 428 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. Sabrina Boll (UKE); Dipl.-Psych. Patrick Bruns (FB Psychologie); MS Tian Gan (NATS); Dipl.-Ing. Christian Graf (WSV); Dipl.-Inform. Sascha Jockel (TAMS); Dr. med. Dipl.-Bioinf. Jens Kleesiek (UKE), Dipl.-Psych. Mario Maiworm (FB Psychologie); Dipl.-Psych. Andreas Marschner (UKE); MLitt Patrick McCrae (NATS); MS Inga Schepers (UKE); Dipl.-Inform. Martin Weser (TAMS)

Laufzeit des Projektes

Seit 4/2006; (zweite Förderungsphase wird 2009 beantragt: 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; Jockel, Sascha; 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 acts of remembering as well as of 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. We expect that a representation that fuses potentially incomplete and distorted single sensor data into a unified framework 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

Laufzeit des Projektes:

5/2006 bis 4/2009

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 if close sensor-control loops are applied. On the other hand, it is well accepted that abstract planning that foresees the effects of the intended course of actions is necessary to achieve flexible goal-directed behaviour. In this project, we develop a two-layered architecture that integrates robot control and AI-planning into one coherent system.

In the last year a set of atomic robot actions was defined and control programs for action execution were implemented. This set of actions forms a procedural memory of low level behaviour and provides basic building blocks for deliberative planning. An established AI-planner (Shop2) is used to parameterize the control programs and sequentially execute robot actions to achieve high-level objectives. Arrangements to include learning of both atomic robot actions as well as abstract strategies to guide the planner are made by adopting a modular approach.

Schlagwörter:

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

Subprojekt 2: Autobiographical Episodic Memory

Jockel, Sascha

Laufzeit des Projektes:

7/2006 bis 6/2009

Projektbeschreibung:

If the human mind stores each item - e.g. an abstract concept, attitudes, etc. - in a single memory location, the retrieval of a past (subjective) experience from memory triggered by current sensings needs to be identical to those sensings of the memorised experience. This kind of organisation would be useless, since the space of all possible experiences is so vast that no two experiences are ever exactly identical. The human mind would have to possess more memory locations than the number of particles in the universe to store all the permutations of sound, colour, and so forth that the senses are capable of detecting. The underlying principle of the brain involved in sensory information processing is that information is represented by a relatively small number of simultaneously active neurons out of a large population and is commonly referred to as sparse coding.

The aim of this project is the research and development of a memory that uses the sparse coding principle to improve action categorisation, prediction and learning in the context of service robotics and manipulation. A system design for our robot will make it possible to compare current sensings with past experienced robot actions – e.g. to successful and comparable prior operations – and to adapt and apply them to the current situation.

Schlagwörter:

Artificial intelligence; memory; sparse distributed memory; action planning; memory architecture

Projekt High-level information fusion for speech and language

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

Laufzeit des Projektes:

4/2006 bis 9/2010

Projektbeschreibung:

This project sets out to investigate cross-modal facilitation 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 of speech and language 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 difference in abstraction 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 faculties, establishing the prerequisites for a synergy which contributes considerably to the rapid and robust behaviour of human multi-modal communication thus setting 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***Laufzeit des Projektes:*

7/2006 bis 6/2009

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. So far, using articulatory information is investigated in two levels, feature level and class level. In the feature level, Articulatory Neural Networks (ANN) are used for performing audio and video articulatory feature classification. Then we use HMM to build the word recognizer. In the class level, we applied 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 decision fusion.

Furthermore, Dynamic Bayesian Networks (DBN) are expected to be applied for better modelling asynchrony among different articulatory channels. This will cause an information fusion in the model level.

Schlagwörter:

Speech recognition, bimodal; speech recognition, automated; image processing, Hidden Markov Models (HMM); Dynamic Bayesian Network (DBN); features, articulatory

Subprojekt 2: Cross-modal compensation in language processing

McCrae, Patrick

Laufzeit des Projektes:

8/2006 bis 7/2009

Projektbeschreibung:

During sentence processing cross-modal semantic influences often result in an (early) disambiguation, whenever there are several competing interpretations that correspond to individual syntactic structure and 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 for disambiguation. 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. This project will study approaches for integrating extra-sentential context from the sensory modalities into the decision processes to obtain 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. The purpose of the context integration component is to query a representation of visual context and to perform a set of inferences needed to propagate the contextual information back into the parser. Feedback from the context integration component is integrated into the parsing process via additional constraints.

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 such since structures corresponding to highly implausible interpretations can be ruled out at an earlier stage in the parsing process. Insights gained from this research can provide useful input to the improvement of applications in areas of machine translation, natural language understanding, and automated information retrieval.

Schlagwörter:

Language processing, natural; dependency parsing; context modelling; information fusion; language-vision interface

Projekt: Interaction of representational modalities in communication

Habel, Christopher, Prof. Dr.; Eschenbach, Carola, Dr., Acartürk, Cengiz; assoziiert: Mohammed Elmogy

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

Laufzeit des Projektes:

7/2006 bis 6/2009

Projektbeschreibung:

Information graphics are visual representations for the display of quantitative information. Contrary to numerical representations of data, graphs visually represent data characteristics such as patterns, trends, comparisons, and interactions. Since visualizations are usually not self-explaining they are generally presented together with accompanying linguistic constituents, thus constituting multimodal representations. Our purpose is to experimentally investigate interaction between linguistic entities and graphical entities in different types of multimodality concerning the interaction between text and graphics. Among the different types of multimodality, two are investigated in this project. The first is graph-internal multimodality, which includes the interaction between graphical entities (such as line of a line graph) and linguistic entities on the graph space (such as annotations and axis labels); the second type includes the interaction between graphical entities and linguistic entities in accompanying paragraphs. This investigation is accompanied by several research questions: Concerning the analysis of co-referenced relations, what types of systematic patterns are used in comprehension of constellations of graphical entities and linguistic entities? How do the 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?

All investigations are based on the analysis of cognitive processes concerning language-graphics interaction, specifically establishing co-reference and coherence in multimodal documents. Based on a proposal for cognitive architecture for integrated processing of multimodal documents, several case studies are performed. Concerning the graph-internal multimodality, linguistic annotations were investigated in two dimensions. The first dimension is the role of annotations and graph lines in line graphs for depicting change and events; the second is the attribution of causality induced by temporal relations that are represented by the graph line. Concerning the interaction between graphical entities and linguistic entities in accompanying paragraphs, the bridging role of annotations between paragraphs and the graph is investigated. A further analysis on graphical entities is performed with respect to the analysis of scale types that are used in psychophysics and scalar concepts in conceptual semantics. The interaction between linguistic descriptions and graphical representations of scalar relations is experimentally investigated.

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 assistance systems to assist learners to interpret, create and manipulate graphs and annotations. 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 and verbal assistance for the haptic exploration of graphs.

Schlagwörter:

Comprehension, multimodal; graph comprehension; text comprehension; modalities, representational

Subprojekt: Multimodal Representations in Communication: Interaction of representational modalities in communication

Graf, Christian

Laufzeit des Projektes:

1/2008 bis 12/2009

Projektbeschreibung:

In this dissertation, we investigate audio-tactile representations (tactile maps accompanied by verbal assistance) and how they can be orchestrated to provide survey knowledge to the human user. The use of multimodal representations to convey survey knowledge of an environment is assumed to outperform monomodal representations. An information model is hypothesized that incorporates information to automatically generate representations of an environment in a multimodal way, that is, by using tactile-spatial and propositional concepts. Tactile-spatial concepts are currently investigated in experiments to guarantee that spatial concepts have realisations in the tactile domain that convey spatial information. Work in the linguistic domain has already shown that language provides appropriate concepts to describe spatial information. The information model contains two qualitatively distinct parts: the vocabularies and the rules. For each representational modality and for each spatial concept, we have to store whether the concepts can be realized in that modality, and how the abstract description in the model can be transformed into concrete instances that are usable in a mode-dependent realisation. The approach to finding spatial concepts for each modality and rules for the transformation between modalities is to gain insight into how people realize spatial concepts in different modalities and how they do the transformation between them. The information model provides a basis to systematically build audio-tactile representations that convey survey knowledge. The results of the dissertation are expected to serve as one module in a workbench that provides the opportunity to navigate through a virtual tactile map accompanied by verbal instructions that assist the user in acquiring survey knowledge of the modelled environment to independently navigate through the world.

Schlagwörter

Communication, multimodal; maps, audio-tactile; modalities, representational; assistance, verbal;

Wissenschaftliche Zusammenarbeit

Tsinghua University, China

Finanzierung

Geldgeber:	DFG
Laufzeit der Förderung:	4/2006 bis 9/2010
Sachmittel (Anteil Informatik):	€519.175
Personalmittel (Anteil Informatik):	€614.790

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 2008 in Hamburg