MUMIN

A Nordic Network for MUltiModal INterfaces

Patrizia Paggio*, Kristiina Jokinen and Arne Jönsson*****

* Center for Sprogteknologi, Copehagen
email: patrizia@cst.dk
** University of Art and Design, Helsinki
email: Kristiina.Jokinen@uiah.fi
*** University of Linköping
email: arnjo@ida.liu.se

Abstract

This paper reports on a recent initiative undertaken under the language technology programme of the Nordic Research Training Academy (NORFA) to create a network of Nordic research institutes working with multimodal interfaces. In the paper we present the objectives of the network and give an overview of multimodal research and resources in the Nordic countries.

Keywords: multimodal integration, cognitive and usability studies, multimodal dialogue, multimodal research and resources in the Nordic countries

1. Motivation and objectives

In the Nordic countries as elsewhere, both the research and the industrial communities are showing a growing interest in multimodal interfaces. Therefore, there is a need to channel the efforts of individual organisations and countries in joint activities to establish a common research agenda and to define relevant standards and generic methodologies. The aim of the Nordic Network for MultiModal Interfaces (MUMIN) under the NORFA programme, is to stimulate Nordic research in this area and increase its visibility in the international research community. MUMIN, which was established in January 2002 and has funds for two years of activity, has the following goals:

- encouraging joint activities in building generic models and architectures as well as defining standards for the integration and development of multimodal interaction;
- encouraging investigations on the use of multimodality in various practical applications;
- providing a forum for sharing resources and results, and by encouraging network participants to make their research results available via the network's web site;
- organising PhD courses and research workshops on issues related to multimodal interaction;
- creating a network of contacts and a pool of shared knowledge that can be taken advantage of for the definition of collaborative research projects and for product development.

It will also be an important objective of the network to support investigation of the use of multimodal interaction in non-expert environments, and the accessibility of disabled people to IT technology. The network will thus contribute to the Nordic countries' social objectives and help them advance in their vision of building democratic and equal societies for everyone. This also conforms to the EU objectives of creating a user-friendly information society, with accessibility of IT benefits and services for all citizens.

A whole range of research issues, some of which have already received attention from the institutes engaged in the network, are relevant to this overall objective, and constitute topics of interest around which the network's activities will be organised.

2. Multimodal integration

A central issue is that of multimodal integration, which will be approached from different perspectives. A promising approach put forward by several researchers is that of using techniques known from NLP. A similar distinction to that made in NLP between grammar rules and parsing algorithms can be made between a multimodal grammar and an algorithm for applying the grammar to input from multiple modalities. By upholding this separation of process and data, the process of merging inputs from different modalities can be made more general, as the representation becomes media-independent. Furthermore, defining algorithms for modality integration independent of the specific

modalities used in a particular application, increases the chances that components of the system can be extended and reused. For example in the Danish research project Staging, Center for Sprogteknologi (CST) has developed a multimodal dialogue interface to a virtual environment (Paggio *et al*. 2000) where speech, keyboard and gestural inputs are merged by a feature-based parser. Relevant to this issue is also the work carried out by the Speech and Multimedia Communication (SMC) group from Center for PersonKommunikation, Aalborg, which also has extensive research and teaching experience in the area of multimodality, complemented with expertise in speech and image processing (Larsen & Brøndsted 2001).

Another promising approach to modality integration represented in MUMIN is the use of different machine learning techniques, in particular neural networks. As has been the case with many other application domains, also for multimodal integration hybrid systems mixing rule-based approaches with machine learning algorithms, may well provide the most interesting results. Although rule-based methods in general work reasonably well, it is a well-known problem that an explicit specification of the steps, i.e. rules that are required to control the processing of the input, is a difficult task, and when the domain becomes more complex, the rules become more complex too. Often the correlation between input and output is difficult to specify. This is the case e.g. with multimodal interfaces, and thus approaches which are both robust and able to adapt to new inputs are needed. Expertise in this domain is brought to the network by the Media Lab at the University of Art and Design in Helsinki (UIAH), and especially its Soft-Computing Interfaces Group which is devoted to designing adaptive interfaces and developing tools for human-machine interaction, relying on nature-like emergent knowledge that arises from subsymbolic, unsupervised processes of self-organizing nature (see e.g. Jokinen *et al*. 2001).

In a similar way as the rule-based integration of modalities can be enhanced using machine learning techniques, results obtained through pure probabilistic analysis methods may well be boosted by the addition of symbolic rules. An example relevant to multimodal interfaces are the algorithms for character and word prediction used in connection with eye-tracking, where the system tries to guess what the user is "typing with the eye". Although the performance of the probabilistic approaches implemented in current systems is promising, language technology techniques seem to constitute a valuable add-on. This is an issue that the IT University of Copenhagen is working on.

3. Neurocognitive basis and usability studies

To fully exploit multimodality in various interfaces, it is important to know how the neurocognitive mechanisms support multimodal and multisensory integration. In comparison to that devoted to single sensory systems, there has been very little research on the integration mechanisms of information received via different senses. However, the research group of Cognitive Science and Technology at the Helsinki University of Technology is using various methods to uncover the neurocognitive principles of multisensory integration with the purpose of developing mathematical models of this integration. The group is also developing a Finnish artificial person – a talking and gesturing audiovisual head model – which will serve as a well-controlled stimulus for neurocognitive studies (Sams *et al.* 1998).

A related issue is that of the impact of multimodality on the users. Relevant questions are which input and output modalities should be used for which task, which are the best combinations, and how different modalities are used by or for users with different degrees of expertise. There are in general two ways to use multimodal input: to react directly to the user's intentional input and to observe the user's unconscious use of certain modalities (e.g. eye-gaze). The former method is based on direct control and has been used in earlier conversational interface prototypes. Observing the user and understanding their intentions and mental states has not been extensively studied, and would add valuable information to the design of multimodal systems. The Computer Human Interaction group at the University of Tampere (TAUCHI) will bring to the network its extensive expertise in the design and use of innovative user interfaces and in usability testing, as well as the agent-based development platform Jaspis (Turunen and Hakulinen 2000). The Natural Interactive Systems Laboratory (NISLab) at the University of Southern Denmark, has also made pioneering contributions to the theoretical understanding of unimodal input/output modalities, and of the multiple conditions which determine the usability of individual modalities and their combinations (Bernsen 2001).

4. Multimodality and dialogue

A third issue, which encompasses a great deal of research work carried out by several of the network members, is that of multimodal dialogue. In this respect, it is interesting to note that the growing interest in multimodal interaction is opening a new perspective to Nordic research on dialogues, which is already acknowledged internationally. Several institutes in the Nordic countries have in fact contributed substantially to dialogue research, and developed dialogue models as well as implemented dialogue systems.

The Department of Linguistics at the University of Göteborg has extensive experience in corpus collection and dialogue management. They have developed tools for spoken language analysis and coding which can be applied to the collection and analysis of multimodal dialogues, thus providing empirical basis and insight for research on multimodal interaction: how different modalities are used in human-human communication (Allwood, 2001). NISLab has a strong background in dialogue management, dialogue systems evaluation, and spoken dialogue corpus coding from a number of EU projects. NISLab is currently addressing best practice in the development and evaluation of natural interactivity systems and components; surveying data resources, coding schemes and coding tools for natural interactivity;

and building a general-purpose coding tool for natural interactive communicative behaviour. The natural language processing research group (NLPLab) at the University of Linköping has for almost two decades conducted research on dialogue systems and now has a platform for the development of multimodal dialogue systems for various applications to be developed further towards an open source code repository (Degerstedt & Jönsson, 2001). Current focus is on integrating dialogue systems with intelligent document processing techniques in order to develop multimodal dialogue systems that can retrieve information from unstructured documents, where the request requires that the user, in a dialogue with the system, specifies their information needs (Merkel & Jönsson, 2001). Finally, the Centre for Speech Technology at the Royal Institute of Technology in Stockholm (KTH) has developed several multimodal dialogue systems with the motivation of studying speech technology as part of complete systems and the interaction between the different modules that are included in such systems. Their first system, Waxholm, was a multimodal system exploring an animated agent (Carlsson & Granström, 1996) **.** Current work and interests involve research on multimodal output using animations and also to some extent multimodal input using both speech and pointing (Gustafson *et al.* 2000).

5. Participating groups

The groups participating in MUMIN are shown in Table 1 below. Currently, three Nordic countries are represented, but the network is interested in welcoming participants from other Nordic countries, as well as in cooperation with non-Nordic countries.

Denmark	Finland	Sweden
Center for Sprogteknologi, Copenhagen	University of Art and Design Helsinki , Media Lab, Helsinki	Linköpings Universitet, Institutionen för datavetenskap, Linköping
Syddansk Universitet , NISLab , Odense	University of Tampere, Department of Computer and Information Sciences, Tampere Unit for Computer-Human Interaction (TAUCHI), Tammerfors	KTH, CTT, Centre for Speech Technology, Stockholm
Aalborg Universitet , Center for PersonKommunikation, Aalborg	Helsinki University of Technology , Laboratory of Computational Engineering, HUT Espoo	Göteborgs Universitet, Institutionen för lingvistik, Göteborg
IT-Højskolen, Eye Gaze Research Team, Copenhagen		

Table 1: The groups participating in MUMIN

6. Conclusion

The MUMIN network is expected to play an important strategic role in the establishment of a common research agenda for Nordic researchers working with multimodal interfaces, but also to relate its activities to those of the international community, and to contribute to the general progress in the area. Therefore, it is highly relevant for MUMIN to participate in this workshop, and to provide a Nordic contribution to the discussion of a multimodal roadmap.

7. References

Allwood, J. (2001) Dialog Coding - Function and Grammar, Göteborg Coding Schemas, *Gothenburg Papers in Theoretical Linguistics, GPTL 85*. Göteborg University, Department of Linguistics.

Bernsen, N. O. (2001) Multimodality in language and speech systems - from theory to design support tool. Chapter to appear in Granström, B. (Ed.): *Multimodality in Language and Speech Systems.* Dordrecht: Kluwer Academic Publishers.

Carlson, R., Granström, B. (1996) The WAXHOLM spoken dialogue system. *Acta universitatis Carolinae philologica* 1, pp. 39-52.

Degerstedt, L. and Jönsson, A. (2001) A Method for Iterative Implementation of Dialogue Management , *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Seattle.

Gustafson, J, Bell, L, Beskow, J, Boye, J, Carlson, R, Edlund, J, Granström, B, House, D & Wirén M (2000) AdApt - a multimodal conversational dialogue system

in an apartment domain, In *Proc of ICSLP 2000*, Beijing, 2:134-137

Jokinen, K., Hurtig, T., Hynnä, K., Kanto, K., Kaipainen, M., and Kerminen, A. (2001) Dialogue Act classification and self-organising maps. In *Proceedings of the Neural Networks and Natural Language Processing Workshop* , Tokyo, Japan.

Larsen , L.B. and Brøndsted, T. (2001) A Multi Modal Pool Trainer. In *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue IPNMD-2001*. Verona, Italy 14-15 December 2001, pp. 107-111.

Merkel, M. and Jönsson, A. (2001) Towards multimodal public informations systems, *Proceedings of 13th Nordic Conference on Computational Linguistics, NoDaLiDa '01*, Uppsala, Sweden.

Paggio P., Jongejan B. and Madsen C.B. (2000) Unification-based multimodal analysis in a 3D virtual world: the Staging project. In *Proceedings of the CELE-Twente Workshop on Language Technology: Interacting Agents*, pp. 71-82.

Sams, M., Manninen, P., Surakka, V., Helin, P. and Kättö, R. (1998) Effects of word meaning and sentence context on the integration of audiovisual speech. *Speech Communication*, 26, 75-87.

Turunen, M. and Hakulinen J. (2000) JASPIS - a Framework for Multilingual Adaptive Speech Applications. In P*roceedings of the 6th International Conference of Spoken Language Processing. ICSPL 2000.*