Spoken dialogue systems for IntelliMedia 2000+[*]

Paul Mc Kevitt and Paul Dalsgaard

Center for PersonKommunikation (CPK)
Fredrik Bajers Vej 7-A2
Institute of Electronic Systems (IES)
Aalborg University
DK- 9220, Aalborg, DENMARK.
E-mail: pmck@cpk.auc.dk

Abstract

The area of Intelligent MultiMedia (*IntelliMedia*) involves the real-time computer processing and understanding of perceptual input from speech, textual and visual sources and contrasts with the traditional display of text, voice, sound and video/graphics with possibly touch and virtual reality linked in. This is the newest area of MultiMedia research which has seen an upsurge over the last two years and one where many universities internationally do not have, or have not integrated, such expertise.

The Institute of Electronic Systems (IES) at Aalborg University, Denmark has initiated a programme in IntelliMedia under the Multi-modal and Multi-media User Interfaces (MMUI) initiative (see *WWW: http://www.cpk.auc.dk/CPK/MMUI/* for further details). The initiative includes implementation of educational MMUI (M.Sc.Eng., Ph.D.), production of real-time MMUI demonstrators, and establishment of a strong technology-based group of MMUI experts. The Institute has a strong tradition within research activities which fit into the real-time processing of IntelliMedia systems. Such initiatives will ensure the position of Denmark and the European Union (EU) in the construction of the future of SuperinformationhighwayS.

1 Introduction

The area of MultiMedia is growing rapidly internationally and it is clear that it has various meanings from various points of view. MultiMedia can be separated into at least two areas: (1) (traditional) MultiMedia and (2) Intelligent MultiMedia (*IntelliMedia*). The former area is the one that people traditionally think of as being MultiMedia, encompassing the display of text, voice,

---

sound and video/graphics with possibly touch and virtual reality linked in. However, the computer has little or no understanding of the meaning of what it is presenting. This area can typically be found in Computer Science and Arts Departments and Aalborg has strengths in both Computer Science and Humanities (Communication, Humanistic Computing) which focus on this.

IntelliMedia, which involves the computer processing and understanding of perceptual input from speech, text and visual images and reacting to it is much more complex and involves research from Engineering, Computer Science and Cognitive Science. This is the newest area of MultiMedia research which has seen an upsurge over the last two years and one where most universities internationally do not have expertise. The Institute of Electronic Systems at Aalborg University has expertise in this area.

Aalborg has already initiated a Multi-modal and Multi-media User Interfaces (MMUI) initiative. The initiative involves the implementation of educational MMUI, the production of a number of real-time MMUI demonstrators, and the establishment of a strong technology-based group of MMUI experts.

Hence, Aalborg's strengths in traditional MultiMedia (Humanities and Computer Science) and in IntelliMedia (Electronic Systems) establish Denmark's position in international MultiMedia development.

2 Visiting Professor

The Center for PersonKommunikation (CPK) acts as a host for Dr. Paul Mc Kevitt who is a Visiting Professor in IntelliMedia and Language and Vision integration and started here in February, 1996. Dr. Mc Kevitt has edited four books (see Mc Kevitt 1995a, 1995b, 1996a, 1996b) comprising work in many aspects of IntelliMedia and Language and Vision processing. He has three papers in those on his work to solve the philosophical Symbol Grounding and the Chinese Room problems, the use of MultiMedia in intelligent interfaces and the use of language and vision processing for the analysis of medical data in the form of language and image input. Dr. Mc Kevitt also has a strong background in the development of natural language dialogue systems where he has developed a system called OSCON (Operating System COnsultant) which will answer English questions about computer operating systems (UNIX, MS-DOS). This work has focussed very much on context and pragmatics which is important for language and vision integration. Dr. Mc Kevitt is also a British EPSRC Advanced Fellow for five years in the area of speech, language, and vision integrated processing.

3 Institute of Electronic Systems (IES)

The Institute has four research groups which focus on spoken language processing (CPK), image/vision (LIA), medical informatics (MI), and computer science (CS). The first two groups focus on the two main components of IntelliMedia processing while the other two provide a strong basis with demonstrator

projects in medicine and platforms and tools for integration from computer science.

3.1   Center for PersonKommunikation (CPK)

Research at the CPK is focused within the following three areas: Spoken Language Dialogue Systems, Data Communications and Radio Communications. CPK is an engineering research center funded by the Danish Technical Research Council.

The research within Spoken Language Dialogue Systems has for a long time been focused on human-computer interfacing and interaction and to a large extent been developed in connection with ESPRIT and nationally funded projects. The results obtained so far are of high relevance to many foreseen practical MultiMedia applications and to Framework IV of the European Union (EU), and they may advantageously be utilised as partial basis for all activities of the MMUI initiative.

CPK has an already developed a Dialogue Specification, Design and Management tool called Generic Dialogue System (GDS) (Dalsgaard and Baekgaard 1994, Baekgaard 1996) which is an appropriate platform for IntelliMedia research, and which from the very beginning may be used in various specialisations and student projects.

The research so far has been focused on the engineering design and development of IntelliMedia for speech and language in the context of professional use. The research is now ready to be further extended into the subsequent research paradigm which is based on the use of a number of available user interface components such as pen-based character recognition, optical character recognition, bar code readers, speech recognition, images and text and by combining these into an integrated MultiMedia interface (e.g. report generation, Personal Data Assistants (PDAs)).

A basic position taken in this research is that the separate interfacing technologies have already reached a stage of development where it will be possible to use them, with specific and identifiable extension of capabilities, to create an integrated multi-modal user interface featuring a spoken human-computer dialogue. It is expected that such dialogue engineering research will form the basis for many future computer systems.

3.2   Laboratory of Image Analysis (LIA)

The research at LIA is directed towards three areas: systems for computer vision, computer vision for autonomous robots, and medical and industrial application of image analysis.

Research within all three areas is sponsored by national and international (EU ESPRIT) research programmes. The main emphasis has been development of methods for continual interpretation of dynamically changing scenes. Example applications include surveillance of in-door and out-door scenes, vision-guided navigation, and interpretation of human and machine manipulation.

Research projects concern extraction of features for description of actions in an environment (i.e. the movement of people, fish, and blood cells) and utilising these descriptions for recognition, monitoring and control of actuators such as mobile robots (safe movements in a dynamically changing environment). This includes recognising and tracking dynamically changing objects, such as hands and human bodies, which has applications in IntelliMedia systems.

So far the research has referred to sensory processing using single modalities, but it seems obvious that the available methods may be integrated into multi-modal system, where a major objective is coordination and optimal use of available modalities. New IntelliMedia systems may also include much more flexible modes of interaction between computers, including both speech, body movements, gestures, facial expressions and sign language.

3.3 Medical Informatics (MI)

The research in the Medical Decision Support System group is centered around medical knowledge-based systems and the development of general tools to support complex decision making.

The research is building on a theory for representing causal dependencies by graphs (Bayesian networks), and uses these to propagate probability estimates. The group has developed several successful medical decision support systems, including sophisticated human-computer interaction issues. A central part of the theoretical development of this paradigm, seen in a global perspective, has taken place at Aalborg University, mainly within the research programme ODIN (Operation and Decision support through Intensional Networks) (a Danish PIFT (Professionel Informatik i Forskning og Teknologi) framework project).

The knowledge-based system technology based on Bayesian networks allowing for a proper handling of uncertain information has shown itself to be usable in creating intelligent coupling between interface components and the underlying knowledge structure. This technology may be integrated in IntelliMedia systems. The Bayes network paradigm, as developed in Aalborg, is already in practical use in user interfaces such as in Intelligence, a user and environment context sensitive help system in the major word processing and spreadsheet products from Microsoft.

It is foreseen that IntelliMedia systems will play a central role in the dissemination of information technology in the medical informatics sector. Systems representing complex knowledge, models and data structures e.g. advanced medical diagnostics system, virtual operation room, the telemedical praxis and so on, will require use of knowledge-based techniques for efficient interfacing.

3.4 Computer Science (CS)

The research at Computer Science includes computer systems and the design/implementation of object-oriented programming languages and environments. The scientific approach covers the formally logical, the experimentally constructive, as well as the empirically descriptive.

Of particular interest for MultiMedia are the following subjects: principles for hypermedia construction, theories of synchronisation and cognition, distributed systems and networking, high volume databases, and the design and use of language mechanisms based on conceptual modelling. Furthermore, the CS has a strong research tradition within the interplay between humans, organisations and information systems, and also within the subject of decision support systems and communicating agents, which is highly relevant for emerging research on models for user/system interaction.

CS contributions include experiments for performance evaluation of the available technology (e.g. high speed networking) and experiments on the methodology for design of MultiMedia systems. These contributions will be based on existing research activities, which includes networks, distributed models (Topsy), and prototype hypermedia environments.

In the long term perspective, the CS will contribute with models for intelligent human-computer interfaces and fundamental understanding of languages/dialogues, graphic elements, etc. based on conceptual understanding, and with implementations of these models. Such models are indispensable for the construction of efficient MultiMedia systems. Also, contributions will be made on efficient techniques for storing of high-volume MultiMedia data. Cases will include remote interactive MultiMedia teaching based on existing remote teaching.

4   Demonstrator CHAMELEON

The MMUI initiative Working Group has had a number of meetings and has come to a conclusion on the development of a single demonstrator.

The demonstrator platform will have a general architecture of communicating agent modules processing inputs and outputs from different modalities and each of which can be tailored to a number of application domains. The demonstrator will be developed in both a top-down and bottom-up manner making sure it is general enough for multiple application domains but at the same time keeping particular domains in mind.

It has been decided that the demonstrator will be a single MMUI platform (called CHAMELEON) demonstrating that existing platforms for (1) image processing, (2) spoken dialogue processing, (3) expert systems and (4) microphone arrays can be interfaced to the single hub platform and act as communicating agent modules within it. The platform will be independent of any particular application domain and the intention is that it will be possible to run it over different server machines. We will consider programming languages such as C++, C, and maybe Java for implementation. The Generic Dialogue System (GDS) of CPK implemented in C and C++ is being considered as an initial platform tool.

The hub platform will demonstrate that (1) it is possible for agent modules to receive inputs particularly in the form of images, spoken language, expert system queries/decisions and present outputs particularly in the form of speech, images, and graphics; (2) individual agent modules within the platform can

produce output in the form of semantic representations to show their internal workings; (3) the semantic representations can be used for effective communication of information between different agents for various applications; and (4) various means of synchronising the communication between agents can be tested to produce optimal results. The platform will be applied in the domain of IntelliMedia VideoConferencing involving the spoken control of cameras, interpretation of gestures and speaker identification and localisation.

5 Education

We have established a new International Masters degree (M.Sc.Eng) education in Intelligent MultiMedia which is 1 and 1/2 years in duration over 3 semesters. Each semester has a theme associated with it and involves both project work and courses. Semester I focusses on Basic methods, Semester II on Advanced methods and III on a Master's Thesis in Intelligent MultiMedia. The latter semester has no courses.

Most courses involve students working on project work in groups in the unique Aalborg style. Here, each semester the students work together in groups of three to four on self-chosen projects and this has proven to give students better opportunities after their education. Approximately 50% of the courses have individual examinations and all courses can be examined as part of an oral examination based on the prepared project report.

The Masters course is open for non-Danish and Danish students. All courses are given in English and the thesis can be written in English or Danish. Each student is graded according to internationally recognised grading schemes.

A Lifelong Learning course is given each August for returning students of Aalborg University who wish to continue their education. This course will involve minimodules on Human Computer Interfaces, IntelliMedia and Virtual Reality.

The courses prepare students for project work and for the growing demand for IntelliMedia in commerce and industry. Already we have begun a proposal to integrate IntelliMedia teaching within the EU SOCRATES 1 year M.Sc. programme in Language and Speech and other SOCRATES programmes in MultiMedia and distance learning (EuroPACE 2000). It is important that these courses give students a sound basis for conducting Ph.D. research in this area. We are also interested in integrating our more technical MultiMedia training with the arts based courses in Humanities at Aalborg.

The emphasis on group oriented and project oriented education at Aalborg University is an excellent framework in which IntelliMedia, an inherently interdisciplinary subject, can be taught.

6 Conclusion

Aalborg University is well equipped in terms of research expertise and education to be able to contribute to IntelliMedia 2000+ in Denmark which will be important for the future of international computing and media development.

We believe IntelliMedia will also throw light on the numerous developments in Computer and Cognitive Science (CS) (see Ó Nualláin 1995 and Ó Nualláin et al. 1997).

We have already begun an investment in the area with the MMUI initiative (see WWW: http://www.cpk.auc.dk/CPK/MMUI/) and believe that this will be important for the future of Denmark's links to SuperinformationhighwayS and international focus on IntelliMedia.

7   References

Baekgaard, Anders (1996) Dialogue Management in a Generic Dialogue System. In *Proceedings of the Eleventh Twente Workshop on Language Technology (TWLT) Dialogue Management in Natural Language Systems*, 123-132, Twente, The Netherlands.

Dalsgaard, Paul and A. Baekgaard (1994) Spoken Language Dialogue Systems. In *Prospects and Perspectives in Speech Technology: Proceedings in Artificial Intelligence*, Chr. Freksa (Ed.), 178-191, September. München, Germany: Infix.

Mc Kevitt, Paul (Ed.) (1995a) *Integration of Natural Language and Vision Processing (Volume I): Computational Models and Systems*. Dordrecht, The Netherlands: Kluwer-Academic Publishers.

Mc Kevitt, Paul (Ed.) (1995b) *Integration of Natural Language and Vision Processing (Volume II): Intelligent Multimedia*. Dordrecht, The Netherlands: Kluwer-Academic Publishers.

Mc Kevitt, Paul (Ed.) (1996a) *Integration of Natural Language and Vision Processing (Volume III): Theory and Grounding Representations*. Dordrecht, The Netherlands: Kluwer-Academic Publishers.

Mc Kevitt, Paul (Ed.) (1996b) *Integration of Natural Language and Vision Processing (Volume IV): Recent Advances*. Dordrecht, The Netherlands: Kluwer-Academic Publishers.

Ó Nualláin, Seán (1995) *The search for mind: a new foundation for cognitive science*. Norwood, New Jersey: Ablex Publishing Corporation.

Ó Nualláin, Seán, Paul Mc Kevitt and Eoghan Mac Aogáin (Eds.) (1997) *Two sciences of mind: readings in cognitive science and consciousness, "Advances in Consciousness Research" (AiCR 9)*. USA: John Benjamins.