Intelligent MultiMedia

Paul Mc Kevitt*
Center for PersonKommunikation (CPK)
Fredrik Bajers Vej 7-A2
Institute of Electronic Systems (IES)
Aalborg University
DK-9220, Aalborg
DENMARK, EU.
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). Such initiatives augment 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 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.

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.

*Paul Mc Kevitt is also a British Engineering and Physical Sciences Research Council (EPSRC) Advanced Fellow at the University of Sheffield, England for five years under grant B/94/AF/1833 for the Integration of Natural Language, Speech and Vision Processing.

2 Visiting Professor

The Center for PersonKommunikation acts as a host for Paul Mc Kevitt who is a Visiting Professor in IntelliMedia and Language and Vision integration. Dr. Mc Kevitt has edited four books (see Mc Kevitt 1995/1996). 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. He 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.

3 IntelliMedia 2000+

Aalborg University, Denmark has already initiated IntelliMedia 2000+ which involves research with the production of a number of real-time demonstrators

showing examples of IntelliMedia applications and to set-up a new education with a new Master's degree in IntelliMedia and a nation-wide MultiMedia Network which is concerned with technology transfer to industry. More details can be found on WWW: http://www.cpk.auc.dk/CPK/MMUI/.

Four research groups exist within the Faculty of Science and Technology in the Institute of Electronic Systems, each of them covering expertise which together is necessary for building up Intelligent MultiMedia systems. The four research groups are Computer Science (CS), Medical Informatics (MI), Laboratory of Image Analysis (LIA) and Center for PersonKommunikation (CPK) each of them contributing knowledge to platforms for specification, learning, integration and interactive applications, expert systems and decision taking, image/vision processing, and spoken language processing/sound localisation.

3.1 Computer Science

The research at CS 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, 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, 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.

Finally, CS will be able to contribute, with technology they have developed, to synchronize both multiple media streams and their content. It is worth noting that several major actors in IntelliMedia such as those in Germany and Japan have identified synchronisation of processes as the central technical problem. The CS can supply this technology to our programme's advantage.

3.2 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.

3.3 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. This motivates/reinforces the research in interpretation of manipulation and description of dynamically changing objects. Issues of research include also use of the combination of live images and computer graphics for creation of enhanced reality systems, which for example may be used in medical informatics systems on for example medical MRI images of brain tissue, tele presence systems, and tele manipulation.

3.4 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) (see Dalsgaard and Baekgaard 1994) which is being upgraded 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)).

Two scenarios are envisaged: (1) for use in offices, an integrated multi-modal user interface will be investigated; (2) aspects of this multi-modal user interface will be considered for incorporation in a handheld computer system (e.g. the Personal Data Assistant) permitting direct data capture, as a computer version of a notebook. A major goal of the research is to transfer aspects of the multi-modal interface to mobile handheld computer systems.

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.

4 Demonstrator CHAMELEON

The results from the research groups have hitherto to a large extent been developed within the groups themselves. However, there is no doubt that the establishment of future and widespread use of IntelliMedia systems requires collaborations among the groups in order to integrate their results in new user-friendly applications. Some of the results may be integrated within a short term perspective as some of the technologically based modules are already available, others on the longer term as new results become available.

In general, applications within IntelliMedia may conceptually be divided into a number of broad categories such as intelligent assistant applications, teaching, information browsers, database-access, command control and surveillance, and transaction services (banking).

Examples of applications which may result within a short term perspective are enhanced reality (e.g. li-

Figure 1: IntelliMedia Workbench

The system keeps a database record of offices and their functionality/tenants. An advanced scenario would involve multiple speakers planning building and institution layout.

The application involves the integration of a distributed processing and learning platform, the HUGIN decision taking tool, image processing of plans and pointing, and spoken dialogue processing and microphone arrays using a specification and design hub platform which can take MultiMedia input and conduct a semantic resolution of the individuals' inputs.

Examples of interesting problems to be solved as part of this application are resolution of ambiguity where a user says "Who's office is this?" but where the pointing-gesture is ambiguous since the person points sloppily between two rooms rather than into one. The system can then ask the user for a clarification as to which office he/she means. Other interesting interactions are "Point to Paul's office" and "Who's in the office beside Paul's?"

The hub platform demonstrates that (1) it is possible for agent modules to receive inputs particularly in the form of images, pointing gestures, spoken language, and sound sources and respond with required outputs (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 modules for various applications; and (4) various means of synchronising the communication between agents can be tested to produce optimal results.

Let's take a look at the internal frame semantics for a given interaction with the system. For example a spoken language query such as "Who's office is this?" would be processed by the spoken dialogue system to give the associated frame semantics to be stored on a blackboard:

USER: "Who's office is this?"

[SPEECH
INTENTION: query?
LOCATION: office (person)
REFERENT: this
TIME: timestamp]

and an associated image interpretation of a pointing-gesture as:

USER: (pointing-gesture)

[GESTURE
INTENTION: pointing
LOCATION: (X,Y) coordinates
TIME: timestamp]

Further processing would be carried out over the blackboard entities and the query matched to a domain model of stored domain-model database (of coordinates/names) to determine an integrated representation as follows which could in turn be used by a speech synthesizer for giving an answer to the user:

[SPEECH+GESTURE
INTENTION: query+pointing
LOCATION: office (IPKE)
REFERENT: (X, Y) coordinates
TIME: timestamp]

Another example of a spoken language instruction is "Point at Ipke's office" which will be represented as follows:

USER: "Point at Ipke's office"

[SPEECH
INTENTION: instruction!
LOCATION: office (IPKE)
TIME: timestamp]

There is no associated image interpretation from input but the instruction is matched to the domain-model database to determine the coordinates of Ipke's office and then the system directs a laser pointer to point at his office. We can also have 'learning':

USER: "No, that's Tom's office"

[SPEECH
INTENTION: instruction!
LOCATION: office (TOM)
REFERENT: that
TIME: timestamp]

and through dialogue processing (reference resolution) solves 'that' = office (IPKE) from previous dialogue, so we have:

[SPEECH
INTENTION: instruction!
LOCATION: office (TOM)
REFERENT: office (IPKE)
TIME: timestamp]

A decision-taking model such as HUGIN would notice a clash here between LOCATION: office (TOM) and REFERENT: office (IPKE) and this would have to be resolved by either the system asking a request-for-confirmation or assuming the user is right!

Mobile computing aspects of this demonstrator become evident if we consider the user walking in the building represented by the plans/model with a wearable computer (see Bruegge and Bennington 1996, Rudnicky et al. 1996, and Smailagic and Siewiorek 1996) and head-up display. This research could eventually be incorporated into more advanced scenarios involving multiple speakers in a VideoConferencing environment.

5 Education

Teaching is a large part of the IntelliMedia programme and three courses have been initiated: (1) Graphical User Interfaces, (2) IntelliMedia Systems and (3) Readings in Advanced Intelligent MultiMedia. Graphical User Interfaces is a more traditional course involving teaching of methods for the development of optimal interfaces for Human Computer Interaction (HCI). The course brings students through methods for layout of buttons, menus, and form filling methods for interface screens and has hands on experience with the XV windows development tool.

IntelliMedia Systems involves the new and innovative topics of speech, language and vision processing. Here, minimodules are given on methods for recognising and interpreting spoken language in dialogue situations and speech and audio representation. The Dialogue Description Language (DDL) tool and Generic Dialogue System of CPK are explained and demonstrated. Vision minimodules are given on relationships between audio analysis and image analysis, 2D model based recognition of static gestures (hand signals), 3D model based tracking of human motion (limbs), and recognizing/'understanding' human motion patterns. There are minimodules on Natural Language Processing (NLP) and pragmatics. The course is augmented with videos and live demonstrations. Hence, this course is true IntelliMedia involving speech, vision and language processing. A guest lecture can be given as part of this module.

The course on Readings in Advanced Intelligent MultiMedia is innovative and new and involves active learning where student groups present research papers and then the whole class can have a general discussion of them. The presentations would include four aspects: (1) who the group is and what their project is, (2) a summary and critical analysis of the papers, (3) how the papers relate to their project and (4) how do the papers and their project relate to IntelliMedia 2000+. Then, the whole class discusses the readings and group presentation. One group member writes up the minutes of our discussions for posterity. The idea here is that it will not only develop the students' presentation skills but also their ability to assimilate, analyse critically and use recent research in the field. The papers are chosen from a selection of books which have just been published on the latest research.

A new international Master's Degree (M.Sc.) has been established and incorporates the courses just mentioned as core modules of a 1 and 1/2 year course taught in English on IntelliMedia. More details can be found on WWW: http://www.kom.auc.dk/ESN/ A Lifelong Learning course is given in August for returning students of Aalborg University who wish to continue their education. This course is a compression of the core IntelliMedia courses.

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. Groups can even design and implement a smaller part of a sys-

tem which has been agreed upon between the groups.

6 MultiMedia Network (MMN)

Aalborg University has initiated a number of Networks to link to industry and to conduct technology transfer. A MultiMedia Network (MMN) has been established which integrates IntelliMedia 2000+ from the engineering and computer science faculties with the humanistic faculty and already we have made a number of links to companies. It is also a natural forum with respect to conducting joint research and projects for student groups. The board of the network meets regularly to discuss joint activities between the University and industry. More details can be found on WWW: http://www.auc.dk/nc/

7 Conclusion

Intelligent MultiMedia will be important in the future of international computing and media development and IntelliMedia 2000+ at Aalborg University, Denmark brings together the necessary ingredients from research, teaching and links to industry to enable its successful implementation. Particularly, we have research groups in spoken dialogue processing, image processing, and radio communications which are the necessary features of this technology. Our application demonstrator which focusses on giving help on building usage is an ideal one for testing integration of various modules.

The emphasis on groupwork and project based education at Aalborg University enables better multidisciplinary integration which can normally be a problem. Our new Master's education will help to produce graduates who are proficient in the necessary theories and tools. The MultiMedia Network is helping to not only bring together the engineering, computer science and humanistic faculties but also to build links to industry. All of these ingredients are necessary for the success of the future of Intelligent MultiMedia which is part of the increasing Global Information Society and links to SuperinformationhighwayS.

References

Bruegge, Bernd and Ben Bennington (1996) Applications of wireless research to real industrial problems: applications of mobile computing and communication. In *IEEE Personal Communications*, 64-71, February.

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.) (1995/1996) *Integration of Natural Language and Vision Processing (Vols. I-IV)*. Dordrecht, The Netherlands: Kluwer-Academic Publishers.

Rudnicky, Alexander I., Stephen D. Reed, Eric H. Thayer (1996) SpeechWear: a mobile speech system. In *Proceedings of ISSD-96, International Symposium on Spoken Dialogue (ISSD 96), October 2-3, Wyndham Franklin Plaza Hotel, Philadelphia, USA*, Fujisaki, Hiroya (Ed.), 161-164. Tokyo, Japan: Acoustical Society of Japan (ASJ).

Smailagic, Asim and P. Siewiorek (1996) Matching interface design with user tasks: modalities of interaction with CMU wearable computers. In *IEEE Personal Communications*, 14-25, February.