Handbook of Research on Socio-Technical Design and Social Networking Systems

Brian Whitworth
Massey University-Auckland, New Zealand

Aldo de Moor
CommunitySense, The Netherlands

Volume I

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey · New York

Director of Editorial Content: Kristin Klinger
Director of Production: Jennifer Neidig
Managing Editor: Jamie Snavely
Assistant Managing Editor: Carole Coulson
Typesetter: Michael Brehm
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on socio-technical design and social networking systems / Brian Whitworth and Aldo de Moor, editors.
p. cm.
Includes bibliographical references and index.
Summary: "Every day throughout the world, people use computers to socialize in ways previously thought impossible such as e-mail, chat, and social networks due to emergences in technology. This book provides a state-of-the-art summary of knowledge in this evolving, multi-disciplinary field"--Provided by publisher.
ISBN 978-1-60566-264-0 (hardcover) -- ISBN 978-1-60566-265-7 (ebook)
1. Online social networks. 2. Internet--Social aspects. 3. Information technology--Social aspects. I. Whitworth, Brian, 1949- II. Moor, Aldo de.
HM742.H37 2009
303.48'33--dc22
2008037981

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter XXVIII
Creating Social Technologies to Assist and Understand Social Interactions

Anton Nijholt
University of Twente, The Netherlands

Dirk Heylen
University of Twente, The Netherlands

Rutger Rienks
University of Twente, The Netherlands

ABSTRACT

In this chapter the authors discuss a particular approach to the creation of socio-technical systems for the meeting domain. Besides presenting a methodology this chapter will present applications that have been constructed on the basis of the method and applications that can be envisioned. Throughout the chapter, illustrations are drawn from research on the development of meeting support tools. The chapter concludes with a section on implications and considerations for the on-going development of social technical systems in general and for the meeting domain in particular.

Assimilation into the Borg Collective might be inevitable, but we can still make it a more human place to live.

—Pentland, 2005

INTRODUCTION

Socio-technical computing inherits the complexity related to software engineering and system integration whilst embedding the human in the loop. It also inherits the difficulties of understanding and modeling human-human and human-computer interaction in the context of a changing environment (see Clancey, 1997). In this chapter we will outline an approach to the development of Social Technical Systems, with the focus on meeting support. This approach can be characterized as theory-informed data-driven. In essence the method consists of the following four steps.

Step 1: Collection of a multimodal corpus of social activity signals
Step 2: Description of a myriad of aspects of system relevant activities (annotation) in the collected material
Step 3: Discovery of interdependencies between recorded signals and annotations, annotations and annotations, and signals and signals (e.g. by means of machine learning.)
Step 4: System creation based on knowledge obtained from the previous steps

In the collection and annotation steps, the process relies heavily on the insights provided by the social sciences; in particular sociology, social psychology and linguistics. In return, the annotated collection and the machine learning effort may provide important insights for social theorizing as the annotated corpus provides the researcher with statistics about the occurrence and distribution of certain phenomena and interesting correlations. Increased insight into how people behave can point out problems they encounter in their activities that may be relieved by technologies that are based on this understanding of their activities as derived through Steps 1 to 3. This means that these steps can be viewed both as a way into requirements engineering and as providing the basic data and algorithms to build the tools that can solve some of these problems.

Technology that inherits these possibilities can be said to be social for three reasons. The first is in the way in which the system supports social activities. The second relates to the way the technology can provide insight into social processes which occurs when correlations between phenomena are found. The third reason in which the qualifier *social* relates to the term *technical system* is in how social theories are at the basis of the construction of the technical applications. Given theories on how humans 'operate', technology is equipped with the manual in order to understand and support their operating.

As example case for this chapter our focus is on small business meetings. Currently several projects worldwide are investigating the way technology can support the needs of people in meetings and how it can relieve them of some of the frustrations that meetings seem to impose upon them. Examples in this chapter will be drawn mainly from studies in a series of European projects on meeting analysis and meeting support: M4, AMI, and AMIDA. These projects investigated how human-centred computing techniques can detect and interpret activities of participants in smart meeting rooms and how these techniques can be used to design tools that support meeting participants in their encounters and activities.

This chapter discusses a variety of methodological issues and charts several results showing the rationale behind the scientific drive to develop technological support for social gatherings and events. The chapter also contains a short discussion on ethical issues and potential pitfalls on the road ahead.

MACHINE INTERPRETATION OF HUMAN ENCOUNTERS

When humans interact, they use their natural skills to sense and interpret signals in the environment in such a way that specific behavioural responses result. In any social encounter, including meetings, every person displays both consciously and unconsciously a pattern of verbal and nonverbal behaviour, which

when recognized, reveals his view of the situation and shows information about his internal assessment of the other participants (Goffman, 1955). Recognition and retention of behavioural regularities and patterns identifies opportunities, and can be turned into new insights, a competitive advantage, and a profitable business. The emergence of social patterns forms the basis for automatic detection, analysis and for the retrieval of its components. The main challenge here, of course, is to know how to let machines distinguish patterns of interest, and to let machines make sound and reasonable inferences and decisions, not forgetting the technological opportunities for exploitation.

Although the *automatic* observation and interpretation of human interactions (e.g. on large multimodal corpora) has only recently become an established domain for human computing research, the study of human interactions both computationally (in the field of natural language understanding, for instance) and within the humanities is well established. Social psychologists, for example, have been actively engaged in the development of explanatory (Smith, 1942) and descriptive (Bales, 1950) models of behavioural patterns for over 60 years. All the theories of group behaviour and interaction research can, when operationalized, potentially be used as input for social technical systems in the way we will describe in the next sections. They provide valuable insights that can be exploited for the creation of the quantitative and mathematical models suited for machine perception.

The (business) meeting domain is a relevant and practical domain for the analysis and support of humans and their activities. We cannot think of a world without meetings, and although sometimes we wish we could, they play an important part in our daily lives. Meetings are hard to avoid and everywhere. The domain embodies the comprehension of a subset of people's everyday activities, working and living, that moves beyond the individual. In multi-party interaction, messages are exchanged between individuals in various flavours and melodies, thereby exposing the full gamut of human communication abilities.

The way people meet and interact with each other has altered significantly in the last decade through new telecommunication technologies. Everyday conversations have, by means of technology, more and more been replaced by e-mail, conference calls, and shared data access. A high speed Internet connection, a webcam, a microphone and a few speakers nowadays offer employees access to almost all the resources they need. Technology has altered the notion of a meeting in a way that, instead of physically sharing the same environment, the opportunity to mentally share the same environment has become a more frequent condition for people to interact.

In 1987, Richman predicted that software systems could one day change the way groups of people work together by means of *comprehending* the ongoing group process (Richman, 1987). Although the state of the technology was far from actual recognition, the field of socio-technical computing and interaction augmentation by means of technology started to gain increasing momentum (for a summary, see Rienks, 2007). In the 1990's the idea of autonomous software agents was introduced aiming to assist humans in their everyday task. The agents were assumed to be able to adapt their actions to the environment depending on their understanding of the environment. It is this sort of system -*that can adapt its actions to the interpretation of the sensed environmental information*- and that is central in the remainder of this chapter.

This type of socio-technical system can be decomposed into three parts, the sensing ability, the reasoning ability and the acting ability. These three are depicted in Figure 1.

Sensor information is gathered and depending on the system's abilities, to a certain extent, analyzed and interpreted. Given the systems' interpretations models of the environment are fed with, and possibly adapted by, this information. If a model decides to plan an action based on the input, the acting ability of the system is subsequently triggered to execute the plan that maximizes the system's performance, possibly by using both its physical and its environmental conditions. Knowledge of the environment into which the system is to be applied can be an essential point.

For the meeting case socio-technical systems are likely to understand at least parts of the human-human communication process, before it can begin to provide adaptive and active support. But how do humans communicate (what is there to be sensed) and what are the technical abilities to interpret, reason and act (how can we act)? This makes a fundamental case to successful development.

The obvious initial question that one is to ask on the way towards the creation of a social technical system is about the goal of the system. What should the system do and where is its added value is expected? What do end users really want from it systems, and in our social event case: How can social events be improved by technological means? What is to be understood from the environment (a gathering of people) to make the system successful? What is in scope and what is out of scope? Does the system consider a single conversational partner or the group as a whole? Does it confine itself to just the conversation? Which input modalities are to be sensed: verbal and nonverbal, both, or none?

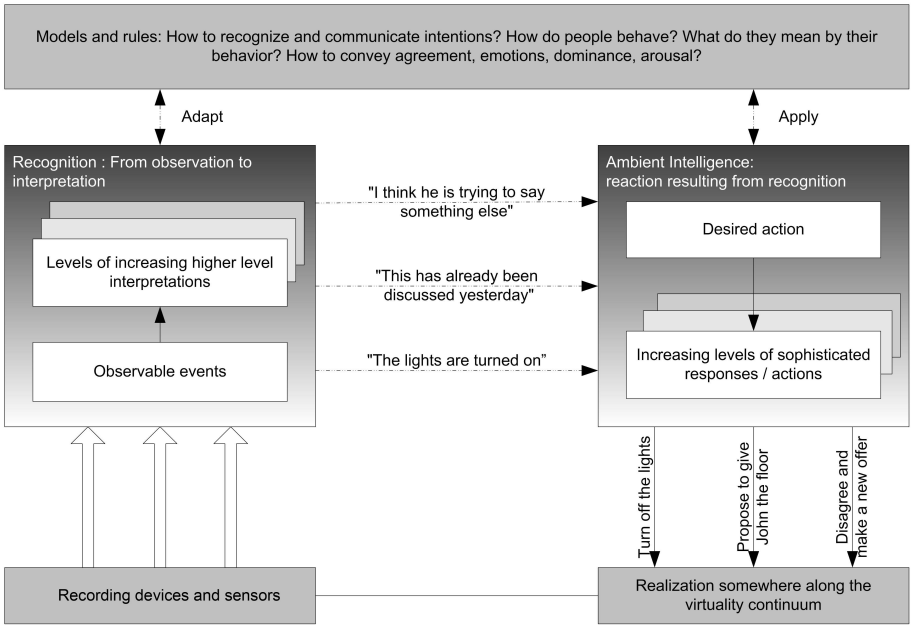
METHODOLOGY

The previous figure showed the idea behind the social technological systems that process audio and video data, interpret what is happening and then react in various ways. The process of developing such systems generally starts by collecting a large number (a corpus) of recordings resembling the phenomena or the situations of interest (to the system). Starting from this collection of signals, manual or preferably *automatic,* recognition processes are then to apply predefined models or coding schemes. The resulting observations that systematically describe the data (annotations) are then to be used as input for further recognition, reasoning and acting, be it for either on-line and/or off-line (hindsight) support.

The construction of models that allow for the interpretation, annotation and derivation of human behaviour are central. For this construction an iterative loop of four steps is generally used.

- A representative corpus should be collected from which the behaviour (consisting of objects and events) that is to be modelled or detected emerges.

Figure 1. Three steps to action generation



- Initial coding schemes need to be devised to facilitate (statistical) inferences and correlations inspired by events or objects that are contained within the data.
- The coding scheme should be mapped correctly onto this data before inferences can be made.
- Machine learning algorithms are to be trained on the extended corpus for successful automatic replication of the coding scheme applied. Examination of the classification results then provides information on how to alter the coding schema in a subsequent iteration

The choice of models stems from research objective or foreseen applications and/or can be derived from corpus investigations (e.g. clustering techniques).

Collecting a Corpus

To learn things, one has to gain insight into what is going on and in the case of machine learning to obtain this insight one needs a lot of examples. Progress in data driven approaches to human computing research therefore requires a large data set that allows for empirical observations of the phenomena of interest. A large dataset that comprises a collection of recorded signals that represent a preferably representative sample of a particular phenomenon is also known as a *corpus*. A corpus enables the validation of domain related rules and hypotheses on empirical grounds. It also provides the opportunity for scientific explorations and hypothesis testing. As a corpus typically contains labels, tags, or annotations that signify occurrences of particular phenomena, it can in this way be used to check for the coexistence of certain phenomena within particular contexts and for the correlation of particular signals and events in a (semi-)automatic manner.

In a corpus that contains just data such as text, one could for example extract word combinations either to create a model that predicts the next word for any word from the text, or to validate such a model in terms of correct predictions. However, if this same corpus also contains a Part-of-Speech tag (such as `Noun' or `Verb') for each word, models can be built that predict the Part-of-Speech tag given a word (see e.g. Brants, Skut, & Uszkoreit, 2003). These models that explicate patterns in the data, and that transform data into information, can in turn also be validated either on other corpora, on parts of the corpus that were not used for training the model, or on new samples.

Machine learning techniques use statistic inferencing to deduce more complex observations from aggregations of features describing the signals. Focusing on multiple signals helps to disambiguate observations and therefore (theoretically) also allows for better recognition. Multi-modal signal collections or multimodal corpora are therefore usually collected to study social phenomena in which one wants to study higher level phenomena such as for the meeting case: agreement, rapport, dissonance, and group performance, that are not directly manifest through one unique behaviour but may show through a combination of features of various kinds of behaviour.

For the research that was conducted within the AMI project over one hundred hours of meetings that followed a similar scenario were recorded. The corpus comprised 120 different meetings in total. The signals that were recorded of these meetings were captured in meeting rooms equipped with many sensors. Typical sensors used for capturing the data were cameras (recording global and close-up views), lapel microphones, microphone arrays, a whiteboard and smart pens. But also meta-information such as the seating arrangement and the (PowerPoint) presentations that were used were collected. In the end the recorded data also included manually created transcripts, dialogue acts and summaries[1].

This corpus was analyzed by means of tools to discover regularities in annotated human behaviour and to construct consecutive models and hypotheses. These models in turn were evaluated, for example, by using the corpus itself, but also by means of simulations and user studies (See section on tools).

Annotation Schema Creation

Annotations are used to codify judgments of observers in relation to an annotation model or schema. They are the tangible result that captures, organizes, and conveys observed information in a structured manner.

Annotation schemas are created to fulfil a certain need; be it either answers to the question of the researcher or, as in case of a system, to fulfil part of its goals. Any resulting model, to put it more generally, should obtain sensible and interpretable distinctions from the data. For human computing applications, the annotation schemas are often inspired by social psychological hypotheses that try to describe human-human interaction. The models from Bales for example distinguish task-based and process-based participants whilst given a set of features that were to be recognized by the observers. He showed with his model that face-to-face interactions contain formal similarities that occur irrespective of the individual participants and their locations. In the AMI schema for dialogue acts, some of the typical categories used by Bales in his Interaction Process Analysis scheme were incorporated.

The annotation schemes in AMI stretch from the description of more easily observable features such as speech, gestures, and focus of attention to more semantic information: dialogue acts, topics discussed and perceived level of dominance.

As mentioned before, these annotations can be used for a number of tasks. They can be used to evaluate hypotheses in the area of social psychology, as examples for machine learning algorithms that strive for automatic model application on unseen data, and for the validation and re-design of the annotation schemas themselves.

The Annotation Process

To be able to apply an annotation scheme accurately, observers should make judgments about what they observe. This is not always a trivial task. Making adequate judgments requires observers to understand the `culture' of the observed interaction and to possess a certain (social) 'sensitivity' that includes the ability to empathize with the observed interactions. To quote Bales: 'We consider ourselves fortunate when we have roughly comparable rates of incidence of a series of phenomena. When these rates are based on data gathered in a comparable way and data conform standard definitions, we are able to make more definite comparisons' (Bales, Strodbeck, Mills, & Roseborough, 1951).

Thus a high agreement between observers means that observers highly agree on the chosen categories from the annotation schema for particular sections of the observations. A high agreement is beneficial as the observations now generalize across observers and become more easily reproducible (Cohen, 1960). However, there is a trade-off here between the amount of training that is required for the observers and the desired level of agreement. The more training needed for the observers, the harder it will be for others to apply the same set of categories with any assurance of obtaining similar results (see Bales et al., 1951).

Many projects face the challenge of manually annotating a large amount of data for various signals and modalities. The process of creating the annotations by itself is, even without focusing on the training of the observers and reliability of the resulting annotations, a tedious and expensive task. If annotations have to be performed manually, one can develop tools that allow for the efficient creation of annotations. Currently there are several tools available for free that all offer a similar functionality and interface; examples are Anvil,[2] or Elan.[3] The Nite-NXT toolkit[4] has the advantage that the interface can be easily adapted with a minimum of programming allowing the creation of an annotator-friendly interface depending on the kind of annotation.

So the way the annotations are created, the way the annotation schema is devised and the way the data is gathered are all relevant aspects to consider when one wants to create algorithms that are to replicate the human annotations on unseen data. For more elaborate information about annotations and issues related to their obtainment see (Reidsma, to appear).

Schema Validation

Annotation schemas can be evaluated in order for them to be improved. These improvements can sometimes be necessary to realize an easier schema application for the observers, or a better fit with the data. This can happen in the case where particular categories that could describe the observations are missing, or if some are indistinguishable because there is too much overlap. Confusion matrices generated from annotations by various observers and/or algorithms can provide valuable insights in this respect. On the other hand, the applied annotations can be used in simulation environments to see how well they fulfil the goals of the designers. One way to do this is to re-create the events in a virtual environment using the annotations as a script to run a scene. One can then visually compare the video as it was originally recorded in parallel with the virtual scene and look for discrepancies. For our studies on the AMI corpus we built a replica of the meeting rooms for this kind and other kinds of studies (see picture further below).

Machine Learning

Automatic recognition of human behaviour and events, in the data-driven approach boils down to automating annotation of higher level phenomena using aggregates of lower level (more straightforward observable) features, where the automatic procedure is derived from examples created by hand. If we want to investigate how this can be achieved, we enter the world of machine learning. The field of machine learning is concerned with the question of how to construct computer programs that can learn from examples and that can adapt to their environment. Machine learning provides the technical basis of data mining, that is, it enables the extraction of implicit previously unknown and potentially useful information from the data. To be more precise, we want the machine learning algorithms to learn to reproduce the annotations that have been created on top of recorded 'sensor' data. If the corpus has been carefully selected and annotated, the resulting algorithms should be able to produce good results on new data as well, as long as this is sufficiently comparable to the data the algorithms were trained on.

To deduce the higher level phenomena we need classifiers that learn from the annotated data how to combine those features that are able to describe the categories defined by the annotation schemas with the highest accuracy. One needs to select the labels that one wants to enter in the algorithm: which combination of values for a series of phenomena can predict the outcome of the value of some other phenomenon. Machine Learning Toolkits allow one to investigate this and similar questions. Features are aspects that describe phenomena and a certain combination of features can be used to differentiate between phenomena. They check a single property of the classification instances, that is, the phenomena that are to be distinguished. For every phenomenon that is to be distinguished from any other phenomenon by a classifier, the same set of features needs to be available.

To know the appropriate set of features that is able to make the distinction amongst the phenomena that one is after is always a big challenge that is to be resolved. From all the features and their values that are available in the corpus machine learning algorithms are able to distil models by means of for instance rule miners that are able to predict a class label given a set of feature values. An example of such a resulting model is shown in the figure below.

For each of the meetings in the AMI corpus, we had several annotators decide who they found to be the most influential or dominant participant. We found that the agreement on this issue was quite high. We selected several features that can be fairly easily obtained by studying the speech transcripts and used several techniques to find out whether we could predict the same scores for dominance/influence based on these features. If one divides the participants and feature values in high(3), normal(2) and low(1) influential and uses the features "how many times did a participant take the turn", "how many interruptions by the

speaker were successful", and "how often did the participant attempt to grab the floor", one can build an algorithm that judges correctly that a participant is in the same dominance/influence category that the human judges did 85% of the time.

The model depicted is able to give an influence label to instances of the feature set{Turns, Successful Interruptions, Floor grabs}. The feature values in this case are integers collected from the behaviour of one person in range from 1 (low) to 3 (high). From the model one could, for instance, distill that the observation {2,2,2} would obtain the class label 'Normal Influential'.

With the use of classification algorithms like this, one can start to craft off-line and on-line applications.

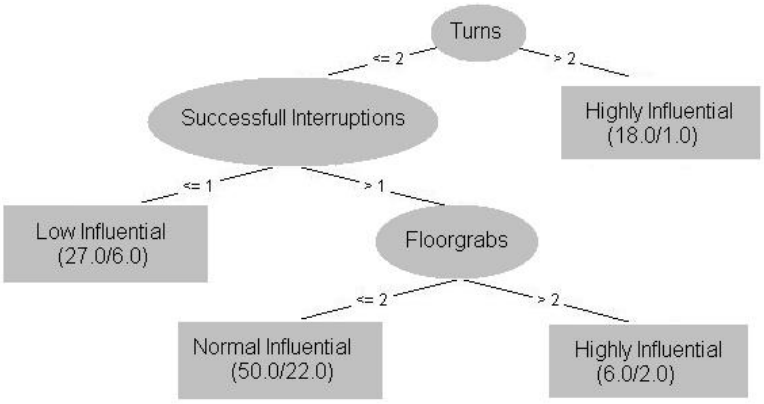
TOOLS AND APPLICATIONS

One application that was developed based on the influence detection showed the influence levels of participants over the course of a meeting. If this information were available in real time, a chairman could alter his style of leadership in order to increase the meeting's productivity (DiMicco, 2004). Combined with other information, systems could be created that directly suggest how to change the leadership style. One could even imagine a virtual chairman who is able to lead a meeting all by himself, maintain a good balance, give turns and keep track of a time-line.

Another implementation has been realized in a Virtual Meeting Room (VMR), (Nijholt, Rienks, Reidsma, & Zwiers, 2006). This VMR was particularly developed for schema validation, signal replay, as a remote conferencing application, and to serve as a test environment for software agents. This virtual meeting room can be augmented with the relative influence levels, as in this case depicted in Figure 4 by the size of the black balls shown in front of the participants. The domes surrounding the participants' heads provide information about their gaze behaviour.

One of the other results of our work that has been executed on the corpus is that a tentative profile has been constructed of how influential participants, as experienced by actual meeting participants, distinguish themselves by means of verbal behaviour from less influential participants. Our results here show that if a participant raises issues, elicits solutions, evaluates these solutions and then steers towards a choice amongst the possible solutions, this is indicative for a person who is highly influential, and who controls the course of a discussion (which intuitively also seems correct). On the other hand, it appeared that if someone provides options, back-channels a lot to others and resorts to shorter contributions in the decision phase of a discussion an (understandable) profile of a less influential participant appears.

Figure 2. A resulting decision tree to determine if a participant is of a particular influence category



IMPLICATIONS AND CONSIDERATIONS

Ongoing developments in the area of progressing meeting technology and socio-technical computing could result in far reaching ramifications for human life and human well-being. The advent of the networked society has permitted people to interact with each other remotely in a fashion unprecedented in history. This, on the one hand, has brought about enormous benefits and convenience, whilst on the other hand, it has extended a dark side where a new technology is abused or disrupts human relations (Nishida, 2007).

It is however not unlikely that the introduction of new technologies in the meeting domain will, for example, pose difficult challenges for participants and their supervisors. Although a participant's access to remote participants all over the globe, for instance, may theoretically increase his or her productivity, ubiquitous connections to others comes along with temptations for distraction and the wasting of time. Not to mention the temptation that will emerge for supervisors to implement automated supervision techniques. How useful would it be for an employer to gain automatic insights into the performance of his or her participants over the previous meetings? And what would the participants think of this? It

Figure 3. An example of a meeting browser

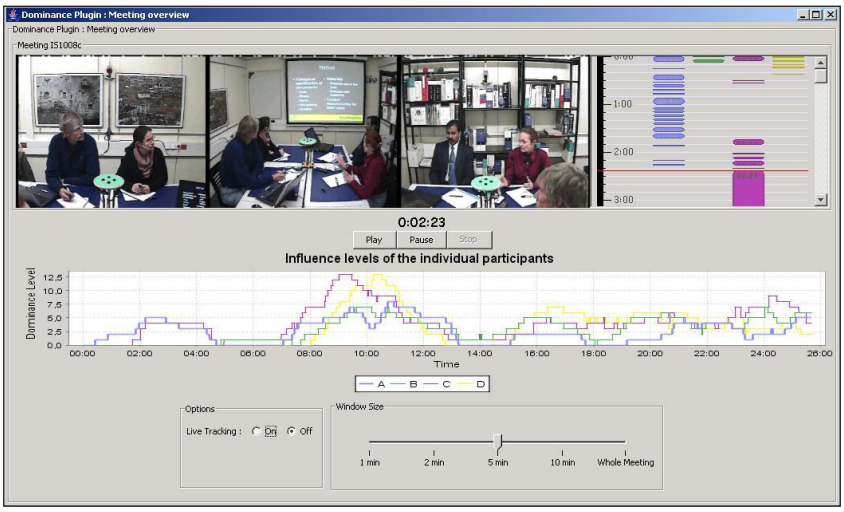
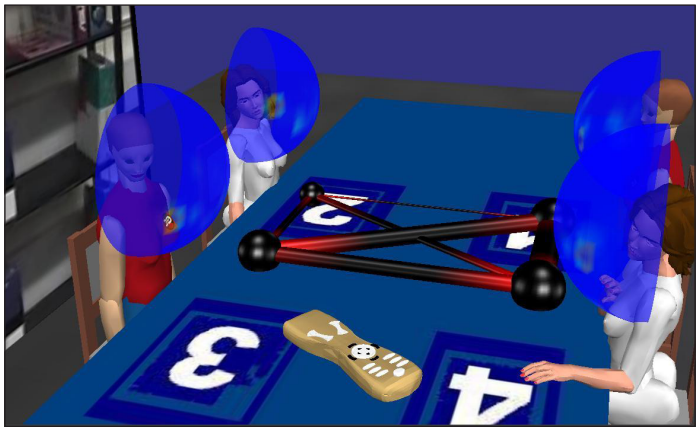


Figure 4. Visualizing gaze and dominance information in a virtual representation



seems not unimaginable that these `monitoring' techniques could lead to tension, distrust, and resentment. So what could seem beneficial and an advantage at first sight, might turn out to be a disadvantage in the end.

Another potential danger that lies enclosed in emergent technologies is over reliance on systems that are not flawless and that are trained on a specific domain. Over reliance on automatic systems, especially without knowledge of the rationale behind the systems could lead to annoying situations in which high expectations can turn out to become nasty dampers. The impact of faulty meeting technology will perhaps not be as large as that of an earthquake warning system that makes a mistake, but for a business meeting where high interests are at stake the risks can be serious. We assume it would be better to at least think twice and to always refrain from blindly following a system's proposals, and rather consider its advice as suggestions that could be taken into account. Of course the level of authority and autonomy that is given to the system plays a part in this. Also, as the technologies have been trained for a specific domain, the risk exists that they are put into practice in different domains.

Challenges

The characteristics of emerging socio-technical systems imply new approaches to usability engineering as well as associated evaluation and testing techniques. Emerging systems that are devised to support, and to a certain degree also understand, social events as they naturally occur require the ability to comprehend messages emitted through various social signals, including voice, gestures, gaze and facial expressions. When allowing humans to communicate naturally with the input devices, these systems should be able to distil, within this gamut of signals, all the items that are of interest to the system.

Despite considerable research efforts in the field of multi-modal fusion (see e.g. Oviatt, 2003), knowledge about how humans combine different channels is still limited. Not to mention the recognition of the behaviour of the group as a whole. Furthermore, the system should also be sufficiently prominent, because a lack of a prominence might result in users who are unaware of the system's existence (Nijholt, Rist, & Tuijnenbreijer, 2004).

Data that is automatically sensed from sensors, such as microphones and cameras, needs to be sensed by sufficiently accurate sensors. The subsequent recognition module that transforms the perceived data into information should, in turn, also be sufficiently reliable for its task.

It is often mentioned that social behaviour is to be interpreted in a given context. For example, a smile in an everyday conversation can be a sign of appreciation, whereas, during negotiation, it can be a sign of disagreement. So, for the reliable interpretation of human behaviour, it is important for human sensing systems to be aware of the context of the situation. To date, there is no consensus on what context precisely is, or on how we should specify this. Without a good representation for context, developers are left to develop *ad hoc* systems for storing and manipulating this key information (see e.g. Abowd & Mynatt, 2000). Sometimes the major components of context are referred to as the 5 W's: who, what, where, when, why (Pantic, Pentland, Nijholt, & Huang, 2007). It is difficult to automatically assess the values for most, if not all, of these properties. As a consequence it is therefore recommendable that these socio-technical supportive systems are to be used as suggestive, rather than pro-active.

CONCLUSIONS

Social behaviour is an extremely complex phenomenon where many aspects of everyday life play a part and come together. Systems that are able to perceive and understand what is going on in any social setting pertain to the emergent human computing paradigm in which adaptive systems respond in accordance to their perceived (human) environment.

The methodology of corpus based research investigates the possibilities for this technological trend to sense higher level concepts after a clever

combination of more direct observations. This methodology requires a model that describes the phenomena that should be recognized as well as a carefully chosen example domain on which this model should be manually applied. After manual application machine learning algorithms can be trained in order to replicate the human observations from a set of features that are both easily observable and expected to relate to the phenomena under consideration.

Blind reliance on current state of the art technological performance might lead to erroneous decision making and entails the temptation of abuse, which in turn can lead to nasty privacy and responsibility issues. In our opinion, at this moment in time, socio-technological systems can, hinging on their performance, in the best case be used as suggestive or informative guides. This is by itself not a bad achievement, especially when we realize that decisions concerning higher level human-human communication phenomena, such as those that occur in social encounters, are of a highly subjective nature on which humans themselves often disagree.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers of an earlier version of this paper for their suggestions and discussions. This work is supported by the European IST Programme Project FP6-033812 (Augmented Multi-party Interaction, publication AMIDA-137). This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

REFERENCES

Abowd, G. D., & Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction*, *7*(1), 29–58.

Bales, R. (1950). *Interaction Process Analysis*. Cambridge: Addison-Wesley.

Bales, R., Strodbeck, F., Mills, T., & Roseborough, M. (1951). Channels of communication in small groups. *American Sociological Review*, *16*, 61-468.

Brants, T., Skut, W., & Uszkoreit, H. (2003). Syntactic annotation of a German newspaper corpus. In Abeillé, A. (Ed.), *ATALA sur le Corpus Annotés pour la Syntaxe Treebanks* (pp. 69-76). Dordrecht: Kluwer.

Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations.* Cambridge, Mass.: Cambridge University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *1*(20), 37–46.

DiMicco, J. (2004). Designing interfaces that influence group processes. *In Doctoral Consortium Proceedings of the Conference on Human Factors in Computer Systems (CHI'04),* Vienna, Austria 1041-1042.

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry: Journal of Interpersonal Relations*, *18*(3), 213-231.

Nijholt, A., Rienks, R., Reidsma, D., & Zwiers, J. (2006). Online and Off-line Visualization of Meeting Information and Meeting Support. The *Visual Computer International Journal of Computer Graphics*, *22*(12), 965-976.

Nijholt, A., Rist, T., & Tuijnenbreijer, K. (2004). Lost in ambient intelligence? In Extended abstracts on Human factors in computing systems (CHI'04), Vienna, Austria, 1725-1726.

Nishida, T. (2007). Social Intelligence Design and Human Computing. In M. Pantic, S. Pentland, A. Nijholt, & T. Huang (Eds.), *Artificial Intelligence for Human Computing* (pp. 190-214). Lecture Notes in Artificial Intelligence 4451, Berlin: Springer-Verlag.

Oviatt, S. L. (2003). Multimodal interfaces. In Sears, A., & Jacko, J.A. (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 286-304). Mahwah, NJ: Lawrence Erlbaum Associates.

Pantic, M., Pentland, S., Nijholt, A., & Huang, T. (2007). Human computing and machine understanding of human behaviour: A survey. In M. Pantic, S. Pentland, A. Nijholt, & T. Huang (Eds.), *Artificial Intelligence for Human Computing* (pp. 47-71). Lecture Notes in Artificial Intelligence 4451, Berlin: Springer-Verlag.

Pentland, A., (2005). Socially Aware Computation and Communication. *IEEE Computer, 38*(3), 33-40.

Richman, L. (1987). Software systems that catch the team spirit. *Fortune*, *115*(12), 125-136.

Rienks, R. (2007). *Meetings in smart environments: Implications of progressing technology* (Ph. D. Thesis, University of Twente, the Netherlands).

Smith, M. (1942). An approach to the study of the social act. *Psychological Review*, *49*(5), 422-440.

KEY TERMS

Machine Learning: Machine Learning: This subfield of artificial intelligence is concerned with the design, analysis, implementation and applications of programs that learn from experience. The discovery of general rules from large data sets using computational and statistical methods is an important application area. Such large data sets can, for example, be corpora that contain audio and video recorded human-human or human-computer nteraction.

Corpus-based Research: Traditionally a corpus is a collection of language examples: written or spoken examples of words, sentences, phrases or texts. Nowadays a corpus can be any collection of examples, for example, human-human interactions, protoin interaction, video fragments, maintenance information, etc. A corpus is collected in order to learn from it, that is, to extract domain-specific information. Examples can be analysed and rules and models underlying the examples can be discovered. Machine learning algorithms are used to extract relationships between examples. Manual structuring of such data (annotation) allows the integration of human preferences and knowledge in machine learning algorithms.

Annotation Process: A corpus of examples, whether these are language or interaction examples (distinguishing between different kinds of interaction) can be annotated with human knowledge that makes it possible to distinguish characteristics of these examples. Machine learning algorithms can be guided and supported by such annotations and machine learning results provide feedback about our intuition and heuristics concerning which features of the examples help to distinguish them into classes. To support human annotators, tools are developed that visualize and otherwise emphasize characteristics of the examples in the corpus.

Multimodal Interface: Interface to a computer system (from a mobile device to a smart environment) that allows multiple modes of interaction. Among the modalities can be speech, touch, gaze, or gestures. Modalities can supplement one another, but also complement one another.Combining different input modalities is called fusion. It allows a system to disambiguate user input in order to get a more complete understanding of a user's commands or behavior.

Smart Meeting Room: A smart meeting room uses multi-modal sensors to detect and capture the verbal and nonverbal behavior of meeting participants. This is done in order to provide real-time support to these participants and to record meeting activity for off-line intelligent browsing and retrieval of meeting activities. Modeling multi-party human-to-human interaction, e.g. by using machine learning approaches, helps to recognize important activities and events during a meeting.

Nonverbal Behaviour: Nonverbal behaviour not only supports verbal communication. By observing nonverbal behavior, the observer, whether it is a computer system or a human observer, can learn about the intentions, the attitudes and the feelings of its human partner. Nonverbal behavior includes gaze behavior, facial expressions, body posture, gestures, and prosodic information, but it can also include physiological information. Hence, supporting verbal communication, issuing nonverbal commands, and allowing our human or computer partners to learn about our feelings, intentions, and preferences are the main reasons for needing to detect and interprete nonverbal behavior.

Sensor Information: Sensors in smart environments provide us with information about its inhabitants, their activities, and their interactions. Cameras and microphones allow audio-visual processing of perceived activity. Proximity and pressure sensors tell us about the location of inhabitants. Such sensors allow us to track the inhabitants and their activities in the environment. Devices that measure physiological information, including brain activity, can provide detailed information about the affective state of a user.

ENDNOTES

1 See http://corpus.amiproject.org
2 http://www.anvil-software.de/
3 http://www.lat-mpi.eu/tools/elan/
4 http://www.ltg.ed.ac.uk/NITE/