Issue N° 15 / August 2008

Editor : Valérie Devanthéry, valerie.devanthery@idiap.ch

AMI CONSORTIUM

Newsletter

www.amiproject.org

AMI c/o IDIAP Research Institute, Centre du Parc,
Rue Marconi 19, P.O. Box 592, CH-1920 Martigny
info@amiproject.org - www.amiproject.org

Contents

News

MLMI 2008

5th Joint Workshop on Machine Learning and Multimodal Interaction

8-10 September 2008
Utrecht, The Netherlands

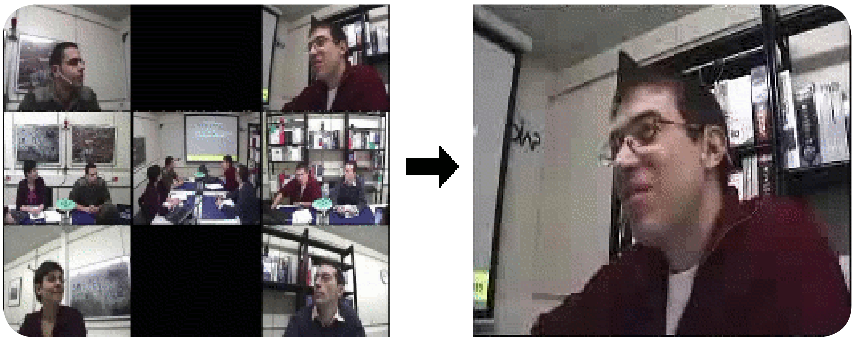
Cover Story

Video Editing in Meetings

A COOPERATION OF TECHNISCHE UNIVERSITÄT MÜNCHEN (TUM) AND BRNO UNIVERSITY OF TECHNOLOGY (BUT)

If you think of today's business meetings where a single participant is connected remotely to a meeting room, you have a problem with representing activities in the meeting room to a remote participant. One camera for the room is normally not enough because some of the important details would not be recorded.

If the meeting room is equipped with several cameras, current video conference systems show all video streams to the remote participant. This is useable if someone is sitting in front of a huge screen where all cameras can be shown simultaneously. But does it make sense to watch all the cameras at the same time? This problem is getting worse if a remote participant is using a small screen or even a mobile phone.



The video editing system from TUM and BUT solves this problem. All meeting rooms in the AMI corpus are equipped with different cameras. There are close-ups for each participant, as well as some cameras, which are recording a overview of the meeting room.

Depending on what is currently happening in the meeting, the system selects the camera which represents the meeting in the best way. For example, a person who is shaking the head or nodding is possibly more important than the current speaker. For this task an event detection system was developed at TUM which recognises events like standing-up, sitting down or taking notes from video and audio sources. These events with additional information from the audio channels are sent to the BUT's editing system.

There, a camera is selected depending on certain rules, which an editor has to take into account when cutting a video. The user has total control over the system, so that it is possible, that only a special participant is shown, or a selected camera is used more often.

The current system can be used for the efficient recording of past meetings, because only one video stream has to be stored. On the other side the system could be used for online video conferencing. Video editing would help to reduce the work load of the remote participant and the bandwidth of the data connection.

Contact:

Benedikt Hörnler
Lehrstuhl für Mensch-Maschine-Kommunikation
Technische Universität München
Arcisstrasse 21
80333 Munich
Germany

Augmented Multiparty Interaction Distance Access (AMIDA) is an Integrated Project funded by the EC's 6th Framework Program, jointly managed by Idiap (CH) and the University of Edinburgh (UK).

Issue N° 15 / August 2008

Editor : Valérie Devanthéry, valerie.devanthery@idiap.ch

AMI CONSORTIUM

Newsletter

Hybrid Multi-Step Disfluency Correction

Human speech production and articulation is full of syntactical and grammatical speech errors. The rate of these errors which are often referred to as **disfluencies** typically lies between 5% - 10% of our speech [1] and tend to increase with the speaker's cognitive load [2] (e.g., in business meetings).

While humans are able to filter disfluencies instantly, natural language processing (NLP) systems are often developed for written, grammatical text and thus not able to deal with speech disfluencies. Therefore, an intermediate system that at best should repair or at least mark the disfluent words before the speech gets processed is desirable. An example of an output of such a system is given in the above picture. There, the red marked words are superfluous and can be ignored by the reader while the black words form the speaker's real intended sentence.

This is where GroDi - Get rid of Disfluencies - steps in. GroDi is a hybrid disfluency detection and correction system that has been developed using the classification scheme developed on the AMI corpus (see [3]). The system is able to detect and correct a broad field of different disfluency types, such as repetitions, hesitations, stutterings, etc.

A thorough investigation of the annotated material (so far 40 meetings), showed a heterogeneity of the disfluencies with respect to how the different types can be detected and let to our hybrid detection approach: Easily detectable disfluencies are identified by a simple lexical rule-matching approach while the remaining disfluencies are tackled by sophisticated machine learning techniques. In fact, our system is able to detect more than 56% of all occurring disfluencies and works faster than real-time. Our next planned step, is to implement a «Hub-version» of GroDi to support the upcoming AMIDA demos with the disfluency correction.

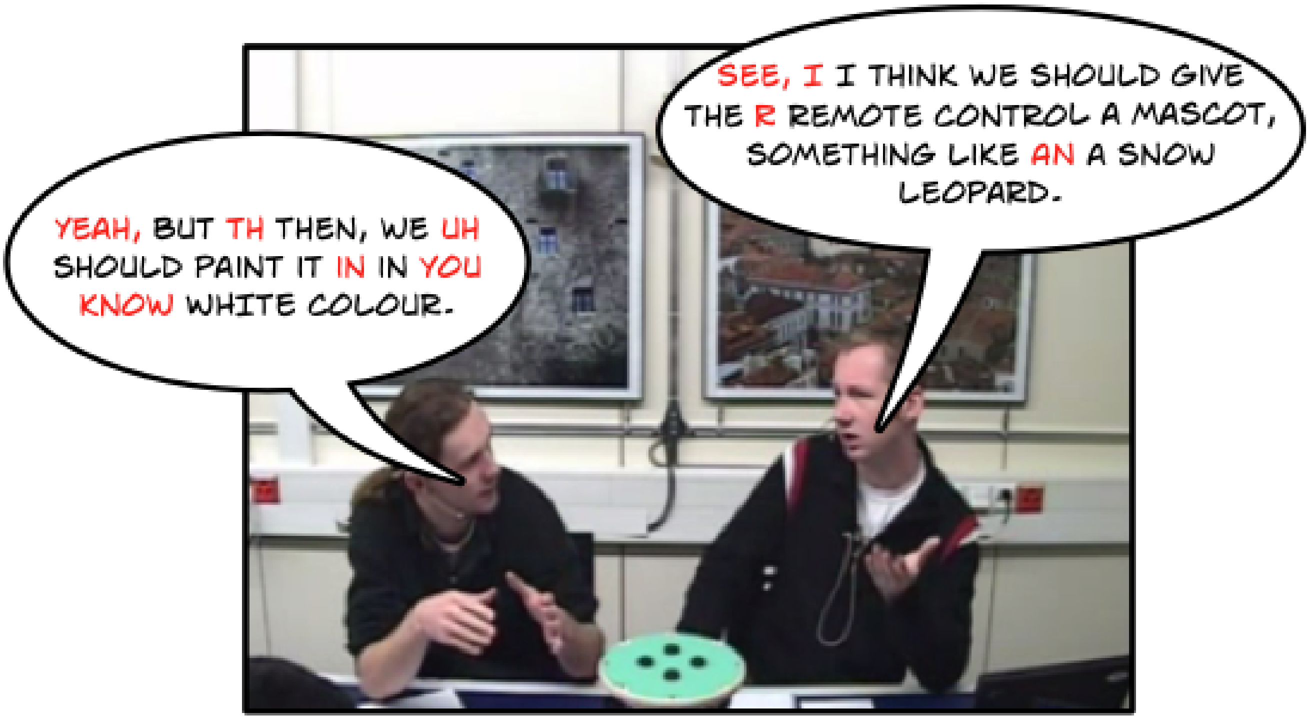
REFERENCES

[1] Shriberg, E.: «Phonetic consequences of speech disfluency», In: International Congress of Phonetic Sciences (ICPhS-99), Volume I., pp. 619622, San Francisco, 1999.

[2] Müller, C., Gromann-Hutter, B., Jameson, A., Rummer, R., and Wittig, F.: «Recognizing Time Pressure and Cognitive Load on the Basis of Speech: An Experimental Study», In UM2001, User Modeling: Proceedings of the Eighth International Conference, pp. 24 - 33, New York - Berlin: Springer, 2001

[3] Besser, J.: «A Corpus-Based Approach to the Classication and Correction of Disfluencies in Spontaneous Speech», Bachelor's thesis, Computational linguistics institute, University of the Saarland, Germany, 2006

Contact:
Sebastian Germesin
DFKI GmbH
Intelligente Benutzerschnittstellen
Stuhlsatzenhausweg 3
D-66123 Saarbruecken
Germany


SEE, I I THINK WE SHOULD GIVE THE R REMOTE CONTROL A MASCOT, SOMETHING LIKE AN A SNOW LEOPARD.
YEAH, BUT TH THEN, WE UH SHOULD PAINT IT IN IN YOU KNOW WHITE COLOUR.

AMI CONSORTIUM

Issue N° 15 / August 2008

Editor : Valérie Devanthéry, valerie.devanthery@idiap.ch

Newsletter

Use of AMI technology in assessing Clinical Network Meetings

The CSIRO Australian E-Health Research Centre joined the AMIDA project in August 2007. Since then the project team has undergone a transformation and the new team has started on the research related to the recording and analysis of clinical network meetings. Researchers known to the AMIDA team, Iain McCowan and Darren Moore (both previously at IDIAP), have transferred to new labs within CSIRO.

The new research team members are Simon Locke (Post Doctoral Fellow, funded under the Australian Government International Sciences Linkage scheme), Hazel Harden (PhD student, funded by Griffith University and CSIRO) and David Wang (PhD student, funded by Queensland University of Technology and CSIRO) and David Hansen (as project manager).

The new team has started the research of using AMI technologies to support clinical network meetings. Clinical network meetings are used to lead health system change. They involve participants, mainly clinicians, from a variety of backgrounds, addressing a particular issue or need in health care. The meetings, which have shared leadership where organisational structure is irrelevant, are vital to changes in the delivery of health care. Hence, early identification of those teams which are likely to be successful can greatly enhance the final outcome of a project.

Hazel's PhD is focusing on social psychology aspects of team interaction, which will involve manual encoding of the video recording of team meetings. Hazel will be using the methodology of interaction process analysis, a technique for studying interaction in small groups. Coding schemes classify communication acts and deliver positivity/negativity, inquiry/advocacy, and other/self ratios. Simon will focus on automating this interaction encoding as far as possible from recorded audio, either directly encoding or by looking for proxy signals in the conversation.

The team has purchased eight Apple iPod nano (3rd Gen.) equipped with Belkin TuneTalk amplifier and Sony ECM-C115 lapel microphone to record audio from each individual during meetings, for automatic encoding. As well as helping Hazel to manually encode video, Simon has synchronised the multiple audio streams using mutual information and is examining how to identify who is speaking from the multiple audio streams.

Future work by Simon will involve more complex analysis of the multiple audio streams. David Wang, during his PhD, will be researching speaker segmentation and clustering from a single audio stream, with help from the other team members.

The team will also investigate automated structuring and summarisation of medical concepts in edited speech transcripts using natural language processing techniques.

As well as our new researchers joining the team, the CSIRO Australian e-Health Research Centre has just moved to its brand new site at the University of Queensland, Royal Brisbane and Women's Hospital. For further information visit: http://aehrc.com.

Contact:

CSIRO
E-Health Research Centre
Dr David Hansen
UQ CCR Building 71/918
Royal Brisbane and Women's Hospital
Herston, Queensland 4029
Australia

ACM SIGIR Conference

20-24 JULY 2008, SINGAPORE

SIGIR Conferences
SIGIR is the major international forum for the presentation of new research results and for the demonstration of new systems and techniques in the broad field of information retrieval (IR).

The Conference included:

- Technical Sessions comprising plenary sessions and talks with topics of interest covering original research contributions relating to IR.
- Posters as an opportunity for researchers to present late-breaking and new ideas in a relaxed, interactive setting.
- Panels will consist of discussions on timely and controversial topics.
- Technical Demos will include leading edge work in every area of IR technology and its application.
- State-of-the-Art Tutorials by leading experts will precede the technical program. The full- and half-day offerings will span a wide variety of topics.
- Day-long Workshops on topics of great current interest to members of the IR research community will take place after the technical program.
- The Doctoral Consortium is a venue for doctoral students to present their research and receive feedback from members of the IR research community.

More information at: http://ilps.science.uva.nl/SSCS2008

Contact:
Dr Wessel Kraaij
TNO Information and Communication Technology
PO BOX 5050
2600 GB Delft

Issue N° 15 / August 2008

Editor : Valérie Devanthéry, valerie.devanthery@idiap.ch

Newsletter

News and Upcoming Events

MLMI 2008

8-10 September 2008, Utrecht, The Netherlands http://www.mlmi.info

MLMI 2008 - 5th Workshop on Machine Learning and Multimodal Interaction

The MLMI series brings together researchers from the different communities working on the common theme of advanced machine learning algorithms applied to multimodal human-human and human-computer interaction.

Several events are also associated to MLMI 2008:

- The AMI Career Day will provide an opportunity for young scientists to talk to representatives of companies working on meeting technology and prepare the next steps of their careers.
- A special session on user requirements and evaluation of multimodal meeting browsers/assistants.
- A workshop on the evaluation of automatic speech recognition systems for Dutch.
- An interproject meeting on the evaluation of space-time audio processing
- A student poster session.

The workshop website has more information about the program, venue and satellite events: *http://www.mlmi.info.*

Looking forward to welcoming you in Utrecht:

- Andrei Popescu-Belis, Idiap research institute (Programme Co-chair)
- Rainer Stiefelhagen, University of Karlsruhe (Programme Co-chair)
- David van Leeuwen, TNO (Organization Chair)
- Anton Nijholt, University of Twente (Special Sessions Chair)

Selected publications

AdaBoost Engine.
P.Zemcik and M.Zadnik
In International Conference on Field Prog Logic and Applications, FPL 2007, pp 656-660, IEEE Computer Society, NL, 2007.

Adaptive Beamforming with a Maximum Negentropy Criterion.
K.Kumatani, J.McDonough, D.Klakow, P.N.Garner, W.Li
In Proceedings of The Joint Workshop on Hands-free Speech Communication and Microphone Arrays, 2008.

Adaptive Beamforming with a Minimum Mutual Information Criterion.
K.Kumatani, T.Gehrig, U.Mayer, E.Stoimenov, J.McDonough, M.Wolfel
In IEEE Trans on Audio, Speech and Language Proc., vol. 15, pp. 2527-2541, 2007.

Analysis of feature extraction and channel compensation in GMM speaker recognition system.
L.Burget, P.Matejka, P.Schwarz, O.Glembek, J.Cernocky
In IEEE Transactions on Audio, Speech, and Language Processing, volume 15, number 7, pages 1979-1986, 2007.

Annotating Subjective Content In Meetings,
T.Wilson
In Proceedings of the Language Resources and Evaluation Conference, Springer, LREC-2008, Marrakech, Maroc, 2008.

Application of CMLLR in narrow band wide band adapted systems.
M.Karafiat, L.Burget, T.Hain and J.Černocky
In 8th Annual Conference of the International Speech Communication Association, pages 4, International Speech Communication Association, Antwerp, Belgium, 2007.

Audio-based unsupervised segmentation of multiparty dialogue.
P-Y. Hsueh
In IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp 5049-5052, Las Vegas, 2007.

Automatic Meeting Segmentation using Dynamic Bayesian Networks.
A.Dielmann, S.Renals
In IEEE Transactions on Multimedia, volume 9, number 1, pages 25-36, 2007.

Evaluation and comparison of tracking methods using meeting omnidirectional images.
I.Potucek, V.Beran, S.Sumec, P.Zemcik
In Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI), pages 12, Brno, CZ, 2007.

Evaluation of Automatic Video Editing.
S.Sumec, I.Potůček
In Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI), pages 12, Brno, CZ, 2007.

Experiencing-in-the-World: Using Pragmatist Philosophy to Design for Aesthetic Experience.
D.Vyas, D.Heylen, A.Nijholt, A.Elien
Proceedings of the 2007 conference on Designing for User eXperiences, Chicago, Illinois, p 16, 2007.

Exploring Contextual Information in a Layered Framework for Group Action Recognition.
D.Zhang, S.Bengio
In 2007 IEEE International Conference on Multimedia and Expo, pages 2022-2025, Beijing, 2007.

Filter Bank Design Based on Minimization of Individual Aliasing Terms for Minimum Mutual Information Subband Adaptive Beamforming.
K.Kumatani, J.McDonough, S.Schacht, D.Klakow, P. N. Garner, W.Li
In Proceedings International Conference on Acoustics, Speech and Signal Processing, IEEE, 2008.

Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006,
N.Brümmer, L.Burget, J.Cernocky, O.Glembek, F.Grezl, M.Karafiat, D. van Leeuwen, P.Matejka, P.Schwarz, A.Strasheim
In IEEE Transactions on Audio, Speech, and Language Processing, volume 15, number 7, pages 2072-2084, 2007.

Hardware Acceleration of AdaBoost Classifier
J.Granat, A.Herout, M.Hradis, P.Zemcik
In Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI), pages 1-12, Brno, CZ, 2007.

Hierarchical Pitman-Yor Language Models for ASR in Meetings
S.Huang, S.Renals
In Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on, Kyoto, pages 124-129, 2007.

How do I address you? Modelling addressing behaviour based on an analysis of multi-modal corpora of conversational discourse
R. op den Akker, M.Theune
In AISB 2008 Symposium on Multimodal Output Generation (MOG 2008), pages 10-17, Aberdeen, UK, 2008.

Interpretation of Multiparty Meetings: The AMI and AMIDA Projects
S.Renals, T.Hain, H.Bourlard
In Hands-Free Speech Communication and Microphone Arrays 2008 (6-8 May), pages 115-118, Trento-Italy, 2008.