

# Meeting recordings at Brno University of Technology

Petr Jenderka, Igor Potůček and Stanislav Sumec  
Dept. of Computer Graphics and Multimedia,  
Faculty of Information Technology, Brno University of Technology  
Email: jenderka|potucek|sumec@fit.vutbr.cz

The goal of creating VUT Brno meeting room is to provide M4 and follow-ups with more data, complementary to ICSI and IDIAP. We aim at recording different speakers with different accent (especially heavily accented "CZEnglish"). The setup of a secondary M4 meeting room is low-cost, mobile, built from off-the-shelf products, and easy to install and operate. The specialty of our meeting room is the 360 degree image capturing using hyperbolic mirror, with the following image-processing.

**Setup** In order to create more data for the project, but also to be able to demonstrate its results outside IDIAP facilities, we aimed at designing a cheaper, lighter, and "buyable" room. In the current setup, we are able to record up to 4 persons using 4 lapel mikes (no mike arrays). The video is captured by one commercial camera with hyperbolic mirror allowing for almost 360 degree view. Audio/video are both recorded on one standard notebook PC.

**Hardware** The *audio* is captured by 4 lapel microphones Sennheiser MKE 2 P-C (same as in IDIAP). 2 mixers Behringer EURORACK UB1204-PRO are providing gain control and phantom power supply (actually 90% of mixers are not used, but they are cheaper than special mike pre-amps with phantom feeders). The notebook PC is equipped by two Hi-Fi PC-MCIA sound-cards VX-pocket. *Video* is taken by digital camera SONY CLIP 345. The hyperbolic mirror by NEOVISION (Prague) is used to transform the 360 degree view for the camera. The PC is connected through IEEE1394 (FireWire) interface. The camera also provides 2 additional (global) audio channels. The total *cost* of the hardware (except the notebook) is  $\approx 4.2$  kEUR. The hardware is easily transportable by one person the only bulky thing being the stand for the camera.

**Software** The recording is performed using a home-made (Stanislav Sumec) tool, based on Direct-show libraries, under MS-Windows. All audio and video streams are stored in real time: two audio wav files with four audio channels and one DV-compressed video avi file with video and additional stereo audio stream from camera. The audio and video channels are acquired synchronously using the Direct-X interface. To ensure minimum time-shift between the channels, low-level layers were used and on top of this, the time differences of the channels are monitored to allow possible further resampling of the output channels. A post-processing tool is required to convert the acquired video sequence. Such software was written (by Igor Potůček) for unwrapping video to panoramic format of view [1, 2].

**Collection and Annotation** The annotation of data from VUT Brno is performed on two levels. For the basic dialogue act annotations, we use "annotation editor" (written by Stanislav Sumec). It works with audio and video, and allows user to define hotkeys to mark starting and ending points of events in real-time. The software allows for annotation of arbitrary number of audio and video streams in multiple AVI type files. The annotation output is stored in XML in the format used throughout the M4 project. This annotation is done internally by pre- and post-graduate students at the Department of Computer Graphics and Multimedia of FIT VUT. Orthographic speech transcriptions are done using ETCA "Transcriber" with MultiChannel extension (developed at ICSI for the Meeting Recorder project). Annotators were selected among students (a test segment of a meeting was distributed and the results were evaluated). The quality of speech transcriptions will be assessed in quality check cycles developed during the creation of large Czech corpora collections [3].

## References

- [1] Svoboda, T., Central Panoramic Cameras Design, Geometry, Egomotion, Center for Machine Perception, Faculty of Electrical Engineering, Czech Technical University, September 30, 1999.
- [2] Nayar, S., Baker, S., A Theory of Catadioptric Image Formation, Department of Computer Science, Columbia university, Technical Report CUCS-015-97.
- [3] Pollák P., Černocký J., Orthographic and phonetic annotation of very large Czech corpora with quality assessment, accepted to Language Resources and Evaluation Conference (LREC), May 2004, Lisbon, Portugal.