Detection and Identification of Rare Audiovisual Cues

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than things you do expect) Plautus (ca 200 (B.C.))

Project no: 027787

DIRAC

Detection and Identification of Rare Audio-visual Cues

Integrated Project
IST – Priority 2

DELIVERABLE NO: D6.8
Database of Incongruent Human Locomotion

Date of deliverable: 31.12.2009
Actual submission date: 09.02.2010

Start date of project: 01.01.2006 Duration: 60 months

Organization name of lead contractor for this deliverable: FRA

Revision [0]

Project co-funded by the European Commission within the Sixth Framework Program (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

DIRAC
Detection and Identification of Rare Audiovisual Cues

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than things you do expect) Plautus (ca 200 (B.C.))

D6.8 – Database of Incongruent Human Locomotion

Fraunhofer Institute of Digital Media Technology,
Project Group Hearing, Speech and Audio Technology
(FRA)

Abstract:
One of the objectives of WP6 is "recording audio, visual, and audio-visual databases that would support DIRAC research thrusts".
For this purpose the partners have defined two application domains: 1. the security and surveillance domain, 2. the in-home monitoring of elderly people. For both application domains scenarios have been developed to demonstrate the application of the methods developed in DIRAC.
A detector model capable of detecting incongruencies in human locomotion has been developed within WP3.
The purpose of this deliverable is to aggregate recordings containing examples for incongruent human locomotion. These recordings have been requested by WP3 to be used for testing the so-called articulated tracker. The recordings contain both outdoor and indoor scenes. All recordings have been delivered to the partners in WP3.

DIRAC
Detection and Identification of Rare Audiovisual Cues

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than
things you do expect) Plautus (ca 200 (B.C.))

Table of Content

1. Introduction

In the Technical Annex submitted at 12th of February 2009, two application scenarios were defined for the DIRAC technology, namely the security market with its high demand for automated and intelligent surveillance systems (scenario 1), and the in-home care market with its need for monitoring elderly people at home (scenario 2). It was concluded that both domains would benefit considerably from the technology developed in the DIRAC project.

For both application domains, scenarios have been developed to show the potential of the DIRAC theoretical framework and the techniques developed in the various work packages, while attempting to address realistic and interesting situations that can not be handled properly by existing technology.

One of the detectors developed in WP3 for detecting incongruent events is the so-called articulated tracker, a detector model capable of detecting incongruencies in human locomotion.

The purpose of this deliverable is to aggregate additional recordings containing examples for incongruent human locomotion. These recordings have been requested by WP3 to be used for testing the articulated tracker. The recordings contain both outdoor and indoor scenes and utilize the AWEAR-II recording platform.

2. Recording Platform Hardware and Software

For the recordings in this deliverable, the AWEAR-II mobile recording platform has been used. The platform has been designed by the partners in DIRAC to increase portability and usability in comparison to the AWEAR-I platform.

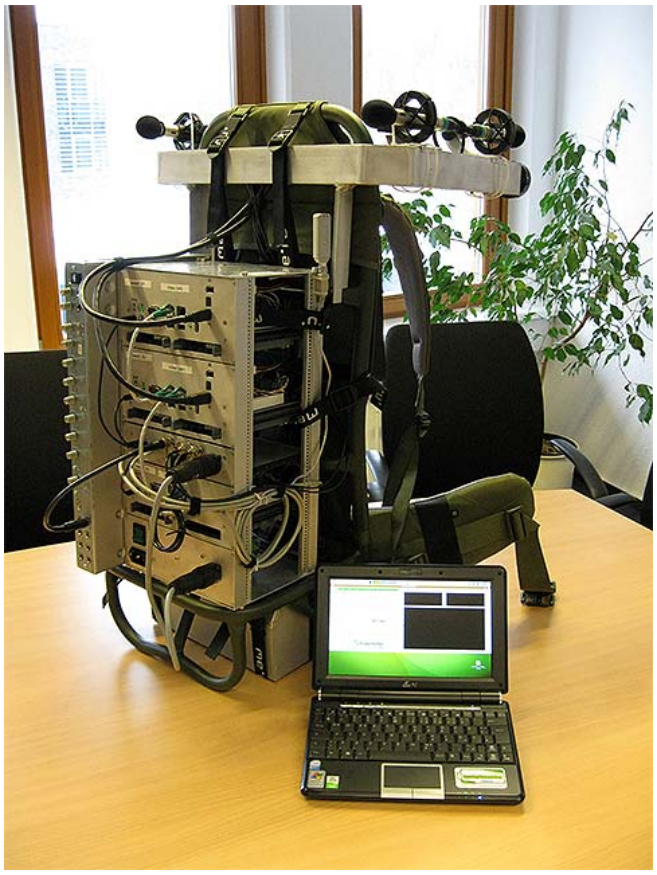

Figure 1. AWEAR-II platform: record cradle and remote control net book.

2.1 AWEAR-II platform

The AWEAR-II recording platform consists of 3 mini-atx personal computers, 2 FireWire cameras, 4 microphones, 1 FireWire audio capturing device, additional trigger electronics, a battery pack and a power distribution box. The hardware is

mounted on a wearable backpack frame. An additional net book is used as remote-control-pc for both systems (see Figure 1).
A multitude of custom parts, including the triggering system, were developed by KUL, who also handled construction and hardware-testing of the AWEAR-II. Weighing a total of ~20 kgs the platform can be carried around. Its battery capacity allows for ~3 hrs of autonomous use.
FRA has two AWEAR-II systems at its disposal, one is equipped with batteries only, and the other has an additional AC current supply, which allows for unlimited processing/backup time.

2.2. Recording software

FRA developed a graphical user interface (GUI) application to simplify the procedures necessary to enable the record mode on an AWEAR-II device. The GUI automates all steps necessary to start and stop the recording mode, and thus provides the user with a one-button application. The GUI application runs on the net book pc and is used for preparation of recordings, for hardware adjustments (camera settings); to receive test pictures from the cameras, and to display performance indicators, e.g. next trigger signal or wireless connection strength. The GUI enables the user to easily set the video-parameter for every camera at once (parameters: white, shutter, gain and gamma). Preset values are also available for common, recurrent conditions. There is the feasibility to take single pictures with the actual settings preliminary to the real recordings, to test the settings on the actual conditions.

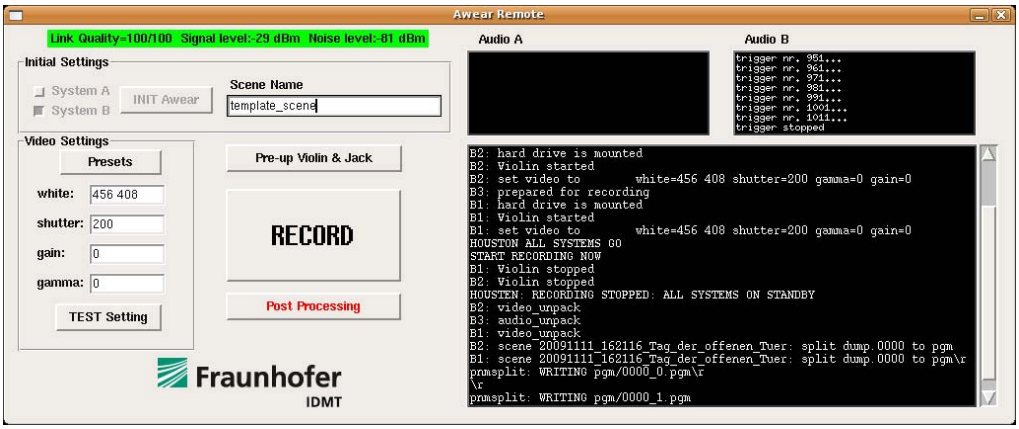

Figure 2. AWEAR-II remote control graphical user interface (displaying running post processing after record session).

The easiest way to start recording with the GUI is to press the Record button, but manual control of the process remains possible for the advanced user. The GUI checks automatically whether or not all required software parts are running properly. Any malfunction of the underlying hardware/software parts signaled to the user. Additionally, an input level meter for the audio channels is displayed.

2.3. Processing of recorded data

Once a session is recorded, the generated audio and video data generated by the session has to be processed. De-packeting of the video stream and mp3 audio generation (only for trimming of the recordings) is done with the AWEAR-II computers using the (changeable) data hard disks connected to each pc. The employment of all 3 computers significantly speeds up the de-packeting of the video data.
After synchronizing and trimming the data, the data is annotated and preview videos are generated. Annotation data and a preview video are generated to facilitate a quick search of specific incidents, and to get an impression of the content recorded without having to use the full AWEAR-II pre-processing chain. The annotation data consists of:

- a specific name for the recording session,
- date of recording,
- location,
- equipment and technical parameters chosen,
- comments,
- data entries: (start time | end time | keywords | short description).

A special list of keywords gives information like number of persons, walking style, speech/non-speech. These keywords facilitate a quick search over different data entries for specific scenes. Additionally, a preview video with reduced resolution is generated to let the user of the database get an impression of the recording without having first to pre-process the raw video and audio data.

2.4 Distribution of recorded data

Up to this point in the processing of the recorded data, all data is stored on hard disks of a local machine. In order to disseminate the material, a web page in the DIRAC Wiki has to be generated, and the recorded and processed data has to be uploaded to a server accessible by all partners.

For an easy overview of the material, the DIRAC Wiki is used. It is accessible via web browser using a secure internet connection and a login for each registered user. In the Wiki, a web page for each recording session is generated containing the most important information and the preview video, together with a link to the processed data on the data server.
The processed recording data itself has to be uploaded to the data server. On the server, a generic directory tree structure is used to easily localize the data of different recordings sessions (see Figure 3):

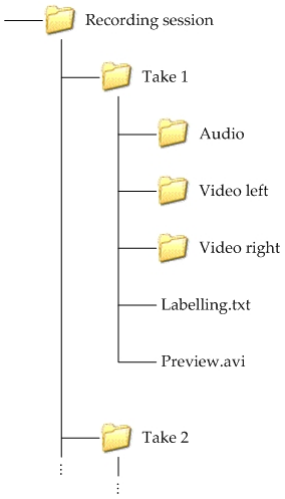

Figure 3. Directory structure template for storing the recordings

3. Recording Sessions

In the following, for each recording session a short description of each scene is given, together with a picture taken from the recorded material. In a table, the date of recording, the name of the recording session (as appearing on the data server and in the DIRAC Wiki), the number of takes and a short description is collected.

3.1 Recording session on Feb 5, 2009

The recording date for this session was Thursday, 5th February 2009. We have chosen the parking area of the department store "Bruno Kleine" nearby the House of Hearing in Oldenburg as location. Two AWEAR-II systems were used, one rack was fixed positioned on a table (rack B) and the other was carried (rack A). Only pictures of right cameras are available, with a frame rate of 12 fps. Five subjects and four variations (walking, hesitating, falling and running) have been recorded.

Figure 4. Person suddenly stopping

date	name	#	short description
2009-02-05	Walking Bruno	32	Five different actors walking, hesitating, falling, running

3.2 Recording session on June 11, 2009

On June 11, 2009, an indoor recording session has been done with the AWEAR-II. The location has been a corridor inside the House of Hearing in Oldenburg. All recording have been done with a frame rate of 14 fps.
In all eight takes of the recorded scene, two persons are walking towards each other and act in different usual or unusual manner.

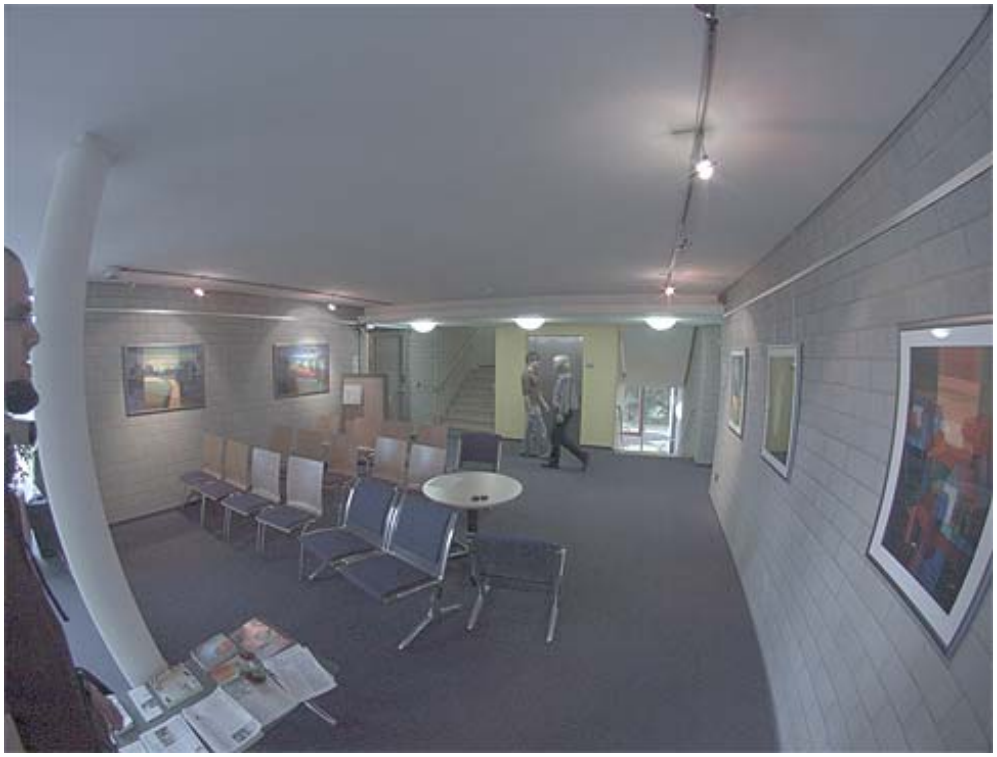

Figure 8. Two persons walking by, greeting

date	name	#	short description
2009-06-11	WalkingAnteroom	08	Guys walking towards on each other, passing by or greeting

3.3 Recording session on Dec 8, 2009

This recording session has been made indoor on a corridor just before the FRA office in front of a slightly grey wall for better contrast. The data has been recorded with the AWEAR-II platform in stationary use. Three different scenes have been recorded; each in different variations, 23 takes in total.
In the first scenario, different walking styles have been recorded. Two different actors walk, run, limp, stumble, flee and fall. Each action is repeated two times.
In the second scene, three different patterns were recorded: the actors walk towards each other, greet each other and pass by, or don't greet each other and pass by, or greet each other and walk backwards. Each of theses scenes has been recorded under the following four conditions: complete silence, one person talking, both persons talking and cross talking (active speaker switches in the middle of the each scene).

Figure 14. A pedestrian suddenly stopping

Figure 15. Two pedestrians greeting each other

date	**name**	#	**short description**
2009-21-08	WalkingStyles	15	Two persons walking in different variations
2009-21-08	SpeakerIdentification	08	Two persons walking towards on each other, passing by or greeting

4. Conclusion

The purpose of this deliverable has been to provide material which can be used by WP3 to evaluate the articulated tracker model, a detector model to detect incongruencies in human locomotion.
In three recording sessions, four scenes with a total number of 63 recorded takes have been recorded, processed and distributed to the partners in WP3 via upload to the DIRAC data server and Wiki.