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: D6.9
Database of Recordings of Scenario 1

Date of deliverable: 31.12.2009
Actual submission date: 04.02.2010

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.))

D6.9 – DATABASE OF RECORDINGS OF SCENARIO 1

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.
The purpose of this deliverable is to aggregate recordings containing examples for typical situations correlating with the scenario 1 "security and surveillance". The recordings have all been carried out with the AWEAR-II recording platform which has been developed inside DIRAC, too. All recordings have been documented and made available for the project partners using the DIRAC server.
Recordings made for the scenario 2 "in-home care of elderly people" have been delivered to the partners in another deliverable D6.10 and thus will not be considered here.

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.

In both domains there is a need for 24/7, unobtrusive, autonomous, and therefore intelligent, monitoring systems to assist human observers. The use of sound to augment camera surveillance has recently been introduced in the security market (van Hengel & Andringa, 2007) and is planned in in-home monitoring (van Hengel & Anemüller, 2009). Spotting and properly responding to unforeseen situations and events is one of the crucial aspects of monitoring systems in both application domains.

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. To make these developed methods and techniques in DIRAC applicable in both scenarios, and to learn about their capabilities and restrictions, it was decided to record example situations for both scenarios, to build a separate database for each scenario and to have the task of recording each database formulated in two deliverables: D6.9 for scenario 1 (lead contractor: FRA, this deliverable) and D6.10 for scenario 2 (lead contractor: OHSU). Therefore, in what follows, only recordings for the scenario 1 are considered.

Prior to recording example situations, the content and preliminaries of each scene to record had do be discussed with the partners involved in order to match the specific needs of the detectors and to plan the recording setup, e.g. location, background, camera distance, people involved, plot. All relevant partners were involved by e-mail or phone conference to develop a so-called storyboard for each scene to be recorded. The recordings have been primarily audio-visual of outdoor scenes, where the use of the AWEAR II mobile recording device provides the required flexibility. After recording a scene, it was pre-processed on the AWEAR-II platform, video and audio were synchronized, the material was uploaded on the DIRAC server and all partners were asked to use it and to give comments on it.

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 netbook.

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 netbook is used as remote-control-pc for both systems (see Figure 1). The operating system on both the mini-atx

pc's and the netbook is the ubuntu linux. A list of the specific hardware used for the AWEAR-II platform is given in Appendix A.
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 has been programmed in TCL/TK, a scripting based programming language for controlling text-based programs. In combination with the Expect extension of the programming language, it was possible to applications via several (network) connection protocols, including SSH.
The GUI application runs on the netbook 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.
To prepare a recording session, the user chooses a name for the scene, and both the information and data belonging to this scene will be automatically named on the different recording computers of the AWEAR-II platform. The scene name also includes a time stamp to clearly define the recordings.

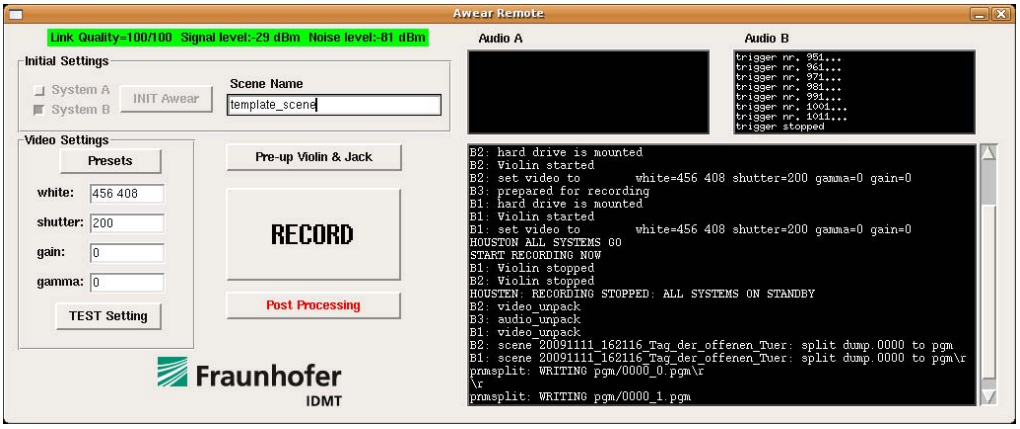


Figure 2. AWEAR-II remote control graphical user interface (displaying running post processing after record session).

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.

For synchronization and trimming of the recorded audio and video data, the data disks are plugged off the AWEAR-II system and connected to a separate pc via eSATA hard disk cradles. Thus, no extra time for copying the data is needed (a spare set of hard disk allows for continuous use of the AWEAR-II system). The synchronization utilizes the hardware synchrony impulse (recorded on audio track 5) to map the sequence of video frames to the corresponding audio tracks. Additional software developed by FRA is used for trimming begin and end of the recorded data, splitting the recording is also possible.

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. An example of such an annotation file is given in Appendix C.
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):

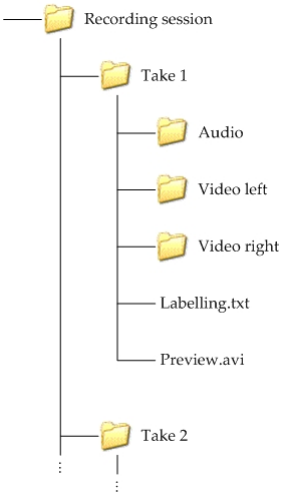


Figure 3. Directory structure template for storing the recordings

3 Recording Sessions

In the following, for each recording session a short description of the different scenes is given, together with a picture taken from the recorded material. In a table, the date of recording, the name of the recording session, the number of takes and a short description is collected.

3.1 Recording session on 18 March 2009

Recordings made on Wednesday, 18th March 2009, with three different scenarios. Location is in front of the House of Hearing on the parking area. Some scenes are made on the nearby pavement and side road. The AWEAR-II was used with a frame rate of 14 fps as a moving platform carried along the pavement. Several subjects are crossing the wearer and one of them stumbles on the street.

Figure 4. Person stumbles

date	name	#	short description
2009-03-18	PedestrianFalling	01	A Pedestrian falls with two other pedestrians passing by

3.2 Recording session on April 4, 2009

This session took place on Saturday, 4th April 2009. Two different scenes been recorded near the House of Hearing. The AWEAR-II was used both moving and fixed, with a frame rate of 14 fps.
The first shoot contains a long walk form the House of Hearing to the Wechloy campus of the University of Oldenburg. In the second shoot a pedestrian is falling in front of the cameras and several other peoples are crossing by.

Figure 5. Long walk

Figure 6. Person falling, pedestrians walking by.

date	**name**	#	**short description**
2009-04-04	LongWalkGPS	01	Long walk with a GPS-transmitter from the HSA to university
2009-04-04	PedestrianFallingII	01	A Pedestrian falls with two other pedestrians passing by

3.3 Recording session on April 6, 2009

The Session date is Monday, 6th April 2009. Two scenes have been recorded with the AWEAR-II at a frame rate of 14 fps. As location a street and parking lot nearby the House of Hearing has been chosen.
The first recording contains cars driving past and pedestrians crossing the screen. In the second recording, a camera failure is simulated by holding a cloudy foil in front of a camera lens.

Figure 7. Pedestrians and cars passing by

date	name	#	short description
2009-04-06	TrafficScene	01	Recording of several cars and pedestrians passing by
2009-04-06	PedestriansCamCover	01	One camera is temporarily covered while meeting pedestrians

3.4 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.

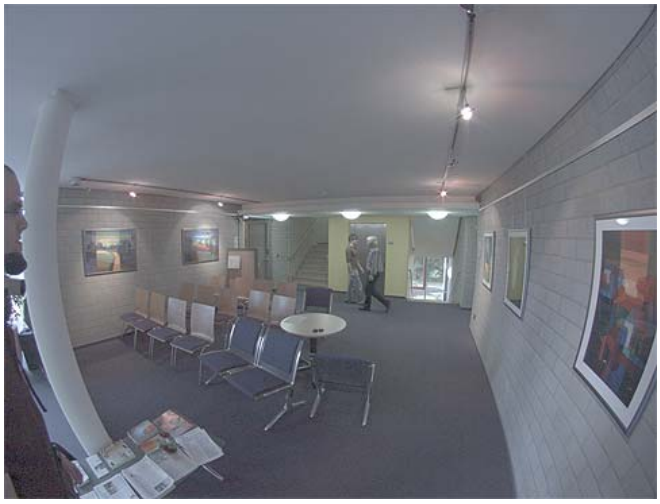


Figure 8. Two persons walking by, greeting

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

3.5 Recording session on July 14, 2009

This outside recording was made on July 14, 2009 at a road crossing in front of the House of Hearing. In order to have each scene recorded with both a body-worn and a fixed AWEAR-II two systems have been used simultaneously. The basic concept for this recording has been to separate dangerous traffic situations in everyday life from those who are innocuous or generally uncommon. Thus, three versions of a comparable setting were recorded: a "danger", a "no danger" and an "extraordinary behaviour" scene. As the scenes were quite complex, they were recorded in short blocs before the complete scene was videotaped. For this reason, 32 takes have been recorded in total.

In the "danger"-version a pedestrian approaches the AWEAR-II, stops and starts a conversation. Another person (outside the picture) shouts a warning and the vehicle (car or bicycle) approaches the AWEAR-II. A pedestrian comes close and screams.
In the "no danger"-version a pedestrian approaches the AWEAR, stops and starts a conversation. Another person (outside the picture) greets. A vehicle passes by and a pedestrian jogs towards the AWEAR-II.

In the "extraordinary"-version a pedestrian approaches the AWEAR, stops and starts a conversation, but now uses many OOV words. A second person (outside the picture) greets using OOV words. A vehicle drives by, and a pedestrian walks backward.

A detailed script for this recording including the walking routes can be found in Annex B. As the video material shoed a green cast (probably due to wrong camera settings) use of the video data was restricted. The scene has been redone on July 28, 2009.

Figure 9. Picture taken at the set for "collision course"-scene (danger version with a car approaching the AWEAR)

date	name	#	short description
2009-07-14	CollisionCourse	32	Three different versions of an almost crash with a car/bicycle

3.6 Recording session on July 28, 2009

This recording session is a remake/modification of the “collision course”-scenario from July 14, 2009. The location has been changed to a parking lot next to the University of Oldenburg. Again, each scene was recorded with two AWEAR-II systems, one moving (body-worn) and the other stationary (on a small table).
Minor adaptations to the plot of the scenes have been made, compared to the Session n July 14, 2009: the routes of the acting persons have been fitted to the new location, and to vehicles (a car and a bicycle) have are used simultaneously in one scene.

Figure 10. Image taken from set of the scene "collision course" (danger version with a car approaching the AWEAR and a pedestrian warning)

date	name	#	short description
2009-07-28	CollisionCourseRem	16	Remake/modification of the scene CollisionCourse

3.7 Recording session during summer school August 26, 2009

Two scenes were recorded from PhD students during the summer school in Leuven on August 26, 2009. For both recordings, the AWEAR-II platform was used (fixed, standing on a small table). Frame rate was 14fps.
The first scene has been recorded on a busy street in Leuven using an additional headset microphone to pick up an improvised monologue of one person in the scene to facilitate the out of vocabulary detection processing.
The second scene has been recorded on the same busy street in Leuven with different people walking by, running, falling over, cars and bicycles passing.

Figure 11. Interview scene

Figure 12. Busy street scene

date	**name**	**#**	**short description**
2009-08-26	Interview	01	Recording of improvised interview with oov words, busy city
2009-08-26	MovingObjects	01	People (walking, falling, passing), bicycles and cars passing

3.8 Recording session on Sept 10, 2009

These outdoor recordings were taken with a moving AWEAR-II device in the inner city of Oldenburg from the "Lange Strasse" over the "Lappan" and back over the "Achternstrasse" (see Figure 13 on the right).

This recording does not contain any plot performed on purpose by any actor, but is a recording of a typical scene in a pedestrian zone. Besides the usual pedestrians and bicyclists passing by, a police car with a siren passing by has been recorded. The recorded scene has a length of about 40 minutes and was therefore cut into six separate takes.

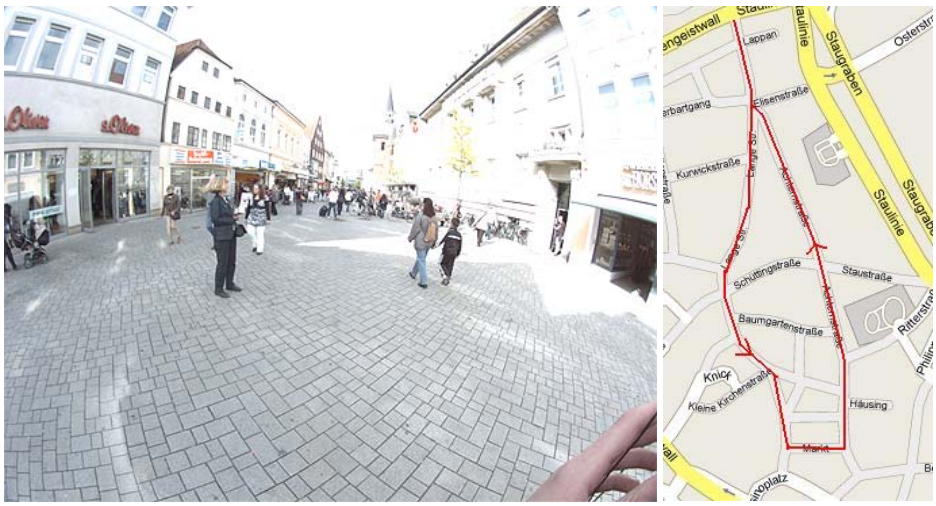


Figure 13. Image taken from the Oldenburg City Sequence (right: map of Oldenburg with marked route)

date	name	#	short description
2009-09-10	OldenburgCity	06	Recordings of the inner city of Oldenburg

3.9 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, 25 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).

The third scene implies an attempted bag robbery. In a first version, a pedestrian carries a bag, and a thief approaches. The thief snatches the bag away from the man and escapes with it. In a second version, the thief approaches the pedestrian from behind. In a third version, the pedestrian and the thief walk towards each other. The thief tries to steal the bag again, but is put to flight by the pedestrian.

Figure 14. A pedestrian suddenly stopping

Figure 15. Two pedestrians greeting each other

date	name	#	short description
2009-21-08	WalkingStyles	15	Two guys walking in different variations
2009-21-08	SpeakerIdentification	08	Guys walking towards on each other, passing by or greeting
2009-21-08	StealingBag	02	Three different versions of a bag-robbery

3.10 Recording session on Dec 16, 2009

This outside recording session was made in front of a house next to the House of Hearing in Oldenburg. The data was recorded with the AWEAR-II platform in stationary use. Two different scenes have been recorded, each in three different variations and two different constellations of actors resulting in12 takes.
The first scene implies an attempted book robbery. In a first version, a pedestrian carrying a book and the thief walk towards each other. The thief snatches the book away from the pedestrian and escapes with it. In a second version, the thief approaches the pedestrian from behind. In a third version, the pedestrian and the thief walk towards each other. The thief tries to steal the book, but is put to flight by the pedestrian.

The second scene of this recording deals with different variations of aggressions between two pedestrians. In a first version, two persons hustle each other, roar at each other and pass by. In a second version, one person pushes the other back roaring. The other one acts defensively and passes by after the aggressor lets him go. In a third version, person 1 (aggressive) suddenly hits person 2 in the stomach, person 2 breaks down and limps away.

Figure 16. Image taken from the recording session with two pedestrians hustling each other

date	**name**	#	**short description**
2009-12-16	StealingBook	06	Three different versions of a book-robbery
2009-08-26	AggressionScenes	07	Two guys hustle each other roaring (different variations)

4 Conclusion

The purpose of this deliverable has been to provide material which can be used by the DIRAC partners to test the techniques developed in the other work packages on realistic and interesting situations related to the security scenario (named "Scenario 1"). With the 16 different scenes recorded in 10 sessions with an overall number of 112 recorded takes, this deliverable indeed accumulated a wide variety of different situations for evaluation. Intensive use has been made of the AWEAR-II recording platform which had been developed in the DIRAC project. All recorded data has been presented to the partners via a database on a file server together with a separate description reachable through secure http access on a Wiki hosted on the DIRAC server.
The evaluation of the database by the DIRAC partners showed the potential of the detectors and principles developed so far in DIRAC, but nevertheless clearly indicated that there are many circumstances under which the application could fail: changing recording conditions (light to low/high, shadows, external sound sources) and scene complexity (to many actors, action too quick/complex) have been the main reason for detectors not to handle the recorded material properly.

5 References

Jörn Anemüller, Jörg-Hendrik Bach, Barbara Caputo, Michal Havlena, Luo Jie, Hendrik Kayser, Bastian Leibe, Petr Motlicek, Tomas Pajdla, Misha Pavel, Aki Torii, Luc Van Gool, Alon Zweig, Hynek Hermansky, "The DIRAC AWEAR Audio-Visual Platform for Detection of Unexpected and Incongruent Events", Proc. International Conference on Multimodal Interaction (ICMI) 2008, pp. 289-293.

Peter van Hengel and Tjeerd Andringa. Verbal Aggression detection in complex social environments. Proceedings of the 2007 IEEE International Conference on Advanced Video and Signal based Surveillance. 2007.

Appendix A: AWEAR-II hardware list

List of the components used for the AV recording platform AWEAR-II:

Video:

AVT Stingray camera	2
Fujinon FE 185CO86HA1 fish-eye lens	2

Audio:

T-bone EM700 stereo mic set	2 (4 mics in total)
FOCUSRITE Saffire PRO 10	1
Sennheiser EH 350 headphones	1

PCs:

Siemens D2703-S mini ITX boards	3 (2 video, one audio & control)
2.5” Hard disks, 320 GB each	6 (3 operating system, 3 data)

Frame:

Tatonka Lastenkraxe backpack	1
Camden 12V gel batteries	4

Appendix B: Storyboards

scenario: collision_course_1	
location:	date:
required material: AWEAR*2 + Laptop, bycicle	
required persons: Steffen (+S), AWEAR-Wearer (A_B), driver (D), 3 pedestrians (P_1, P_2, P_3)	

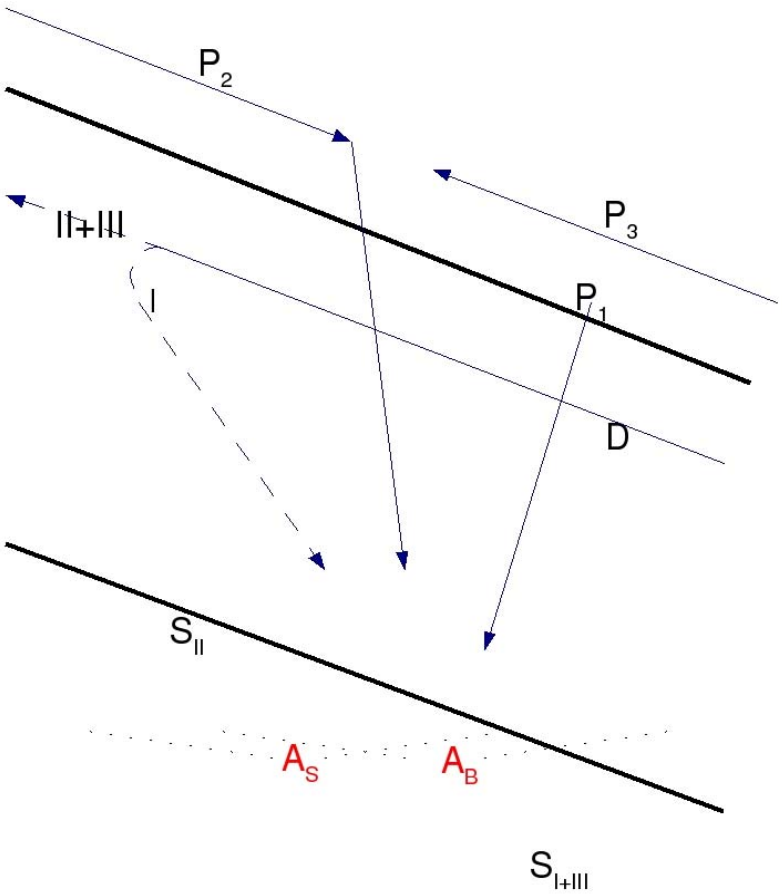
General comments
dialogs in english, loud and clear
arms held close to the body
A_B with sometimes more, sometimes less movement

Take	a_interlocutor	b_pedestrians	c_driver	d_all
Participants	P_1 and S	P_2 and P_3	D	all together

Session1: danger:
P_1 approaches AWEAR (A_B)
P_1 stops about 2 meters before AWEAR and starts conversation
P_2 and P_3 walk towards each other und stall face to face for greeting
D drives by bicycle on straight line
S shouts (outside the AWEAR-visual range) a warning / P_2 and P_3 look frightened to shouter (respectively driver)
D drives a curve and is heading for a collision with A_B
P_2 warns A of danger and runs ahead the AWEAR screaming
Session2: nodanger:
P_1 approaches AWEAR (A_B)
P_1 stops about 2 meters before AWEAR and starts conversation
P_2 and P_3 jog towards each other und stall face to face for greeting
D drives by bicycle on straight line
S shouts (inside the AWEAR-visual range) a greeting
D passes by, still on straight line
P_2 goes on jogging towards A_B and passes by

Session3: extraordinary behaviour:
P_1 approaches AWEAR (A_B)
P_1 stops about 2 meters before AWEAR and starts conversation (including many OOV words)
P_2 and P_3 walk towards each other und stall face to face for greeting
D drives by bicycle a winding course
After passing the bycicle, P_2 and P_3 go straight on and pass each other
S shouts (outside the AWEAR-visual range) a greeting (using OOV word)

P_2 and P_3 walk towards each other, P_3 backwards though
P_2 spins around and goes on walking towards the AWEAR (P_3 still goes backwards)



scenario: collision_course_2	
location:	date:

Required material:	AWEAR*2 + Laptop, car, (remote-controlled airplane)
Required persons:	Steffen (+S), AWEAR-wearer (A_B), driver (D), 3 pedestrians (P_1, P_2, P_3), disposer of the airplane (F)

General comments
dialogs in english, loud and clear
arms held close to the body
A_B with sometimes more, sometimes less movement

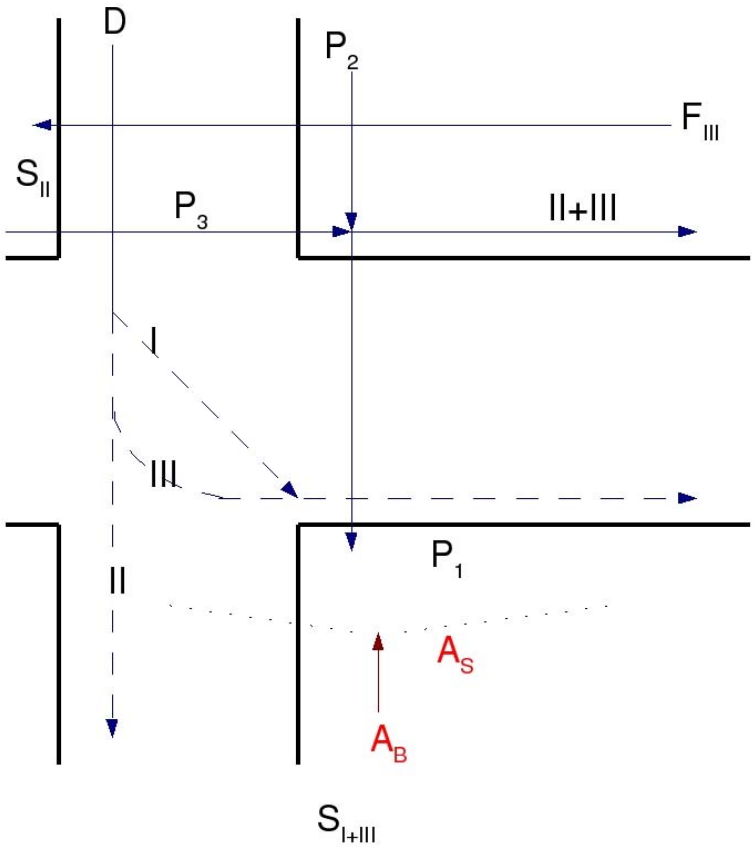
Take	a_interlocutor	b_pedestrians	c_driver	d_all
Participants	P_1 and S	P_2 and P_3	D	all together

Session1: danger:
P_1 (about 2 meters before AWEAR) starts conversation AWEAR (A_B)
P_2 and P_3 walk towards each other und stall face to face for greeting
D drives by car on straight line
S shouts (outside the AWEAR-visual range) a warning / P_2 and P_3 look frightened to shouter (respectively driver)
D drives a curve and is heading for a collision with A_B
P_2 warns A of danger and runs ahead the AWEAR screaming

Session2: nodanger:
P_1 (about 2 meters before AWEAR) starts conversation AWEAR (A_B)
P_2 (jogging!) and P_3 walk towards each other und stall face to face for greeting
D drives by car on straight line
S shouts (inside the AWEAR-visual range) a greeting
D passes by, still on straight line
P_2 goes on jogging towards A_B and passes by / P_3 goes straight forward

Session3: extraordinary behaviour:
P_1 (about 2 meters before AWEAR) starts conversation AWEAR (A_B) (including many OOV words)
The airplane (F) passes by in the background
P_2 (limping!) and P_3 walk towards each other und stall face to face for greeting
D drives by car a winding course
S shouts (outside the AWEAR-visual range) a greeting (using OOV word)
D drives a curve und along the AWEAR
P_2 goes on limping towards A_B and passes by / P_3 goes straight forward

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.))



Appendix C: Annotation example

Example for the annotation data of a recorded scene:

Recording: AggressionScenes
Scene : 0_AggressionScenes_verC_RobertStephan
Date : 2009-12-16
Location : house front next to HdH (FRA Oldenburg)
Equipment: AWEAR 2.0b (fixed)
Framerate: 12 fps
Comments : shades visible

start time (in sec) | end time (in sec) | key words | short description;
0000 | 0001 | persons_2,walking | Two persons walk towards each other;
0002 | 0003 | persons_2,walking,speech | The defensive person begins conversation;
0003 | 0005 | persons_2,persons_interact,shouting | The aggressor suddenly hits the other one in the stomach;
0005 | 0010 | persons_2,falling,fleeing,shouting | Person breaks down, aggressor flees;
0010 | 0017 | persons_1,limping | Person limps away;

Appendix D: List of keywords used for annotation

List of key words used for labeling annotation data:

walking styles:
- standing
- lying
- running
- limping
- stumbling
- arresting
- fleeing
- falling
- backwards

objects (video):
- persons_1
- persons_2
- persons_3
- persons_N
- car
- bike

audio signals (audio):
- speech
- shouting
- noise
- loudspeaker
- asig_iv
- asig_ov
- oov

Other:
- dropout
- persons_interact
- overlapping