Information Society Technology

Perception On Purpose

POP

Specific Targeted Research Project

Call identifier: FP6-2005-IST-4

Date of preparation: March 21, 2005

Coordinator:
Radu Horaud
INRIA Rhône-Alpes
Radu.Horaud@inrialpes.fr
fax: +33 476 615 454

	Coordinator:	
1	Institut National de Recherche en Informatique et Automatique (France)	INRIA
	Partners:	
2	University of Osnabrück (Germany)	UOS
3	University Hospital Hamburg-Eppendorf (Germany)	UKE
4	Fac de Ciencias et Tec. Univ Coimbra – ISR (Portugal)	FCTUC-ISR
5	The University of Sheffield (United Kingdom)	USFD

Contents

List of Tables

List of Figures

Proposal summary page

Proposal full title: Perception On Purpose

Proposal acronym: POP

Strategic objective addressed: 2.4.8 Cognitive Systems

Proposal abstract: The ease with which we make sense of our environment belies the complex processing required to convert sensory signals into meaningful cognitive descriptions. Computational approaches have so far made little impact on this fundamental problem. Visual and auditory processes have typically been studied independently, yet it is clear that the two senses provide complementary information which can help a system to respond robustly in challenging conditions. In addition, most algorithmic approaches adopt the perspective of a static observer or listener, ignoring all the benefits of interaction with the environment. This project proposes the development of a fundamentally new approach, perception on purpose, which is based on 5 principles. First, visual and auditory information should be integrated in both space and time. Second, active exploration of the environment is required to improve the audiovisual signal-to-noise ratio. Third, the enormous potential sensory requirements of the entire input array should be rendered manageable by multimodal models of attentional processes. Fourth, bottom-up perception should be stabilized by top-down cognitive function and lead to purposeful action. Finally, all parts of the system should be underpinned by rigorous mathematical theory, from physical models of low-level binocular and binaural sensory processing to trainable probabilistic models of audiovisual scenes. These ideas will be put into practice through behavioural and neuroimaging studies as well as in the construction of testable computational models. A demonstrator platform consisting of a mobile audiovisual head will be developed and its behaviour evaluated in a range of application scenarios. Project participants represent leading institutions with the expertise in computational, behavioural and cognitive neuroscientific aspects of vision and hearing needed both to carry out the POP manifesto and to contribute to the training of a new community of scientists.

1 Scientific and technological objectives of the project and state of the art

1.1 Objectives, approach, and original contributions of the proposal

Whenever researchers attempt to investigate and build an intelligent system, they are confronted with the problem of modelling the interactions between the system and its physical environment. The world that surrounds artificial and/or biological agents has three dimensions and its geometrical/physical structure varies over time. Therefore it is not surprising that one of the primarily goals of perception is to create an internal spatio-temporal representation of the external environment. One reason why both the spatial and temporal dimensions of sensory processing are so important is because internal representations of space and time are needed to guide behaviour such as movements and therefore they allow the system to interact with the physical world.

Some of the major difficulties stem from the fact that the sensorial stimuli have different physical formats (visual, auditory, tactile, etc.) and are too vast, too complex, and too ambiguous to be analyzed in a simple manner and properly interpreted. Among the possible sensorial modalities we choose to concentrate on vision and audition. The visual stimuli are two-dimensional and they are formed by beams of light that reflect on surfaces before reaching the retina, and therefore the visual data do not directly encode the geometry, the size, the shape, the structure, and the meaning of the objects in the world.

Similarly, due to the additivity of acoustic sources, the auditory system is faced with a complex mixture of both direct and reverberant reflections together with contributions from other sound sources present. For each individual source, location is not encoded in a simple way, and for sources such as talkers, information-bearing properties such as pitch, vocal tract length, rhythm and phonetic content are overlaid in the waveform.

In the recent past, visual and auditory processes have been studied almost independently. Cross-modal integration of visual and auditory data and interaction between auditory and visual processes are beneficial because each modality provides partial information about different aspects of world objects and events, and combination of these modalities can be very important in understanding how they complement each other to provide unambiguous world information, how they transform their inputs into knowledge and meaning, and how they control behavior.

> *The objective of this project is to put forward the modelling of perception (visual and auditory) as a complex attentional mechanism that embodies a* ***decision taking*** *process. The task of the latter is to find a trade-off between the reliability of the sensorial stimuli and the plausibility of prior knowledge.*

For example, the system should be able to localize in space, identify, and track over time an object that can be seen and heard simultaneously, i.e., crossmodal integration. From the point of view of auditory analysis, **binaural cue detection** may be used for sound localization and for building a spatial map of auditory sources. From the point of view of visual analysis, visual cues such as **stereoscopic vision** may be used for producing an observer-centered depth map of the 3-D layout. However, crossmodal integration cannot be performed at that level because there is no obvious way to associate depth and sound sources. Temporal processing of both sound and vision enhance these spatial descriptions. Sequential processing of audio data may **group events into streams** and produce sets of distinct sounds perceived as arising from a single source. Similarly, **optical flow** may be extracted from image sequences and the visual data can be segmented into static and dynamic, into fast and slow motions, and into rigid and deformable motions. Therefore we have additional audio and visual descriptors, yet this is not sufficient to associate audio and visual events. Crossmodal integration, followed by decision and action, necessitates a higher level association process, beyond the physical and geometrical formats that describe sound and light stimuli. We firmly believe that the perception-action cycle needs some form of cognitive modelling.

The integration of visual and auditory cues with control of behavior will be one of our central topics of investigation. *In particular, we believe that the interactions between the sensory level processes (bottom-up) and the cognitive level processes (top-down) reside at the core of a cognitive system and these interactions distinguish the latter from a classical computer system.*

The approach that will be advocated in this project and implemented within its associated work-packages relies on the following principles:

- **Auditory and visual cues must be integrated both in space and time**. We will encapsulate various perceptual processes for enabling the fusion of the processed data. We will develop theoretical and algorithmic models based on image and signal processing, on geometric fusion, on probabilistic modelling of both the input and the processed data, and on statistical modelling.
- **Physical inputs must be confronted with prior knowledge**. This data-to-meaning association problem will be thoroughly addressed and its computational complexity will be modelled and investigated. Probabilistic models (Bayesian decision theory and hidden Markov models, in particular) will be combined with efficient optimization techniques.
- **Visual sensors must actively explore the world.** The field of view of a visual sensor is inherently limited. Therefore a static sensor cannot build a complete and reliable representation of the world. Furthermore, when the attention of an observer must focus onto a specific region, the narrowness of the field becomes a desirable feature. Therefore, active vision plays a crucial role at all levels of information processing: there is a strong link between attention and sensory-motor control. We will thoroughly investigate these issues from the view point of control theory and in synergy with neurophysiology and psychophysics.
- **Computational models for active listening will be thoroughly investigated.** The freedom to orient and move the location of the auditory sensors has tremendous potential for improving the signal-to-noise ratio of the attended sound source. Orientation allows for directional listening and blocking of unwanted sources by head-shadowing, while movement enables the listener to avoid hard reflective surfaces, move closer to the target, and get an unobstructed view of information-bearing features such as lip movements.
- **We will build a proof-of-concept robotic platform** which will include the following elements: the design of an active audiovisual head mounted onto a mobile robot, algorithms and software that implement visual and audio cue extraction, cross-modal integration, end sensory-motor decisional loops.
- **Biological attention mechanisms will be quantitatively studied** using the most modern eye-tracking devices in conjunction with advanced brain imaging techniques. In particular we will study the relation between cross-modal stimulus integration and attentional selection at a behavioral level. Key areas in the brain where crossmodal integration occurs must be identified. These experiments will guide the implementation of the active audio-visual robotic head.
- **Neurophysiological and psychophysical findings will be cast into an abstract computational model that is implementable with today's computers**. The task of modelling

the brain from a computational point of view is tremendous. It is possible to model groups of neurons as a dynamic system using partial derivative equations. But these models cannot yet be generalized to complex brain functions such as those involved in perception. We believe that an intermediate reachable goal is to consider an abstract model based on modern statistics. This type of approach has been successful in many cognitive-system related fields such as artificial intelligence, computer vision, speech recognition, etc.

We believe that this proposal has the following contributions:

- In the past, **computational models for vision and hearing have been studied independently** by distinct research communities with almost no interaction. **We will adopt a multi-disciplinary approach** involving a broad range of methodologies from computational and biological vision, robotics and control, computational auditory scene analysis, speech recognition, psychophysics, and psychology.
- **The simultaneous study of visual and auditory processes and their relationship with attention and motor control** is a novel scientific endeavour. We feel that the time to combine research results from all these disciplines has come.
- More precisely, we will address the difficult problems of **integrating spatial and temporal audio-visual stimuli** using a geometrical and probabilistic framework and we will attack the problem of **associating sensorial descriptions with representations of prior knowledge**. Moreover, we will **design an audio-visual active robotic head** and we will **build models for sensory-motor control.**
- **The fact that computer scientists are trying to build an artificial system that models biological systems made out of neurons is often ignored**, and until recently there was no answer to the question of biological plausibility: how a given algorithm may be implemented in a biological brain? Equally important is the issue of computational neuroscience: how one could implement with hardware and software recent neuro-physiological findings which model the way our brain functions?
- Nevertheless, **the structure of today's computers and their common programming languages is very far removed from the brain's architecture**. Moreover, there is a lack of formalism and of common theoretical and practical methodologies allowing the above mentioned disciplines to work together and to produce common research results. This gap (and its impact onto both computational and biological models) may be reduced in the near future but it is obvious that this cannot be done within a single research project.
- **The contribution of the POP project will be to address a specific research topic and to set up a common set of formal models** (upon which both communities agree), and **to combine a broad range of methodologies** in order to both (i) *implement these models with computers* and (ii) *verify them through the measuring and interpretation of human-brain activity.*

1.2 The state of the art

1.3 Recent European projects

2 Relevance to the objectives of the IST priority

The primarily objective of the POP project is to address a specific research issue, namely the problem of modelling, implementing, and testing an artificial system that gathers sensorial information on purpose – according to the task at hand – and that controls its attention and behaviour. Therefore we will address scientific and technological issues associated with an *artificial cognitive system* that interacts with the physical world. In our specific case these interactions will be materialized through the use of sensors and actuators.

We will consider both visual and auditory sensors and we will address the problems of how these sensorial modalities integrate and complement each other, how they interact with task-dependent higher-level knowledge, and how they are combined with actuators within sensory-motor feedback loops. We firmly believe that the development of systems able to transform perception of sensorial stimuli into meaning is at the core of future cognitive systems. The in depth understanding of the processes allowing to map light and acoustic **physical signals** into **symbolic descriptions** is the keystone for the development of intelligent agents able to communicate with people and to take decisions.

Neuroscience researchers have recognized for a long time that perception is a high-level cognitive function. For example, experimental neurophysiology formally established that as much as half of the human-brain's neural circuitry is concerned with vision. Although we have already accumulated a great deal of knowldege about the brain's anatomy, physiology, and neural mechanisms, this knowledge is not nearly enough to determine analytic equations that describe large systems of neurons. Even if precise knowledge of neural dynamics is available, it would yield only partial understanding of how brain functions. Therefore, direct approaches based on studying the brain **must be augmented** with information processing approaches that attempt to model perception (and intelligence in general) at a more abstract level.

The POP project will therefore contribute to establish a unifying theory of auditory and visual perception that is quantifiable (i.e., mathematical formulation, computer algorithms and software, and thorough experimental validation), that can provide a theory for human perception, and that can be used to validate rigourous neurophysiological and psychophysical experiments.

Let us exemplify and analyse this statement from the viewpoint of the specific research issue that is addressed in this project, i.e., **the perception-action cycle of attention**. When viewing and analyzing complex natural scenes, humans do not process all the sensorial information homogeneously and simultaneously. Attention is rapidly directed to small parts of the scene that are processed in greater detail. *Overt attention* involves *movement* of the sensory organs towards the stimulus of interest, is it visual, auditory, or a combination of both. This attentive behaviour, however, goes well beyond simple stimulus/actuator loops. A decision has to be made concerning which part of the environment should be attended to, and subsequently, appropriate movements of the sensors have to be made (these movements raise difficult inverse-kinematic and dynamic issues that are worth to be studied in their own right). Thus, attention involves a complete perception-decision-action cycle in closed loop with both the internal representations and the environment.

In contradiction with the biological process just described, current computerized visual and auditory scene analysis systems operate in a much more static way. **Snapshot vision has been the major paradigm studied worldwide** and researchers exhaustively investigated geometric, photometric, and statistical methods for recovering world information from a set of fixed cameras. Similarly, **computational hearing has been dominated by speech recognition**, which has typically been studied in the absence of competing sources that characterize everyday listening situations.

Thus we believe that time has come to study and implement computerized perception systems that act like humans. The POP project will fill in an important gap by attempting to reduce the distance that separates current audio/visual processing paradigms, on one side, from higher-level skills, on the other side, such as the ability of an artificial system to take decisions, to interact with the world, and on a longer term, to communicate with people at an abstract level.

The POP partners will be committed to foster interdisciplinary research within a set of work-packages that interleave mathematical and biological approaches to perception.

To conclude, the POP objectives, approach and methodology are fully relevant to the strategic objective "Cognitive systems" of the IST Work Programme 2005-2006, whose objectives are:

> *To develop artificial systems that can interpret data arising from real-world events and processes (mainly in the form of data-streams from sensors of all types and in particular from visual and/or audio sources); acquire situated knowledge of their environment; act, make or suggest decisions and communicate with people on human terms, thereby supporting them in performing complex tasks.*

And with focus on:

> *Focus is on research into ways of endowing artificial systems with high-level cognitive capabilities, typically perception, understanding, learning, knowledge representation and deliberation, thus advancing enabling technologies for scene interpretation, natural language understanding, automated reasoning and problemsolving, robotics and automation, that are relevant for dealing with complex realworld systems. It aims at systems that develop their reasoning, planning and communication faculties through grounding in interactive and collaborative environments, which are part of, or connected to the real world.*

3 Potential impact

The POP project will address a number of **challenging and unsolved scientific topics** and will investigate **critical technological issues** that will allow to transform some of these scientific outcomes into a working hardware/software laboratory prototype. The outcomes of POP will open new possibilities for a **wide range of technologies and their associated applications**. Ambitious **training activities** will be carried out **in cooperation with on-going IST projects and Marie Curie actions** (one integrated project, one network of excellence, one early stage training action and one research training network action) as well as with a network of excellence to be submitted to the IST call 4 (cognitive systems). This planned networking with related FP6 projects and actions will ensure that POP will have an important **European dimension**.

From a scientific point of view, the main challenge of POP is to address the specific problem of modelling attentional mechanisms by integration of three ingredients: interactions between prior knowledge and sensorial data, cross-modal integration of visual and auditory stimuli, and the coordination between sensorial information processing and motor control.

The research will be carried out through cross-fertilization between five disciplines: speech and hearing analysis, computer vision and image understanding, robotics and control theory, computational cognitive neuroscience, and neurophysiology.

Multi-disciplinary research is more difficult to achieve than single-topic research and, if successful, it will have a greater impact onto scientific findings than the sum of independent and isolated investigations, projects, and studies.

In order to support the need of inter-disciplinary research and its associated impact, let us give two examples.

Recent experimental findings in neuroscience (through brain imaging analysis) are in contradiction with the classical paradigm used by computer vision researchers. The classical view refers to two distinct stages that do not interact, a feed-forward stage that is able to process the sensorial stimuli and a feed-back stage that processes the information preferentially. The modern neurobiological and psychophysical model postulates that the human visual system works on the basis that there are massive interactions between top-down and bottom-up pathways: the system checks whether its own internal representations are consistent with the messages captured and sent by the retinal ganglion cells. In other terms, we see *on purpose* and it is quite tempting to try to understand these hypotheses from a thorough computational and algorithmic point of view. Whatever the answer is, if we get an answer, we will significantly advance the state of the art in both disciplines.

Another example is the collaboration between auditory and vision researchers. The domain of *computational auditory scene analysis* aims at building computer models of auditory functions. Likewise, the domain of *computer vision and image understanding* aims at a similar goal of building artificial vision systems. It is interesting to notice that, although the two research communities share a large body of mathematical knowledge, they have barely collaborated in the past. The question of how these two sensorial modalities should cooperate for providing richer perceptual information has not been properly addressed. We are aware of methods trying to combine speech recognition with visual lip reading, and of methods trying to improve the spatial localization of a speaker using visual head tracking. Nevertheless, a general purpose thorough formulation of the problem of cross-modal integration remains to be investigated. As with the previous example, cross-fertilization with neuroscience and psychophysics is crucial and it will be one of the essential impacts of our project. Both neuro-imaging and psychophysical experiments revealed the importance of visual cues for guiding auditory attention. An interesting discovery hypothesizes that there is no auditory detection of motion and therefore the human auditory system relies on visual detection of motion. POP will study and test methods and algorithms that implement these findings.

The POP partners will be committed to build a common technological platform. The latter will basically consist in both hardware and software components. It will gather an audiovisual active head – a stereoscopic camera system and a binaural microphone system – mounted onto a mobile robotic platform, as well as software that will implement both existing algorithms and algorithms to be developed by the POP partners. This platform will be designed by the POP partners. The software libraries and packages will be open source code. We will exploit both our hardware design and our open-source software developments (through patents and software licenses) and we will disseminate them such that they impact onto technology transfer. Moreover, the raw and processed data sets produced by this platform will be disseminated as well in order to allow other researchers to test their own methods and/or for bench-marking purposes.

The POP partners will also be committed to launch an ambitious interdisciplinary training programme. The POP themes and teams will be attractive to talented doctoral and postdoctoral researchers. However, very few will possess the required interdisciplinary skills from the outset. The application of bio-inspired approaches to engineering problems requires the integration of ideas from a number of disciplines, ranging from cognitive science and psychology to computation, mathematics and neurobiology. Few individuals can claim expertise in more than one of these areas, and even within a single discipline such as neurobiology, far too little work crosses traditional boundaries. Auditory and visual processing are typically tackled in isolation by different communities, and hardly any engineering work straddles the boundary between hearing and vision. A solution to the problems which will be tackled in POP will require integration of work in hearing, vision and other modalities at every level of the project, from low-level salience to high-level scene understanding. Consequently, **all project researchers will be given the opportunity for cross-training in disciplines in which project partners have expertise**. They will be encouraged to develop the technical and collaborative skills needed to solve complex problems. The training impact of POP will be amplified through the European networking that is planned (see below).

Therefore we foresee that the inter-disciplinary synergies developed within the POP project will foster the emergence of new and original scientific knowledge and will considerably advance the state of the art both at the theoretical and practical levels – a tight coupling between science and technology.

POP's outcomes will also be likely to have **an important impact on a wide spectrum of application developments** and this impact is detailed below.

There are increasing needs for **human activity recognition** based on video and audio analysis. These needs span from **security and safety applications** (traffic, airport, and building surveillance) to **medical monitoring** of people with special needs. In order to cope with some legal regulations and to protect the private life of people under monitoring, it is crucial not to communicate images but higher-level abstract descriptions. Therefore, a cognitive approach is crucial.

Another challenging application domain on which POP will have impact is the domain of **human-to-human and human-to-machine communication**. Video conferencing systems are prominent examples of systems that heavily rely on visual and audio processing and interpretation. These systems are in their primitive stage of development. Current prototypes and commercial products use fixed cameras and microphones. In the near future we predict that the audio-visual capturing device will (and should) act like a *autonomous and intelligent cameraman robot* that selects a speaker among several people, and purposively moves towards this speaker to avoid occlusions, to increase the signal to noise ratio for reliable speech recognition, and to zoom onto the speaker's face from the right viewpoint.

Another example where an audio-visual cognitive system that selectively processes and interprets the sensorial information is important, is the example of a **driving assistant**. In the near future car driving is likely to drastically change. European car manufacturers have always pioneered technological innovations. In the long term, all European car manufacturers foresee that visual sensors and their associated software will become part of standard equipment. Cameras' fields of view will span both outside and inside the car. The system will be able to detect and recognize obstacles and to verbally alert the driver. Other audio and visual sensors will be dispatched inside the car for the recognition of the driver's state and behavior. Clearly, these driving assistance and monitoring systems will play a major role for safer cars and the outcomes of the POP project will have an important impact on such technologies.

The methodological and algorithmic developments to be carried out within POP's work-packages will allow integration of vision and hearing in complex situations. It will be possible, for example, to observe a scene with several people (moving around and speaking) and to associate a voice signal to each one of the persons in the scene. This will have a **direct and practical impact onto the development of hearing aids**. State-of-art hearing aids give good understanding of individual speakers but fail in cluttered and multi-source environments. For this reason, the acceptance of hearing aids by patients is still poor. The ability to focus the auditory attention onto a target that is visually selected could solve this fundamental problem.

The **European impact of POP** will be effective through collaborations with other projects. The POP partners and researchers are active within ongoing FP6 projects:

- The coordinator (partner 1, INRIA) is currently involved in a Marie Curie Early Stage Training action (VISITOR, MEST-CT-2004-008270). VISITOR is a single-partner action that started in December 2004 for 4 years and that is coordinated by the INRIA Rhone-Alpes computer vision group involved in POP (Edmond Boyer). Within EST VISITOR there will be 8 full-time PhD students and 8-10 PhD visiting students (for 3-6 month periods of time). VISITOR's objectives and topics of research (perception, learning, and autonomy) are strongly related to the POP project.

- The Research Training Network action VISIONTRAIN (MRTN-CT-2004-005439), coordinated by INRIA (Radu Horaud) and that involves 10 other partners from 10 countries, will start in May 2005 for 4 years. RTN VISIONTRAIN's objectives and topics of research are to establish cognitive and computational models for vision systems, and therefore VISIONTRAIN is fully relevant to the POP project.

- The University Hospital Hamburg-Eppendorf (partner 3) is involved in coordinating the EU Network of Excellence Neuro-IT (IST-2001-35498), which provides a European platform for co-ordination and dissemination of research in the field of neuroengineering, biorobotics and computational neuroscience. Andreas Engel is acting as a steering committee member in the board of Neuro-IT.

- The University of Sheffield (partner 5) participates to the AMI integrated project (Augmented Multi-Party Interaction) that started in January 2004 for a duration of 4 years. The University of Sheffield's speech and hearing group is the AMI training coordinator. AMI is composed of 9 academic partners and 5 industrial partners from 7 European countries.

- The University of Sheffield currently coordinates the submission of a network of excellence proposal – WhatNext (WN), which gathers an European/international consortium and which is intended to complement and reinforce the research and development efforts carried out within the AMI project by adding a substantial cognitive dimension.

Within this European research networking environment, **the POP project will play a pivotal inter-disciplinary and training role**. Indeed, none of the above mentioned projects (VISITOR, VISIONTRAIN, AMI, and the WN proposal) covers all the scientific disciplines spanned by POP. For instance, there is no robotics, computer vision, and neuroscience research in AMI and there is no robotics, neuroscience, speech and auditory research in VISIONTRAIN. **POP will act as a link between these large projects and will contribute to their training and dissemination activities**. Moreover, it will add a cognitive and neuroscience dimension to these IST projects.

3.1 Contribution to standards

The activity of POP is not expected to contribute to standards directly, but the project's results will be likely to be useful in standardization work. In the area of semantic interpretation of visual contents, POP will use and expand many relevant items from the MPEG-7 standard, and develop technologies which may be relevant to future calls for profiles, amendments or addenda. Another example is the video object-based coding in MPEG-4 that requires advanced means for image interpretation, such as motion estimation of deforming objects.

A POP researcher contributed to MPEG-7 standard in the past: Rémi Ronfard from INRIA (partner 1) took an active part and acted as France's national representative. Dr. Ronfard continues to be informed of MPEG activities and he will keep the POP researchers updated of the latest MPEG standardization actions that are relevant to the project.

There is another pre-standardization facet: The cognitive systems research community could benefit from shared experimental procedures including standard stimuli sets, data sets, procedures evaluating results, etc. The ECVision network IST-2001-354540, an FP5 project sponsored under *Cognitive Vision Systems*, made successful efforts in this direction.

We will take up the ECVision recommendations whenever we will disseminate data sets and software packages meant for benchmarking.

4 The consortium and project resources

4.1 Collective expertise of partners

The composition of the consortium has been driven by three guidelines: excellence, cross-fertilization between several disciplines, and optimal integration of research results. The consortium is built around 5 academic partners and 6 research teams from 4 member countries:

- **INRIA, partner 1:** The computer vision group and the statistics group of INRIA Rhône-Alpes (Montbonnot, France): INRIA researchers are internationally recognized for their contributions to geometric computer vision. Recently they developed new methodologies able to reconstruct 3-D complex and dynamic scenes from multiple cameras. In collaboration with the statistics group they will complement this geometric layer with a probabilistic/statistical framework for multiple cue integration and for sensory/knowledge interactions.

- **UOS, partner 2:** The Institute of Cognitive Science's neurobiopsychology group of the University of Osnabrück (Germany): UOS will focus on studying cognition under natural conditions. Peter König's researchs focusses on the processing of sensory stimuli under natural conditions with psychophysical and theoretical techniques. These include high-precision binocular eye-tracking, recording of natural visual stimuli, and computer simulations of cortical dynamics. These techniques lend themselves for advanced analysis of experimental data provided by UKE.

- **UKE, partner 3:** The Institute of Neurophysiology of the University of Hospital Hamburg-Eppendorf (Germany): UKE will contribute with state of the art physiological techniques to study the mammalian and the human brains. These include invasive micro-electrode recordings as well as non-invasive EEG, MEG and fMRI techniques. These techniques are used to investigate processing in sensory systems with an emphasis on plasticity and sensorimotor integration. Furthermore, Andreas Engel has considerable experience studying physiology and pathophysiology of cognitive processes in the human brain

- **FCTUC-ISR, partner 4:** The Institute of Systems and Robotics of the University of Coimbra (Portugal): ISR has several important contributions to the study of coordination of perception and action in Robotics. In particular the ISR researchers involved in POP developed robust visual-based robot control methods based on optical flow. ISR will contribute to the aspects related to sensory-motor coordination and to the design of the audiovisual robotic platform.

- **USFD, partner 5:** The speech and hearing research group (SpandH) of the University of Sheffield's department of computer science (United Kingdom): USFD has a unique expertise in auditory scene analysis that is recognized worldwide. USFD will lead the development of algorithms for acoustic cue extraction in noise and robust speech decoding in the presence of multiple talkers, and models of auditory attention.

4.2 Individual role and skills of partners

Partner 1: INRIA

The "Institut National de Recherche en Informatique et Automatique" is one of the largest institutes in the world whose research activities are fully dedicated to computer science, applied mathematics, and control theory. INRIA's headquarters are in Rocquencourt near Paris. INRIA is composed of six research units spread over the French territory. Each research unit has strong relationships and collaborations with universities, their associated research laboratories, and their PhD programs. The two research teams involved in this project are associated with INRIA Rhône-Alpes in Montbonnot (near Grenoble).

The computer vision group. Named the **MOVI group – MOdels for computer VIsion**, the computer vision group is lead by Dr. Radu Horaud. The group also includes five senior researchers (doctors Peter Sturm, Edmond Boyer, Rémi Ronfard, Fréderic Devernay, and Hervé Mathieu), as well as 10 PhD students, 2 post-doctoral fellows and 3 development engineers. The group was founded in 1990 both by professor Roger Mohr and by Radu Horaud, and by 2002 the group counted 8 researchers and 2 professors. In 2003 Cordelia Schmid, William Triggs, and Roger Mohr founded an independent group, the LEAR group (Learning and Recognition in Vision).

The MOVI group carries out research in vision-related topics of interest: camera modelling and calibration, geometric modelling, feature extraction and matching, learning and recognition, 3D point-based and surface-based reconstruction from binocular stereo and from several cameras, motion segmentation, motion understanding, and motion capture, spatio-temporal tracking in long video sequences, human gesture recognition, object recognition, visual-based robot control, etc. The general philosophy is to (i) establish solid theoretical foundations, (ii) develop methods and algorithms based on mathematical modelling and numerical analysis, and (iii) validate the results through proof-of-concept experiments and software development. In the recent past, the group participated to a large number of European projects such as the VIGOR project (Esprit-IV reactive LTR project 26247), the CUMULI project (Esprit-IV reactive LTR 21914), the VISIRE project (IST-1999-10756), the EVENTS project (IST-1999-21125) and the VIBES project (IST-2000-26001).

Currently, the MOVI group is involved in the following French and FP6 European projects:

1. The HOLONICS STREP project (partner, IST-2004-511318), 2004-2007, *Holograhic and Action Capture Techniques*, carried out with three industrial partners;

2. The VISITOR Marie Curie EST action (coordinator, MEST-CT-2004-008270), 2004-2008, *Interaction, Learning, and Autonomous Systems*, aimed at training 8 full-time PhD students and 10 visiting PhD students. Two full-time PhD students have an appointment in the MOVI group.

3. The VISIONTRAIN Marie Curie RTN action (coordinator, MRTN-CT-2004-005439), 2005-2009, *Computational and Cognitive Vision Systems*, aimed at networking 11 academic partners from 11 European countries. There will be 11 full-time PhD students and 12 post-doctoral fellows funded by the VISIONTRAIN network and who will work collaboratively.

4. Real-time stereo for identifying static and moving obstacles (contractor), 2004-2007, project funded by Renault SA.

5. Human motion capture with a small number of cameras (partner), 2004-2006, project funded by the French RIAM initiative and carried out in collaboration with 2 multi-media companies and the University of Rennes.

6. A multi-camera interactive platform for augmented-reality applications (partner), 2004-2006, project funded by the French RNTL programme and which is carried out in collaboration with 2 industrial companies.

The statistics research group. Named the **MISTIS group**, its main topics of research are the modelling and inference of complex and structured stochastic systems. More precisely, the group aims at developing statistical methods for dealing with complex systems, complex models, and complex data. The applications that we tackle consist mainly of image processing and spatial data problems, as well as applications in biology and medicine. Our approach is based on the statement that complexity can be handled by working up from simple local assumptions in a coherent way, defining a structured model, and that is the key to modelling, computation, inference and interpretation. The methods we consider involve mixture models, Markovian models, and more generally hidden structure models on one hand, and semi- and non-parametric methods on the other hand. Currently we investigate the two directions of research:

- *Dealing with complex phenomena.* This is twofold, we have to deal with complex models and with complex data. To account for complex phenomena, we propose to use structured models and methods allowing easy interpretations. At the models level, we propose to develop model selection and approximation techniques for complexe structure models. At the data level, we study dimension reduction techniques based on non linear data analysis.
- *The theoretical and practical behaviour of methods.* We focus on approximations justifications, asymptotic behaviour and convergence analysis.

The MISTIS group is led by Dr. Florence Forbes and it includes a total of 6 members: two senior research scientists, one associate professor and three PhD students.

Laboratory resources and facilities. INRIA Rhône-Alpes has recently developed a laboratory fully dedicated to real-time multi-media and interactive applications – the GrImage platform. The laboratory is hosted in a 80 m^2 room and is composed of a PC cluster (27 dual-processor modern computers interconnected through a dual-gigabit Ethernet network), 20 digital video cameras (12 colour cameras, 8 grey-level cameras, as well as the associated hardware for synchronized multiple-camera video acquisition), 16 video projectors arranged and calibrated to display a high resolution image (4096×3072 pixels) onto a 2×2.7 meter display screen, and 14 light projectors for controlling the photometric environment. The PC cluster runs under Linux and the software libraries and packages that are available on-site include: single- and multi-camera calibration software and devices, syncrhonized multiple-video acquisition, modern computer vision algorithms (image and video segmentation, stereo reconstruction, background subtraction, real-time 3D reconstruction of articulated objects, human body tracking, etc.), 2D and 3D support for multi-projector visualization (XDMX, VNC-wall, Chromium, VR Juggler, OpenSG), standard parallel programming tools and libraries (MPI, PVM), and dedicated libraries (FlowVR). There are 2 in-house engineers and 1 in-house technician whose full-time duties are to insure that the GrImage platform runs smoothly and to assist researchers and PhD students whenever they intend to prepare and run an experiment.

A proposal to use GrImage as an European transnational platform was recently submitted (FP6-2004-Infrastructure-5). If successful, (currently the proposal is under evaluation after being submitted on March 3, 2005) this single-contract project intends to provide funding and to host European teams wishing to run experiments onto the GrImage platform. Therefore, it is expected, if granted, that there will be an important synergy between the GrImage users's group and the POP partners. This laboratory will host the audiovisual robotic platform to be built by the POP partners.

Contribution to the POP project. There will be two research groups involved in the POP project – the computer vision group and the statistics group. The computer vision group will provide the scientific and technological expertise needed to integrate various vision modules into a working system, i.e., everything from camera acquisition and camera calibration software, to methods for recovering depth using stereo, motion segmentation methods, and object tracking methods. The statistics group will provide theoretical insights for the development of robust statistical methods. These methods will be used for the 3-D reconstruction and tracking algorithms that must operate in the presence of noisy, bad, or partially missing data.

The expertise on real-time 3-D reconstruction of humans from image contours will be used within the context of the active stereo sensor to be developed by the POP partners. The immense expertise on camera calibration using generic camera models will be used for the calibration of the ative stereo head.

Both the statistics and the vision groups will contribute to the development of a theoretical and methodological model of perceptual attention, in particular they will investigate together the most robust and efficient way to model interactions between high-level knowledge and low-level sensorial information. The statistics group is particularly competent in the domain of Bayesian decision theory and therefore will have a major contribution to model the integration of auditory and visual cues.

Our experimental and development platform/laboratory will be made available to POP researchers. One in-house development engineer (Hervé Mathieu, member of the vision group) will supervise the software development activities of POP in order to integrate and build the POP's experimental platform.

Florence Forbes holds a position of senior research scientist at INRIA and she is the scientific leader of the MISTIS group. Florence Forbes obtained her PhD in applied probabilities from the "Université Joseph Fourier", Grenoble, France in 1996. Then she spent one year as a post-doctoral fellow with the department of statistics of the University of Washington, Seattle, where she worked in Adrian Raftery's group. In December 1997 she joined INRIA and in 2003 she has been appointed head of the statistics group at INRIA Rhône-Alpes. Florence Forbes' topics of interest and research activities include Bayesian decision theory applied to image analysis, Markov processes, Markov random fields, hidden structure models, as well as the development of methodological and algorithmic statistical methods such as Expectation-Maximization (EM) techniques. Florence Forbes has 10 journal publications in major international journals.

Radu Horaud holds a position of director of research at INRIA; he is the scientific leader of the MOVI group and he is in charge of European relationships at INRIA Rhône-Alpes. Radu Horaud is born in Rumania and became a French citizen in 1975. After obtaining his PhD in 1981, he started his career in 1982 at SRI International (Menlo Park, CA USA) as a post-doctoral fellow in the Artificial Intelligence Center. In 1984 he joined the Centre National de la Recherche Scientifique. In 1998 he joined INRIA. Radu Horaud pioneered work on object recognition using range data and on stereo matching using graph representations and heuristic-search techniques based on geometric constraints. He has important contributions to the problems of visual sensor calibration, 3-D reconstruction, and image-based robot control. Radu Horaud has 100 scientific publications mostly in international journals and conferences. In 1995 he co-edited a special issue on *High-level vision* for the "Artificial Intelligence Journal" and in 2001 he co-edited a special issue on *Image-based robot control* for the "International Journal of Computer Vision". Currently he is an area editor for the "Computer Vision and Image Understanding" journal and a member of the editorial boards of the "International Journal of Robotics Research" and of the "International Journal of Computer Vision". In 2001 he served as programme co-chair of the IEEE International Conference on Computer Vision (Vancouver, Canada) and coordinated the David Marr prize for the best ICCV paper. So far he directed 20 PhD students. His current research interests include analysis of dynamic scenes, human motion and gesture recognition, augmented and mixed reality applications of computer vision, visual-motor coordination, and real-time vision algorithms. Radu Horaud participated to a large number of European and national projects. He coordinated the Esprit-IV Reactive LTR project VIGOR which successfully developed and implemented vision-based robot control in cooperation with Odense Steel Shipyards (Denmark), with the University of Cambridge, and with the University of Karlsruhe. Currently he is the coordinator of the Marie Curie RTN action VISIONTRAIN (2005-2009). He will be the coordinator of POP.

Peter Sturm holds a position of senior researcher at INRIA. Peter Sturm is born in Germany where he graduated from the University of Karlsruhe in 1994. He moved to France in 1994. He obtained his PhD in 1997 from the Institut National Polytechnique de Grenoble. He started his professional career at the University of Reading where he spent two years as a post-doctoral fellow. He joined INRIA and the MOVI research team in 1999. In 1998 Peter Sturm got the SPECIF PhD award (he was the first one to receive this prize). His research interests cover camera calibration, 3D reconstruction with a moving camera, and photometric-based reconstruction. Peter Sturm has 30 scientific publications.

Edmond Boyer is assistant professor (maître de conférences) in computer science at Université Joseph Fourier and a member of the MOVI research team. He obtained his PhD from the Institut National Polytechnique de Lorraine in 1996. He started his professional career as a research assistant at the University of Cambridge's Department of Engineering. His interests and topics of research cover both computer vision and computational geometry. He has internationally recognized research contributions in the domain of 3D surface reconstruction from multiple cameras. Edmond Boyer has over 25 scientific publications.

Remi Ronfard holds a position of senior researcher at INRIA. Remi Ronfard graduated from Ecole des Mines de Paris in 1986 and obtained his PhD in 1991, also from Ecole des Mines de Paris. He worked as a software engineer for Dassault Systemes (CATIA) and Computervision (CADDS) and as a researcher at Institut National de l'Audiovisuel. He was invited twice at the IBM T.J Watson Research Center in Yorktown Height, New York as a post-doc in 1992, and as a visiting scientist in 2000. His research interests have evolved over the years from deformable contours and geometric modelling to video processing, analysis, and indexing.

Frédéric Devernay holds a position of senior researcher at INRIA. Frédéric Devernay is a former student and PhD student of Ecole Polytechnique (1992 and 1996). He started his professional career with Istar Ltd. He joined INRIA in 2000. His research interests cover computer vision, computer-assisted robots for surgery, augmented reality systems, and motion capturing systems. Frédéric Devernay has 14 scientific publications.

Significant publications. [43], [24], [25], [42], [74], [44], [77], [38], [36], [54], [16], [80], [10], [66], [78], [52], [55], [50], [51], [56], [26], [53], [75], [35], [64], [76].

Partner 2: University of Osnabrück

The Institute of Cognitive Science is the pioneer in Germany offering a study program in Cognitive Science covering all levels from Bachelor, Master to Doctorate studies. It houses seven research groups (Computational Linguistics, Cognitive Psychology, Artificial Intelligence, Neuroinformatics, Theoretical Computer Science, Philosophy of Cognition and Neurobiopsychology) covering a wide range of different methodologies. All groups focus on key cognitive capabilities of humans such as high level sensory processing, production and understanding of language. In the research group Neurobiopsychology we study human cognition with an emphasis on the underlying physiological mechanisms and implementations of these concepts in artificial agents:

1. Processing of natural visual stimuli in the mammalian brain;
2. Control of human overt attention under natural conditions;
3. Interaction of bottom-up and top-down signals in the visual hierarchy, and
4. Stimulus statistics and unsupervised learning in neuronal networks.

For this purpose we (i) perform psychophysical and physiological experiments; (ii) use advance mathematical tools for data analysis and develop quantitative models; (iii) implement the resulting concepts in real-world agents to compare their performance again with that of biological systems. This interaction of studies of cognition, mathematical analysis and modelling and implementation in behaving artefacts has proven highly synergetic. The group has made contributions to our understanding of the dynamics in neural networks for image segmentation, adaptation of receptive field properties in primary and higher visual cortices, interaction of bottom-up and top-down signals, the role of low level features in the control of human attention. Furthermore, the group has extensive experience utilizing the synergies of such complementary techniques. Previous interactions with physiological experiments (own and of other partners) have proved to unfold large synergies and have been highly successful.

The research group of Neurobiopsychology at the Institute of Cognitive Science was founded little more than a year ago and is rapidly expanding. Currently it is composed of 1 professor, 1 post-doctoral researcher, 3 PhD students and 4 master students. The composition of the group includes researchers from 3 different European countries.

Research is performed in cooperation with a number of other universities and companies. Currently running projects include:

- Collaboration and support by Future Technology Research Division HONDA R&D EUROPE (Deutschland) GmbH.
- EU-Project AMOUSE Call Identifier IST-00-4-1A.
- The German Israeli Foundation For Scientific Research And Development (G.I.F.) - Processing of natural images by population activity in primary visual cortex.
- Neuro-IT.net Thematic network Contract no.: IST-2001-35498.

Laboratory resources and facilities. In the last year a lab for stereoscopic presentation of still images combined with eye tracking was set up. Importantly, both eyes are tracked simultaneously with high precision allowing determination of vergence movements. Additionally a dual-Purkinje based eye-tracker is available for ultra-high precision tracking of a single eye. Furthermore, we have developed a calibrated dual-camera system for recording stereoscopic still images and short video clips suitable for presentation in the setup described. This lab is fully functional and first results were presented at the German Neuroscience Conference this spring. In a separate lab a 32-channel EEG recording system is setup. This system is now functional and will be integrated with the eye-tracking setup as part of the present proposal. These EEG recordings lead to massive amounts of data, testing even modern computer systems. Workstations with suitable hardware and software are available for efficient data analysis. We have developed a small-scale real-world environment (2m * 3m) with 2 Kephera based robots. These robots have either directional or omni-directional cameras. They are interfaced with a cluster of PCs running under Linux. We use software developed as a part of running EU-projects (Amouse & Amoth) for simulation of large-scale neural networks and online interaction with the two robots. The development of the software is based on rapid prototyping using a MatLab toolbox, and then implemented in a C++ class library for highly efficient usage of computational resources. With this setup we are able to achieve short turn around times and still process the camera images with 32 frames per second. Using an algorithmic pre-processing step leads frame rates well above 100 Hz. (4) In collaboration with our colleagues at the Institute of Cognitive Science we have access to a laser scanning device, currently used on a mobile platform in the rescue league. Software for complete interchangability of visual stimuli recorded by robots for measuring eye-movements, or feeding stereoscopic videos into the neural network simulation is available.

Contribution to the POP project. The partner at the Institute of Cognitive Science at the University of Osnabrück will contribute psychophysical experiments on human overt attention under natural conditions; advanced tools for data analysis; develop models of sensory processing and sensori-motor coupling in the mammalian brain and interface with the development of behaving artefacts. Firstly, we will study the contribution of low-level features and high-level tasks to the guide of human overt attention under natural conditions. Controlled modifications of stimuli allow dissecting purely correlative effects from causal influences that can be described by quantitative mechanistic models. This is complemented by an investigation of effects of context and task.

The investigation of the interaction of such top-down effects with the bottom-up signals forms the second contribution. It will be based on advanced mathematical tools like auto-regressive techniques and an information theoretic analysis. The results will be matched to an implementation suitable to guide a behaving artefact.

Thirdly, the adaptation of receptive field properties to the properties of natural stimuli complements the former studies. Using techniques of unsupervised learning over the space of natural visual and auditory stimuli we study the emergence of cross-modal integration.

Peter König holds a position of the head of the department of Neurobiopsychology at the Institute of Cognitive science at the University of Osnabrück. Peter König studied physics (Diplom) and medicine (Staatsexamen & Approbation) in Bonn/Germany. In 1987 he moved to the Max-Planck-Institute of Brain Research in Frankfurt. In that time he made important contributions to the fast dynamics and synchronization of neuronal activity in the mammalian visual cortex. 3 years later he obtained the PhD at the University of Würzburg. In 1995 he took on a position as a senior fellow at the Neurosciences Institute in San Diego/California. Here he started the behavioural work and the investigation of context dependent processing of sensory stimuli. In 1997 he moved to the University/ETH Zürich and worked for several years on processing of natural visual stimuli and the adaptation of neuronal response properties to the statistical properties of the sensory input. At the end of 2003 he accepted on offer of the University of Osnabrück. His research interests cover cognition with a view onto a firm theoretical understanding of underlying mechanisms. He is member of the editorial board of Neural Computation and of Biological Cybernetics.

Anita Schmid enrolled in physics at the Swiss Federal Institute of Technology (ETH) Zürich in 1995. She focussed on experimental physics and attended several courses in Neuroinformatics. Getting increasingly fascinated by this subject, she performed her diploma project on visual psychophysics at the Institute for Neuroinformatics (ETH Zürich/University of Zürich) in 2000. Shortly afterwards, she started my PhD project with Prof. Kevan Martin, director of that institute. She performed electrophysiology experiments in the visual cortex of anesthetized cats and monkeys. In addition, she developed models of the corresponding neuronal responses. In 2004 the ETH Zürich accepted my PhD thesis 'Processing of Feature Discontinuities in Cat and Monkey Primary Visual Cortex'. Beginning of 2005, I moved to Osnabrück, Germany, and started a postdoc position in the Neurobiopsychology lab of Prof. Peter König. I am in the process of building a lab for chronic multi-electrode recordings in awake behaving mice studying the neurophysiology of learning.

Hans-Peter Frey enrolled at the University of Konstanz/Germany in October 1998 in order to study Psychology with Information Engineering as minor subject. After the pre-diploma he did a half-year practical at the Institute of Neuroinformatics (Swiss Federal Institute of Technology/ University of Zürich, Switzerland), where he conducted psychophysical experiments with cats. This project turned out to be highly successful and was published in the Journal "Neuroscience" with Hans-Peter Frey as first author. Back in Konstanz, he worked as a students researcher at the Department of Clinical Psychology, developing software for MEG/EEG analysis. At the Institute of Neuroinformatics he conducted his diploma project "The influence of colour on overt visual attention". Since 2004 he is a PhD student in the Neurobiopsychology laboratory of Prof. Peter König at the University of Osnabrück, Germany. Here he is working on cross-modal attention and overt visual attention using psychophysical and eye-tracking methods, as well as EEG.

Selim Onat received his high school diploma from French high school Saint Benoit in (Istanbul, Turkey) in 1997. In the same year he took up Cellular Biology and Physiology at the graduate school of the University of Rennes (France). From 2000 to 2001, he studied at the University of Paris XI and received his bachelor degree in Cellular Biology and Physiology with specialization in Neurosciences. The following year he accomplished the master degree at the University of Paris VI. His master thesis in the laboratory of Neuromodulation and Cognitive Processes (Institute of Neuroscience, Paris) was supervised by Dr. Susan Sara and its results published in the prestigious Journal of Neuroscience. In 2002 he started my PhD at the Institute of Neuroinformatics/ETH Zürich in the Laboratory Peter König. His research group was relocated in 2003 to the University of Osnabrück and where he is continuing his PhD project. He is supervising several Bachelor and Master students and working on a project in electrophysiological and psychophysical aspects of natural 3D vision.

Significant publications. [72], [15], [22], [57], [1], [79], [21], [23], [65].

Partner 3: The University Hospital Hamburg-Eppendorf

The Institute of Neurophysiology and Pathophysiology (head: Prof. Dr. Andreas K. Engel) is part of the medical school of Hamburg University. Research activities at the institute focus on the level of integrative and behavioral neurophysiology.

Central research topics addressed in current projects are:

- information processing in sensory systems;
- integration of sensory and sensorimotor signals;
- neuroplasticity and maturation of sensory systems;
- the relevance of neural dynamics for cognitive processes;
- pathophysiology of sensory and motor functions.

Investigation of these issues is carried out in a variety of different neural systems, including visual, tactile, nociceptive and motor structures. One of our major goals is to synergistically combine approaches in humans and in animals. The human studies are carried out on both normal subjects and on patients with neurological disorders, using techniques for the recording of electrical (EEG) signals from the brain in combination with functional neuroimaging (fMRI). Studies in animals focus on three different species, namely, ferrets, rats and mice. Animal studies use two different approaches. Recordings of neural signals are either performed under anesthesia or, alternatively, in awake behaving animals after chronic implantation of microwires in cortical or subscortical structures.

The central focus of the research of the Institute of Neurophysiology is the dynamics of neuronal populations and, specifically, temporal correlations between different neurons leading to the formation of coherent cell assemblies. Since 1987, Andreas Engel's group has investigated these phenomena in sensory and motor systems by means of extracellular recording techniques in carnivores and rodents. Several years ago, the group has begun to address similar question in the human brain. Currently available data suggest that the temporal dynamics of neural activity may be essential to understanding the neural foundations of perceptual, cognitive and motor functions. The projects of the Institute of Neurophysiology therefore investigate neural synchronization and oscillatory response patterning in the context of sensory processing, attention, cross-modal and sensorimotor integration, working memory and conscious awareness.

Currently, Andreas Engel's group is involved in two European projects: "Artificial Mouse"; EU-IST; 2001-2006; Prof. Dr. A. Engel acts as coordinator and "Thematic Network NEURO-IT"; EU-IST; 2002-2006; Prof. Dr. A. Engel is member of the Steering Committee of the network.

Laboratory resources and facilities. The institute is equipped with 6 in-vivo laboratories for studies in anesthetized and behaving animals. To assure transfer of the results to the human brain, we have established a lab for high-density (128 channel) EEG recordings and a setup for psychophysical studies. The EEG measurements are combined with fMRI studies carried out both at the University Hospital Eppendorf as well as in cooperation with several other groups. Moreover, we are performing invasive microelectrode recordings in patients with neurological disorders. These studies are carried out as part of a cooperation with neurosurgery and neurology groups in Hamburg.

Contributions to the POP project. Within POP, the UKE research team will contribute in the area of neurophysiology and carry out experimental work on visual-auditory interaction and the relation of such interactions to attention and scene segmentation. Experiments will be carried out using EEG and fMRI in humans. Complementing the human experiments, microelectrode recordings will be performed in animals, using the ferret as a model system to study cellular mechanisms of visual-auditory interactions. In addition, the UKE team will contribute theoretical knowledge from the field of neurophysiology and neurobiology to the modelling and robotics work of the other POP partners.

Andreas K. Engel is full professor of physiology and director of the Institute of Neurophysiology and Pathophysiology. He studied medicine and philosophy and graduated in medicine at the Technical University Munich (Germany) in 1987. From 1987 through 1995, he worked as a staff scientist at the Max-Planck-Institute for Brain Research (Frankfurt, Germany), where he developed a long-standing interest in dynamics of distributed sensory representations, inter-modal and sensorimotor integration, and computational theories of perception and representation. In 1996, A. K. Engel established an independent research group at the Max-Planck-Institute for Brain Research which was funded by the Heisenberg Program of the German Research Council (Deutsche Forschungsgemeinschaft). From 2000 to 2002, he has been leading the "Cellular Neurobiology Group" at the Research Center Julich, Germany. In fall 2002, he was appointed as a director of the Institute of Neurophysiology and Pathophysiology at the University of Hamburg, Germany. A. K. Engel has extensive experience with in-vivo multielectrode recording from humans, cats and rodents, as well as with human EEG recordings. The main focus of his work has been in the visual system.

Andrej Kral is associate professor of physiology and deputy director of the Institute of Neurophysiology and Pathophysiology. He has studied medicine and graduated from the University of Bratislava (Slowakia). His work focusses on neurophysiology and neuroplasticity in the auditory system, and the development of cochlear implants.

Stefan Debener is a psychologist and head of the EEG group at the Institute of Neurophysiology. He graduated from the Technical University of Dresden (Germany). His work focusses on mechanisms of attentional modulation of visual and auditory stimulus processing.

Gerhard Engler is a biologist and heads the Behavioural Neurobiology Lab at the Institute of Neurophysiology. He graduated from the University of Freiburg (Germany). He is interested in acute and chronic multielectrode recordings from visual, somatosensory and motor system of ferrets and rodents, studying dynamic neural interactions and their dependence on sensory stimulation and behavioural context.

Significant publications. [48], [40], [67], [46], [73], [18], [19], [58], [59], [17], [39], [41], [45], [60], [47], [61], [33], [62], [68], [20], [34], [63].

Partner 4: University of Coimbra

The **Institute for Systems and Robotics** (ISR) is a leading research institute in robotics in Portugal. It has three poles located in Coimbra, Lisbon and Porto. ISR–UC (Institute for Systems and Robotics at the University of Coimbra) has 20 Ph.D.s (from Mathematics and Engineering) who are permanent researchers. In addition about 45 Ph.D. and Master students are integrated in the research projects of the Institute. ISR-UC has activities in the areas of mobile robotics, computer vision, intelligent control, sensor data fusion, sensors and actuators for automation, and control theory. The ISR-UC has been involved in several projects funded by FCT (Portuguese Science and Technology Foundation), the European commission, NATO, as well as by several Portuguese industrial companies.

ISR–UC is located in the new engineering campus of the University of Coimbra, occupying the first level of the building where is also located the Department of Electrical and Computer Engineering. The Computer and Robot Vision Lab (CRVLAB) is part of the ISR–UC. The CVRLAB of ISR has focussed its research activities on three main lines of research: development of real time active and robot vision systems, mathematical modeling of cameras, and interpretation of human action.

Laboratroy resources and facilities. FCTUC-ISR has a total area for laboratories of 1200 m2. The Computer Vision Lab has an area of 100 m2. FCTUC-ISR has been involved in several national and international research projects in Robotics. Most of the research developed at the Computer Vision Lab concerns vision applied to Robotics. The resources available include several manipulators (Puma 560, Adept and ABB 2000). The manipulators have been used in projects dealing with haptics and also computer vision, namely calibration. Several mobile robots are also available (1 Nomad 200, 5 Nomad Scouts, 1 Robuter). The mobile platforms have been used in research dealing with navigation, map building, obstacle detection and localization. Significant part of the research effort of the Lab in the last few years has been devoted to Active and Omnidirectional Vision. As a result the Computer Vision Lab has two active vision platforms, designed and built in the Lab. One of them (MDOF) has 9 mechanical degrees of freedom and 6 optical degrees of freedom. The second platform has 6 mechanical degrees of freedom. For research in omnidirectional vision the Lab has 1 parabolic video camera, 1 parabolic digital camera and 2 hyperbolic video cameras. The Lab has also several stereovision systems from Videre Design and Point Grey Research.

FCTUC-ISR has access to a mechanical workshop that is shared with the Department of Electrical and Computer Engineering of the University of Coimbra. The Lab has also a network of 20 PCs integrated in the University network. All the Labs of ISR are also covered by a wireless network.

Contributions to the POP project. ISR will contribute to several of the scientific topics of the project and, in particular, to the aspects related to sensory-motor coordination and design of the robotic devices. The Computer Vision Lab at ISR has developed several active vision systems, two of which are currently being used to study perception-action coordination in Robotics. The Computer Vision Lab has made several relevant contributions in active vision namely be developing visual control methods based on optical flow and by integrating them with mobile robots. The systems that will be developed within the framework of this project will be active and will be used to develop and test models of focus of attention. That will imply that the mechanical devices will be oriented towards the features deemed to be sources of attention. The scientific problems to be solved involve the integration and coordination of the visual and auditory measurements and the control of the mechanical devices. Since the mechanical devices have multiple degrees of freedom they have to be synchronized and coordinated so that they generate the expected behavior. The ISR team will contribute to the design and development of the robotic devices and also to the algorithms to coordinate the motions and activities of the mechanical systems.

In addition ISR will contribute to the design and development of algorithms for extracting the visual cues, and to their evaluation as well as to aspects related to the integration and fusion of the visual and auditory cues. The processes and algorithms for visual and auditory information extraction will be tightly coupled to the control algorithms.

ISR will also contribute to the development of a platform enabling the test of the biological models and algorithms on the robotic devices. This platform will allow the use, on the robotic devices, of the images and sequences of images tested in the psychophysical experiments. The mechanical response of the mechanism will be recorded and as a result the performance of the biological models and algorithms can be evaluated.

Helder Araújo is an Associate Professor at the Department of Electrical and Computer Engineering of the University of Coimbra. He has a background in Electrical and Computer Engineering, with an Engineering degree (University of Coimbra) and a Ph.D. (University of Coimbra) in Computer Vision and Image Processing. He has been doing research in computer vision and robotics for the last 15 years, resulting in the publication of over 80 refereed journal and conference papers. He has done work in active vision, camera calibration, visual surveillance, active tracking and robot navigation. He has participated in European projects in robotics and computer vision and has been principal investigator of several nationally funded projects. He is currently principal investigator of the FCT project OMNISYS – Omnidirectional Vision for Control and Navigation.

Joao Barreto is currently an Assistant Professor at the Department of Electrical and Computer Engineering of the University of Coimbra. He has a background in Electrical and Computer Engineering, with an Engineering degree (University of Coimbra) and a Ph.D. (University of Coimbra) in Computer Vision. His research focusses on Geoemtrical modelling of central catadioptric camera systems, active vision and visual servoing.

Jorge Batista is currently an Assistant Professor at the at the Department of Electrical and Computer Engineering of the University of Coimbra. He has done work on camera calibration, active vision and visual surveillance. He has been principal investigator of several research nationally funded projects on computer vision. He is currently the principal investigator of the project AI-FI – Advanced Interaction – using Facial Information, funded by the FCT. He is also the principal investigator of a project aiming at introducing automatic visual surveillance in several motorways of Portugal. This project is fully funded by the private company Brisa.

Paulo Peixoto is an Assistant Professor at the Department of Electrical and Computer Engineering of the University of Coimbra. He has a Ph.D. (University of Coimbra) in Computer Vision. He has been doing research in computer vision and robotics for the last 7 years, resulting in the publication of over 30 refereed journal and conference papers. He has done work in active vision, visual surveillance, and human computer interaction. He has participated in several European and national projects in robotics and computer vision and is currently the principal investigator of the FCT project HCOGICOS - Human Computer Gesture Interfaces for Collaborative Scenarios.

Significant publications. [12], [11], [13], [37], [6], [14], [7], [8], [9], [70], [5], [71], [4], [27], [3].

Partner 5: University of Sheffield

The Department of Computer Science at the University of Sheffield is a grade 5 research-rated department. The **Speech and Hearing Research Group** (SPandH) has an international reputation as a leading group in the multi-disciplinary field of speech and hearing research, and is one of the largest such teams in the UK with 3 Professors (Martin Cooke, Phil Green, and Roger Moore), 4 faculty staff (doctors Jon Baker, Guy Brown, Yoshi Gotoh, and Tom Hain), 5 research associates and 12 research students. The group has pioneered 'computational auditory scene analysis' (CASA) and has linked this work to speech technology through the 'missing data' approach to robust automatic speech recognition (ASR). SPandH has recently extended the statistical theory of ASR to general auditory scene analysis through the implementation of a novel decoder that processes speech in the presence of other sources without requiring models for the background or a knowledge of the number of competing sources. SPandH also has major interests in large-vocabulary ASR as well as audio-visual integration and generation. The Machine Learning Group has expertise in areas such as Bayesian methods in neural networks and learning classifiers from sloppily labelled data.

The Speech and Hearing Research group of the University of Sheffield participated and participates to the following projects and contracts:

- Recognition of speech by partial information techniques (**RESPITE**) EC ESPRIT Long term research (with Daimler-Benz, INP Grenoble, IDIAP, MATRA, ICSI, Mons), finished 2002.
- **Dipsy** (funded by the UK Engineering and Physical Science Research Council), finished in 2003. Supported behavioural experiments on the perception of narrowband speech.
- **SPHEAR** (TMR; 1998-2003).
- **Multisource** (funded by the UK Engineering and Physical Science Research Council), due to finish in August 2005. Supports work on the decoding of multiple, simultaneous speech sources.
- **HOARSE** (coordinator, Funded by the Marie Curie programme), 2002-2006,
- **AMI** (FP6 integrated project; 2004-2008), SpandH acts both as a major academic partner and as the AMI's training coordinator.

Laboratory resources and facilities. The Speech and Hearing Group at Sheffield University has a well-equipped laboratory with facilities for the recording and playback of high-quality audio data. It contains several binaural mannekins, a sound-attenuating chamber, state-of-the-art AD/DA converters, preamplifiers and calibrated microphones. It also has a signficant quantity of specialist software for running behavioural experiments, training and testing automatic speech recognition systems and separating speech from noise. Numerous large speech corpora are also available.

Contributions to the POP project. The University of Sheffield will contribute expertise in speech, hearing and audiovisual speech processing, both at the level of computational modelling and behavioral experiments. Drawing on an existing collection of software tools, audio and audiovisual corpora, Sheffield will lead the development of algorithms for acoustic cue extraction in noise and robust speech decoding in the presence of multiple talkers, and models of auditory attention. Sheffield will contribute to work on multimodal integration of auditory and visual processes at all levels of the project, including the design of behavioral experiments, development of the theory, algorithmic implementation and integration with hardware and software for an active audiovisual head.

Sheffield will lead a detailed comparison of auditory and visual attention and produce recommendations for modelling work on both auditory attention and multimodal attentional integration. Sheffield will also extend existing algorithms for cue extraction and integrate them with visual cues in a principled fashion. A multi-source decoder, which allows bottom-up and top-down information to be combined within a single probabilistic framework, will be augmented to deal with visual inputs. Sheffield will work on active listening strategies for improving the robustness of machine cognition, and contribute to algorithms which combine auditory and visual information in a synergistic manner.

Martin Cooke is a professor in the Computer Science department and works on speech perception in cluttered acoustic backgrounds (Cooke, 1993; Cooke, 2001a) and the application of missing data theory to robust automatic speech recognition (Cooke, 2001b). His recent work has been to construct a computational theory of early processes in human speech perception based on glimpsing the target source in a noisy background (Cooke, 2003) and to extend the statistical theory of speech recognition to handle the presence of other sound sources (Barker, Cooke and Ellis, 2005).

Dr Jon Barker (Computer Science, Sheffield) has interests in noise-robust speech recognition and audio-visual speech processing. His work has involved developing statistical techniques for recognizing speech in the presence of non-stationary sound sources (Barker, Cooke and Ellis, 2005). His recent focus has been the challenge of speech recognition in simultaneous speaker conditions, leading to the development of systems, inspired by auditory process, for recognizing simultaneous speakers in anechoic monaural conditions, (Coy and Barker, 2005) and in highly reverberant binaural conditions (Pallomaki, Brown and Barker, 2004; Harding, Barker and Brown, submitted). His current work aims to extend the techniques presented in (Barker, Cooke and Ellis, 2005) to exploit information from the visual domain.

Significant publications. [2], [30], [31], [29], [28], [32], [49], [69].

5 Project management

5.1 Management structure and role of different bodies

The management will be organized as follows. First of all a **Steering Committee** will be created (SC). The SC will be the consortium's main decision making and arbitration body. The SC will have one representative from each contractor and will be chaired by a representative of the coordinator – **The project coordinator**.

The Steering Committee is the decision making body of the project. All contractual issues, changes in technical specifications of workpackages, IPR issues, risk management, etc. will be discussed and decided by the Steering Committee. The decisions will be taken with a majority vote, one vote per partner.

The SC will hold a video-conference meeting every month and a physical meeting every 6 months in order to monitor the progress of the project, anticipate difficulties, and take decisions, through the following tasks:

- Appoint the **site-** and **workpackage managers**;
- Decide revisions and their use of project resources for work-packages;
- Be in charge of risk management, monitoring the progress of the project and ensuring that any changes are discussed and implemented in a timely fashion;
- Take actions in case a contractor makes default;
- Anticipate and suggest solutions in case of a conflict;
- Elaborate the rules governing the Intellectual Property Rights (IPR);
- Communicate with the European Commission;

Under the direction of the **Steering Committee** there will be the **site Manages** and the **work-package managers** (or work-package leaders. Since each partner will have the leadership of at least one workpackage, the site managers will be the same physical persons as the work-package managers.

The **Project coordinator** appointed by the coordinating partner and by the SC will be responsible for the overall day-to-day project administration and financial management. He/she will be the manager of **Work-package 6**, page 51 and in particular he/she will:

- Organise the relation between the Contractors and the European Commission acting as the contact-point for the project;
- Collect, monitor and integrate all the technical, administrative, and financial data from the partners and prepare appropriate documents for the European Commission and for the auditors: management reports, progress reports, final report, cost statements, financial statements, audit reports, etc.;
- Organise the technical audits, and;
- Chair the **site managers'** and **work-package managers'** sessions and meetings;
- Play a pivotal role between the **Steering Committee** on one side and the project researchers and engineers on another side, both top-down and bottom-up.

The **Work-package managers** will be responsible of the scientific and technical advancement of the project. They will work in close collaboration with the **Project coordinator** and with the **site managers**. In particular the **work-package managers** will:

- Hold a video-conference meeting on a monthly basis or as often as it is required for advancing the work;
- Hold a technical meeting every 6 months;
- Coordinate the work carried out within the individual work-packages and tasks and ensure its adherence to the pre-defined work-plan and timetable;
- Anticipate deviations from the planned work and monitor actions to correct these deviations;
- Report to the **Project coordinator** and to the **site managers** any foreseen problems and suggest practical solutions to overcome these problems in order to allow smooth execution of the planned tasks;
- Coordinate the preparation of project deliverables and assist the **Project coordinator** for the preparation of the technical documents to be delivered to the European Commission;

The **Work-package managers** will also act as work-package leaders and will coordinate the collaborative work within work-packages WP1 to WP5. They will ensure that the work planned within each work-package task is produced on time. They will be responsible for the deliverables of each work-package.

5.2 Exploitation, training, and dissemination

The manager of work-package 5 will act as the **Exploitation-Dissemination Manager** and will coordinate all the exploitation, training, and dissemination activities of the project. His/her specific responsibilities include:

- Coordinate the promotion of the results of the project through the **POP website** (see description below);
- Promote the preparation and presentation of POP papers at international journals, conferences, and specialized workshops.
- Organize tutorials and thematic schools both at the project level and in collaboration with other parties (University departments, European networks, Marie-Curie actions, etc.);
- Promote and establish bi-lateral contacts with other IST projects and instruments relevant to POP;
- Assist project researchers for the organisation of workshops and tutorials aimed at promoting the project and at collaborating with other similar European consortia;
- Establish and maintain a list of European companies interested in the outcomes of the project.
- Periodically advertize the project achievements through the dissemination of an **POP electronic newsletter**;

- Promote the results and outcomes of the project through the organisation of **Industrial workshops**;
- Promote the results of the project towards standardization and patenting committees and offices (see below); Invite representants of these committees and offices to project meetings whenever appropriate.

POP website. The **Project coordinator** assisted by the **Exploitation-Dissemination Manager** will be responsible of the maintainance of a project website hosted by the coordinating partner. **The coordinator will be committed to make available the necessary human and computer resources needed to create and maintain a website which should be active through the duration of the project and at least three years after the end of the contract**. A website technical master and a content-editor will be appointed by the coordinator.

The public website will contain the following information:

- Links to project partners and to research teams involved in the project;
- Links to similar projects and consortia;
- Access to scientific publications and technical reports arising from the project;
- Access to software packages and to data sets;
- Links to demonstration systems;
- Announcements of POP events, such as tutorials, thematic schools, workshops, etc.
- Access to the POP Newsletter;

5.3 Risk, problem, and conflict management

The ambitious objectives of this project and its associated work-plan have been thought out by a team of senior and experienced researchers who have accumulated a large body of scientific and technological knowhow and who have a great deal of expertise in carrying out collaborative projects. Nevertheless, risks, problems, and conflicts cannot be ignored.

Under the leadership of the project coordinator, the work-package managers will be responsible for smoothly carrying out the work during the duration of the work-packages. Whenever needed they will take the initiative to organize management meetings at the work-package levels in order to anticipate any risks, problems and/or conflicts:

- Detect scientific and technical difficulties in achieving a work-package or a work-task and provide alternatives and solutions to the problems encountered.
- Anticipate problems with the termination of a work-package and offer suitable solutions before the situation results in delays or dead-ends.
- Foresee fluid information sharing between the partners (background research, outcomes of a work-task, etc.) and avoid technical and bureaucratic barriers to information access.
- Anticipate the end of contract of a researcher, changes in personnel, reallocation in human resources by a partner, leave of absence of a researcher, temporarily absence of a researcher, conflicts between researchers, etc.

If a satisfactory solution cannot be found on a rapid and cordial basis, the work-package manager will refer to the project coordinator and to the site managers who will ask the Steering Committee to meet and to take a decision.

5.4 Management of knowledge and intellectual property rights

Intellectual Property Rights (IPR) will be handled in line with the general policy of the European Commission regarding ownership and exploitation of rights and confidentiality. Management of IPR issues will have two aspects: management of IPR internal to the Consortium and management of IPR in relation with external actors. While the protection of IPR against external actors is in most of the cases the competence of the individual organisations, the Consortium will offer several initiatives in that respect. The **Steering Committee** will be responsible for the design of the overall IPR protection policy of the project. The general policy will include, among others: rules on security, procedures for information exchange, recommendations for safe dissemination of results, advice for IPR protection, definition of use cases and procedures, etc. The **Steering Committee** will advice individual partners on the best IPR protection method at every stage of the project development. The Committee will also take the responsibility to lead and coordinate protection of results jointly owned by several of the partners in the POP consortium.

In a first approach, **background information** and **background patents** will be made available to the consortium members on favorable conditions, if they are necessary to perform the research in this project and if no major business interest of the owner of the background information opposes the disclosure or grant of licences for such patents or information. **Foreground information** and **foreground patents** are owned by the contractor generating such information. Each contractor shall make available its foreground information, on a royalty-free basis, to other contractors to the extent that such information is necessary for the execution of their own research within the project. If proprietary information is made available, the information shall be duly marked as confidential and the recipient will preserve its confidentiality. Non-disclosure agreements will be prepared if so requested.

IPR issues will be detailed more in the Consortium Agreement that will be signed by all members of the Consortium before the start of the project. The **Steering Committee** will be responsible for developing those principles through the project extension. With respect to IPR affecting third parties, the SC will develop policies to assure the protection of methods, new discoveries, software, hardware, and in general any outcome of the project susceptible to be protected by any legal means according to the legislation applicable in the different countries. For that purpose, all available legal resources will be employed when relevant: patents, copyright of software, trademarks or contents, secret knowledge, etc.

Particular emphasis will be done on the **coordination of dissemination and IPR policies** in order to make compatible ample diffusion of the project results with the necessary protection of rights.

In the particular case of the POP project, the consortium is formed of academic partners. Therefore, there will be no IPR restrictions concerning the publications of scientific articles and reports.

6 Detailed implementation plan

6.1 Implementation plan introduction

WP1: Cognitive mechanims of attention.

WP2: Integration of visual and auditory cues.

WP3: Sensory-motor coordination.

WP4: Development of methodological and experimental platforms.

WP5: Exploitation, training, and dissemination.

WP6: Project management.

6.2 Work planning

WP	Name	Start/End	Partners	Leader	Deliverables
WP1	Cognitive attention	1/30	1, 2, 3, 4, 5	UKE (3)	
WP2	Audio/visual integration	6/30	1, 2, 3, 4, 5	USFD (5)	
WP3	Sensory-motor coord.	6/30	2, 4, 5	UOS (2)	
WP4	Platforms and experiments	12/36	1, 2, 3, 4, 5	FTCUC-ISR (4)	
WP5	Expl., train., diss.	1/36	1, 2, 3, 4, 5	USFD (5)	
WP6	Management	1/36	1, 2, 3, 4, 5	INRIA (1)	

Table 1: *The start and end months, the participating partners, the leader and the list of deliverables for each work-package.*

Work-package	INRIA (1)	UOS (2)	UKE (3)	FCTUC-ISR (4)	USFD (5)	Total
Research & innov.						
WP1	25	25	30	10	20	110
WP2	30	20	30	25	20	125
WP3	-	25	6	30	20	81
WP4	20	5	6	40	20	91
WP5	3	3	3	3	5	17
Sub-total 1	**78**	**78**	**75**	**108**	**85**	**424**
Management						
WP6	5	3	3	3	3	17
Sub-total 2	**5**	**3**	**3**	**3**	**3**	**17**
Total	**83**	**81**	**78**	**111**	**88**	**441**

Table 2: *The break-down of efforts (in person months) for each partner and for each work-package.*

6.3 Graphical presentations of the components

The graphical presentation of the work-plan components, i.e., work-packages, tasks, deliverables, and milestones, are shown on figures 1 and 2 on the next two pages.

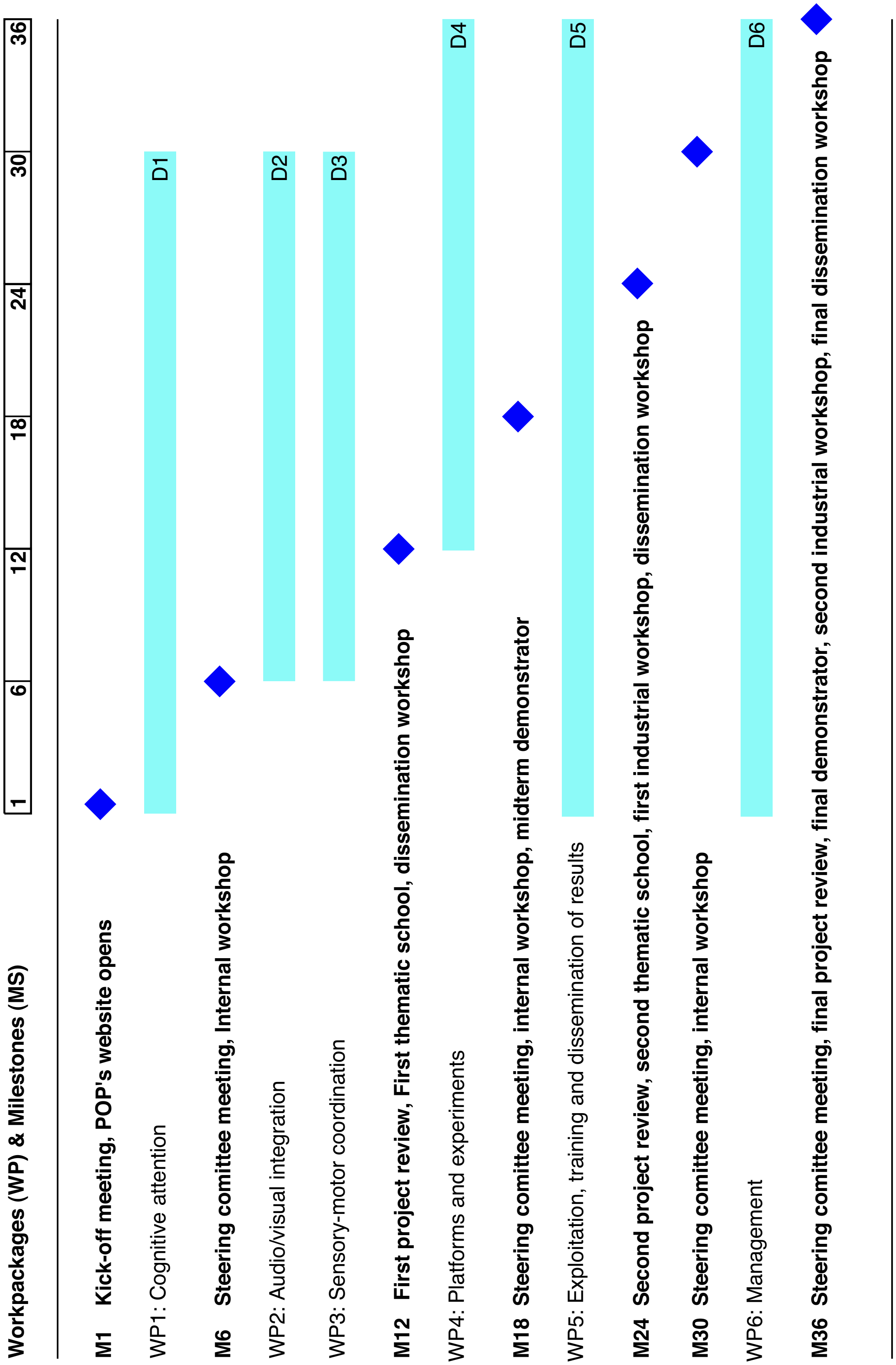

Figure 1: Gant chart.

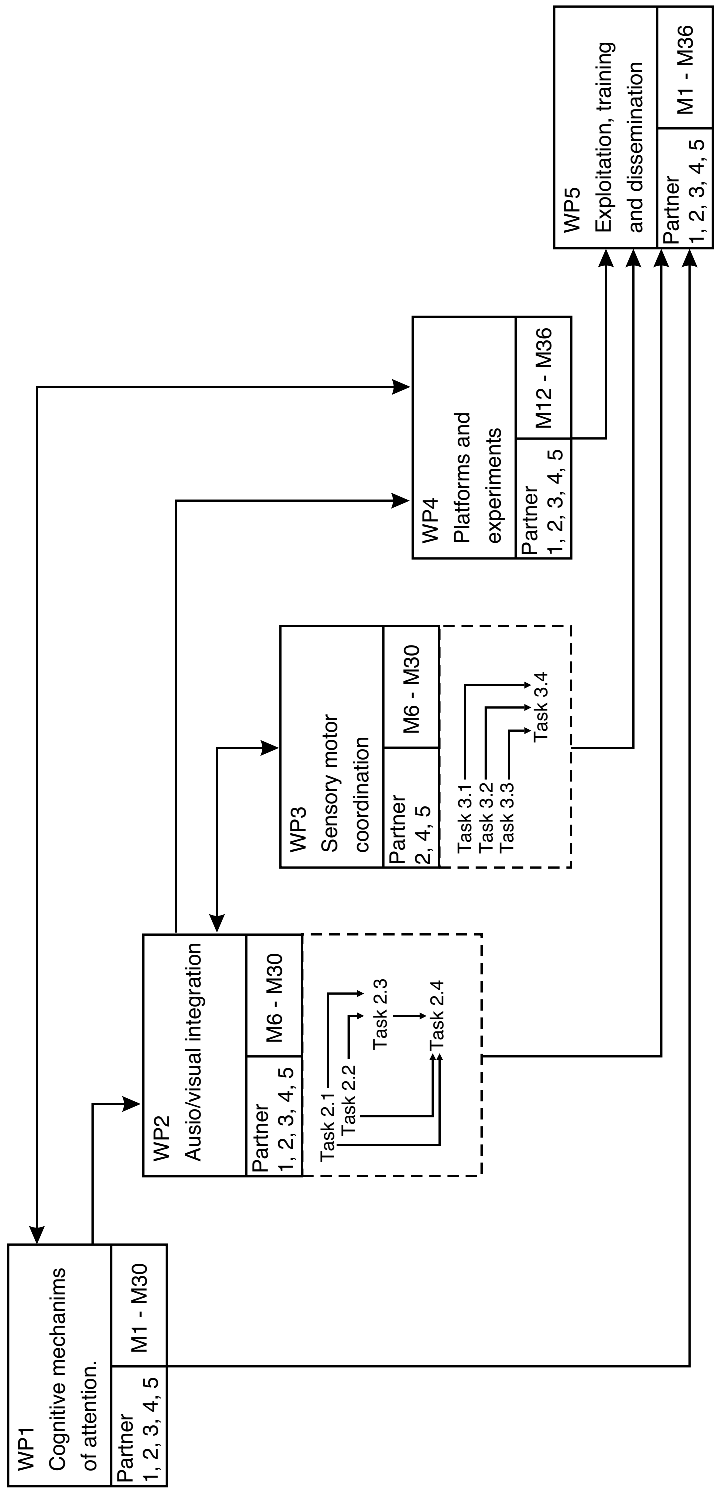

Figure 2: Pert chart.

6.4 List of milestones

Milestone	Month	Description
M1	1	Kick-off meeting POP's website opens First POP tutorial session WP1, WP5, WP6 start
M6	6	Steering committee and technical meetings WP2 and WP3 start Second POP tutorial session First internal workshop
M12	12	Steering committee and technical meetings First project review First thematic school Dissemination workshop The partners specify the project's demonstrator WP4 starts
M18	18	Steering committee and technical meetings Second internal workshop
M24	24	Steering committee and technical meetings Second project review Midterm demonstrator available Second thematic school First industrial workshop Dissemination workshop
M30	30	Steering committee and technical meetings Third internal workshop WP1, WP2, WP3 completed
M36	36	Steering committee and technical meetings Final project review Final demonstrator WP4, WP5, WP5 completed Second industrial workshop Final dissemination workshop

Table 3: *The milestones of the POP project.*

6.5 List of deliverables

WP/Task	Deliverable	Month	Description	Type/Diss.
WP1	D1			
T1.1				
T1.2				
T1.3				
T1.4				
T1.5				
T1.6				
T1.7				
WP2	D2			
T2.1				
T2.2				
T2.3				
T2.4				
T2.5				
T2.6				
WP3	D3			
T3.1				
T3.2				
T3.3				
T3.4				
WP4	D4			
T4.1				
T4.2				
T4.3				
T4.4				
T4.5				
T4.6				
T4.7				

Table 4: *The list of project deliverables: WP1, WP2, WP3, and WP4.*

WP/Task	Deliverable	Month	Description	Type/Diss.
WP5	D5			
T5.1 T5.2 T5.3 T5.4 T5.5 T5.6 T5.7				
WP6	D6			
T6.1 T6.2 T6.3 T6.4 T6.5 T6.6				

Table 5: *The list of project deliverables: WP5 and WP6.*

6.6 Detailed description of work-packages

WP1: Cognitive mechanisms of attention

WP1	INRIA	UOS	UKE	FCTUC-ISR	USFD	Total
Person-months	25	25	30	10	20	110
Start/End	1/30					
Leader	UKE (partner 3)					
Deliverables	D1					
Milestones	M1, M12, M24, M30					

Description of work. The general objective of this workpackage is to study a model of perceptual attention at a fundamental level (behaviroual and cognitive) using the most modern investigation techniques and to propose a computational paradigm for audiovisual attention. More precisely, we will investigate the contribution of major visual and auditory features, interaction of features, cross-modal interactions and the nitration of bottom-up and top-down signals in the control of overt attention. Using the combined expertise in the consortium we will apply psychophysical, electrophysiological, and imaging methods to the same paradigms and use theoretical methods to arrive at a quantitative computational model of this central human cognitive ability.

Task 1.1. Attentional selection in the visual and auditory domain by single features.

This task will characterize the contribution of visual (luminance, colour, motion, orientation, spatial frequency and stereoscopic depth) and auditory features in guidance of overt attention. This covers all visual features where strong experimental evidence is available that they contribute to the guidance of attention in humans performing pop-out tasks. In most models of overt attention luminance contrast is a major contribution to saliency. With respect to colour careful calibration and relevant natural stimuli are an important aspect. Although the outstanding importance of motion cues in capturing attention it is well recognized, little data is available. Here, we will use natural motion stimuli and apply the algorithms developed in WP2 for determination of motion vectors. The feature of orientation will be investigated with well established procedures. All these features will be considered at different spatial scales in a resolution hierarchy. Binocular disparity is an important cue for depth estimation and particularly effective at close distances, i.e. in the range where object can be manipulated. A comparison with the selection of fixation points under normal conditions allows separating the contribution of disparity information and pictorial depth cues. Importantly, absolute disparity, disparity contrast and depth information available from the laser scan data will be evaluated and compared. Very little behavioural or computational work has investigated the nature of attention in audition. We will investigate whether attentional processes in the two modalities behave similarly, and what the relevant auditory features are guiding overt attention.

Task 1.2. Multi-feature interactions during attentional tasks.

The interaction of different features is investigated by information theoretic methods. We will determine the influence of any single feature on the relation of another feature to gaze movements in terms of the conditional probabilities and mutual information. This model free approach facilitates comparison to experimental data (Tasks 1.5 and 1.6) and technical implementations (WP4). We will factor out correlations of features in natural stimuli by computing the partial correlations. Finally, matching the psychophysical data against the image statistics will lead to a quantitative description of the probability to fixate a region as a function of local features.

Task 1.3. Attention on purpose.

This task investigates the context dependence of overt attention and its relation to conscious perception. Firstly we study the modulation of bottom-up signals to the guidance of overt attention by a set of different task instructions. We have already identified a number of different tasks that result consistent behaviour across subjects. We will compare global statistics of eye movements, such as average saccade latency and length and area covered, and the relation to local features in different contexts. An emphasis will be place on the role of depth cues like binocular disparity in the interaction with task instructions of manipulating objects within reach. This will elucidate the degree of semantic processing before visual information feeds a saliency map. Secondly we exploit the setup for stereoscopic presentation to induce binocular rivalry with simultaneous measurements of eye movements. We will determine the contribution of the suppressed image to the selection of fixation points and compare this with physical superpositions of images. In this way we determine the degree of semantic processing of visual stimuli when contributing to the control of overt attention.

Task 1.4. Multimodal interactions in overt attention.

Here we study the integration of simultaneously presented auditory and visual cues for the control of human overt attention. We will present multi-modal stimuli developed in WP4 in congruent, concongruent and isolated conditions. This allows comparing and evaluating different hypotheses pertaining to the kind of interaction of multi modal signals (additive, multiplicative, maximum). Integrating the results we will construct the first multimodal model of attention. The combination of auditory cues to new events with visual cues will help to produce a more robust attentional component which is capable of behaviours such as preemptive visual orienting based on unseen but heard auditory cues as well visual identification of novel auditory stimuli.

Task 1.5. Neuronal interactions during overt attention.

Here we study the integration of simultaneously presented auditory and visual cues for the control of human overt attention. Attentional effects will be tested using 128-channel EEG recordings. Data analysis will put particular emphasis on so-called induced rhythms, i.e., EEG activity components that are not phase-locked to stimulus transients. The hypothesis will be tested that cross-modal interactions during perception and attentive stimulus processing occur specifically in the gamma-band. Auditory and visual stimuli will be implemented using commercially available stimulus-presentation packages. Prior to EEG measurements, perceptual and attentional effects of the stimuli will be tested in behavioural measurements. While in part of the experiments, subjects will view stimuli with central fixation; in later series of measurements we will allow free viewing and perform a saccade-triggered analysis of the EEG data. This is of particular interest when moving from more simple to complex visual stimuli. This will require methodological preparatory steps, since a combined setup allowing EEG recording and simultaneous eye tracking needs to be developed (WP4).

Task 1.6. Identification of brain regions involved in multisensory attentional control using fMRI.

Using the same stimuli as in previous tasks, we will carry out MR measurements to map location and degree of activation of low-level unimodal cortical areas as well as polymodal association areas. In addition, we will test which sites are undergoing attentional modulation. This approach will clarify where saliency maps are located in the brain that subserve attentional control, and whether such candidate sites are involved in attentional control irrespective of sensory modality. Moreover, these experiments can be used to guide source localization for the EEG data obtained in task 1.5.

Task 1.7. Computational model of visual and auditory attention

Attention serves two roles in perception. First, it is required to direct processing capacity towards new events. Second, it helps to track an ongoing acoustic or visual source such as the utterances of a speaker in a background of competing sources of both visual and auditory nature. This task will develop a computational model capable of both responding to new visual/acoustic stimuli and following existing ones in a mixture. The bottom-up component of the model will utilize an "old+new" principle in determining whether incoming information should be considered as part of an ongoing hypothesis or as the start of a new event. For the top-down component, an existing probabilistic decoder capable of tracking single visual and audio sources in a mixture will be extended to handle attentional focus via the dynamic loading of detailed prior models for the attended source together with a more generic model for the background.

WP2: Integration of visual and auditory cues

WP2	INRIA	UOS	UKE	FCTUC-ISR	USFD	Total
Person-months 30	20	30	25	20	125	
Start/End	6/30					
Leader	USFD (partner 5)					
Deliverables	D2					
Milestones	M6, M12, M24, M30					

Description of work. Extraction and integration of visual and auditory cues are fundamental building blocks for understanding perception. Cues must be properly extracted from the raw data and their individual merit and reliability must be characterized and quantified. Eventually they must be combined together within a probabilistic framework based on their geometric and temporal properties. We will concentrate on both visual and auditory cues and, moreover, on the interaction of spatial and temporal stimulus features. First, we will study stimuli individually, second we will propose theoretical models for spatial fusion and for temporal fusion, third we will devise a statistical method for sensory integration, and fourth we will study the biological mechanisms of sensory integration. Regarding possible biological mechanisms, this WP aims at testing the hypothesis that coherence of neural signals can change during the formation of cross-modal conjunctions or during cross-modal interference in human subjects. We assume that changes of neural synchrony between areas processing different stimulus modalities will occur in instances of cross-modal feature binding and, moreover, that temporal patterning will be influences by cross-modal shifts of attention.

Task 2.1. Extraction of visual cues.

Partner 1 and Partner 4 will make available algorithms and software for extracting visual cues. In particular Partner 1 has recently developed a stereo method that computes a dense depth (disparity) map under the form of a 3D set of surface patches. Partner 4 has recently developed an optical flow method that computes a 2D velocity field in real time. These approaches are embedded into a probabilistic framework such that a reliability measure of the depth and motion maps will be provided as well.

The computing power available today allows, in principle, real-time implementations (at 30 frames per second) of these algorithms. Therefore, particular emphasis will be but on achieving and delivering efficient software libraries.

The novel work also to be carried out within this task will be to combine a 3D depth map with a 2D velocity field in order to hypothesize a 3D velocity field. This *instantaneous* depth-velocity temporal/geometric representation, with its associated measure of confidence, will be used to build a primary stimulus-based visual saliency map which in turn will be used as input by tasks 2.3 and 2.4.

Task 2.2. Extraction of auditory cues.

Starting from an existing model of peripheral auditory processing, the goal of this task is to form a description of the incoming binaural signals in terms of dominant spectro-temporal elements tagged with their spatial position. Interaural time and level differences will be used to localise the acoustic energy at a given time and frequency. However, these cues are unreliable in natural reverberant conditions, and robust estimates can only be achieved by integrating across time and frequency. Cues such as onset time and harmonicity will be used to group energy into larger time-frequency regions which can be robustly localized. Prior knowledge of the location of sources, gained by tracking the sources in the audio-visual domain, will be fed back into the primitive processing stages to reduce ambiguity and improve the clarity of acoustic event descriptions. Attention processing and sensory-motor coordination (described in WPs 1 and 3) will ensure that for much of the time the sound source of interest will be at a known location with respect to the head (typically at an azimuth of 0 degrees). Active audiovisual strategies to improve the acoustic signal-to-noise ratio for the sound source of interest will feed into this task.

Task 2.3. Geometric and probabilistic fusion of spatial visual and auditory cues.

The goal of this task will be to represent both visual and auditory cues in a common sensor-centered reference frame and to hypothesize associations between visual events (such as a moving object) and auditory events (such as the presence of a sound source). This will consist in an optimal estimation (using the maximum likelihood estimator) of the geometric transformation between the 3D visual space and the 3D auditory space, given the initial estimates provides by Tasks 2.1 and 2.2. This will allow to associate sound sources with a visual depth map and to predict spatial locations where both motion cues and auditory cues are likely to be simultaneously present.

Task 2.4. Synchronization and fusion of temporal visual and auditory cues.

This task will take input from tasks 2.1, 2.2, and 2.3 and will take from granted the fact that visual and auditory information has been spatially integrated. The goal of this task is to investigate methods and algorithms allowing to correlate temporal auditory events with visual motion detection. If several sources of sound are present in the scene, then the task of grouping auditory events emitted by the same source into a coherent temporal auditory signal (such as a speech signal) is a tremendously difficult problem. On one side, there is evidence from neurobiological data and from psychophysical experiments that *auditory attention* can be positively biased by visual motion detection. On the other side, the visual information is evenly and continuously distributed in space and it is difficult to decide whether a visual object is more relevant than another visual object; this is therefore a case where localization and identification of auditory sources may help vision.

It is therefore crucial to develop a formalism allowing the synchronization between time-varying motion events and auditory events. The output of this task will be an audiovisual "tracker" – able to provide the 3D trajectory of an audiovisual object.

Task 2.5. Development of statistical methods for cross-modal integration.

An appealing approach for cross-modal integration could be to rely on statistical methods only. This has been a successful approach in speech recognition where 1D hidden Markov models (HMM) have extensively been used to associate sound signals to identities of phonemes. 2D HMMs have also been applied, with some success, to visual data in order to associate "labels" and visual cues, where labels may refer to object parts, human gestures, facial expressions, etc. Researchers in computational neuroscience believe that statistics underlie many mental processes, such as perception, thinking, and acting.

In more detail, cross-modal integration can be achieved by modelling the correlated dynamics of the acoustic and visual features of an event. An approach based on factorial HMMs, developed for modelling acoustic mixtures, will be extended. Here, an independent HMM is employed to model the energy contribution of each sound source. By employing primitive processes to segment the time-frequency plane into a small number of source fragments (task 2.2), the decoding search space can be controlled. These techniques have been demonstrated in simultaneous-speaker speech recognition tasks and work well when the speakers are acoustically distinct. However, performance may be greatly enhanced if the acoustic models are supplemented with features from the visual modality. The associations between modalities that are learnt by training on isolated speakers mean that the acoustic fragments are bound to the source with the matching visual model. Although techniques for the visual parameterization of speech are well developed, it is less clear how other sounds can be described visually. One of the challenges in this task is to generalize to non-speech events.

These developments should lead to such paradigms as *seeing a sound* which form the bases of the interactions that exist between visual attention and auditory scene analysis.

Task 2.6. Biological mechanisms for multi-sensory integration.

Objective of this task is to identify biological mechanisms underlying multisensory integration and to derive relevant algorithmic principles than will be used in the technical implementation. Cross-modal integration will be investigated at the cellular level using multielectrode recordings in anesthetized animals. Ferrets will serve as a model system. The experiments will address putative saliency map sites, corresponding to the superior colliculus (SC) and the anterior ectosylvian cortex (AEV), a multimodal cortical area. Multielectrode recordings with visual-auditory stimuli will be performed to characterize neuronal receptive fields, response properties and dynamic neural interactions in these putative saliency maps. The stimulation paradigms will include unimodal stimulus presentations as well as cross-modal settings where the spatial and temporal relation between the visual and auditory stimulus are varied. Properties of the multi-modal integration process will be investigated with information-theoretic methods. As a baseline, we will use correlation analysis and autoregressive models that are linear methods. As a next step we use non-linear methods including entropy and mutual information. These will be applied to the physiological data and the signals of both modalities. Importantly, we will partition the information content with respect to rates, signal correlations and noise correlations. These measures allow a characterization of the integration of information while keeping the modality specific information accessible.

WP3: Sensory-motor coordination

WP3	INRIA	UOS	UKE	FCTUC-ISR	USFD	Total
Person-months	-	25	6	30	20	81
Start/End	6/30					
Leader	UOS (partner 2)					
Deliverables	D3					
Milestones	M6, M12, M24, M30					

Description of work. The goal of this work-package is the design and development of principles, methods and algorithms to allow the coordination of the motor activities with the sensor measurements. The rationale is to be able to direct attention towards scene areas with activity relevant to the task at hand. In this project we deal with two sensor modalities, namely vision and hearing. The control of the positions and orientations/poses of the robotic device will be performed using the results of work-packages 1 and 2. The information used to locate the focus of attention will be extracted from both sounds and videos. Since the motion of the cameras and microphones is known, it will be used to predict the positions and orientations of the relevant scene elements in the images. Such predictions are an essential element in processes such as smooth pursuit. Prediction in this type of systems can be regarded as performing a kind of information feed-forward. Active listening will be used to improve the robustness of the motor coordination. Gaze and oculomotor control will be performed using saccades, smooth pursuit and vergence. As a result of the work developed in this work-package the robotic device will be able to respond to both the visual and the auditory stimuli. Coordination is obtained by synchronizing the image and sound measurements with the motor control.

Task 3.1. Sensor, geometric, and dynamic calibration.

To perform the control of the robotic device the sensors have to be calibrated. That means the estimation of the intrinsic parameters of the cameras and their relative positions and orientations. The control of the active system requires the knowledge of the inverse kinematics. The estimation of the inverse kinematics implies the geometric calibration of all the system. The overall system has also to be dynamically calibrated so that the estimates for the overall bandwidth can be computed. The outputs of this task are: –Software/algorithms for geometric calibration of the cameras (including intrinsic parameters and relative pose) and –The inverse kinematics of the robotic device;

Task 3.2. Oculomotor control.

The oculomotor control will be performed using outputs of work-package 2. Feedback and predictive feedforward control will be used. The motions of the audiovisual head will be made up of three visual behaviors: –Saccadic motions; these are fast responses to the detection of an event and are performed in open loop, i.e., visual information will not be used to control motion. At low-level the motor bandwidths will be used to define maximum speeds and accelerations; –Smooth pursuit; these are motions that use visual measurements to define the positions and velocities errors on a feedback loop. On the other hand to compensate for the delays due to the visual processing predictive control will be used to update the positions and velocities of the motors; –Vergence; the goal of this behavior is to have the event or element that is the source of attention located in the center of both images in a binocular active system. Position (stereo) and velocity (motion) disparities will be used to guarantee vergence on the same visual element. Therefore control is performed using both measured positions and estimated velocities.

The outputs of this task will be the implementations of the visual behaviors described above.

Task 3.3. Audio-motor control.

Active listening has enormous potential to improve the robustness of both auditory and visual scene analysis. Within the constraints imposed by the active audiovisual head, this task will investigate, both behaviorally and computationally, a range of hypothesized active listening strategies. Possible strategies include: moving to a location which maximizes the signal-to-noise ratio (SNR) (e.g. moving towards a source of interest); orienting the head to a position which maximizes SNR (e.g. using head shadow to attenuate the strongest competing source, or moving the head to face the source in order to reduce ambiguities in cues to location); moving to get an un-occluded view of the target source (e.g. for speech reading); and moving away from walls and hard surfaces to reduce the effect of reverberation. Since relatively little is known about the strategies used by humans to improve the identifiability of a target in noise, the behavior of participants in adverse environments such as crowded meetings will be observed and annotated in an attempt to better understand the types of active strategies used and the situations in which they are employed. Findings from this investigation will feed into algorithm development for active listening which in turn will be implemented and demonstrated in the active audiovisual head.

The outputs of this task will be the development and implementations of: –active listening strategies; –a set of auditory behaviors for directing the head towards the source of interest;

Task 3.4. Cross-modal motor control.

In this task the results of work-packages 1 and 2 as well as the results of tasks 3.1, 3.2, and 3.3 will be used to develop methods for motor control using information from both vision and audition. In this task the control algorithms will be designed taking into account the specific nature of the cross-modal measurements. Problems to be dealt with in this task are:

- Identification of the behaviors for actively directing the attention of the robotic device using both vision and audition. These behaviors can be based on the visual behaviors adaptively changed to use audition. Attention can be directed to a visually occluded target as well as to a silent target. When measurements from both modalities occur and are associated to the events feedback control will use the measurements estimated from both modalities;

- Development of the feedback control whose task is to direct the attention towards the positions and velocities of selected world events.

- Feedforward strategies to account for processing delays. Since these delays will be highly variable adaptive predictive techniques will be used;

WP4: Development of methodological and experimental platforms

WP4	INRIA	UOS	UKE	FCTUC-ISR	USFD	Total
Person-months	20	5	6	40	20	91
Start/End	12/36					
Leader	FCTUC-ISR (partner 4)					
Deliverables	D4					
Milestones	M12, M18, M24, M30					

Description of work. This work-package will integrate the results of WP 1, 2, and 3 and will demonstrate these results. There will be a robot prototype designed, developed, and built by the partners. This prototype will integrate both neurophysiological findings (tested on special-purpose platforms) and computational developments. A mid-term (month 24) and a final demonstrator (month 36) will be developed as well.

Task 4.1. Mid-term and final demonstrators.

Under this task we will specify the mid-term and final project demonstrators based on WP1, WP2, and WP3 and which will be implemented onto the hardware and software platforms described below. First we will conduct a survey in order to target the set of applications that are likely to take advantage of the POP's results. The purpose of the POP's partners is to target a proof-of-concept demonstration.

Task 4.2. Development of an audio-visual robot platform.

The platform will be composed of hardware – an audio-visual "head" (cameras, microphones, and actuators) mounted onto a standard mobile robot, and of software – developed within WP2 and WP3 and combined with existing software libraries and packages from the POP partners.

A complete robot platform (hardware and software) will be available at INRIA (partner 1). The audiovisual head will be built by FCTUC-ISR (partner 4) and the technical specifications will be provided to the other partners such that they can easily duplicate the prototype. USFD (partner 5) will specify the auditory equipment for the other partners and will provide the software packages necessary to operate this equipment. INRIA (partner 1) will specify the visual equipment for the other partners and will provide the correspondinc software packages. UOS (partner 2) and UKE (partner 3) will provide specifications for the type of data sets that such a platform should provide for their experiments (see task 4.5 below).

Task 4.3. Methodological platform for investigation of biological mechanisms.

In this task the techniques for simultaneous eye tracking and EEG recording are developed. We will modify the mechanical setup of the head-mounted eye tracking system and apply electrical shielding to make it compatible with multichannel EEG recordings. As eye movements induce large signals into the EEG recordings, in a second equally important step techniques have to be developed to separate these from the EEG signals proper. These will be based on on-line feedback of the tracked eye-position to the electrophysiological recordings, independent component analysis and time-delayed auto regressive models.

Task 4.4. Development of a stimulus database for multisensory studies in electrophysiology, psychophysics and robots.

The stimulus materials used in WPs 1 and 2 comprise photographs, videos and sounds of isolated objects and naturalistic scenes. Important technical aspects are faithful color calibration, inclusion of disparity information by stereoscopic images and validation by laser range scans. For the later experiments are ratings of the emotional valence, association with different possible actions and subjective similarity for each record in the database, and the option to use congruent and incongruent multi-modal stimuli. "Normalized" stimuli are matched in 2nd order statistics, but retain differences in higher order structure.

Task 4.5. Development of a platform for testing biologically-based algorithms.

The models and algorithms hypothesized as a result of the psychophysics and neurophysiology will be tested with the audiovisual head (task 4.2). For that purpose the robot platform will be able to respond to sequences of images and sounds recorded during psychophysical experiments. These recored data will be input to the system and the system response (the motor encoders) will be recorded. As a result, the effects of the algorithms can be evaluated and compared to the psychophysical data. The output of this task will be a package of software enabling the operation of the robotic device with recorded data (images and sounds).

WP5: Exploitation, training, and dissemination of results

WP5	INRIA	UOS	UKE	FCTUC-ISR	USFD	Total
Person-months	3	3	3	3	5	17
Start/End	1/36					
Leader	USFD (partner 5)					
Deliverables	D5					
Milestones	M1, M6, M12, M24, M36					

Description of work. This work-package implements the exploitation, training, and dissemination activities through the following tasks.

Task 5.1. Creation and maintenance of a website.

See paragraph 5.2, page 29 for a detailed description.

Task 5.2. Consortium publication activities.

The exploitation-dissemination manager will coordinate the publication efforts of the consortium by making sure that the scientific results go out for publication and that papers are submitted to appropriate journal and conferences. The work-package 5 manager will also have the responsibility to verify that the material to be published is consistent with the IPR policy as agreed by the consortium.

Task 5.3. Industrial liaison.

During the lifetime of the project two industrial meetings will be organized in order to promote the results and outcomes of the project through demonstrations and seminars specifically designed for an industrial audience.

This task will also coordinate the transfer of knowledge and of technology from the partners towards companies that manifest their interest in the projects' outcomes. The work-package manager will make sure that these partner/company relationships are in line with the management of knowledge and intellectual property rights.

Task 5.4. Links with other projects and consortia.

The dissemination and training activities will not be carried out in isolation but in collaboration and coordination with other EC and national projects.

A list of EC projects has already been identified, i.e., section 3, page 10. The work-package 5 manager will contact these projects and consortiae will organize meetings in order to establish bi-lateral and multi-lateral formal activities such as workshops, tutorials, and seminars, and thematic schools.

Task 5.5. Exploitation.

The POP researchers and engineers will develop software, hardware, an integrated robot platform, as well as various laboratory devices for carrying out psychophysical experiments. Proper exploitation and dissemination of these outcomes needs protection of intellectual property rights. The latters will be effective through licensing and patenting and these activities will be coordinated by the exploitation-dissemination manager in coordination with the steering committee and with legal offices of each one of the partners.

Task 5.6. Training.

The POP double-cultural topics of research will be attractive to talented postdoctoral researchers. However, very few will possess the required interdisciplinary skills from the outset. The application of bio-inspired approaches to engineering problems requires the integration of ideas from a number of disciplines, ranging from cognitive science and psychology to computation, mathematics and neurobiology. Few individuals can claim expertise in more than one of these areas, and even within a single discipline such as neurobiology, far too little work crosses traditional boundaries. Consequently, all project researchers will be given the opportunity for cross-training in disciplines in which project partners have expertise. They will be encouraged to develop the technical and collaborative skills needed to solve complex problems. A training programme with the following elements will be implemented:

- Postdoctoral researchers will be expected to spend a significant proportion of the first 18 months outside their nominal home institution in order to transfer their expertise and resources to other partners, and to complement and broaden their existing skills set. It is expected that the prospect of working in several labs will be attractive to applicants
- Doctoral students will have the opportunity to spend part of their time in a partner institution and to attend the thematic schools organized by the POP partners in coordination with the beforementioned IP project, networks and actions.
- Senior scientists involved in POP will be encouraged to spend periods ranging from short visits to longer sabbaticals in partner institutions.
- A series of tutorials will be organised and delivered at the internal and dissemination workshops that are planned. Tutorials will cover foundational elements of the discipline (e.g. auditory scene analysis, visual attention, active stereo) and will offer training in relevant hardware (e.g. active vision head) and software.

In particular there will be 2 thematic schools (at months 12 and 24) that will be organized by the POP coordinator in conjunction with the thematic schools planned within the RTN action VISION-TRAIN (coordinated by the same person as POP).

Task 5.7. Workshops

The work-package 5 manager will coordinate the organization of a number of internal, dissemination, and industrial workshops: 3 internal workshops (at months 6, 18, and 24), 3 dissemination workshops (at months 12, 24, and 36) and 2 industrial workshops (at months 24 and 36).

Both the dissemination and the industrial workshops will be organized in collaboration with the IP AMI, with the RTN action VISIONTRAIN, and with the NOE WhatNext (if granted).

WP6: Management

WP6	INRIA	UOS	UKE	FCTUC-ISR	USFD	Total
Person-months	5	3	3	3	3	17
Start/End	1/36					
Leader	INRIA (partner 1)					
Deliverables	D6					
Milestones	M1, M6, M12, M18, M24, M30, M36					

Description of work. The overall management structure of the project is described in detail in **section 5 at page 28**. The management activities will be carried out under the following tasks:

Task 6.1 Communication with the EC.

Collect, monitor and integrate all the technical, administrative, and finacial data from the partners and prepare appropriate documents for the European Commission: management reports, progress reports, final report, cost and financial statements, deliverables, etc.

Task 6.2. Organisation of EC reviews and audits.

The project coordinator will be responsible for organizing the audits concerning all the aspects of the project: technical audits including the annual project reviews, financial audits, as well as any other audits that the EC wishes to organize.

Task 6.3. Organisation and preparation of internal meetings.

The project coordinator, in coordination with the site managers and with the work-package managers will prepare and organize the Steering Committee meetings and the technical meetings.

Task 6.4. Legal, financial, and administrative management.

The project coordinator with the manager of work-package 6 will be jointly responsible of this task: receive payments from the EC, transfer payement to the partners, prepare the consortium agreement, obtain the audit certificates when required, etc.

Task 6.5. Technical management.

The work-package managers, supervised by the project coordinator will be in charge of the monitoring, the coordinating, and the controlling of the scientific and technical progress of the project. They will be responsible for preparing the technical deliverables of the project as well as of the annual technical reports and final report of the project.

Task 6.6. Risk, problem, and conflict management.

Under the leadership of the project coordinator, the work-package managers will be responsible for smoothly carrying out the work during the duration of the work-packages. Whenever needed they will take the initiative to organize management meetings at the work-package levels in order to anticipate any risks, problems and/or conflicts:

- Detect scientific and technical difficulties in achieving a work-package or a work-task and provide alternatives and solutions to the problems encountered.
- Anticipate problems with the termination of a work-package and offer suitable solutions before the situation results in delays or dead-ends.
- Foresee fluid information sharing between the partners (background research, outcomes of a work-task, etc.) and avoid technical and bureaucratic barriers to information access.
- Anticipate the end of contract of a researcher, changes in personnel, reallocation in human resources by a partner, leave of absence of a researcher, temporarily absence of a researcher, conflicts between researchers, etc.

If a satisfactory solution cannot be found on a rapid and cordial basis, the work-package manager will refer to the project coordinator and to the site managers who will ask the Steering Committee to meet and to take a decision.

6.7 Mobilisation of resources

In order to accomplish its objectives, the POP project will require certain resources. The break down of these resources for each partner and for each category (RTD and innovation, demonstration, management) is detailded in form A3. In more detail, these resources can be categorized in personnel, consumables, durable equipment, travel, and management.

The EC and the project participants (according to their respective cost models) will jointly fund these resources. The funding effort of partner 1 (INRIA) is equal to the requested EC funding (full costs). The funding efforts of partners 2, 3, 4, and 5 (UOS, UKE, FCTUC-ISR, and USFD) do not explicitly appear on form A3 since these partners adopted an additional cost model. Their funding efforts sum up to a total of 142 person-months.

The grant that POP requests from the EC is 1,916,000 euros. Personnel resources will take the lion share of the budget and will be basically aimed at RTD and innovation activities. At the other end, the management budget, 110,000 euros, represents 5.7% of the requested budget. In detail:

- **Salaries** will represent 72% of the requested EC funding. POP will hire 2 PhD researchers and 6 post-doc researchers for the whole duration of the project, i.e., 288 person-months. In addition, There will be 11 person-months directly allocated to the exploitation and training activities.
- **Consumables** will represent 6% of the requested EC funding. These resources will be used for the development of the robotic platform, for interfacing the audio and video sensors with the computers and with the actuators, and for fMRI sessions (subject costs, maintainance, micro-electrodes).
- **Durable equipment** will represent 11% of the requested EC funding. These resources will be used for the development of an audio-visual head, the acquisition of a mobile robot, the acquisition of an eye tracker for use with an existing EEG and for psychophysics experiments, and computers.
- **Travel** will represent 5.4% of the requested budget. These resources will be used for cross-visits between the partners, project meetings and audits, participation to thematic schools and to training activities organized by other EC projects, and for attending workshops and conferences.
- **Management** will represent 5.7% of the requested budget and these resources will be used for the activities described in WP6.

7 Other issues

In-vivo data on neuronal dynamics will be used to develop algorithms for interface programming and robot control. Studying the interactions between neurons processing visual and auditory information will provide critical insights into mechanisms underlying multisensory integration. WP 2 (Integration of visual and auditory cues) will conduct in-vivo experiments with anesthetized ferrets. These experiments will be conducted by partner 3 (UKE). All experimental procedures will be conducted in accordance with the latest revised version (12 April 2001) of the German Animal Protection Law. Experiments will comply fully with European Community guidelines (EUVD 86/609/EEC) regulating the care and use of laboratory animals. Every in-vivo experiment will be submitted for approval to the local authority (Regierungspräsidium Hamburg) who will assess (i) whether the anticipated benefits justify the use of animals; (ii) the number of animals used in each experiment; (iii) the procedures adopted to ensure that animal suffering is minimised. Most experiments will be performed using procedures that are similar to those already in use at University Hospital Hamburg-Eppendorf, and that have already been approved by the local authorities in Hamburg. We certify that we will inform the EC of local authority approval before the start of the in-vivo experiments.

As part of WP1 and WP4, measurements are performed on human subjects using EEG, fMRI and psychophysical methods. All techniques are non-invasive and do not involve any risk for participants. All participants are volunteers. They are informed about the goals and procedures of the study and give their written consent. Participants always have the possibility to terminate the experiment without given reasons. All studies are performed in accordance with the ethical standards laid down in the 1964 declaration of Helsinki (most recent pass: Edinburgh, Scotland, October 2000).

A. Proposers are requested to fill in the following table

Does your proposed research raise sensitive ethical questions related to:	YES	NO
Human beings	×	
Human biological samples		×
Personal data (whether identified by name or not)		×
Genetic information		×
Animals	×	

B. Proposers are requested to confirm that the proposed research does not involve:

Research activity aimed at human cloning for reproductive purposes,

Research activity intended to modify the genetic heritage of human beings which could make such changes heritable

Research activity intended to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer,

Research involving the use of human embryos or embryonic stem cells with the exception of banked or isolated human embryonic stem cells in culture[1]

Confirmation:	YES	NO
The proposed research involves none of the issues listed in section B	×	

[1]Applicants should note that the Council and the Commission have agreed that detailed implementing provisions concerning research activities involving the use of human embryos and human embryonic stem cells which may be funded under the 6th Framework Programme shall be established by 31 December 2003. The Commission has stated that, during that period and pending establishment of the detailed implementing provisions, it will not propose to fund such research, with the exception of the study of banked or isolated human embryonic stem cells in culture.

References

[1] Benucci A, Verschure PFMJ, and König P. On the existence of high-order correlations in cortical activity. *Phys Rev E Stat Nonlin Soft Matter Phys.*, 68:041905, 2003.

[2] J.P. Barker, M.P. Cooke, and Ellis. D.P.W. Decoding speech in the presence of other sources. *Speech Communiation*, (45):5–25, 2005.

[3] J. Barreto and H. Araújo. A general framework for the selection of world coordinate systems in perspective and catadioptric imaging applications. *Int. Journal of Computer Vision*, 57(1):23–47, 2004.

[4] J. Barreto, J. Batista, and H. Araújo. Model predictive control to improve visual control of motion: Application in active tracking of moving targets. In *ICPR'2000-15th Int. Conf. on Pattern Recognition*, Barcelona-Spain, September 2000.

[5] J. Barreto, J. Batista, H. Araújo, and A. Almeida. Control issues to improve visual control of motion: Applications in active tracking of moving objects. In *Proc. of AMC'2000-6th Int. Workshop on Advanced Motion Control*, pages 13–18, Nagoya-Japan, March-April 2000.

[6] J. Barreto, P. Peixoto, J. Batista, and H. Araújo. Control performance issues in a binocular active vision system. In *Proc. of IROS'98–IEEE/RSJ Int. Conf. on Intelligent Robot and Systems*, pages 886–891. IEEE Press, 1998.

[7] J. Barreto, P. Peixoto, J. Batista, and H. Araújo. Improving 3d active visual tracking. In *ICVS99–First Int. Conf. on Computer Vision Systems*, pages 412–431, 1999.

[8] J. Barreto, P. Peixoto, J. Batista, and H. Araújo. Tracking multiple objects in 3d. In *IROS'99-IEEE/RSJ International Conference on Intelligent Robots and Systems*, Kyongju, Korea, October 17–21 1999.

[9] J. Barreto, P. Peixoto, J. Batista, and H. Araújo. Evaluation of the robustness of visual behaviors through performance characterization. In M. Vincze and G. Hager, editors, *Robust Vision for Vision-Based Control of Motion*, chapter 12, pages 145–161. IEEE Press, 2000.

[10] Joao Barreto, Frederick Martin, and Radu Horaud. *Visual Servoing/Tracking Using Central Catadioptric Images*, chapter VI, pages 245–254. Springer Tracts in Advanced Robotics 5. Springer Verlag, 2003.

[11] J. Batista, P. P. Peixoto, and H. Araújo. Real-time vergence and binocular gaze control. In *IROS97–IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Grenoble, France, September 1997.

[12] J. Batista, P. Peixoto, and H. Araújo. Real-time visual behaviors with a binocular active vision system. In *ICRA97–IEEE Int. Conf. on Robotics and Automation*, New Mexico, USA, April 1997.

[13] J. Batista, P. Peixoto, and H. Araújo. Visual behaviors for real-time control of a binocular active vision system. *IFAC Journal on Control Engineering Practice*, 5(10):1451–1461, 1997.

[14] J. Batista, P. Peixoto, and H. Araújo. Real-time active visual surveillance by integrating peripheral motion detection with foveated tracking. In *Proc. of the IEEE Workshop on Visual Surveillance*, pages 18–25, 1998.

[15] Körding KP Betsch BY, Einhäuser W and König P. The world from a cat's perspective - statistics of natural videos. *Biol Cybern*, 90:41–50, 2004.

[16] Edmond Boyer and Jean-Sébastien Franco. A hybrid approach for computing visual hulls of complex objects. In *Computer Vision and Pattern Recognition*, pages 695–701, June 2003. Madison, Wisconsin, USA.

[17] M. Brecht, R. Goebel, W. Singer, and A. K. Engel. Synchronization of visual responses in the superior colliculus of awake cats. *Neuroreport*, 12(1):43–7, 2001.

[18] M. Brecht, W. Singer, and A. K. Engel. Correlation analysis of corticotectal interactions in the cat visual system. *J Neurophysiol*, 79(5):2394–407, 1998.

[19] M. Brecht, W. Singer, and A. K. Engel. Patterns of synchronization in the superior colliculus of anesthetized cats. *J Neurosci*, 19(9):3567–79, 1999.

[20] N. A. Busch, S. Debener, C. Kranczioch, A. K. Engel, and C. S. Herrmann. Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clin Neurophysiol*, 115(8):1810–20, 2004.

[21] Kayser C, Körding KP, and König P. Learning the nonlinearity of neurons from natural visual stimuli. *Neural Computation*, 8:1751–1760, 2003.

[22] Kayser C and König P. Population coding of orientation in the visual cortex of alert cats - an information theoretic analysis. *NeuroReport*, (22):2761–4, 2004.

[23] Kayser C, Einhäuser W, and König P. Temporal correlations of orientations in natural scenes. *Neurocomputing*, 52-54:117–123, 2003.

[24] G. Celeux, S. Chrétien, F. Forbes, and A. Mkhadri. A component-wise em algorithm for mixtures. *Journal of Computational and Graphical Statistics*, 10:699–712, 2001.

[25] G Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition*, 36(1):131–144, 2003.

[26] S. Christy and R. Horaud. Euclidean shape and motion from multiple perspective views by affine iterations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1098–1104, November 1996.

[27] S. Chroust, M. Vincze, J. Barreto, and H. Araújo. Solutions for visual control of motion: Active tracking applications. In *Proc. of ICAR'2001-*10^{th} *Int. Conf. on Advanced Robotics*, Budapest-Hungary, August 2001.

[28] M.P. Cooke. *Modelling Auditory Processing and Organisation*. Cambridge University Press, 1993.

[29] M.P. Cooke. Glimpsing speech. *Journal of Phonetics*, 31:579–584, 2003.

[30] M.P. Cooke and D.P.W. Ellis. The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, (35):141–177, 2001.

[31] M.P. Cooke, P.D. Green, L. Josifovski, and A. Vizinho. Robust automatic speech recognition with missing and uncertain acoustic data. *Speech Communication*, 34:267–285, 2001.

[32] A. Coy and J.P. Barker. Recognising speech in the presence of a competing speaker using a 'speech fragment decoder'. In *Proc. ICASP 2005*.

[33] S. Debener, C. S. Herrmann, C. Kranczioch, D. Gembris, and A. K. Engel. Top-down attentional processing enhances auditory evoked gamma band activity. *Neuroreport*, 14(5):683–6, 2003.

[34] S. Debener, S. Makeig, A. Delorme, and A. K. Engel. What is novel in the novelty oddball paradigm? functional significance of the novelty p3 event-related potential as revealed by independent component analysis. *Brain Res Cogn Brain Res*, 22(3):309–21, 2005.

[35] D. Demirdjian and R. Horaud. Motion-egomotion discrimination and motion segmentation from image-pair streams. *Computer Vision and Image Understanding*, 78(1):53–68, April 2000.

[36] G. Dewaele, F. Devernay, and R. Horaud. Hand motion from 3d point trajectories and a smooth surface model. In T. Pajdla and J. Matas, editors, *8th European Conference on Computer Vision*, volume I of *LNCS 3021*, pages 495–507. Springer, May 2004.

[37] J. Dias, C. Paredes, I. Fonseca, H. Araújo, J. Batista, and A. Almeida. Simulating pursuit with machines: Experiments with robots and artificial vision. *IEEE Trans. on Robot. and Automat.*, 14(1):1–18, 1998.

[38] Y. Dufournaud, C. Schmid, and R. Horaud. Image matching with scale adjustment. *Computer Vision and Image Understanding*, 93(2):175–194, February 2004.

[39] A. K. Engel, P. Fries, and W. Singer. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci*, 2(10):704–16, 2001.

[40] A. K. Engel, P. Konig, A. K. Kreiter, and W. Singer. Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science*, 252(5010):1177–9, 1991.

[41] A. K. Engel and W. Singer. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci*, 5(1):16–25, 2001.

[42] F. Forbes and N. Peyrard. Hidden markov random field model selection criteria based on mean field-like approximations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(8):1089–1101, 2003.

[43] F. Forbes and A. E. Raftery. Bayesian morphology: Fast unsupervised bayesian image analysis. *Journal of the American Statistical Association*, 94(446):555–568, June 1999.

[44] J.S. Franco, C. Menier, E. Boyer, and B. Raffin. A Distrbuted Approach for Real-Time 3D Modeling. In *IEEE CVPR Workshop on Real-Time 3D Sensors and their Applications, Washington DC*, July 2004.

[45] P. Fries, S. Neuenschwander, A. K. Engel, R. Goebel, and W. Singer. Rapid feature selective neuronal synchronization through correlated latency shifting. *Nat Neurosci*, 4(2):194–200, 2001.

[46] P. Fries, P. R. Roelfsema, A. K. Engel, P. Konig, and W. Singer. Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc Natl Acad Sci U S A*, 94(23):12699–704, 1997.

[47] P. Fries, J. H. Schroder, P. R. Roelfsema, W. Singer, and A. K. Engel. Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *J Neurosci*, 22(9):3739–54, 2002.

[48] C.M. Gray, P. Konig, A.K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337, 1989.

[49] S. Harding, J.P. Barker, and G. Brown. Mask estimation for missing data speech recognition based on statistics of binaural interaction. Submitted to IEEE Trans. on Speech and Audio Processing.

[50] L. Hérault and R. Horaud. Feature Grouping and Figure-Ground Discrimination: A Recursive Neural Network Approach. In *Proc. of the IEEE International Joint Conference on Neural Networks*, pages 2606–2611, Singapore, November 1991.

[51] L. Hérault and R. Horaud. Figure-ground discrimination: a combinatorial optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):899–914, September 1993.

[52] R. Horaud and M. Brady. On the geometric interpretation of image contours. *Artificial Intelligence*, 37(1–3):333–353, December 1988.

[53] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, 14(4):525–532, August 1998.

[54] R. Horaud, D. Knossow, and M. Michaelis. Camera cooperation for achieving visual attention. Technical Report RR-5216, INRIA, INRIA Rhône-Alpes, Montbonnot, June 2004. To appear in Machine Vision and Applications.

[55] R. Horaud and Th. Skordas. Stereo matching through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(11):1168–1180, November 1989.

[56] R. Horaud and H. Sossa. Polyhedral object recognition by indexing. *Pattern Recognition*, 28(12):1855–1870, 1995.

[57] Körding KP Kayser C and König P. Processing of complex stimuli and natural scenes in the visual cortex. *Curr. Opin. Neurobiol.*, 14(4):468–73, 2004.

[58] R. Klinke, A. Kral, S. Heid, J. Tillein, and R. Hartmann. Recruitment of the auditory cortex in congenitally deaf cats by long-term cochlear electrostimulation. *Science*, 285(5434):1729–33, 1999.

[59] A. Kral, R. Hartmann, J. Tillein, S. Heid, and R. Klinke. Congenital auditory deprivation reduces synaptic activity within the auditory cortex in a layer-specific manner. *Cereb Cortex*, 10(7):714–26, 2000.

[60] A. Kral, R. Hartmann, J. Tillein, S. Heid, and R. Klinke. Delayed maturation and sensitive periods in the auditory cortex. *Audiol Neurootol*, 6(6):346–62, 2001.

[61] A. Kral, R. Hartmann, J. Tillein, S. Heid, and R. Klinke. Hearing after congenital deafness: central auditory plasticity and sensory deprivation. *Cereb Cortex*, 12(8):797–807, 2002.

[62] A. Kral, J. H. Schroder, R. Klinke, and A. K. Engel. Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Exp Brain Res*, 153(4):605–13, 2003.

[63] C. Kranczioch, S. Debener, J. Schwarzbach, R. Goebel, and A. K. Engel. Neural correlates of conscious perception in the attentional blink. *Neuroimage*, 24(3):704–14, 2005.

[64] B. Lamiroy, B. Espiau, N. Andreff, and R. Horaud. Controlling robots with two cameras: How to do it properly. In *Proc. of IEEE International Conference on Robotics and Automation*, pages 2100–2105, San Francisco, CA, April 2000.

[65] Siegel M and König P. A functional gamma-band defined by stimulus-dependent synchronization in area 18 of awake behaving cats. *J Neurosci*, 23:4251–60, 2003.

[66] Frederick Martin and Radu Horaud. Multiple camera tracking of rigid objects. *International Journal of Robotics Research*, 21(2):97–113, February 2002.

[67] M. H. Munk, P. R. Roelfsema, P. Konig, A. K. Engel, and W. Singer. Role of reticular activation in the modulation of intracortical synchronization. *Science*, 272(5259):271–4, 1996.

[68] G. Nase, W. Singer, H. Monyer, and A. K. Engel. Features of neuronal synchrony in mouse visual cortex. *J Neurophysiol*, 90(2):1115–23, 2003.

[69] K.J. Palomaki, G.J. Brown, and J.P. Barker. Techniques for handling convolutional distortion with 'missing data' speech recognition. *Speech Communication*, (43):123–142, 2004.

[70] P. Peixoto, J. Batista, and H. Araújo. Integration of information from several vision sensors for a common task of surveillance. *Robotics and Autonomous Systems*, 31:99–108, 2000.

[71] P. Peixoto, J. Batista, and H. Araújo. Real-time human activity monitoring exploring multiple vision sensors. *Robotics and Autonomous Systems*, 35:221–228, 2001.

[72] Salazar RF, Kayser C, and König P. Effects of training on neuronal activity and interactions in primary and higher visual cortices in the alert cat. *J Neuroscience*, 24:1627–1636, 2004.

[73] P. R. Roelfsema, A. K. Engel, P. Konig, and W. Singer. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*, 385(6612):157–61, 1997.

[74] Remi Ronfard. Reading movies - an integrated dvd player for browsing movies and their scripts. In *ACM Conference on Multimedia*, 2004. New York, New York.

[75] A. Ruf and R. Horaud. Projective rotations applied to a pan-tilt stereo head. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 144–150, Fort Collins, Colorado, June 1999. IEEE Computer Society Press.

[76] A. Ruf, F. Martin, B. Lamiroy, and R. Horaud. Visual control using projective kinematics. In John M. Hollerbach and Daniel E. Koditschek, editors, *Robotics Research, The Ninth International Symposium*, pages 95–104. Springer, 2000.

[77] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, volume 3022 of *Lecture Notes in Computer Science*, pages 1–13. Springer-Verlag, May 2004.

[78] Peter Sturm. Structure and motion for dynamic scenes - the case of points moving in planes. In *European Conference on Computer Vision, Copenhagen, Denmark*, volume 2, pages 867–882, May 2002.

[79] Einhäuser W and König P. Does luminance-contrast contribute to a saliency map of overt visual attention? *Eur J Neurosci*, 17:1089–97, 2003.

[80] Marta Wilczkowiak, Gilles Trombettoni, Christophe Jermann, Peter Sturm, and Edmond Boyer. Scene modeling based on constraint system decomposition techniques. In *Proceedings of the 9th International Conference on Computer Vision*. IEEE, IEEE, October 2003.