SIXTH FRAMEWORK PROGRAMME
PRIORITY 2
INFORMATION SOCIETY TECHNOLOGY

PERCEPTION ON PURPOSE (POP)
A STREP PROJECT

`http://perception.inrialpes.fr/POP/`

Final Report

Project start/end dates: **1/1/2006 – 31/12/2008**
Contract number: **IST-2005-27268**

Coordinator:

Radu Horaud, INRIA
Radu.Horaud@inrialpes.fr

Contents

1 Project statistics

- POP was a 3-year project (January 2006 - December 2008) granted by the European Commission under Cognitive Systems. The EC project number is FP6-IST-2004-027268. In EU jargon, POP was a STREP.
- POP partners:
  1. INRIA Grenoble Rhone-Alpes, France (coordinator), contact: Radu Horaud (Radu.Horaud@inrialpes.fr);
  2. University of Osnabrück (UOFM), Germany, contact: Peter Koenig (pkoenig@uos.de);
  3. University Hospital Hamburg-Eppendorf (UKE), Germany, contact: Andreas Engel (ak.engel@uke.de);
  4. Fac de Ciencias e Tec. Univ Coimbra-ISR (FCTUC), Portugal, contact: Helder Araujo (helder@isr.uc.pt), and
  5. The University of Sheffield (USFD), United Kingdom, contact: Jon Barker (jon@dcs.shef.ac.uk).
- Website: `http://perception.inrialpes.fr/POP/`
- Information available on the website:
  - Partners and their profiles, `http://perception.inrialpes.fr/POP/partners.php3`
  - Research topics, `http://perception.inrialpes.fr/POP/research.php3`
  - On-line publications, `http://perception.inrialpes.fr/POP/publications.php3`
  - Public data sets, `http://perception.inrialpes.fr/POP/results.php3`
  - Software, `http://perception.inrialpes.fr/POP/article.php3?id_article=6`
  - POP's robotic prototypes, `http://perception.inrialpes.fr/POP/prototypes.php3`
  - Detailed workshop programs, `http://perception.inrialpes.fr/POP/events.php3`
  - Demonstrators, `http://perception.inrialpes.fr/POP/demos.php3`

2 Overview

One of the major challenges to which researchers are faced, whenever they attempt to build artificial intelligent systems, is the problem of modeling the interactions between an "agent" and its physical environment. One of the main goals of the POP project has been to make steps towards the understanding and modeling of these interactions from two points of view: biological and computational. In particular, the project concentrated onto two sensorial modalities: vision and hearing.

In the recent past, these two perceptual modalities were studiend almost independently. Cross-modal integration and the interaction between auditory and visual processes are beneficial because each modality provides only partial information about different aspects of the world's objects and events. In POP we investigated the combination of these modalities.

In practice, the POP partners studied the followings:

Cognitive mechanisms of attention. We studied models of attention at a fundamental level using the most modern investigation techniques. Multisensory control of attention was studied using fMRI and EEG/MEG source localization. The goal has been to identify brain regions that are involved in audio-visual information fusion. We also studied eye movements: we developed a computational model for eye movements and their role in 3D perception, i.e., stereopsis. We developed, validated and applied a method to remove eye-movement artefacts from EEG and MEG recordings. This method could open the door to a new kind of EEG / MEG research that studies the electrophysiology of the healthy human s visual system under natural conditions. We started to build on the recent *gist* paradigm and extended the work on visual gist to auditory gist. In particular we studied the perception of very brief speech segments.

Integration of visual and auditory cues. Extraction of auditory and visual cues as well as their integration are fundamental building blocks for perceiving the environment. We implemented algorithms that represent visual and auditory stimuli in terms of spatial cues, namely visual disparity maps (dense and sparse), interaural time differences (ITD) and interaural level differences (ILD). These cues are treated as observations in a probabilistic framework. The auditory and visual cues (disparities and ITDs) are clustered using the expectation-maximization algorithm. Each cluster corresponds to a audio-visual object, e.g., a speaking person visible in the field of view of the cameras. In parallel, we studied the biological mechanims of audio-visual integratin using simple audio and visual stimuli and single-neuron multi-electrode recordings on animals (ferrets).

Sensory-motor coordination. The goal of this research has been the design and development of principled methods and algorithms allowing the coordination between motor activities (eye movements and head turns) with sensor observations. Thus, it should be possible to understand the biological and psychophysical principles of attention, and use these findings to propose a computational model of bottom-up attention, i.e., saliency. A model of visual attention was proposed, implemented and tested using POPEYE, the POP's robotic platform. In parallel we studied audio-motor loops in order to go beyond the static-perceiver assumptions prevailing in the past. The rationale here has been to propose active listening strategies based on spatial auditory cues (ITD and ILD). We also started to investigate the notion of saliency maps to audition and to perform cross-modal control using JAMF framework and the POPEYE robot.

Development of methodological and experimental platforms. The theoretical findings just mentioned and their associated methods need thorough validations. In POP we conducted several experiments and demonstrators. We designed and built several audio-visual robotic head prototypes. We demonstrated, visual, auditory, and audiovisual attention. We implemented real-time algorithms for active visual tracking, active listening, and audio-visual speaker detection and localization.

3 Scientific contributions: The legacy of the POP project

The work carried out in POP by the partners has been of strong interdisciplinary nature. This resulted in novel contributions across several disciplines, such as computational vision, psychophisics, robotics, auditory scene analysis, etc. In particular, the following contributions were achieved:

- UKE (neurophysiology) and UOFM (cognitive neuroscience) developed, validated and applied a method to remove eye-movement artefacts from EEG and MEG recordings.
- UKE (neurophysiology) and UOFM (cognitive neuroscience) have been able to provide evidence in several EEG studies in humans that coherence of neural signals may be an essential mechanism supporting multisensory perception. The POP data indicate that coupled oscillatory activity, in particular at frequencies in the gamma-band, may serve to link neural signals across uni- and multisensory regions and to express the degree of crossmodal matching of stimulus-related information.
- USFD (speech and hearing) and UOFM (cognitive neuroscience) explored the new idea of auditory gist finding support from new experiments on the perception of very brief segments of speech. Moreover, we suggested ways of modelling auditory gist.
- USFD (speech and hearing) developed a method and its associated algortihms and software for real-time robust sound-source localization. This method is based on the fusion between interaural time difference (ITD) and pitch cues using source fragment methods.
- USFD also considered the possibilites arising when the audio perceiver is active and a new class of active listening behaviours are considered: listeners (biological or artificial) can use planned motions.
- INRIA (computer vision and statistics) developed a method for audio-visual spatial clustering. The method relies on a mixture model that clusters observations grasped with binocular vision and binaural perception.
- INRIA developed a computational model for eye movements in relation with stereopsis. It also developed a method for estimating a dense disparity field associated with complex 3D surfaces, based on principled statistical methods.
- FCTUC (robotics) developed three different configurations of an audio-visual robot head (POPEYE) and built four prototypes. POPEYE is used by three other POP partners (INRIA, UOFM, and USFD).

- INRIA, FCTUC, and USFD developed real-time algorithms for audiomotor, oculomotor, and audiovisual-motor integration.
- UOFM developed JAMF – a flexible modelling environment which allows non-experts to develop models of sensory processing and attention. It is interfaced with POPEYE.

3.1 Human electrophysiology during eye-movements

Humans are equipped with a sophisticated visual system that involves continuous eye-movements in order to represent the scene location of interest on the fovea, the region of the retina with the highest density of receptors. EEG and MEG, the only means by which neuroscientists can investigate electrophysiological phenomena in the healthy human brain face a severe problem. These techniques are effected by eye-movements induce electric and magnetic artefacts that overlay the signals originating in the brain. In practice this lead to experiment designs where subjects are forced to fixate. Thus, today EEG and MEG do not allow to study the visual system under its natural functioning. Within the POP project we developed, validated and applied a method to remove eye-movement artefacts from EEG and MEG recordings. This method could open the door to a new kind of EEG / MEG research that studies the electrophysiology of the healthy human s visual system under natural conditions.

3.2 Crossmodal binding by neural coherence

Picture yourself on a crowded sideway with people moving around. The acoustic and visual signals generated by people provide you with complementary information about their location and motion. It is not well understood how such inputs from different sensory channels are integrated. In the POP project, we have been able to provide evidence in several EEG studies in humans that coherence of neural signals may be an essential mechanism supporting multisensory perception. Data from our studies in POP as well as from other studies indicate that coupled oscillatory activity, in particular at frequencies in the gamma-band, may serve to link neural signals across uni- and multisensory regions and to express the degree of crossmodal matching of stimulus-related information. These results argue for a new view on multisensory processing, which considers the dynamic interplay of neural populations as a key to crossmodal integration. This new view has several implications : first, it allows to develop novel research approaches and experimental strategies ; second, new accounts for abnormal variants like synesthesia can be developed ; third, experiments on multisensory interaction actually provide the first true testing of the "binding hypothesis" with non-invasive methods in humans and, thus, these studies will have profound impact for the validation of one of the most-discussed hypotheses in the field.

3.3 Auditory gist

We have explored the new idea of auditory gist, finding support both from existing psychophysical evidence and from new experiments on the perception of very brief segments of speech. We have also suggested ways of modelling auditory gist. One approach is based on the idea that the most salient auditory features provide a sparse representation of the signal : such a representation is compatible with research into glimpsing. Auditory saliency itself is not well-developed as a research topic, and our proposals for auditory gist modelling will help to take the modelling of auditory saliency further.

We have begun to model auditory attention using the JAMF attentional framework, which will allow models for combining visual and auditory saliency maps to be easily created and tested. Furthermore, the existence of JAMF components to control the POPEYE robot head will allow auditory components to be incorporated into the robot control mechanism. Modelling top-down attentional processes (and auditory gist in particular) is more complex and will require components for recognising sources to be incorporated into JAMF, but we can use techniques and code previously developed at the University of Sheffield to do this. Within POP we have proposed an experiment using EEG which is capable of providing more specific information on the timing of auditory gist formation. We also have plans for further psychophysical experiments to explore the relationship between glimpses and gist perception, and to investigate the potential for gist perception of everyday environmental sounds.

3.4 Real-time robust sound source localization

POP has constructed novel algorithms for the robust localisation of sound sources in multisource environments. These localisation techniques are based on a fusion of interaural time difference and pitch cues using source fragment methods that have been inspired by recent glimpsing models of speech perception. Further, these algorithms have been implemented in real-time software capable of driving a robot platform to localise speakers in real environments. The software has been demonstrated using the POPEYE robot, videos of which will help publicise our methods. The software has been constructed in a modular fashion to encourage reuse outside POP. Indeed, it has already drawn interest from other research labs outside POP (e.g. CAHR, Copenhagen ; MARCS, Brisbane ; FCT-UC/ISR, Coimbra) and will hopefully form the cornerstone of future collaborations.

3.5 Active binaural perception

POP has considered the new possibilities that arise when moving beyond the static perceiver assumptions of previous hearing research. A new class of active listening behaviours are consider in which listeners (whether animal or machine) can use planned movement to aid auditory perception. For example, the performance of localisation and tracking algorithms can be improved by turning the head to ensure that the tracked source remains in the auditory fovea (i.e. the region directly in front of the head where azimuthal location cues are most sensitive). More interestingly, by using planned movement, cues for the judgement of distance become available that are not available in the static case. The POP project has reviewed active listening behaviour in humans and has presented active hearing as a solution to the problem of coping with dynamic acoustic environments. New methodologies have been developed, including the simulations of active listening strategies through the use of statistical Monte Carlo methods, and the study of synchronised binaural recording and head track data sets such as CAVA. The promotion of active listening as a novel research area has attracted significant interest and has potential impact in areas other than robotics such as intelligent hearing aid design.

3.6 Eye vergence and and its role in stereopsis

Eye movements are an essential part of biological scene perception. The observer s gaze direction is directed by visual attention, so that the scene is actively sampled. However, other aspects of binocular vision are complicated by the process of re-fixation. In particular, the geometry of stereopsis is sensitive to changes in the relative orientation of the eyes. We analyzed the precise role of eye movements in binocular vision. Two specific questions were considered here. Firstly, in order to make full use of binocular information, the visual system must match corresponding image-features between the left and right eyes. The physiological mechanisms that are available for this task are somewhat limited ; in particular they are tuned to particular directions and magnitudes of binocular disparity. Both of these variables depend on the relative orientation of the eyes, as well as on the structure of the scene. We have characterized the variability of disparity information with respect to eye movements. The object of this work is to show that the properties of low-level disparity mechanisms can be better understood in relation to the observer s eye movements. The visual system must also interpret disparity information with respect to the relative orientation of the eyes. In particular, the direction and distance of the binocular fixation point must be estimated. We argue that oculomotor signals, concerning the orientations of the eyes, are unlikely to be accurate enough for this task. Instead, we are investigating how the gaze variables might be recovered from the retinal images. We use techniques from computer vision to address these questions.

3.7 3D surface estimation from dense stereoscopic matching

In POP we addressed the stereo matching problem, while moving away from the traditional fronto-parallel assumption. We proposed an algorithm that provides disparities in accordance with the surface properties of the scene under consideration. The algorithm carries out cooperatively both disparity and surface normal estimations by setting the two tasks in a unified Markovian framework. A novel joint probabilistic model is defined through two Markov Random Fields (MRF) to favor both intra-field (within neighboring disparities and neighboring normals) and inter-field (between disparities and normals) consistency. Geometric contextual information is introduced in the pair-wise Markovian regularizing term used in both MRFs. Segmentation and plane fitting procedures, usually performed as final steps to increase the quality results are here explicitly used in one of the MRF data terms. We then design an appropriate alternating maximization procedure based on standard belief propagation. The effectiveness and the performance of the approach have been illustrated on synthetic and real data. The results obtained are comparable to the state-of-the-art and show improvement in many cases.

3.8 Audiovisual detection and localization of multiple speakers

One of the major goals of POP has been to investigate cross-modal attention. We concentrated onto two highly cognitive sensorial modalities : vision and hearing. In particular we addressed the issues of detecting and localizing objects in a scene that are both seen and heard. We capitalized on the benefits of a human-like configuration of sensors (binaural and binocular) for gathering auditory and visual observations. We have shown that the detection and localization problem can be recast as the task of clustering the audio-visual observations into coherent groups. We proposed a probabilistic generative model that captures the relations between audio and visual observations. This model maps the data into a common audio-visual 3D representation via a pair of mixture models. Inference is performed by a version of the expectation-maximization algorithm, which is formally derived, and which provides cooperative estimates of both the auditory activity and the 3D position of each object. We experimented with several setups, i.e., single- and multiple-speaker detection and localization, in the presence of other audio sources.

3.9 POPEYE: An audiovisual robotic platform

In POP, The University of Coimbra developed three configurations (four prototypes) of an audio-visual robot platform, i.e., POPEYE, with four degrees of freedom. The degrees of freedom have high repeatability, i.e., high precision and low backlash due to the use of harmonic drives. The audio-visual heads were built to permit high velocities and accelerations, while minimizing the mechanical complexity. The kinematic configuration is flexible allowing changing the baseline of the "eyes" and fine adjustment of their positions relative to the rotation axes. The control allows the "on the fly" change of parameters and high flexibility. One of the configurations (for which two prototypes were built) allows the use of a dummy head with special acoustic properties enabling the development of binaural auditory algorithms that take into account the acoustic properties of the human head. The design takes into account the requirements to allow the reproducibility of the most important motions of the human audio-visual system and low complexity of the kinematic chain. In addition to the low complexity of the kinematic chain the mechanical transmission was also simplified by avoiding complex solutions such as differential drives, belts and pulleys-only harmonic drives are used. The design is not fully bio-mimetic, implementing only the mechanical degrees of freedom that allow eye-gaze and eye-vergence control. POPEYE is therefore an excellent platform for the development of audio-visual algorithms that study and exploit the most important types of motion involved in human vision and hearing.

3.10 JAMF: An attention model framework

JAMF is a flexible modelling environment which allows non-experts to develop models of sensory processing and attention and yet do not compromise on efficiency of the code. This is a crucial feature leading to an automatic merger of the expertise of computer scientists and those of experts in the application. It is a major development of the University of Osnabrück and is interfaced with POPEYE (the POP s robotic platform) and used by another POP partner (University of Coimbra). The target is to further develop cognitive models based on this hardware and to allow non-technical researchers (reasonably) easy access to real-world interaction of simulated systems. Sheffield is currently working on adding to the many existing modules a set of routines specific for auditory processing. Given these features it is expected that the life time of JAMF will be much longer than competing software solutions.

4 Software development and data sets

The POP partners developed several software packages:

- The "popeye" package (restricted to POP partners) is hosted by http://gforge.inria.fr and it contains software modules and libraries jointly developed by three partners : INRIA, University of Sheffield, and University of Coimbra : kinematic and dynamic control of the robot head, stereo calibration, camera-microphone calibration, auditory and image processing, stereo matching, binaural localization, and audio-visual speaker localization.

  The "popeye" package is controlled under SVN (Subversion). Currently this software package is not distributed outside POP.

- JAMF is an open-source software package available at http://jamf.eu/jamf/ and developed by the University of Osnabrueck which implements an attentional model. It features a graphical interface (the client) that allows users to design simulations as graphs by connecting and parameterizing components, and a server that translates those graphs to C code and executes them.

- A library specifically developed for POPEYE is available at http://www.assembla.com/spaces/pop_coimbra (accessible on request). It includes examples, self-alignment procedures, camera calibration, a graphical user interface, and examples of experiments.

  Additional information about the hardware/software components of POPEYE can be found at http://labvis.isr.uc.pt/wiki_pop/doku.php/start.

CAVA: a dataset for Computational Audio-Visual Analysis. The CAVA database is a unique set of audiovisual recordings using binocular and binaural camera/microphone pairs both mounted onto a person s head. The database was gathered in order to develop computational methods and cognitive models for audiovisual scene analysis, as part of POP. The CAVA database was recorded in May 2007 by two POP partners : The University of Sheffield and INRIA Grenoble Rhône-Alpes. We recorded a large variety of scenarios representative of typical audiovisual tasks such as tracking a speaker in a complex and dynamic environment : multiple speakers participating to an informal meeting, both static and dynamic speakers, presence of acoustic noise, occluded speakers, speakers faces turning away from the cameras, etc.

The CAVA database is freely accessible at http://perception.inrialpes.fr/CAVA_Dataset/Site/ for scientific research purposes and for non-commercial applications.

5 Project demonstrators

The aim of the POP robot head demonstrator (POPEYE) has been to showcase the developed feature extraction, fusion and sensory-motor algorithms as well as to enable work on more cognitive issues such as attention. POPEYE has provided a joint platform for collaborative work and has served a multitude of dissemination purposes by operating 'live' in seminars, talks and student open days and featuring in videos in numerous talks across the world. As such it will continue to attract attention to the research areas and function as an invaluable asset and resource, which researchers and students alike will interact with in the POP teams.

Figure 1 shows a functional block diagram of the demonstrator structure. Through the project period a progression of modules has been developed which are being used in various configurations and have resulted in a family of demonstrators. For each major component (audio input, video input and fusion) two different versions are implemented: (i) a fast, more standard approach approach and (ii) a more computationally demanding, but higher performing approach. Further, it is possible to run the demonstrator in an online (using the robot-head) and an offline (using the CAVA database) mode.

Videos of the POP's demonstrators are available at `http://perception.inrialpes.fr/POP/demos.php3`.

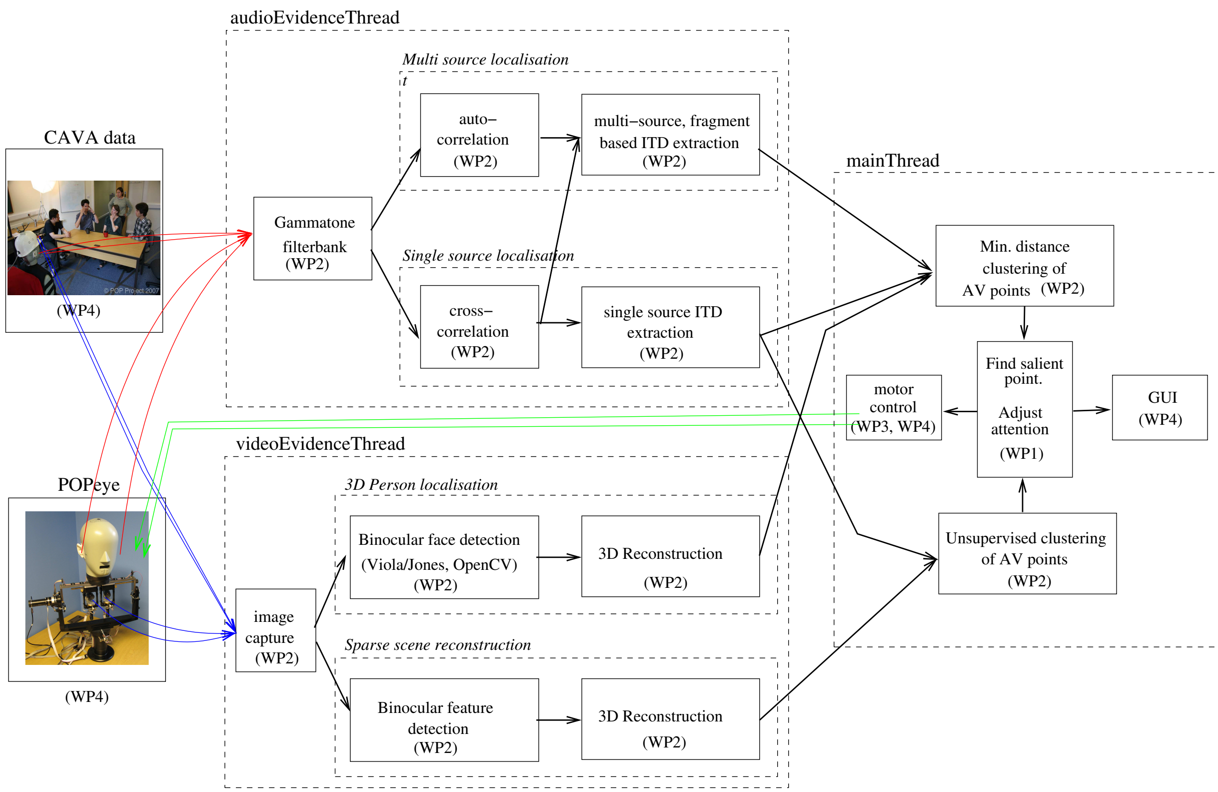

Figure 1: *Logical block diagram structure for the POPEYE audio-visual robot head. The modular architecture allows for flexible development and exploration of various elements.*

6 Dissemination

6.1 Scientific publications

To access the main POP's publications online, please visit http://perception.inrialpes.fr/POP/publications.php3.

INRIA: [1] (abstract), [2], [3], [4], [5], [6], [7], [8], [9].

UKE: [10]. [11], [12], [13], [14], [15], [16].

UOFM: [17], [18], [19], [20], [21], [22], [23]

FCTUC: [22], [24], [25].

USFD: [3], [7], [23], [24], [26], [27], [28] (abstract), [29], [30] (abstract), [31], [32] (abstract), [33].

6.2 Workshops

The POP partners organized the following workshops:

Computational and Neurophysiological Models for Visual Perception. Les Houches Physics School, 25-30 March 2007. Website: http://visiontrain.inrialpes.fr/?page=school3.

Computational and cognitive models for audio-visual interactions. University of Sheffield, 11 March 2008. Website: http://www.dcs.shef.ac.uk/~martin/AV/

Audition, vision, and Their Interactions. INRIA Grenoble Rhone-Alpes, 12 December 2008. Website: http://perception.inrialpes.fr/POP/article.php3?id_article=57

6.3 Dissemination of data, software, and hardware

The CAVA database. Synchronized and calibrated binocular/binaural data sets with head movements. Website: http://perception.inrialpes.fr/CAVA_Dataset/Site/

JAMF. A graphical interface for development of attentional models. Website: http://jamf.eu/jamf/.

POPEYE. The University of Coimbra's audiovisual robot head. Website: http://labvis.isr.uc.pt/wiki_pop/doku.php/start.

References

[1] M. Hansard and R. Horaud. A model of binocular gaze estimation. In *Fourth Computational and Systems Neuroscience (COSYNE) Meeting Abstracts*, Salt Lake City, Utah, USA, February 2007.

[2] M. Hansard and R. Horaud. Patterns of binocular disparity for a fixating observer. In *Proc. BVAI 2007*, volume LNCS 4729 of *Lecture Notes in Computer Science*, pages 308–317. Springer, 2007.

[3] M. Cooke, Y.-C. Lu, Y. Lu, and R. Horaud. Active hearing, active speaking. In T. Dau, J. M. Buchholz, J. M. Harte, and T. U. Christiansen, editors, *Auditory signal processing in hearing-impaired listeners*. Centertryk, 2008.

[4] Miles Hansard and Radu P. Horaud. Cyclopean geometry of binocular vision. *Journal of the Optical Society of America A*, 25(9):2357–2369, September 2008.

[5] Vasil Khalidov, Florence Forbes, Miles Hansard, Elise Arnaud, and Radu P. Horaud. Audio-visual clustering for multiple speaker localization. In *5th International Workshop on Machine Learning for Multimodal Interaction (MLMI'08)*, LNCS 5237, pages 86–97. Springer, September 2008.

[6] Vasil Khalidov, Florence Forbes, Miles Hansard, Elise Arnaud, and Radu P. Horaud. Detection and localization of 3D audio-visual objects using unsupervised clustering. In *ACM/IEEE International Conference on Multimodal Interfaces (ICMI'08)*, Chania, Greece, October 2008.

[7] Elise Arnaud, Heidi Christensen, Yan-Chen Lu, Jon Barker, Vasil Khalidov, Miles Hansard, Bertrand Holveck, Hervé Mathieu, Ramya Narasimha, Elise Taillant, Florence Forbes, and Radu P. Horaud. The cava corpus: synchronised stereoscopic and binaural datasets with head movements. In *ACM/IEEE International Conference on Multimodal Interfaces (ICMI'08)*, Chania, Greece, October 2008.

[8] Ramya Narasimha, Elise Arnaud, Florence Forbes, and Radu P. Horaud. Cooperative disparity and object boundary estimation. In *IEEE International Conference on Image Processing*, October 2008.

[9] M. Hansard and R. Horaud. Cyclo-rotation models for eyes and cameras. To appear in *IEEE Transactions on Systems, Man & Cybernetics* Part B: Cybernetics, 2009.

[10] J Hipp, AK Engel, and M Siegel. Neuronal dynamics of bi-stable cross-modal binding. In *Annual Meeting SFN, San Diego*, 2007.

[11] JF Hipp, AK Engel, and M Siegel. Neuronal dynamics of bi-stable cross-modal binding. In *IMRF*, 2008.

[12] T.R. Schneider, S. Debener, R. Oostenveld, and A.K. Engel. Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage*, 42(3):1244–1254, Sep 2008.

[13] T.R. Schneider, A.K Engel, and S. Debener. Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp Psychol*, 55(2):121–132, 2008.

[14] D. Senkowski, T.R. Schneider, J.J. Foxe, and A.K. Engel. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neuroscience*, 31(8):401–409, August 2008.

[15] D. Senkowski, T.R. Schneider, F. Tandler, and A.K. Engel. Gamma-band activity reflects multisensory matching in working memory. *Experimental Brain Research*, 2009.

[16] M. Siegel, T.H. Donner, R. Oostenveld, P. Fries, and A.K. Engel. Neuronal synchronization along the human dorsal visual pathway reflects the focus of attention. *Neuron*, 2009.

[17] H. Frey, P. Konig, and W. Einhauser. The role of first- and second-order stimulus features for human overt attention. *Perception and Psychophysics*, in press.

[18] S. Onat, K. Tichacek, and P. Konig. Integrating audio-visual information for the control of overt visual attention. *Journal of Vision*, 7:1–16, 2007.

[19] A. Acik, S. Onat, F. Schumann, and P. König. Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. *Spatial Vision*, submitted.

[20] A. Acik, J. Hipp, K. Görgen, P. König, and A. K. Engel. Simultaneous eeg recording and eye-tracking during active viewing. In *Proceedings of Neurizons 2007*, Gottingen, Germany, 2007.

[21] J. Steger, N. Wilming, F. Wolfsteller, N. Höning, and P. König. The JAMF attention modelling framework. In Lucas Paletta and John K. Tsotsos, editors, *Proceedings of the 5th International Workshop on Attention in Cognitive Systems*, Lecture Notes in Computer Science. Springer, Berlin, Germany, 2008.

[22] N. Wilming, F. Wolfsteller, P. König, R. Caseiro, and H. Araújo. Attention models for vergence movements based on the jamf framwork and the popeye robot. In *VISAPP 2009: International Conference on Computer Vision and Applications*, 2009.

[23] C. Quigley, S. Onat, S. Harding, M. Cooke, and P. König. Audio-visual integration during overt visual attention. *Journal of Eye Movement Research*, 1(2)(4):1–18, 2008.

[24] H. Christensen, J. Barker, Y.-C. Lu, J. Xavier, R. Caseiro, and H. Araújo. POPeye: Real-time, binaural sound source localisation on an audio-visual robot-head. Conference on Natural Computing and Intelligent Robotics, 2009.

[25] Luis Perdigoto, Joao P. Barreto, Rui Caseiro, and Helder Araujo. Active stereo tracking of multiple free-moving targets. In *IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, June 2009.

[26] H. Christensen, N. Ma, S. N. Wrigley, and J. Barker. A speech fragment approach to localising multiple speakers in reverberant environments. In *ICASSP'09*, Taipei, Taiwan, 2009.

[27] S. Harding and M. Cooke. Perception of very brief speech segments. *Journal of Experimental Psychology: Human Perception and Performance*, 2009.

[28] S. Harding and M. Cooke. Perception of speech properties from extremely brief segments. *Journal of the Acoustical Society of America*, 123(5):3724, 2008.

[29] S. Harding and M. Cooke. Perception of properties of extremely brief speech segments. In *British Society of Audiology Short Papers Meeting, York*, 2008.

[30] H. Christensen, N. Ma, S. N. Wrigley, and J. Barker. Improving source localisation in multi-source, reverberant conditions: exploiting local spectro-temporal location cues. In *Abstract for Acoust. Soc. Am.*, Paris, July 2008.

[31] H. Christensen and J. Barker. Simultaneous tracking of perceiver movements and speaker changes using head-centered, binaural data. Conference on Natural Computing and Intelligent Robotics, 2009.

[32] Y.-C. Lu and M. Cooke. Strategic listener movement in a model of auditory distance perception. *J Acoust Soc Am*, 123(5):3726, 2008.

[33] Y.-C. Lu and M. Cooke. Binaural distance perception based on direct-to-reverent energy ratio. In *Proc. IWAENC*, Seattle, US, 2008.