

# Multimodal Saliency-Based Bottom-Up Attention A Framework for the Humanoid Robot iCub

Jonas Ruesch, Manuel Lopes, Alexandre Bernardino, Jonas Hörnstein, José Santos-Victor, Rolf Pfeifer

**Abstract**—This work presents a multimodal bottom-up attention system for the humanoid robot iCub where the robot's decisions to move eyes and neck are based on visual and acoustic saliency maps. We introduce a modular and distributed software architecture which is capable of fusing visual and acoustic saliency maps into one egocentric frame of reference. This system endows the iCub with an emergent exploratory behavior reacting to combined visual and auditory saliency. The developed software modules provide a flexible foundation for the open iCub platform and for further experiments and developments, including higher levels of attention and representation of the peripersonal space.

## I. INTRODUCTION

Organisms at all scales continuously select and structure sensory stimulation through their morphologies and sensorimotor interactions [10] [26]. For instance, humans move their eyes and head to focus on specific locations of interest, thereby modulating the input to head related sensory systems. It has been shown that doing so greatly supports subsequent information processing [15] [21]. However, it is not clear where to move head and eyes in order to select valuable regions of interest in an appropriate sequence, especially as different sensory modalities like vision, sound, smell and touch need to be taken into account.

It is known that simple saliency mechanisms in the human visual attention system, trigger multiple reflexive behaviors. For example eye saccades are guided toward visual stimuli with high intensity or characteristic directional features [27], [18], [25]. Similar attentive responses were found for acoustic features [19]. Already at birth, the brain system responsible for reflexive saccadic eye movements is relatively mature [24] and it is assumed that eye saccades in the first 2 to 3 postnatal months are reflexive and controlled by the superior colliculus [13]. The superior colliculus, which is a subcortical structure located in the midbrain, has a topographical map of the spatial environment and produces activity that controls the direction and amplitude of the reflexive saccade. There is evidence that this area is also responsible for merging visual and auditory saliency from modality specific frames of references in order to enable

This work was supported by the European Commission, Project IST-004370 RobotCub, and by the Portuguese Government - Fundação para a Ciência e Tecnologia (ISR/IST plurianual funding) through the POS.Conhecimento Program that includes FEDER funds, and through project BIO-LOOK, PTDC / EEA-ACR / 71032 / 2006.

J. Ruesch and R. Pfeifer are with the Artificial Intelligence Laboratory, Department of Informatics, University of Zurich, Switzerland {ruesch,pfeifer}@ifi.uzh.ch

M. Lopes, J. Hörnstein, A. Bernardino and José Santos-Victor are with the Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal {macl,alex,jhornstein,jasv}@isr.ist.utl.pt

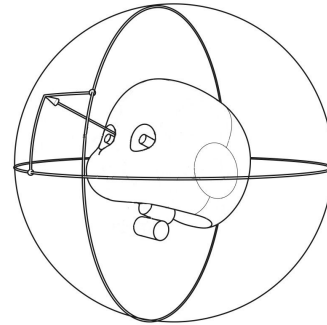


Fig. 1. The iCub ego-sphere: a spherical, torso-based projection surface for egocentric, multimodal saliency information.

a saccadic movement target to be chosen [7], [17], [19]. According to these findings, we focus on two interconnected issues: i) construction of saliency maps by extracting preattentive visual and acoustic features and, ii) fusion of multimodal saliency into a single egocentric representation. We note that this egocentric representation, or ego-sphere, is adequate to represent all kinds of spatially related sensory input.

A computational approach to preattentive visual features has been described in earlier work such as [11], [12]. In [11] an implementation capable of realtime performance for computing visual saliency based on intensity, color, and directional features is described in detail; an approach we follow closely in this work. A system for computing auditory saliency in a time-frequency domain is addressed in [14]. Saliency is assigned to short and long tones as well as to gaps (the absence of frequencies in a broad band noise) and to temporally modulated tones compared to stationary tones. Our approach differs insofar as we are primarily interested in extracting spatial information from a salient acoustic event in order to build an egocentric map. In [4] visual saliency is used to guide a 3-axis binocular vision system. As opposed to the iCub's head and eyes, this setup operates on a map covering a delimited solid angle. Image rectification and mosaicing allow the use of static stereo algorithms on the active platform. [6] and [20] describe the implementation of a complete ego-sphere for a humanoid robot, the sensory ego-sphere (SES). The SES consists of a tessellated sphere which partitions space into hexagonal or pentagonal regions. Each vertex of this mesh corresponds to an index into a database where sensory data related to the node's region is stored. Visual attention using the SES is described in [6] where a

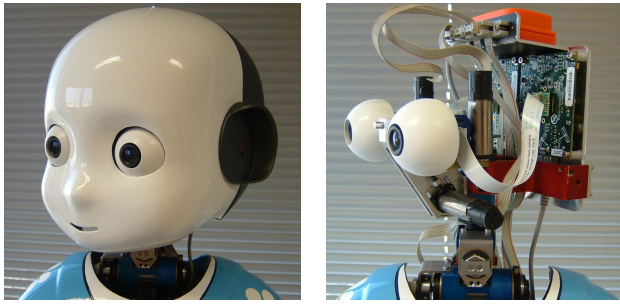


Fig. 2. The iCub head: six degrees of freedom; tilt, swing and pan for the neck; independent pan and joint tilt for the eyes.

cumulative saliency value for each node is computed with the *FeatureGate* algorithm [3] applied to registered images. This approach faces the challenge of integrating attention points calculated on overlapping images. Similar to the iCub ego-sphere, the SES provides a form of short-term memory [20]. By storing data records in the underlying database together with a time stamp the SES allows access to past sensory information.

In this work we present a topologically organized ego-sphere to smoothly combine spatial visual and acoustic saliency maps. In order to capture saliency from the robot's entire surroundings, saliency maps originating from different sensory modalities are aggregated on a continuous spherical surface. Mapping 3D point locations to this ego-sphere corresponds to a projection onto a sphere with infinite radius centered at the robot's body. Visual saliency is composed of the results from preattentive filters for intensity, color and direction in the sense of [11]. Additionally, a saliency filter for motion detection [23] is integrated. Spatial auditory saliency maps are generated using the position of detected sound sources [8]. Based on the final egocentric saliency map, the robot's gaze direction is adjusted toward the most salient location. This incorporates the actuation of neck and eyes with a total of six degrees of freedom (Fig. 2) [1]. Finally, a dynamic inhibition-of-return mechanism acts as the driving force for saliency based exploration of the robot's environment.

The structure of this paper is as follows: Section II describes the implementation of visual and auditory saliency filters. In sections III and IV we show how saliency information from different sensory modalities is integrated into the iCub's ego-sphere and how the selection of a point of interest, together with the inhibition of return mechanism, leads to a dynamic attentive process. Subsequently, after some notes concerning the software implementation, the results of our experiments are presented. We conclude our explanations with a discussion of the obtained results.

## II. SALIENCY FILTERS

The multimodal egocentric map is built from saliency information provided by each sensory channel. This section describes the applied visual and auditory saliency filters.

### A. Visual Saliency

Visual saliency information is calculated by combining the result from several pre-attentive sub-filters into one final map expressed in camera coordinates. The following paragraph provides a brief overview of our approach and points the interested reader to other published work.

Four different sub-filters (filtering intensity, color hue, directional features and motion) contribute to the final visual saliency map. *Intensity saliency* corresponds in a first step to a simple grayscale conversion; subsequently, conspicuous regions are extracted by filtering the map with Mexican hat wavelets of different sizes. To weight and combine the resulting maps we use a method proposed in [11]. The *color saliency* filter follows the implementation presented in [2]. Directional saliency is computed using gabor filters at three different scales  $\sigma \in \{2.0, 4.0, 8.0\}$  in four different directions  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ . The resulting 12 maps are weighted and combined as described in [11]. *Motion saliency* resulting from moving objects is detected according to the Reichardt correlation model [23]. The implementation follows the details given in [9].

### B. Auditory Saliency

The position of a sound source is estimated using interaural spectral differences (ISD) and the interaural time difference (ITD) as described in detail in [8].

To support the localization process, the robot is equipped with two spiral formed ears. The form of the artificial pinna gives spectral notches at different frequencies depending on the elevation of the sound source. A notch is created when the sound wave reaching the microphone is directly canceled by the wave reflected in the pinna. This happens when the microphone-pinna distance is equal to a quarter of the wavelength of the sound (plus any multiple of half the wavelength). The notches, and with it the elevation estimate, can then be found by looking for minima in the frequency spectra.

To obtain the azimuth angle of the sound source, we use the interaural time difference. By looking for the maximum cross correlation between the signals at the right and left ear it is easy to calculate the ITD and the azimuth angle knowing the distance between the microphones and the sampling frequency.

To get a measurement of the uncertainty in the estimated position we divide the samples in several subsets and calculate the azimuth and elevation angle for each subset. We then compute the mean and standard deviation of the estimates.

## III. AN EGO-SPHERE FOR MULTIMODAL SALIENCY

The ego-sphere is a projection surface for spatially related information. In general, such an egocentric map can be used to capture information corresponding to different cognitive levels. This includes low-level stimuli supporting fast reflexive behavior as well as high-level abstractions of the surrounding environment. In our approach, we use the ego-sphere to build a coherent representation of multimodal saliency. The ego-sphere is head-centered and fixed with

respect to the robot’s torso. In this way, rotational or translational corrections are not required as long as the robot does not translate. For efficiency reasons, the egocentric maps are stored as rectangular images, expressed in spherical coordinates (azimuth  $\vartheta$  and elevation  $\varphi$ ). The map’s origin ( $\vartheta = 0^\circ$ ,  $\varphi = 0^\circ$ ) is located at the center of the image. The process of projecting saliency information onto the ego-sphere is as follows:

- 1) convert stimulus orientation to torso-based, head-centered coordinates using head kinematics,
- 2) represent orientation in spherical coordinates,
- 3) project stimulus intensity onto modality specific rectangular egocentric map,
- 4) aggregate multimodal sensory information.

Fig. 3 illustrates a spherical projection of a sound source onto the ego-sphere. This is done by adding the extracted localization information as a normal distribution to the egocentric map. Values for new acoustic observations are combined with past values by selecting the maximum amplitude.

Unavoidably, this projection leads to a deformation. A projected shape is stretched in latitudinal direction by a factor of  $1/\cos(\varphi)$ . To improve the coherence of projected visual saliency, maps originating from camera images are always undistorted based on the camera’s intrinsic parameters.

#### A. Multimodal Saliency Aggregation

After converting the saliency information to a common egocentric frame of reference, it is required to combine the different sensory modalities into a single final map. This is done by taking the maximum value across all saliency channels at each location. Fig. 4 depicts the aggregation process for visual and auditory saliency. While this approach is fast and suited for many occasions, further work is needed to explore different ways of fusing multimodal sensory information onto a single saliency map. In fact, the aggregation function may entail much more complex and context-sensitive information. A talking face is usually considered more salient than a silent one. Similarly a squeaky, colorful toy is probably more interesting than a soundless color

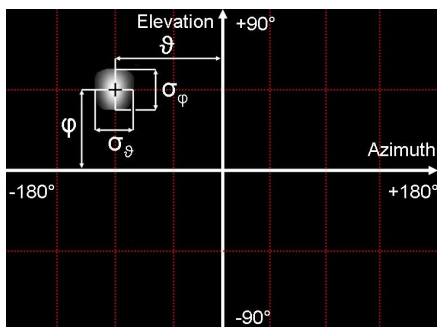


Fig. 3. Projection of a sound source detected at longitude  $\vartheta = -90^\circ$  and latitude  $\varphi = 45^\circ$  with an extension  $\sigma_\vartheta$  and  $\sigma_\varphi$  of  $25^\circ$  corresponding to the calculated uncertainty. The red dotted lines indicate constant latitude and longitude values.

blob. If we imagine an empty room where a human voice is suddenly present, the resulting acoustic saliency should be much higher than that in a face-voice context. These examples suggest that, in the future, the system should be able to learn the saliency aggregation function from its own experience.

#### B. Short-term Memory

The egocentric saliency map corresponds to a short-term memory for salient regions. This allows the robot to behave in a more sophisticated way than a purely reactive agent does. By considering memorized saliency information, the robot can shift attention to previously discovered salient regions. To do so, saliency values mapped to the ego-sphere are maintained and decay at each time step according to a forgetting factor  $d_{mem} \in [0, 1]$ .

### IV. ATTENTION AND EXPLORATION

In the previous section we described how to integrate multimodal saliency information into a unified spatial representation. Based on the saliency ego-sphere, the robot is able to select points of interest in order to explore its surroundings. When referring to exploration, we do not consider an optimal strategy in any sense, but want to investigate if the robot is able to explore its vicinity by using basic attentional mechanisms instead of employing any planning. This exploratory behavior results from the interplay between the attention selection and inhibition mechanisms.

The attention selection process selects a point corresponding to the location with the highest overall saliency value.

The inhibition mechanism attenuates the saliency for locations which have been close to the focus of attention for a certain amount of time. This modulation of the saliency map is motivated by the "inhibition of return" (IOR) observed in human visual psychophysics [22]. To integrate the concept of IOR, two additional egocentric maps are introduced, the Habituation map  $H$  and the actual Inhibition map  $A$ .

The Habituation Maps encodes the way the visual system gets used to persistent or repetitive stimuli, with the system’s attention being drawn toward novel salient regions

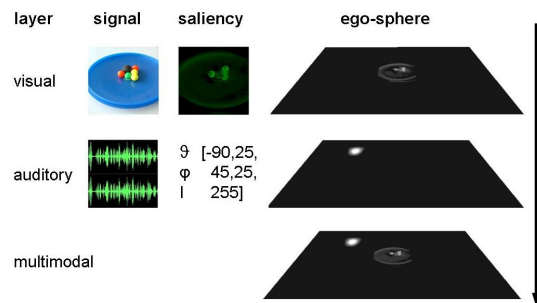


Fig. 4. Multimodal saliency aggregation: saliency computed from sensory signals (left to right); aggregation of different saliency channels on the egocentric map (vertical direction, top down). Spatial auditory saliency is shown as a vector containing center location and uncertainty information in longitudinal and latitudinal direction plus an intensity value.

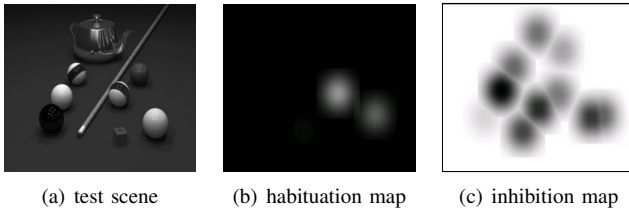


Fig. 5. Exploration of a test scene with fast triggering and relatively slow inhibition decay leading to accelerated and extensive scene exploration. Saliency is calculated by the intensity filter with applied conspicuity calculations.

in the environment. In our system, the Habituation Map is initialized to zero and updated recursively according to a Gaussian weighing function that favors the regions closer to the focus of attention,  $(\vartheta_0, \varphi_0)$ . The amplitude of Gaussian weighing function,  $G_h(\vartheta - \vartheta_0, \varphi - \varphi_0)$  is normalized to 1. The Habituation Map is updated according to:

$$H(t) = H(t - 1) + d_h (G_h(\vartheta - \vartheta_0, \varphi - \varphi_0) - H(t - 1)),$$

where  $d_h \in [0, 1]$  is the habituation gain and  $\sigma_{ior} = 6^\circ$  is the variance of the Gaussian weighing function. While attending to a salient point, the Habituation Map at that location will asymptotically tend to 1, with  $d_h$  determining the speed of convergence.

Whenever the habituation value exceeds a predefined level, the system becomes attracted to novel salient points. The Inhibition Map represents perceptual regions that have already been attended to and that should not capture the system's attention again in the near future, i.e., local habituation leads to inhibition of that location, [22].

The Inhibition Map is initialized to 1.0. Then, when the habituation of a certain region exceeds a threshold  $t_h = 0.85$ , the Inhibition Map is modified by adding a scaled Gaussian function,  $G_a$ , with amplitude  $-1$ , centered at region with maximum habituation and variance  $\sigma_{ior}$ . The resulting effect is that  $G_a$  adds a smooth "cavity" at the respective position in the Inhibition Map. The temporary nature of the inhibition effect is modeled by a time-decay factor,  $d_a \in [0, 1]$ , applied at every location of the Inhibition Map.

$$A(t) = A(t - 1) + d_a(1.0 - A(t - 1)), \quad (1)$$

Finally, for attention selection the multimodal saliency map is weighed (multiplied) by the Inhibition Map, thus combining instantaneous saliency and the memory of recently attended locations.

By combining saliency based attention selection and the described inhibition mechanism, saccadic eye movements toward the most salient locations occur in a self-controlled way. The inhibition mechanism ultimately causes the formation of a bottom-up exploratory behavior. The ability to select and attend to an object during a longer time period can be achieved either by adapting the habituation gain/threshold or by higher order processes.

A snap-shot of the inhibition maps for a test scene exploration are shown in Fig. 5. This behavior models qualitatively

the way people explore and observe pictures with, apparently random, gaze shifts.

## V. IMPLEMENTATION AND RESULTS

In this section we present results obtained by running the described system on the iCub platform. Foregoing, we give a brief overview of the implementation and provide some details on the configuration of the setup.

The software is implemented and available as a portable and open framework in C/C++. The middleware YARP [16] [5] allows the implementation to be split into different modules. A module in the YARP context is an independent process which communicates via network or shared memory with other connected modules beyond machine and platform borders. Modules are designed to be used and connected to each other in a flexible way. The core of the iCub attention system is distributed to the modules: *visual saliency*, *sound localization*, *ego-sphere* and *attention controller*.

Table I lists the user defined parameters and default configuration values. The saliency ego-sphere uses a resolution  $R_{ego}$  of 320 by 240 pixels. This corresponds to a spatial discretization of  $1.125^\circ$  in latitudinal and  $0.75^\circ$  in longitudinal direction. The update time of the entire system cannot be specified in a single frequency value as the different software modules run at different rates. While the motor control module operates at a frequency higher than 20 Hz, the ego-sphere is updated between five to ten times per second. All results presented below were obtained using a setup where the robot's torso is fixed with respect to the world reference frame.

### A. Spherical Mosaic

To visually check the precision of the spherical mapping, the ego-sphere is used to map camera color images instead of saliency maps. While moving the iCub's neck and eyes, undistorted camera images (320 by 240 pixels) are projected onto the ego-sphere. The spherical map is configured to a resolution of 640 by 480 pixels; Fig. 6 shows the resulting mosaic. A slight error in projection is observed if the current joint positions (encoder values) and camera images are not perfectly synchronized. For closed loop movements

TABLE I  
PARAMETERS AND DEFAULT CONFIGURATION SETTINGS

parameter	value	description
$R_{ego}$	$320 \times 240$	resolution of the final multimodal ego-centric saliency map
$R_{cam}$	$128 \times 128$	resolution of images grabbed from the camera
$d_{mem}$	0.95	short-term memory decay of multimodal saliency
$d_a$	0.03	inhibition decay
$d_h$	0.2	visual habituation gain
$t_h$	0.85	threshold for triggering inhibition based on habituation
$\sigma_{ior}$	$6^\circ$	standard deviation of the Gaussian fields $G_h$ and $G_a$ for habituation and inhibition

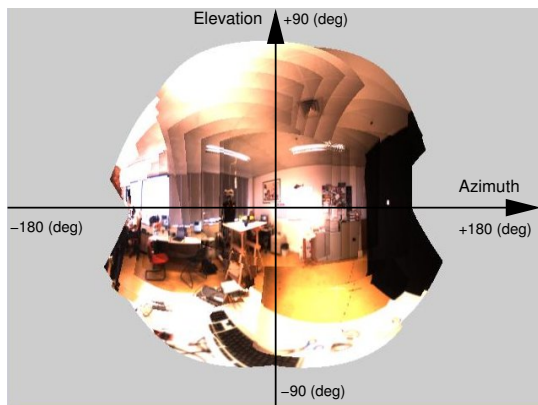


Fig. 6. Experiment V-A: spherical mosaic

like saccades this can be prevented by inhibiting image projection until a "motion done" signal is received from the controlboard.

### B. Visual Exploration

This experiment shows results running the system using visual saliency information only. The robot is located in front of a black curtain where three salient marks are placed. After having focused on the first most salient location, the system continues, driven by inhibition of return, to explore next most salient locations in an emergent sequence. The three salient locations excite the robot to explore the scene by shifting its gaze along a triangular path (Fig. 7). Selected locations are focused by moving neck and eyes. Convergence towards target locations is shown in Fig. 8. We can see from the results that all points are visited the same number of times.

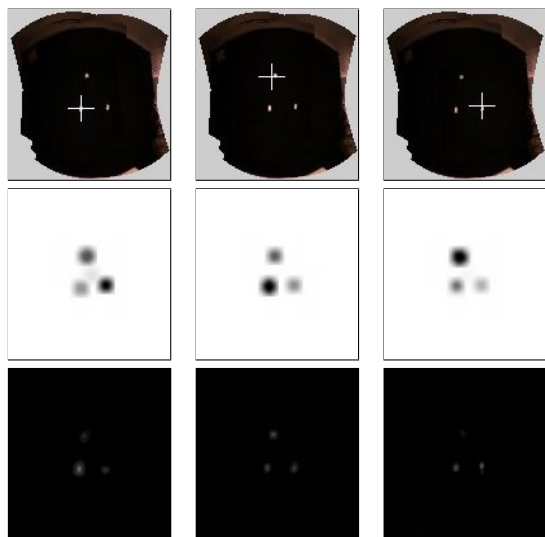


Fig. 7. Experiment V-B (from left to right): Top: selected target locations on the ego-sphere mosaic; middle: corresponding IOR maps (inhibited areas in black); bottom: egocentric saliency map

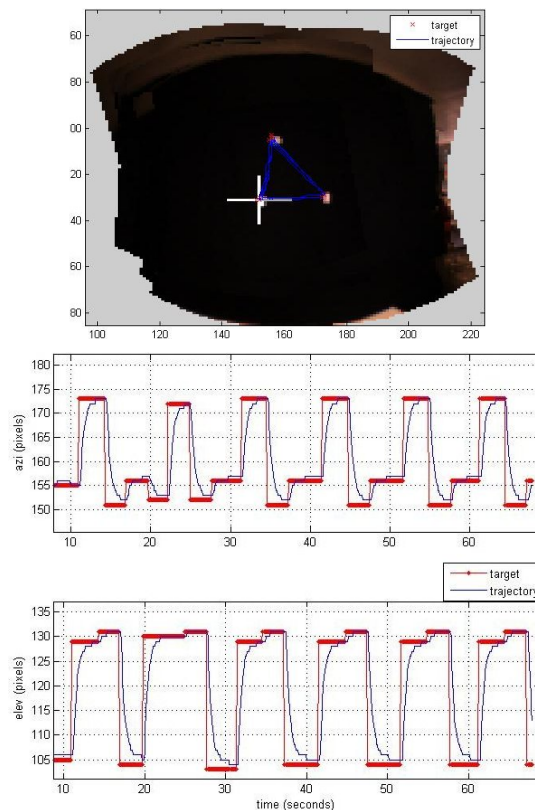


Fig. 8. Experiment V-B: selected target locations and gaze path (target positions are shown in red, the actual gaze path in blue)

### C. Multimodal Exploration

In a second experiment aggregated visual and auditory saliency is used to guide attention. As before, the iCub starts exploring its environment based on visual saliency until it perceives an acoustic stimulus. The top right image in Fig. 9 shows this event: talking to the robot causes auditory saliency being added to the egocentric map and results in a shift in gaze direction. The obtained gaze path is plotted in Fig. 9 (bottom). For this setup  $\sigma_{ior} = 12^\circ$  was used.

## VI. DISCUSSION AND FUTURE WORK

We have presented a multimodal attentional system for humanoid robots, based on bottom-up audio-visual cues. By combining visual and auditory saliency information on a single egocentric map, we are able to drive the robot's attention toward stimulating spatial locations, taking multiple sensory modalities into account.

A dynamic inhibition-of-return mechanism allows the robot to explore salient locations, mainly seeking the regions that have not been attended to in the past. Despite the relatively simple underlying mechanisms, the system shows a rich attentional behavior. The system was implemented on the humanoid platform iCub (only the head was used) and runs in real-time. The experiments in the paper illustrate how the robot reacts to its environment in a rather natural way, the eyes and head being actuated in a non pre-programmed manner. The robot's focus of attention and the sequence of

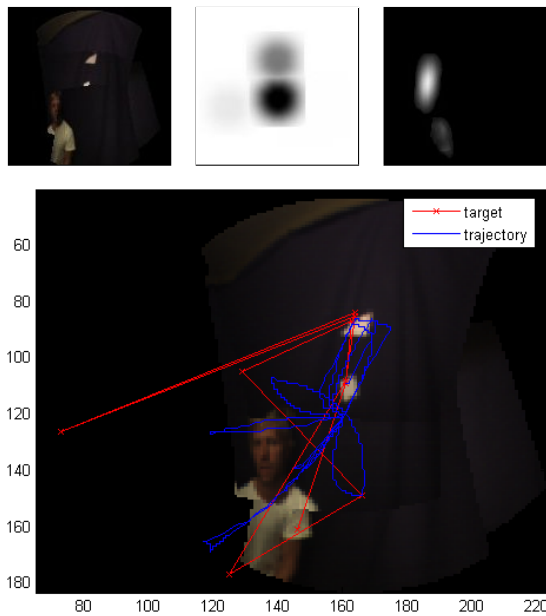


Fig. 9. Experiment V-C: Top: snap-shot of the egocentric mosaic, the IOR map and the multimodal saliency map during an acoustic event; bottom: gaze path visualization for multimodal exploration

gaze shifts are entirely determined by the environment, the iCub automatically adapts its body and sensory systems to the surrounding stimuli.

The ideas for future work are multifold. One research direction is that of automatically adapting the attentional system parameters to display the appropriate exploratory behavior for each situation. We will also investigate how to use the extracted feature/saliency maps to learn cross-modal correlations and further causalities between observed events and perceived sensory stimuli. We hope that these research problems will allow us to further enhance the iCub's attentional capabilities without the need to resort to the introduction of artificial higher-level modules. The software implementation will be publicly available in the iCub software repository.

#### ACKNOWLEDGMENTS

We would like to acknowledge, with thanks, all members of the RobotCub consortium for the fruitful scientific discussions and technical help, in particular to Paul Fitzpatrick, Giorgio Metta, Lorenzo Natale, Luciano Fadiga and David Vernon.

#### REFERENCES

[1] Beira, R., Lopes, M., Praça, M., Santos-Victor, J., Bernardino, A., Metta, G., Becchi, F., Saltarén, R. Design of the Robot-Cub (iCub) Head, *Proc. IEEE International Conference on Robotics and Automation, ICRA*, Orlando, May, 2006.

[2] Breazeal, C. and Scassellati, B. A context-dependent attention system for a social robot, *Proc. of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI99*, Stockholm, Sweden, p. 1146-1151.

[3] Cave, K. R., The FeatureGate model of visual selection. *Psychological Research*, Vol. 62, 1999, pp. 182-194.

[4] Dankers, A., Barnes, N., Zelinsky, A. A Reactive Vision System: Active-Dynamic Saliency, *Proc. of the 5th International Conference on Computer Vision Systems, ICVS*, Bielefeld, Germany, March, 2007.

[5] Fitzpatrick, P., Metta, G., Natale, L. Towards Long-Lived Robot Genes, accepted for publication in *Robotics and Autonomous Systems*, 2007.

[6] Fleming, K. A., Peters II, R. A., Bodenheimer, R.E. Image Mapping and Visual Attention on a Sensory Ego-Sphere, *Conference on Intelligent Robotics and Systems (IROS)*, Beijing, China, October, 2006, pp. 241-246.

[7] Grossberg, S., Roberts, K., Aguilar, M., Bullock, D. A neural model of multimodal adaptive saccadic eye movement control by superior colliculus, *Journal of Neuroscience*, Vol. 17, Month, 1997, pp. 9706-9725.

[8] Hörnstein, J., Lopes, M., Santos-Victor, J., Lacerda, F. Sound localization for humanoid robots - building audio-motor maps based on the HRTF, *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, October, 2006.

[9] Iida, F., Biologically inspired visual odometer for navigation of a flying robot. *Robotics and Autonomous Systems*, Vol 44, No. 3-4, 2003, pp. 201-208.

[10] Ijspeert, A. J., Arbib, M. A. Visual tracking in simulated salamander locomotion, *From Animals to Animats, Proc. of the 6th International Conference on the Simulation of Adaptive Behavior (SAB)*, Paris, France, September, 2000, pp. 88-97.

[11] Itti, L., Koch, C., Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, Nov 1998, pp. 1254-1259.

[12] Itti L. and Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research*, Vol. 40, No. 10-12, May, 2000, pp. 1489-1506.

[13] Johnson, M. H. Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience*, Vol. 2, 1990, pp. 81-95.

[14] Kayser, C., Petkov, C., Lippert, M., Logothetis, N.K. Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map, *Current Biology*, Vol. 15, November, 2005, pp. 1943-1947.

[15] Lungarella, M., Sporns, O. Mapping information flow in sensorimotor networks, *PLoS Computational Biology*, Vol. 2, October, 2006, pp. 1301-1312.

[16] Metta, G., Fitzpatrick, P., Natale, L. YARP: yet another robot platform *International Journal on Advanced Robotics Systems, Special Issue on Software Development and Integration in Robotics*, March, 2006.

[17] Metzger R. R., Mullette-Gillman O. A., Underhill A. M., Cohen Y. E., Groh J. M. Auditory saccades from different eye positions in the monkey: implications for coordinate transformations, *J Neurophysiol.*, Vol 92, October, 2004 Oct, pp. 2622-2627.

[18] Nothdurft, H. C., Saliency and target selection in visual search, *Visual Cognition*, Vol. 14, August, 2006, pp. 514-542.

[19] Onat, S., Libertus, K., König, P. Integrating audiovisual information for the control of overt attention. *Journal of Vision*, Vol. 7, No. 10, July, 2007, pp. 1-16.

[20] Peters II, R. A., Hambuchen, R., Kawamura, K., Wilkes, D. The Sensory Ego-Sphere as a Short-Term Memory for Humanoids, *Proc. IEEE-RAS International Conference on Humanoid Robots*, Tokyo, Japan, November, 2001, pp. 451-459.

[21] Pfeifer, R., Lungarella, M., Sporns, O., Kuniyoshi, Y. On the Information Theoretic Implications of Embodiment Principles and Methods, to appear in *Proc. of the 50th Anniversary Summit of Artificial Intelligence*, Springer-Verlag.

[22] Posner, M. I., Rafal, R. D., Choate, L. S., Vaughan, J. Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, Vol. 2, pp. 211-228, 1985.

[23] Reichardt, W. E. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. *Principles of Sensory Communications*, New York, 1961, pp. 303-317.

[24] Richards, J. E. Cortical Indexes of Saccade Planning in Infants *Infancy*, Vol. 2, No. 2, 2001, pp. 123-133.

[25] Treisman, A. M., Gelade, G. A feature-integration theory of attention, *Cognitive Psychology*, Vol. 12, No. 1, 1980, pp. 97-136.

[26] Vergassola, M., Villermaux, E., Shraiman, B. I. 'Infotaxis' as a strategy for searching without gradients, *Nature*, Vol. 445, No. 7126, January, 2007, pp. 406-409.

[27] Wolfe, J.M. Guided Search 2.0: A Revised Model of Visual Search, *Psychonomic Bulletin & Review*, Vol. 1(2), 1994, pp. 202-238.