Smoke and Mirrors—Virtual Realities for Sensor Fusion Experiments in Biomimetic Robotics

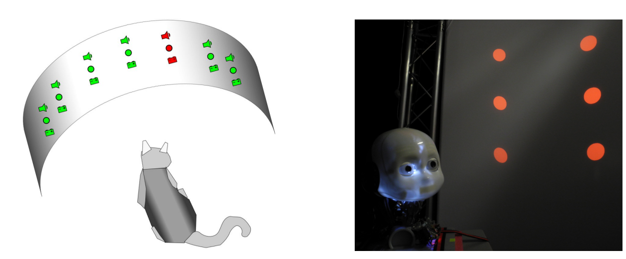
Johannes Bauer, Jorge Dávila-Chacón, Erik Strahl, Stefan Wermter
Department of Informatics
University of Hamburg
Email: {bauer,davila,strahl,wermter}@informatik.uni-hamburg.de

***Abstract*—Considerable time and effort often go into designing and implementing experimental set-ups (ES) in robotics. These activities are usually not at the focus of our research and thus go underreported. This results in replication of work and lack of comparability. This paper lays out our view of the theoretical considerations necessary when deciding on the type of experiment to conduct. It describes our efforts in designing a virtual reality (VR) ES for experiments in biomimetic robotics. It also reports on experiments carried out and outlines those planned. It thus provides a basis for similar efforts by other researchers and will help make designing ES more rational and economical, and the results more comparable.**

I. Introduction

The abstractions and simplifications we use when we design systems often make it impossible to strictly prove their correctness or fitness. Experiments, in the widest sense, are the tool that still allows us to validate or refute a specific idea. In the case of cognitive robotics, examples of general ES are few so far, and best practices are not established. We therefore lay out in this paper our considerations for experiments in this field. As we will see, many of the same general rules apply for experimentally studying artificial and natural cognition.

Researchers in cognitive robotics have thus replicated classical experiments from the cognitive sciences. Ravulakollu et al. [1] replicated a classic neurophysiological experiment, due to Stein and Meredith [2]. They replaced the originally feline subject by a robotic one (see Fig. 1a) to show the similarity of the robotic response to that observed by Stein and Meredith in nature. As we are particularly interested in perception in biomimetic robotic systems, we would like to do similar things. One option would be replicating Stein and Meredith's experiments, where our robots would orient towards flashing lights and sound bursts. Another option is replicating those by Battaglia et al. [3], who used an audio-visual VR set-up to experimentally compare their human participants' performance to two models of multi-sensory integration, maximum likelihood estimation (MLE) and visual capture. Yet another experiment we may replicate is the one by Block and Bastian [4]. They used an interactive VR environment to observe the effect of induced disparity between vision and proprioception in a reaching task.



(a) Experiment Inspired By [2] (b) Experiment with iCub

Fig. 1: Replicating Neurophysiological Experiments with Robots

The rest of this paper is organized as follows: in Sec. II, we relate the degrees of validation one can pursue to the different kinds of experiments that can provide them. Against this background, we describe in Sec. III-A and Sec. III-B the virtual reality (VR) environment we have designed and implemented for multi-sensory robotic experiments. We explain the choices we made and techniques we used in terms of experimental validity on the one hand, and feasibility as well as flexibility on the other. Finally, in Sec. III-C, we describe work we have done in multi-sensory integration in robotics, explain where it fits into our considerations about experiments in general, and how we are going to continue it using our VR environment. This will serve for comparison with other ES used in the field, and thus help establish a clearer understanding of the necessities and best practices for this kind of research activity.

II. Sensory Robotic Experiments

The simplest goals of a robotic experiment are proving the robustness and fitness of a system. In either case, the standard applied depends on the complexity and the capabilities of the system under test. In the case of fitness, for example, this can mean showing that the system accomplishes the needs of a user. It can also mean that it does so better than previously introduced systems, or under different side conditions.

A significantly higher mark to aim for is showing optimality: in some instances of sensory processing, e.g. in simple audio-visual localisation, there are theoretical limits of how well a system can perform ([5, p. 585],[6]). These results

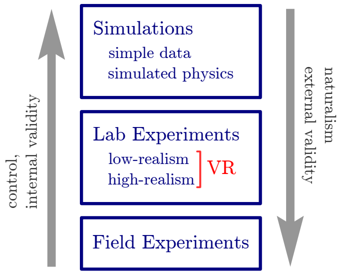


Fig. 2: Continuum of Experiments

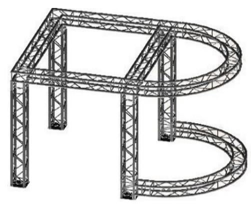


Fig. 3: Aluminium Truss Scaffold

usually come with strong assumptions on the situation, which can affect their applicability to real-world scenarios [7]. Thus, it is often hard to show optimality in a situation resembling the real ‘life’ of a robot.

Another possible goal, which is particular to experiments in biomimetic robotics, is showing that a system behaves like its natural counterpart. This is an important goal because one of the objectives of biomimetic robotics is to validate theories about the functioning of biological systems by modelling them and comparing the behaviour of the model to that of the original.

Let us now turn to the kinds of experiments, in the widest sense, which can be used to assess the performance of a particular system. As we will see, there is a continuum of how much complexity from the real world we allow into our experiments. Which of the goals described above an experiment can accomplish is largely determined by where in this continuum it falls (see Fig. 2).

A. Low Complexity – Simulations

The first kind of experiment we want to consider are simulations. At their most abstract, they are simple programs which generate input data and feed it to the system being tested. The system is often just an isolated algorithm and the data is generated according to the same assumptions, concerning data quality, noise, and world dynamics, that went into the design of the system. Rao, for example, used this kind of simulation to demonstrate the effectiveness of his artificial neural network (ANN) for inference with hidden Markov models [8] and we have used simulations to evaluate our SOM-based model for learning mapping and integration of multi-sensory stimuli [6].

More sophisticated simulations use software physics engines to generate more complex input and evaluate the output more functionally with respect to the environment. Examples include Milford et al.’s simulations showing how their system, called RatSLAM, maintains estimates of its own pose [9], and Stramadinoli et al.’s simulations with a simulated iCub robot grounding language in action [10].

Simulations have some advantages over more naturalistic experiments. One of the greatest advantages is that all the hardware needed is computing hardware. Also, individual subsystems can be tested in isolation from others. Simulations can be perfectly reproducible, and they make it possible to test performance in situations which are hard or impossible to induce in reality. Examples are situations occurring naturally in marine, submarine, and aerospace applications. Finally, they allow us to closely observe both overt and internal system behaviour. Our work mentioned earlier is an example of the former, that of Milford et al. and Rao an example of the latter: while we focussed on the functional aspects of our model, they both used simulations to compare biological neural activations to the neural activations in their models.

However, for conclusive evidence, experiments beyond simulations are needed, as the full complexity of the real world can never be captured in a computer system and thus the results are not guaranteed to be transferable [7].

B. Medium Complexity – Lab Experiments

Laboratory experiments allow to selectively admit real-life complexity into our tests. Results from lab experiments thus validate not only the behaviour of a system but also some of the assumptions.

The standards by which a robot’s performance can be measured in a lab experiment greatly depend on two things: one is the task it is to solve, the other is how well we understand the stimuli and actions available to the robot. If we understand them well enough and we can show that they are very similar in real life and in the experiment, then we can demonstrate robustness and fitness of our systems under natural conditions. If we can additionally observe or manipulate ground truth in the experiment, then it may even be possible to show optimality. In cases where we have behavioural or neurophysiological data from experiments with humans or animals, we can also demonstrate biological realism if stimuli and physics in natural and robotic experiments are sufficiently similar. However, the same observations that are possible on a robot are not necessarily possible on a human or animal. Specifically the kinds of higher-order tasks tested in lab experiments tend to be distributed all over the nervous system [11] and therefore only partially observable neurophysiologically. Thus, only behavioural data is usually available for these tasks for comparison between natural and artificial systems.

When we wrote about admitting real-life complexity, we also hinted at two limitations of lab experiments. The first is the lower bound of complexity let in. It is not always easy or possible to limit interaction of the real world with the test subject exactly as needed for the experiment. Filtering out confounding factors is an art and a challenge not only in robotic experiments. The second limitation is getting *enough* complexity into our ES to ensure realism and therefore external validity.

C. High Complexity – Field Experiments

With no feral robots to observe in their natural habitats, field experiments are the pinnacle of natural complexity and external validity in robotic experiments. All the input impinging on the system in these experiments is real, time constraints are real, and the environment reacts mostly real to the robots’ actions. Bringing robots into their designated field of operation, observing them and attributing their behaviour, their successes, and shortcomings to individual components is what makes field experiments difficult to conduct and ensure their internal validity.

D. Virtual Reality

VR experiments are technically lab experiments. What sets them somewhat apart from classical experiments, however, is the range of possible input stimuli and the control over the reactions of the environment to the robots' actions. VR experiments give us the opportunity to test our robots in circumstances close to those in the environment for which they are designed. At the same time, they allow us to precisely control the stimuli presented to our robots and closely observe their performance without interfering with the test conditions. Thus, they generally have greater external validity than simulations or simpler lab experiments, as at least some of the consequences of embodiment, like time-constraints, sensor and ego-noise, and real physics apply. In summary, the flexibility offered by VR environments, which allows us to tune internal and external validity to our needs, makes them highly attractive. On the other hand, it can be difficult to argue that all relevant aspects of reality are covered in VR environments, just like in regular lab experiments.

Concluding our considerations on experiments in general, we can say that the holy grail in biomimetic robotic experiments is a perfectly controlled field experiment showing optimal behaviour and/or quantitatively the same behaviour and simulated biology as that found in nature. However, not all of these standards can usually be achieved at the same time and often a simulation or a more restricted ES support our points just fine. In fact, a system can first be tested for general fitness in a simulation, then in various stages of a VR experiment, and finally in a field experiment.

III. Multi-Sensory Virtual Reality Lab

The practical requirements for the design of our VR set-up were affordability, ease of operation, and flexibility. Apart from these, the overarching design goal was to give us the maximum range of possibilities with respect to the continuum described in Sec. II. This meant that we wanted to be able to deliver highly controlled stimuli and observe our robot's actions as closely as possible. Also, the VR had to be able to create a rich, complex environment for the robot to behave in. Control and richness of stimuli both needed to be tunable to the needs of the individual experiments. Of course, it did not make sense to invest heavily in creating environmental realism that exceeded our robot's perceptual capabilities. However, we had to ensure that we could produce stimuli which were sufficiently natural to make the results comparable with those in the experiments we were going to reproduce.

On the visual side, this translated to being able to create a sharply focussed, uniformly illuminated picture with high resolution, covering as much of the screen with as little distortion as possible. We wanted the virtual scenes to cover at least $180°$ around the robot, horizontally, and $90°$ vertically in order to allow the robot to react to stimuli e.g. by turning towards them, and still be immersed in the display. For audition, we wanted to be able to generate sounds and precisely control their origin. We plan to compare sound source localization (SSL) of our robots to that of humans. The maximal resolution of human SSL is in the order of magnitude of about $2°$, horizontally, and $3.5°$ vertically [12]. We therefore have to be able to control sound sources with about that precision.

We considered three different types of immersive visual set-ups. One was an array of LCD or LED displays arranged around the robot head. This approach was comparatively simple, from the technical and hardware sourcing perspectives. However it was unclear whether we would be able to place the displays close enough to each other so that the picture would appear seamless. Even more importantly, it would have been very difficult to have sound coming from the precise location of a visual stimulus with a display made up of monitors. We therefore abandoned this approach relatively quickly.

Another idea was a projection scenario in which the robot was to be placed at the center of a half-sphere, or dome. The geometry of the projection would have been comparatively simple, and the distortion was promising to be easily calculated and compensated for. With the actual screen made out of thin fabric, and projection from the inside of the dome, it would have been possible to place speakers anywhere behind it and therefore achieve any desired spatial resolution. Projection could have been done with a single projector and a fish-eye lens or spherical mirror [13].

While we would have been able to simulate a $360°$ scene, horizontally, and $180°$, vertically with this set-up, the downsides of a dome projection outweighed the advantages. Most importantly, we feared a dome-shaped projection screen might have adverse acoustic effects on auditory localisation, the screen and the structure holding it would have had to be custom-built and thus costly, and a dome would have made the space it occupied unusable for anything else.

For these reasons, we decided for a third option, in which the robot sits at the center of a half-cylindrical instead of a half-spherical screen (see Figs. 3 and 1b). The advantages of this solution are that it did not require going through a specialised manufacturer for the structure, projectors, and potentially additional optics and that it offers much greater flexibility and easier handling and manipulation of the robot during the experiments. On the other hand, it requires using multiple projectors and projecting at close distances with overlapping projection areas.

When we designed the metal structure which was to hold screen, projectors, speakers, and robot, we opted for an aluminium truss scaffold filling our entire lab. This gives us greatest flexibility in placing the ES components plus the ability to extend our audio-visual VR by adding real or simulated components for different senses. The screen spans a half circle with a diameter of 2.60 m, and has a height of about 2.2 m. The robot head is fixed at the center of this half circle at about the height of the vertical center of the screen. Looking straight ahead, the projection takes up all of its visual field. It can turn by about $67°$ horizontally in either direction

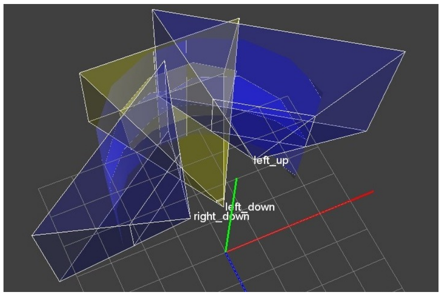


Fig. 4: Array of Projectors

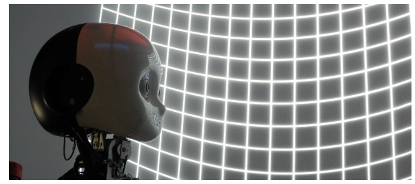


Fig. 5: iCub with Distorted Background

without seeing the borders of the screen.

The robot head which will be used in our experiments is the iCub head (see Fig. 1b). The iCub is a humanoid robot highly suitable for research in artificial intelligence, developmental robotics and embodied cognition. It is state of the art in terms of kinematic design and anatomical similarity to an approximately three-year-old child. The neck has 3 degrees of freedom (DoF) for tilt, swing, and pan movements. There are 3 DoF for oculomotor control: one for the eyes' common vertical orientation, and one for each eye's horizontal orientation. Each eyeball contains a VGA colour camera. The head contains two microphones surrounded by pinnae [14]. Our aim is to perform robotic experiments in biomimetic audiovisual and visuomotor coordination. This makes the iCub head ideal for us, as its design was driven by the idea to mimic human head and eye movement.

A. Projection

a) Choice of Projectors: We used Projection Designer[1], an open-source software which simulates various aspects of non-standard projections, to compare a number of different combinations of projectors and find the solution which matched the competing requirements explained in Section III as closely as possible. In the end, we chose a set-up using four moderately wide-angle Optoma GT 750 projectors located above and below the robot head. Less projectors would have been enough had we been able to use wide-angle projectors rotated by $90°$. This would have decreased the over-all complexity of the set-up. However, the vendors of the projectors in question did not guarantee for the lifespan of projectors operating in this position.

b) Determining and compensating for distortion: Using four projectors to project against the inner surface of a half-cylinder invariably leads to overlap and strong distortion of the individual projections (see Figs. 4 and 5). In theory, the distortion, which is determined by the characteristics of the projectors, the geometry of the screen, and the position of the camera, can be determined mathematically. One can then pre-distort the image such that it will appear even to the camera, and its brightness can be adjusted such that the overlap will be invisible.

Unfortunately, it is very hard to determine or enforce these parameters sufficiently well to do the math and compensate for the distortion. Take for example the projection angles. The scaffolds and anchors holding the projectors would need to be flexible enough on the one hand such that the angles can be set very precisely, and stiff enough on the other hand such that gravity will not change the angles immediately after setting them. The position of the lenses in the projectors would have to be known exactly, and the settings of the projectors themselves for shifting and scaling the image would have to be set to some previously determined configuration. Small aberrations could already lead to projectors not being aligned correctly. All of this makes a 'modelled' approach largely impractical. We therefore chose to pursue a model-free approach, or rather an approach which uses a non-geometric, non-optical model.

In short, what we do is use the iCub's camera and motion capabilities to empirically determine the relationship between pixel positions in the projected image and angles from the robot at which they appear. For this, we first project the horizontal and vertical lines of a white grid, one after another. Whenever a pixel lights up for a vertical line at position i and a horizontal line at position j, we store the horizontal and vertical angles α and β at which it appears to the robot. Let $\mathcal{B} = \{\mathsf{b}_1, \mathsf{b}_2, \ldots, \mathsf{b}_k\}$ be a set of k polynomial and Gaussian basis functions and $(i_n, j_n, \alpha_n, \beta_n)$ be the n^{th} 4-tuple thus collected. Then we construct vectors $A = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ and $B = (\beta_1, \beta_2, \ldots, \beta_N)$, for N the number of 4-tuples, and a $2k \times N$ matrix $\mathbf{X}$ such that the entry $\mathbf{X}_{l,m}$ of $\mathbf{X}$ is

$$\mathbf{X}_{l,m} = \begin{cases} \mathsf{b}_l(i_m) & \text{if } l <= k \\ \mathsf{b}_{l-k}(j_m) & \text{if } l > k \end{cases},$$

for $0 < l \leq 2k$ and $0 < m \leq N$. Finally, we use linear least squares regression to get approximate solutions b_A and b_B to the equations

$$\mathbf{X} b_A = A^T \quad \text{and} \quad \mathbf{X} b_B = B^T.$$

Together with the basis functions $\mathcal{B}$, this gives us the parameters of two mixture models for calculating the horizontal and vertical angles, respectively, to which a given pixel is projected. We pre-calculate the values of these models for every pixel position in the projected image and generate C code which uses the resulting matrix for OpenGL online undistortion of 3D scenes [15]. This procedure allows us to change the position of the robot and the projectors to accommodate every experiment's need for precision and realism without having to fine-tune them mechanically every time.

As general and simple, conceptually, as our approach is for undistortion, it does not give us an easy way to also perform edge blending, i.e. fading out one projector's image

[1] http://orihalcon.jp/projdesigner/

into another in the area where they overlap. While this is not a difficult problem for multiple projectors projecting in parallel against a flat screen [16], it does become non-trivial when the overlap is not rectangular and the projectors' scan lines do not have the same orientation. We thus decided to simply check for every pixel in one projector's image whether it was seen at the same angles as pixels in other projectors' images, and dim it accordingly.

B. Sound

One of the first considerations for designing the sound set-up was the kind of loudspeakers to use. An ideal sound source in an SSL experiment should originate from a single point, so the robot's localisation error can be quantified precisely. Another requirement is to reproduce a broad range of frequencies with high fidelity and enough intensity to cover the background noise generated by projectors, power sources, and the robot itself. In contrast to speaker arrays, coaxial speakers comprise different drivers and membranes for lower and higher frequencies which vibrate parallel to the same axis. Good coaxial speakers thus fulfill both the requirement for crisp localisation and for high frequency bandwidth.

It is possible to create a spatial impression along the azimuth with just two speakers, by varying the time of onset and intensity of sounds. To also create the impression of elevation, a set of so-called head related transfer functions (HRTF) is needed [17]. These HRTFs simulate the effect of a hearer's torso and pinnae. Such a set of HRTFs is valid only for a discrete number of positions of the hearer wrt. the sound sources. This is a problem for experiments in which a robot is to move continuously, often facing directions for which there is no known HRTF. We therefore opted for a much easier approach: in our VR, the robot is actually surrounded by a number of speakers along the azimuth and elevation planes. On top of these consumer-grade, single-membrane speakers, we acquired a pair of high quality coaxial speakers for simple, high-precision localisation scenarios.

For basic experiments with horizontal localisation, it is enough to place speakers e.g. at every $15°$ in the azimuth, from $0°$ to $180°$, and at $0°$ of elevation. When doing experiments with vertical localisation as well, identical speaker line-arrays need to be placed along the elevation plane. Horizontal line-arrays can have $30°$ steps between $\pm 60°$ on the elevation plane, and a more advanced configuration can increase to $15°$ steps. In order to test the limits of SSL near the intersection of azimuth and elevation axes, where SSL accuracy in humans is at its highest, speakers have to be placed very close to each other in front of the robot's head. Such performance has been approached by some biomimetic algorithms [18], [19] and thus flexibility to accommodate these types of experiments is a requirement for the physical infrastructure of the lab (see above).

Sounds with clearly defined parameters, such as pure tones or white noise, can be generated with many numerical manipulation software applications. The relevant part of the generation of stimuli is the possibility to reproduce it on the required speaker at predefined points in time. One alternative is to create as many instances of sound-reproducing modules as the number of speakers in the set-up. Afterwards, it is possible to connect each of them with software applications such as Jack-Audio[2], and instruct the reproduction of the desired sound file at the required time.

C. Audio-Visual Experiments

In previous work, we have developed a SOM-based model for learning mapping and integration of multi-sensory signals as performed by the SC [6]. At the level we modelled this process, it can be shown that, with Gaussian noise in the different modalities, a linear maximum likelihood estimator (MLE) performs optimally at this task [20]. Given the variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ of the Gaussian curves describing the noise in the n modalities whose cues $c_1, c_2, \ldots, c_n$ are to be integrated, a linear MLE computes a weighted average

$$c_{MLE} = \frac{1}{\sum_{i=1}^{n} \sigma_i^2} \sum_{i=1}^{n} \sigma_i^2 c_i .$$

Psychophysics experiments have provided evidence that human behaviour when localising audio-visual stimuli indeed is well-modelled by a linear MLE [21]. In a simulation of the most abstract type described in Section II, we showed that our model was capable of learning how to combine near-optimally noisy stimuli for localisation.

In simple lab experiments using a single speaker, our group experimented with SSL using spiking ANNs [19]. These ANNs modelled how mammals integrate interaural time (ITD) and level (ILD) differences for auditory localisation. The output of this model was used to produce motor commands for the robot to face in the direction of a sound source. First, ITDs and ILDs are extracted from a set of sound frequency components with spiking neural models of the medial superior olive (MSO) and the lateral superior olive (LSO) [22]. Then MSO and LSO outputs were integrated in a model of the inferior colliculus (IC) which provided a more coherent spatial representation across frequencies. The IC model has $j \in \{1 \ldots n_{IC}\}$ neurons for each of the f frequency components that it analyses. The value of n_{IC} equals the total number of azimuth angles θ around the robot where a sound is produced during an experiment. The connection weights from MSO and LSO neurons to IC neurons were estimated using Bayesian inference:

$$p\left(\theta_j | S_f\right) = \frac{p\left(S_f | \theta_j\right) p\left(\theta_j\right)}{p\left(S_f\right)} ,$$

where S_f is the number of spikes produced by MSO and LSO neurons for a given sound. This inference process displayed a robust performance on a robot with high levels of ego-noise. Experimental results showed that the algorithm is capable of differentiating sounds with an accuracy of $15°$.

As a next step, we want to combine our work from auditory localisation and multi-sensory integration, and test the

[2] http://www.jackaudio.org/

resulting system in increasingly more realistic experiments in our VR environment: In Section II, we stated that one goal in experiments with biomimetic robotic systems is showing that these systems behave like the biological systems they model. We also explained that biological realism on the level of neurophysiology can only be shown where comparable data exists on that neurophysiology. This is the case mainly for relatively simple cognitive processes which are somewhat removed from real life. Experiments in psychophysics and neurophysiology like those due to Stein and Meredith [2], Battaglia et al. [3], and Block and Sebastian [4] provide such data. Some of the experiments we are going to conduct in our VR set-up will therefore be modelled after such experiments.

Moving on to the goals of experiments in general robotics, more life-like experiments will be further down the road of our research with the VR environment described in this paper. Single- or multi-speaker recognition in a home or meeting room scenario could be tested. This could be done with our models alone or in combination with models for other, higher-level, sensory processing, like face detection or spoken language comprehension. Also, a vast amount of effort that has gone into creating realistic simulations especially in ego-shooter-type computer games, some of them open source. This effort could be harnessed for scientific purposes. The similarity between localisation and gaze direction on the one hand and aiming and shooting on the other as well as the high demands in speed and precision make this an attractive path, although less martial contents would be desirable. The standards here will either be the performance of other artificial systems, fitness for some purpose, or other metrics depending on the task. In our VR set-up, a system can be tested under the exact same conditions with very different types of stimuli in successive stages of one run of an experiment. This is where VR environments shine.

IV. Conclusion

The considerations laid out in this paper provide reference on multiple levels to anyone designing experiments in cognitive robotics. The continuum described in Sec. II defines the different classes of experiments and what kinds of evidence they can provide. It thus helps identify which types of experiment are needed to validate the capabilities of a specific system. It pays special attention to the role of ESs based on VRs, adding additional relevance for anyone considering this kind of experiment. Sec. III points out requirements and options for building an audio-visual VR. In particular, Secs. III-A and III-B discuss challenges and solutions specific to projection and simulation of localised sound sources, which will be of use for the robiticist designing such a VR. Finally, the description of experiments we have carried out and planned puts all of the above into a practical perspective.

Acknowledgements

This work is funded in part by the DFG German Research Foundation (grant #1247) – International Research Training Group CINACS (Cross-modal Interactions in Natural and Artificial Cognitive Systems).

References

[1] K. Ravulakollu, M. Knowles, J. Liu, and S. Wermter, "Towards computational modelling of neural multimodal integration based on the superior colliculus concept," in *Innovations in Neural Information Paradigms and Applications*, ser. Studies in Computational Intelligence, M. Bianchini, M. Maggini, F. Scarselli, and L. Jain, Eds. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2009, vol. 247, ch. 11, pp. 269–291.

[2] B. E. Stein and M. A. Meredith, *The Merging Of The Senses*, 1st ed., ser. Cognitive Neuroscience Series. MIT Press, Jan. 1993.

[3] P. W. Battaglia, R. A. Jacobs, and R. N. Aslin, "Bayesian integration of visual and auditory signals for spatial localization," *Journal of the Optical Society of America A*, vol. 20, no. 7, pp. 1391–1397, Jul. 2003.

[4] H. J. Block and A. J. Bastian, "Sensory weighting and realignment: independent compensatory processes," *Journal of Neurophysiology*, vol. 106, no. 1, pp. 59–70, Jul. 2011.

[5] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, Dec. 2009. [Online]. Available: http://www.worldcat.org/isbn/0136042597

[6] J. Bauer, C. Weber, and S. Wermter, "A SOM-based model for multisensory integration in the superior colliculus," in *Proceedings of the International Joint Conference on Neural Networks (2012 : Brisbane, Australia)*. IEEE, 2012, to appear.

[7] T. van der Zant and L. Iocchi, "Robocup@ home: Adaptive benchmarking of robot bodies and minds," *Social Robotics*, pp. 214–225, 2011.

[8] R. P. N. Rao, "Bayesian computation in recurrent neural circuits," *Neural Computation*, vol. 16, pp. 1–38, Jan. 2004.

[9] M. J. Milford, J. Wiles, and G. F. Wyeth, "Solving navigational uncertainty using grid cells on robots," *PLoS Computational Biology*, vol. 6, no. 11, pp. e1 000 995+, Nov. 2010.

[10] F. Stramadinoli, M. Ruciński, J. Znajdek, K. J. Rohlfing, and A. Cangelosi, "From sensorimotor knowledge to abstract symbolic representations," *Procedia Computer Science*, vol. 7, pp. 269–271, Jan. 2011.

[11] K. Hartmann, G. Goldenberg, M. Daumüller, and J. Hermsdörfer, "It takes the whole brain to make a cup of coffee: the neuropsychology of naturalistic actions involving technical devices," *Neuropsychologia*, vol. 43, no. 4, pp. 625–637, Jan. 2005.

[12] J. Middlebrooks and D. Green, "Sound localization by human listeners," *Annual review of psychology*, vol. 42, no. 1, pp. 135–159, 1991.

[13] P. Bourke, "Using a spherical mirror for projection into immersive environments (mirrordome)," in *3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, S. N. Spencer, Ed. ACM, Nov. 2005, pp. 281–284.

[14] R. Beira, M. Lopes, M. Praça, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltarén, "Design of the robot-cub (icub) head," *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 94–100, May 2006.

[15] P. Bourke, "Lens correction and distortion," http://paulbourke.net/miscellaneous/lenscorrection/, accessed: May 25, 2012.

[16] ——, "Edge blending using commodity projectors," http://paulbourke.net/texture_colour/edgeblend/, accessed: May 25, 2012.

[17] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge: The MIT press, 1997.

[18] J. Liu, D. Perez-Gonzalez, A. Rees, H. Erwin, and S. Wermter, "A biologically inspired spiking neural network model of the auditory midbrain for sound source localisation," *Neurocomputing*, vol. 74, no. 1-3, pp. 129–139, 2010.

[19] J. Dávila-Chacón, S. Heinrich, J. Liu, and S. Wermter, "Biomimetic binaural sound source localisation with ego-noise cancellation," in *Proceedings of the International Conference on Artificial Neural Networks (2012 : Lausanne, Swiss)*. Lecture Notes in Computer Science, Springer, 2012.

[20] Z. Ghahramani, "Computation and psychophysics of sensorimotor integration," Ph.D. dissertation, Massachusetts Institute of Technology, Sep. 1995.

[21] D. Alais and D. Burr, "The ventriloquist effect results from near-optimal bimodal integration," *Current Biology*, vol. 14, no. 3, pp. 257–262, Feb. 2004.

[22] J. Schnupp, I. Nelken, and A. King, *Auditory neuroscience: Making sense of sound*. The MIT Press, 2011.