Towards an enactive robot audition architecture

Valentín Lunati, Patrick Danès, Claudia Arias, Fernando Bermejo, Fernando González, Juan Rosales, Rodrigo Pérez

ICA 2016 BUENOS AIRES
Buenos Aires – 5 to 9 September 2016
Acoustics for the 21st Century...

PROCEEDINGS of the 22nd International Congress on Acoustics

Acoustic Array Systems: Paper ICA2016-593

Towards an enactive robot audition architecture

Valentín Lunati[(a)], **Patrick Danès**[(b)], **Claudia Arias**[(c,d)], **Fernando Bermejo**[(a,d)], **Fernando González**[(a)], **Juan Rosales**[(a)], **Rodrigo Pérez**[(a)]

[(a)]Centro de Investigación y Transferencia en Acústica - Universidad Tecnológica Nacional, Facultad Regional Córdoba - Unidad Asociada al Consejo Nacional de Investigaciones Científicas y Técnicas (CINTRA - UTN FRC - UA CONICET), Argentina, lunativ@gmail.com
[(b)]LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France.
[(c)]CONICET en CINTRA - UTN FRC - UA CONICET, Argentina.
[(d)]Facultad de Psicología - Universidad Nacional de Córdoba, Argentina.

Abstract

Robots are usually equipped with advanced capabilities in order to autonomously adapt to real and dynamic environments and to interact with humans. Robot Perception is being inspired by new embodied cognition approaches that redefine the notions of perception, cognition and action, basic processes of intelligent behaviour. Enactive approaches consider the perceptual act as a consequence in action of the structural coupling between the organism and its environment in seek of significance. There is a growing interest in the development of robot perception systems based on new architectures to materialize naturally these action-perception functions. In this direction, we propose an evolution of the EAR sensor [4], which fulfills the constraints of mobile robot audition, such as embeddability, synchronous multichannel acquisition and real-time execution. Using Systems-on-a-Programmable-Chip (SoPCs) methodology, this sensor incorporates a novel architecture that offers all the basic calculation blocks necessary to perform most binaural and array auditory functions, and allows to easily develop new functionalities and connections between motor and perceptual modules in order to implement enactive behaviour. Moreover, Microelectromechanical Systems (MEMS) microphones have been studied and implemented, enabling the acquisition of high-fidelity audio on inexpensive and portable devices. In this paper, performance results are presented for sound source detection and localization functions, and progress is shown towards the implementation of MEMS microphones in Human-Machine Interfaces and Robot Audition. Finally, evolutions towards an interdisciplinary design of enactive audition functions are discussed.

Keywords: Robot Audition; SoPC; Microphone Array Processing; Enactive Approach

Towards an enactive robot audition architecture:

1 Introduction

Nowadays, there is a great scientific and technological interest in the development of robotic perception systems which use architectures enabling "natural"—i.e., as a living organisms do in real environments—low level action-perception closed loops named sensorimotor contingencies (SMCs).

Robotics, sustained on active and enactive theoretical perspectives, is achieving promising advances in the autonomy and adaptation of robot agents interacting in real environments. Precisely, one of its main challenges is the development of models for a successful sensorimotor coupling between agent and environment. These models require new computational architectures to implement the SMCs observed in human studies.

The present work takes place within this context. First, section 2 recalls some advances in enactive Robotics, computational models in Robot Audition, as well as the SoPC methodology for their implementation into smart integrated sensors. Using this methodology a new architecture is described in section 3 and results of computational performance of some auditory localization functions are shown. Finally, some conclusions about the application of the architecture to real robots and future works related to implement SMCs are discussed.

2 Embodied approaches to cognition and Enactive Robotics

Approaches from complexity theories are progressively gaining greater explanatory power in cognitive sciences [7]. Particularly, enactive perspectives and the sensorimotor theory contingencies [15, 12], assume that perception is based on the regularities that govern the ongoing coupling between the agent's action and the subsequent sensory changes, i.e., the sensorimotor contingencies, SMCs [15].

The interaction of the agent with his/her environment and/or with other agents constitutes a complex and autonomous dynamic system. The agent must be able to regulate his/her behaviour adaptively according to the dynamic changes (disruptions) that occur in him/herself, in the environment or in the other interacting agents, what implies a continuous process of sensorimotor learning [2].

The study of the role of interaction in the construction of intelligent behaviour, that is, how to perform actions in order to give meaning to each interaction (sensemaking), is a topic of great relevance and interest. In this context, social interaction is defined as a continuous process of reciprocity and interdependence, where the sensorimotor contingencies are redefined.

Robotics is achieving promising progresses based on enactive perspectives in relation to autonomy and adaptive behaviour of robotic agents [20, 9]. One of its main challenges is the development of models that implement successful sensorimotor coupling, what is useful to interact with other agents located in real environments. Such models imply, on the one hand,

the development of a new computer architecture specifically designed to implement SMCs observed in human studies. On the other hand, the use of techniques from stochastic estimation, information theory, and control, becomes an appropriate framework to address the problem of motion control guided by information extracted from the sensorimotor flow.

2.1 Robot audition modelling

Robot audition modelling has been a very active field since more than a decade ago. Some contributions with special interest for the present article are the following:

1) Okuno *et al.* [13] proposed a set of low level auditory functions based on the CASA (Computational Auditory Scene Analysis) approach [5]: sound source localization and tracking, audio stream separation and extraction, identification and recognition of each source.

2) Philipona *et al.* [16] presented a statement of the sensory-motor control loop on audition using a mathematical model to determinate the dimension of the navigation space of the robot, a work continued by Laflaquiere *et al.* [10] using a nonlinear approach.

3) Portello *et al.* [19, 17], developed an audio-motor sound source localization strategy using stochastic estimation techniques, based on the binaural audio stream, the robot motor commands and considering the acoustic scattering of the robot head. On [6] a feedback controller of the sensor motion is proposed so as to reduce the associated uncertainty of the stochastic estimation.

4) Oliveira *et al.* [14] proposed a model for Robot-Human Interaction that integrates a set of low and high level auditory functions for non verbal and verbal interactions.

5) Georgeon *et al.* [9] assumed an enactive robot audition approach which requires a unified treatment of sensorimotor information without prior knowledge of the environment. This implies a dynamic mapping of the auditory scene including, among other parameters: sound sources positions and movements; their spectral contents, intermittent characteristics and semantic connotations; estimation of their own and other agents perceptual and motor actions. The authors proposed the Enactive Markov Decision Process (EMDP) model, that allows an agent to learn satisfactory SMCs to motivationally interact with the environment and other agents.

2.2 SoPC methodology for Robot Audition

The models mentioned above require for their implementation on real robots, a specific computational architecture that fulfills size restrictions, real time execution and energy economy. A System on a Programmable Chip (SoPC) is a processor based system implemented in a programmable architecture like a Field Programmable Gate Array (FPGA) combined with custom hardware modules that perform application specific operations. To implement signal processing algorithms on a SoPC, those can be implemented in hardware or software. The software implementation is executed by a processor and those operations that require a better precision or temporal performance can be directly implemented in custom hardware (Intellectual Property block, IP). This hardware/software partitioning depends of the performance required by the application. Operations can be accelerated in hardware by using application-specific modules

connected directly to the processor internal architecture. These modules—or coprocessors units—speed up specific instructions of the processor and can perform more complex operations while the main processor continues with others tasks. On [11] a SoPC approach was presented in order to implement array functions on an auditory sensor for Robotics. In addition, the electronic devices currently available on the market, have the computational capacity to run complex operating system like GNU/Linux. This improves the development, allowing to use software from a desktop PC easily on the SoPC. A first application of this methodology to Robot Audition was the inclusion of a SoPC to the Embedded Audition for Robotics (EAR) project [1, 4, 11]. This aimed at providing an acoustic sensor which can embed low-level auditory functions, such as source detection, localization, extraction/separation, etc. The first release of this sensor was implemented with a Xilinx Virtex IV FPGA and two homemade mezzanine 8-channel DAQ boards. This hardware, in conjunction with a linear array of N = 8 phase-matched microphones and a host UNIX-based computer, had been used to implement several auditory functions for speaker localization and speech extraction [11].

3 Multichannel audio DSP SoPC for Robot Audition

As an evolution of the EAR SoPC architecture, a new SoPC was developed based on a System On Chip (SOC) Zynq from Xilinx. This new system has functionalities for multichannel audio processing techniques for robot audition functions—like sound source localization and extraction—and for the spatial sound reproduction, using MEMS microphones and loudspeakers arrays. The interface to drive loudspeaker arrays allows the implementation of active auditory functions that requires acoustic interaction with the environment like localization of objects using echolocation [3] or interactions with another agent (verbal or non verbal interactions). Also, the system has the capacity to be connected to other systems through standard interfaces in order to send motor orders for the robot through a Robot Operating System (ROS) network to materialize the SMCs.

This system has been designed to be applied in two directions: to implement and evaluate different binaural and array auditory functions and as a research tool for the instrumentation of behavioural experiments that aims to study perceptual actions from the enactive approach. Figure 1 reports a simplified diagram applied on an experimental setup for a sound source localization test on humans or robots equipped with an array of MEMS microphones.

The system is integrated by several IP blocks (green boxes on the figure) designed on a Hardware Description Language such as VHDL (VHSIC Hardware Description Language), other blocks hardwired on the Zynq SOC—like the ARM Processor core— as well as the communications and I/O modules. The acquisition and reproduction stages—implemented as VHDL IPs on the FPGA section of the Zynq—are connected and controlled by the ARM processor through an AXI interconnect bus. The usage of IPs blocks allows to easily modify the capabilities of the system, by adding or modifying the functionalities and the number and types of I/O devices. In this case, it is easy to use more loudspeaker or microphones by simply adding the respective IPs cores on the SOC and connecting the devices on the adequate pins of the board. Finally, it is possible to perform functions on hardware, like the FFT calculation of each

input channel, beamforming with several kinds of filters, etc. For example, the acquisition stage has an FFT IP chained with a cross correlation IP on the FPGA section of the Zynq device. These IPs are directly connected to the MEMS microphones array input. In this way, the processor core can read the audio, its spectrum and the cross correlation signals as needed. Also the processor can use these IPs to process signals stored in a memory. The only limitation lies in the resources usage of the FPGA on the Zynq device. Table 1 shows the usage utilization of the entire system. As it can be observed, there are still lots of unused resources. So the entire system developed fits into a single component (the Zynq device) that not only is fully re-configurable, but also has a very low power consumption, 1.478 W for this application. This last feature is important, considering that the system can achieve better computational performance that an desktop PC which usually consumes 10 to 40 W and requires a lot of other electronic devices to work.

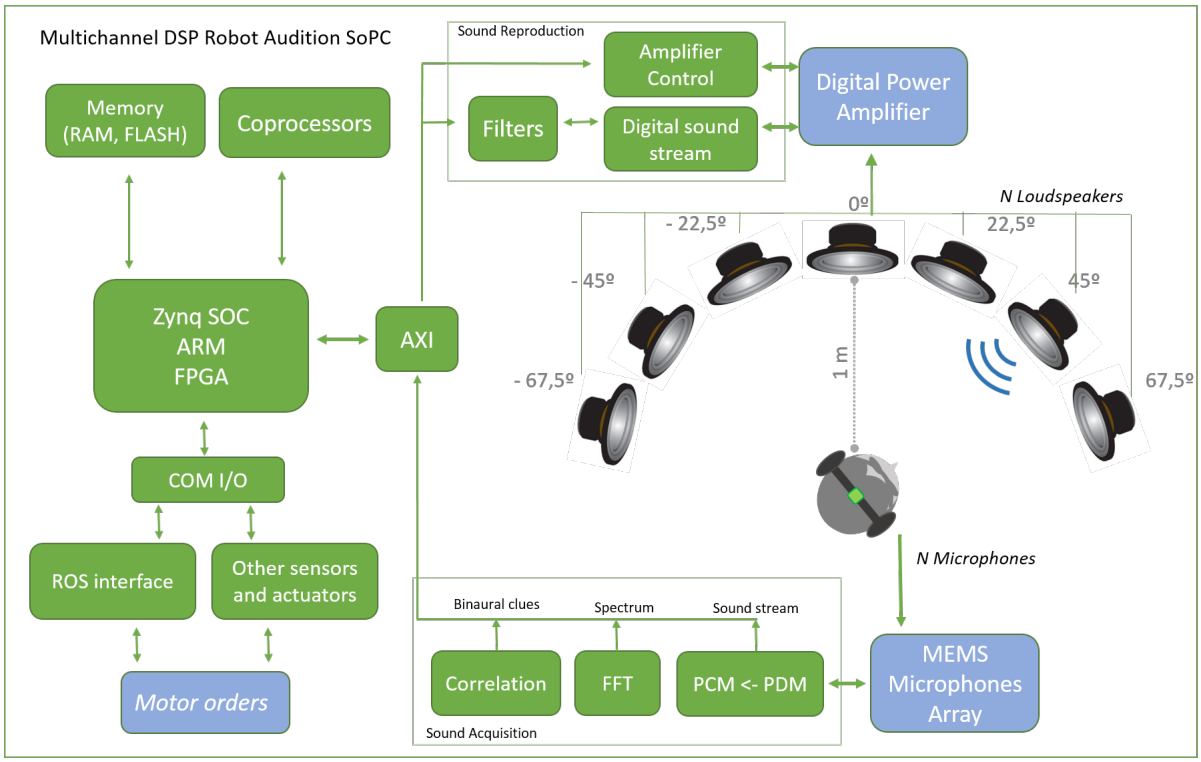

Figure 1: Multichannel DSP Robot Audition SoPC. Green boxes are the implemented IP cores on the Zynq device. Blue boxes on the figure represent external elements: the microphones array, loudspeakers amplifier and the interface to send the motor orders for the robot.

3.1 First results

In order to evaluate the computational performance of this new architecture, a fully software broadband MUltiple SIgnal Clasification (MUSIC) method of [8, 11] was implemented on the ARM processor. In figure 2 (a) and (b) localization spectrum results are shown for simulated

broadband sources (pink noise filtered between 300 and 3000 Hz, Fs=15024 Hz) at different azimuths. Figure 2 (c) shows the localization results from a real source measured with a linear microphone array (N=8 spaced 5.66 cm, signal: pink noise filtered between 300 and 3000 Hz, Fs=15024, SNR of 18dB). In this case the sound source is placed at 90° and is localized by the algorithm at 94° as shows the dot-dashed line. Then applying a statistically learning of the room noise when the source if inactive (as described on [8, 11]), the source is correctly localized as shows the continuous line. All operations were made on double floating point precision (instead of the EAR SoPC implementation [11]), what demonstrates a better performance in noisy scenarios as the last one described. The performance results were similar to the EAR sensor, obtaining an execution time of 20 ms (22 ms on EAR sensor). By using the FFT and others IP a better performance can be obtained, but thanks to the better performance of the ARM processor in contrast the Microblaze of EAR SoPC, a fully software implementation fulfills the real time requirements. The advantages of the hardware acceleration will become important to work with larger arrays and higher sampling frequencies and to execute several auditory functions simultaneously.

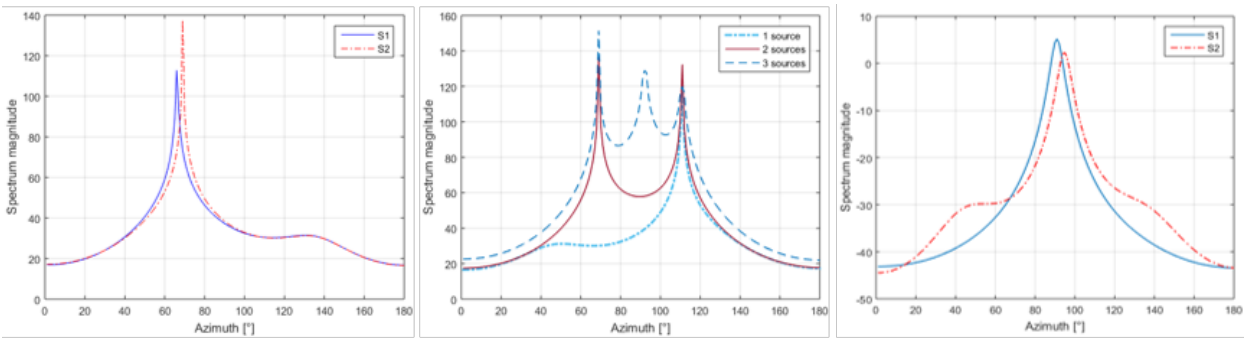

Figure 2: Sound source localization results. The peaks of each spectrum indicate the position (in azimuth) of the source. (a) Two simulated sources S1 at 65° and S2 at 68°. (b) Scenarios with one, two or 3 simulated sources active simultaneously. (c) Localization of a real source from an array of N=8 microphones

4 Conclusions and projections

As can be observed, the new system conceived has the capability to execute auditory functions in real time, this means that has the sufficient computational power to implement intensive DSP calculations with a reduced and known latency. In addition, the Zynq devices fulfills the constraints of power and size for robotics systems because its low power consumption and an adequate package.

The next step on developing and enactive audition sensor will be to implement on this architecture perceptual models of the SMCs from the above mentioned auditory functions integrated with the order motors of the robot and other actuators as the loudspeakers.

Resource	Available	Utilization	%
FF	106400	4383	4.12
LUT	53200	8860	16.65
Memory LUT	17400	62	0.36
I/O	200	17	8.5
BRAM	140	58	41.43
DSP48	220	80	36.36
BUFG	32	7	21.88

Table 1: Resource utilization of the SOC Zynq Z-7020

Acknowledgements

Valentín Lunati is funded by a doctoral grant from the CONICET, Argentina. This project was supported by grants from Universidad Nacional de Córdoba (PIDs N° 05/P130 and 05/P167) and Universidad Tecnológica Nacional (PID N° 1711), both from Argentina.

References

[1] Argentieri, S. and Danès, P. (2007) Broadband variations of the MUSIC high-resolution method for sound source localization in robotics. IEEE/RSJ IROS'2007, pp. 2009–2014.

[2] Barandiaran, X., Di Paolo E., and Rohde M. (2009) Defining Agency: Individuality, Norativity, Asymmetry, and Spatio-Temporality in Action. Adaptive Behavior, 17(5): 367–386.

[3] Bermejo, F., Di Paolo, E. A., Hüg, M. X. and Arias, C. (2015). Sensorimotor strategies for recognizing geometrical shapes: a comparative study with different sensory substitution devices. Frontiers in Psychology vol 6, 2015.

[4] Bonnal, J., Argentieri, S., Danès, P., Manhès, J., Souères, P. and Renaud, M. (2010). The EAR Project Journal of the Robotics Society of Japan (RSJ), 28(1):10-13. (Special issue on Robot Audition)

[5] Bregman, A.S. (1990). Auditory scene analysis : The perceptual organization of sound. MIT Press, Cambridge, MA

[6] Bustamante, G., Danès, P., Forgue, T. and Podlubne, A. (2016). Towards information-based feedback control for binaural active localization. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, Australia

[7] Chemero, A. (2009). Radical Embodied Cognitive Science. Cambridge, MA: MIT Press.

[8] Danès, P. and Bonnal, J. (2010) Information-theoretic detection of broad-band sources in a coherent beamspace MUSIC scheme. IEEE/RSJ IROS'2010, pp. 1976–1981

[9] Georgeon O., Wolf C. and Gay S. (2013) An enactive approach to autonomous agent and robot learning. Proceedings of IEEE 3º JIC EPIROB, Osaka, Japan: 1-6

[10] Laflaquiere, A, Argentieri, Breysse, O., Genet, S., S. and Gas, B. (2012) Non-linear Approach to Space Dimension Perception by a Naive Agent. . Proceedings de IEEE/RSJ IROS, Algarve, Portugal: 3253-3259

[11] Lunati, V., Manhes, J. and Danès, P.: A versatile System-on-a-Programmable-Chip for array processing and binaural robot audition. Proceedings de IEEE/RSJ IROS 2012, Algarve, Portugal: 998-1003

[12] Noë, A. (2004). Action in Perception. Cambridge, MA: MIT Press

[13] Okuno, H.G., Ogata, T., Komatani, K. and Nakadai, K. (2004) Computational Auditory Scene Analysis and its Application to Robot Audition. Proceedings of IEEE ICKS: 73-80

[14] Oliveira, J.L., Ince, G., Nakamura, K., Nakadai, K., Okuno, H.G., Reis, L.P. and Gouyon, F. (2012). "An active audition framework for auditory-driven HRI: Application to interactive robot dancing," RO-MAN, 2012 IEEE , vol., no., pp.1078,1085, 9-13 Sept. 2012

[15] O'regan, J. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. The Behavioral and Brain Sciences, 24(5) 939-973

[16] Philipona, D., O'Regan, J.K., and Nadal, J.P. (2003). Is There Something Out There? Inferring Space from Sensorimotor Dependencies. Neural Computation, 15(9), 2029-2049

[17] Portello, A., Danès, P. and Argentieri, S. (2012). Active binaural localization of intermittent moving sources in the presence of false measurements. Proceedings of IEEE/RSJ IROS, Algarve, Portugal: 3294 - 3299

[18] Portello, A. (2013). Localisation binaurale active de sources sonores en robotique humanoide (Localización binaural activa de fuentes sonoras en robótica). Tesis doctoral: Université de Toulouse III - Paul Sabatier, 10 Décembre 2013, 208p., Président: Y.DEVILLE, Rapporteurs: E.VINCENT, L.GIRIN, Examinateurs: E.MOREAU, R.HORAUD, P.SOUERES, Directeurs de thèse: P.DANES, S.ARGENTIERI , Ndeg 13745

[19] Portello, A, Bustamante, G., Danès, P., Piat, J. and Manhès, J.,(2014) Active localization of an intermittent sound source from a moving binaural sensor. Forum Acusticum, Krakow, 2014

[20] Sandini, G., Metta, G. and Vernon, D. (2007) The iCub Cognitive Humanoid Robot: An Open-System Research Platform for Enactive Cognition. Springer Berlin Heidelberg, Berlin, Heidelberg

[21] Tashev, I. J., (2009) Sound Capture and Processing: Practical Approaches. Wiley Publishing