UNIVERSITÀ CAMPUS BIO-MEDICO DI ROMA
School of Engineering
PhD Course in Biomedical Engineering
(XXII - 2007/2009)

Novel Methods and Technologies for Studying Social Orienting Behavior in Young Children

A thesis presented by
Giuseppina Schiavone

Coordinator

Prof. Saverio Cristina

Tutors

Prof. Flavio Keller

Prof. Eugenio Guglielmelli

Co-Tutor

Ing. Domenico Campolo

April 2010

Novel Methods and Technologies for Studying Social Orienting Behavior in Young Children

Giuseppina Schiavone

XXII-2007/2009

Novel Methods and Technologies for Studying Social Orienting Behavior in Young Children

A thesis presented by
Giuseppina Schiavone
in partial fulfillment of the requirements for the degree of
Università Campus Bio-Medico di Roma
Doctor of Philosophy in Biomedical Engineering
School of Engineering

Coordinator

Prof. Saverio Cristina

Tutors

Prof. Flavio Keller

Prof. Eugenio Guglielmelli

Co-Tutor

Ing. Domenico Campolo

April 2010

To Children

Acknowledgements

This thesis would not have been possible unless the TACT project, thus I would like to thank all the partners of the project, in particular my professors and supervisors, Flavio Keller and Eugenio Guglielmelli, my co-tutor Domenico Campolo, the professor Aude Billard who hosted me in EPFL in Lausanne for an internship of three months, then the energetic Basilio Noris; Gunilla Stenberg and Clara Shmidt authors of the social protocol used for testing in the day-care center the device I developed.

I am grateful to the personnel of *La Primavera del Campus* (Universitá Campus Biomedico, Rome) day-care center who actively collaborated during the experimental sessions.

I would like to show my gratitude to Amalia Scoppa and her family, and all the personnel of the *Fondazione Peppino Scoppa* (children rehabilitation center, in Angri, SA), where I spent a short period for testing the AVV-Cap on behavioral audiometry practise. The daily contact with very young children with neurological problems gave me strongness in continuing my research. Unfortunately the experimentation on application in behavioral audiometry could not be completed within my PhD, thus I hope it could be the subject of a new project.

I would like to thank my colleagues Fabrizio Taffoni, Domenico Formica and Alessandra Tomassetti, with whom I shared ideas and difficulties. Thanks to Marta Calvano and other students who participated to the experiments in laboratory. Thanks to Fabian Chersi for his interest in my work and his contribution in my research.

Special thanks to my mother and my sisters, and all my friends, Luca, Giorgia, Miranda, Massi, Sara, Benoit, Annick who supported and encourage me during the completion of my project. And I am grateful to Ruben Poncia for his patience and for his creative soul which in many ways inspired me.

Giuseppina Schiavone

Abstract

Failure in orienting towards occurring social stimuli represents one of the earliest and most basic social impairments in autism. Although, there are evidences for early signs of autism before three years of age, the available clinical tests for the diagnosis are not suitable for quantifying behavior alterations. There are no laboratory tests to confirm or disprove the diagnosis.

My PhD work focuses on the design and development of a multimodal head-mounted device for assessing children's orienting behavior towards social stimuli in their first three years of life. The aim of the work is to develop new methods and technologies for supporting early diagnosis of neurodevelopmental disorders with quantitative, objective observations.

TACT (Thought in Action) European project guidelines have been followed for the designing of the device. Technological choices are emphasized with respect to un-obtrusive, ecological and multimodal requirements. Also ad-hoc calibration procedures suitable to unstructured environments have been developed and here presented.

Preliminary tests carried out at a local daycare with 12-36 months old infants prove the in-field usability of the proposed technology.

Considerations on the future development of the device stressed the meaningful contribution that such platform can offer to child-robot interaction research.

Contents

Guppa Schiavone

Guppa Schiavone

List of Figures

Gyuffe Shiavone

Gyuffre Schiavone

List of Tables

Introduction

My interest in children's behavior development started about four years ago, when for the first time Professor Flavio Keller introduced me into the TACT (Thought in ACTion) European Union's Project [3].

It was at the beginning of 2005. The TACT Project was just approved by the Adventure/Nest (New and Emerging Science and Technology) of the 6th Framework Programme. TACT was born from the challenging idea of a multidisciplinary group of neuroscientists, engineers and psychologists, with the ambitious aim of designing and developing non-obtrusive, user-friendly technologies and methods to evaluate basic patterns of goal-directed actions in normally developing babies, both under laboratory and naturalistic conditions. The long term goal is establishing standards against which development of infants at risk for neuro-developmental disorders, particularly Autism Spectrum Disorders (ASD), can be measured, with the aim of detecting early signs of disturbed development [18].

As a young student, I was really fascinated to take part to international meetings and to taste the world of research, where scientists constructively and with enthusiasm debate their proposals and contribute together in the development of new scientific knowledge and technology for improving human understanding. Moreover I was catched by the high potential impact of TACT.

Most recent reviews of epidemiology estimate a prevalence of about six per 1000 for ASD [62]; [19] ASD averages a 4.3:1 male-to-female ratio. The number of children known to have autism has increased dramatically since the 1980s, at least partly due to changes in diagnostic practice; the question of whether actual prevalence has increased is unresolved, and as-yet-unidentified environmental risk factors cannot be ruled out [62].

Autism is usually diagnosed at the age of 3 years, in many cases after a period of seemingly normal neurological and behavioral development. The

Giuseppina Schiavone

diagnosis of autism is purely clinical, there are no laboratory tests to confirm or disprove the diagnosis. Observational judgments are then quantified according to standardized protocols, e.g. [79], [36] that are both imprecise and subjective. The broad disagreement of clinicians on individual diagnoses creates difficulties both for selecting appropriate treatment for individuals and for reporting the results of population-based studies [49], [86].

The design and development of user-friendly technologies suitable for observing very young children (less then 3 years old) in familiar environments, such as home and day-care centers, can give a meaningful contribution to early diagnosis and therapeutic assessment of neuro-developmental disorders and increasing the possibility of screening a large number of children. Early intervention can improve long-term function and help the families.

My work on this subject started with my Master thesis (*'Wearable Audio-Visuo-Vestibular Systems for Neuro-developmental Engineering'*) in which I developed a first prototype of a multimodal head-mounted device, designed for systematic assessment of social orienting behavior in children between 12 and 36 months in familiar, ecological scenarios. Social orienting represents one of the variety of social behaviors that are impaired in autism [28]. First tests conducted in day-care center proved the acceptability of the device among normally developed children.

During my PhD I worked on the improvement of the proposed Audio-Visuo-Vestibular (AVV) system.

The principal challenge of my work has been to design and develop a device able to collect a multitude of information from the child's behavior (such as, his/her ability to localize human voice, to shift attention to faces and objects) during his/her daily life in familiar environments, i.e. the playroom at the children's daycare centers. Thus, allowing the children to behave and interact without additional distractions and also allowing the caregivers to be present as advisors. It will be discussed in the present dissertation, the difficulty of working in a poor structured environment, where despite of laboratory setting, noisy conditions (i.e. lightning, disturbing sounds) are not predictable and cannot be controlled, thus, affecting the measurements and limiting the performance of the device.

Another challenge of the present work is to provide a system which processes data automatically, or at least semiautomatically, in order to reduce the time of screening, despite the traditional tedious and time consuming video coding, widely used in experimental psychology practice.

Future developments of the device intend to design an integrated multimodal interface which could be used not only for monitoring children behavior and supporting early identification of altered behavioral patterns but also for improving child-robot interaction research, which in the last twenty years is oriented more and more to therapeutic treatments for children with autism.

This thesis dissertation is structured as follows:

- In Chapter 1 the rationale of the research work is presented by reporting scientific evidences of social orienting behavior impairments in autism. The innovative technological approach adopted within TACT Project for assessing children behavior development is described. The new principles of design for ecological objective observations with the AVV system are defined, despite of the state of the art of the traditional techniques and methods for monitoring children's sensori-motor systems development.

- In Chapter 2 the hardware configuration adopted for the AVV system is described. Modifications respect to the previous prototype are explained together with the resolution of technical issues. The synchronization procedure designed for the acquisition of multimodal signals is presented. Algorithms used for the local data processing system are described.

- In Chapter 3 the ad-hoc calibration procedures designed for the AVV system according to the new technological approach are described. Laboratory and in-field tests for assessing its performances are also reported.

- In Chapter 4 the experimental protocol specifically designed for investigating child orienting behavior and the first results from the experiments carried on children in *La primavera del campus* day-care center are reported. Qualitative evaluation of the performance of the AVV system in detecting relevant behaviors compared to the standard video coding technique is presented.

- In Conclusion the main outcomes of this research work will be summarized and I will outline future research directions, stressing the multiple application fields of the designed device.

Giuppa Schiavone

Chapter 1

Technologies for Behavior Assessment: A Novel Ecological Approach

In this chapter the rationale of the research work is presented by reporting scientific evidences of social orienting behavior impairments in autism. The innovative technological approach adopted within TACT Project for assessing children behavior development is described. The new principles of design for ecological objective observations with the AVV system are defined, despite of the state of the art of the traditional techniques and methods for monitoring children's sensori-motor systems development.

1.1 Social Orienting behavior Impairments in Autism

In DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association, 1994), autism is defined as a pervasive neurodevelopmental disorder, characterized by deficits in three basic domains: social interaction, language and communication, and pattern of interests.

Autism is an important benchmarks to highlight failures within sensory motor mechanisms. Losses in the perception, processing, integration and interpretation of sensory information automatically create serious functional problems and alterations in the development of higher cognitive and complex social abilities.

Failure in orienting towards occurring social stimuli represents one of the earliest and most basic social impairments in autism ([28] - [78] - [14]). Detailed home videos analysis [83] - [43] - [84], on children later diagnosed with autism, reveals that children with autism are clinically distinct from

Gjuppa Schiavone

their peers before the age of two years, and that there are clearly observable behaviors which are important predictors of autistic disorder in pre-verbal children. Children with autism showed subtle problems in sensory attention and arousal, coordinated and synchronized movements.

Main characteristic manifestations of the difficulty in encoding and processing the sensory components of social stimuli in autism are:

- lack of social orienting to other's faces [78]
- altered processing and recognition of socially relevant information from peoples faces;
- unusual patterns of eye-contact and gaze directionality [14]
- alterations of auditory processing;
- difficulty in shifting attention to speech sounds [57] - [20] - [75]
- significant lack of sharing attention and activity [40].

Attentional deficits in autism may be due to dysfunction in the cerebellum that slows down the ability to carry out rapid attention shifts that are required to interpret complex stimuli [22]. Using functional magnetic resonance imaging (fMRI), enlarged volume of the amygdala dimension have been observed in children with autism [61] - [13]. Amygdala is a brain area associated with numerous functions, including the processing of faces and emotion; brain abnormality related to this area appears to be associated with the ability to share attention with others, fundamental ability thought to predict later social and language function in children with autism.

If social attention impairments could be identified early in life and changes made to the manner in which the child attends to his or her social environment, children with autism might be directed back closer toward the path of typical development [29].

The most important senses involved in orienting behavior are:

- vision: how the eyes and the brain work together to intake and organize visual information
- hearing: how the ears and the brain work together to intake and organize auditory information
- vestibular system: orienting of head position and sensory organs responseability to the environment

Giuppe Schiavone

Our responses to social stimuli is mainly supported by perception of the world through these senses and it is expressed with the coordinated control of our movements (i.e. head and gaze orienting). Moreover, head orientation and eye contact appear to be one of the first and more powerful type of non-verbal communication among humans.

Taking into account previous considerations the proposed AVV system has to be able to detect gaze and facial expressions, to sense child's head kinematics, and to localize sound stimuli, in particular human voices, interacting with the child.

Objective and continuous monitoring on the development of head-eye kinematics control and of the sensori-motor coordination system during social interaction in very young children could improve our understanding of important processes such as: the development of the *self* perception and of the interactional dynamics with peer and adults.

1.2 Neuro-Developmental Engineering

The TACT project opens the path for a new and emerging interdisciplinary research field: the Neuro-Developmental Engineering (NDE).

The NDE is defined at the intersection of developmental neuroscience and bioengineering aiming at providing new methods and tools for:

- understanding neuro-biological mechanisms of human brain development;
- quantitative analysis and modeling of human behavior during neuro-development;
- assessment of neuro-developmental milestones achieved by humans from birth onwards.

Main application fields of NDE are:

- New clinical protocols and standards for early diagnosis, functional evaluation and therapeutic treatments of neuro-developmental disorders;
- New generations of educational, interactive toys which can provide adequate stimuli and guidance for supporting the physiological neuro-development process.

Giuppa Schiavone

This technology is expected to be also useful in the long term for developing new tools, e.g. toys, which can sustain, in ecological scenarios, the regular development of motor and cognitive abilities of the child, based on a rigorous scientific approach.

According to NDE, three main principles for the design of the AVV system are:

1. **non-obtrusivity**

2. **minimally structured and ecological operating environments**

3. **multi-modality**

The first one assures suitability for continuous monitoring without being distressful or obtrusive for children; this sets technical constraints for the design of the device, such as small in size, lightweight, and portability.

The second one points out the field of application of the device, that is in unstructured home-like situations, which differs from laboratories and clinical centers environments.

The third one stresses the demand for a complete and integrated analysis of the child behaviors from multiple point of views (i.e. different sensory features) at the same time. The AVV system has to monitor the child's gaze and head orientation and localize voices close to the child.

Huge amount of data can arise from the collection of multimodal information. Current tools ([58] - [68]) used in behavioral observations are based on videos recording. Although there exist many powerful software for videos coding (e.g. The Observer XT from Noldus or ELAN developed at the Max Planck Institute), this procedure still results very tedious and time consuming for the experimenters. Thus, another important challenge of the proposed technology is to provide processing system able to analyze and extract complex behavioral features in limited time and with automatic or at least semi-automatic modality. Therefore, simple and robust algorithms are required both for low level processing of the signals provided by each sensor and for data integration processing.

1.3 Available Technology for Behavior Assessment

Current technologies for orienting behavior assessment can be classified in three main groups: technology for motion tracking, technology for gaze tracking and methods for assessing sound localization abilities. In order to identify the technological solution which better fits with the principles of design above mentioned the main characteristics of available technologies are briefly summarized hereafter.

1.3.1 Technology for motion tracking

Motion tracking can count on a host of different technological solutions, operating on entirely different physical principles, with different performance characteristics and designed for different purposes [87].

Mechanical sensing: typically used for body motion capture; it uses angle and range measurements with the help of gears and bend sensors; very accurate but bulky, often limiting mobility.

Optical sensing: several principles are available, typical systems are camera- based ones; position of markers in 3D space can be estimated very accurately within working volume (typically a few cube meters, depending on the number of deployed cameras); line-of- sight issues (i.e. the fact that body parts or other objects may occlude the visual scene of a camera, losing thus the sight of one or more markers) is a limiting factor; very expensive; often requires highly structured environments, at least when high accuracy is needed.

Acoustic sensing: typically based on time-of-flight of ultrasound pulses between emitters and receivers; sound speed in air (about 340m/s, resulting in sampling periods in the order of a few tens of milliseconds) is slow but still acceptable for sensing human (in particular infants) movements; line-of-sight issues are not as severe as for the optical technology; requires much less structured environments than optical trackers; suitable to be used in ecological conditions (e.g. kindergartens).

(Geo)Magnetic sensing: a first method is based on electromagnetic coupling between a source and several trackers; main drawbacks are that signal decays as $1/d^3$ (where d is the source-tracker distance) and is affected by the geomagnetic field; these devices are quite expensive and require a certain amount of structuring of the environment. A second method is electronic compassing; estimates heading and solely relies on the geomagnetic field, i.e. it does not require any artificial source and is therefore sourceless; measurements can be altered by ferromagnetic influence of surrounding objects.

Inertial sensing: highly miniaturized accelerometers and gyroscopes are used to sense, respectively, acceleration (comprising the gravitational field) and angular velocity; used as inclinometers, accelerometers can sense the gravity vector, i.e. the vertical direction, in this sense they are also sourceless.

In Fig.1.1, a selection chart for the different available technologies is provided. For each available technology (columns) its suitability with respect

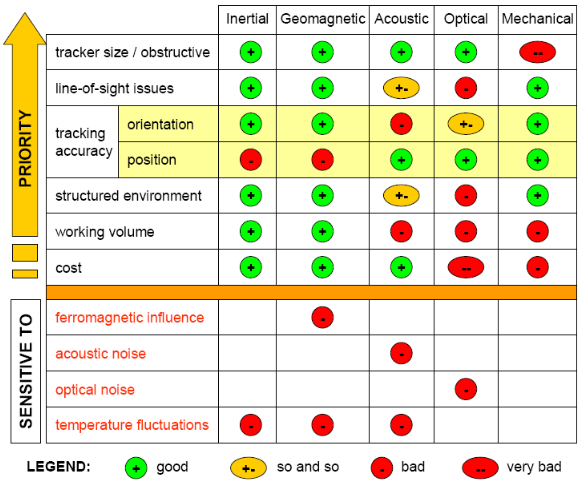

Figure 1.1: Selection chart of different motion tracking technologies [5]

to the performance characteristics of interest (rows) is indicated. Since the main purpose was developing technological device that is wearable by infants the highest priority is given to technologies which are unobtrusive. Thus, solutions involving mechanical trackers are discarded.

The second element considered for selection are the line-of-sight issues, since we are going to deal with infants, it is extremely difficult to perform experiments with technologies that are limited by the line of sight, a peculiarity of the optical technology which is only suitable to experiments with collaborative subjects who are somehow willing to act in front of a camera. Line-of-sight issues are much less severe for the acoustic technology which is thus still appealing for movements analysis in infants.

The third element of the selection criterion is performance with respect to tracking accuracy. Here a distinction is made between tracking positions and tracking orientations. Measurement principles such as the time-of-flight (typically deployed in acoustic measurements) or camera-based tracking are inherently suitable to measure the distance of points (markers) and the origin of the measurement system (e.g. the source of acoustic waves or a camera etc...). Orientations can be inferred indirectly by estimating distances between two or more markers and the source of measurement. The larger the distance between two markers, the better the estimation of orientation. As dimensions shrink, as in the case of infants' head movements, accuracy of indirect orientation measurements also decreases. Other technologies allow a direct measurement of orientations (for example inertial sensors used as inclinometers can sense deviations from the vertical axis while magnetic sensors used as compasses can sense deviations from the horizontal geomagnetic north direction) without requiring the positioning of multiple markers.

As long as orientation is concerned, inertial and magnetic technologies appear to be very appealing since: are highly unobtrusive due the availability of miniaturized off-the-shelf devices; do not suffer from line-of-sight issues; can provide high accuracy in orientation tracking are sourceless: do not require any structuring of the environment; have virtually unlimited working volume; are low-cost.

The bottom half of Fig.1.1 shows, for each technology, the main limiting factors to a correct operation. Besides temperature, which affects any electrical device and that can be compensated in most of the cases, the real limiting factor for the magnetic technology is the presence of ferromagnetic materials. Common ferromagnetic objects such as iron parts of doors, chairs, tables etc... can produce local distortions of the geomagnetic field, causing thus errors in the estimations of orientations. As discussed in [48], some care should be taken, when conducting experiments, to avoid large ferromagnetic objects in the surroundings. This can be easily done in environments such as day-cares where, for safety reasons, all metals are usually avoided and typical materials used with children are wood, rubber and plastic.

1.3.2 Technology for gaze tracking

Devices for measuring eye movements are commonly referred to as *eye trackers*. In general there are two types of techniques for monitoring eye movement [92]:

- *eye-in-head* measurement: the sensing device is fixed on the head and therefore the eye position is measured in craniotopic coordinates
- *point of regard* or *gaze*: the sensors are located in the external environment and the eye position is measured in spatial coordinates.

These two kind of measurements are coincident when the head is kept in a fixed position. When the head is free to move, measurement of the head orientation is also required to derive gaze from craniotopic coordinates. Eye tracking methodologies can be classified in four categories:

1. Magnetic Induction Method (Search Coil)

Gufpe Shave

2. Electro-Oculography (EOG)

3. Photoelectric Methods: Infra-Red (IR) Oculography

4. Video-Oculography (VOG)

Each methodology is characterized by parameters such as range of measurement, sensitivity, linearity, accuracy, discomfort for the subject, interference with the field of view of the subject, tolerance to head movement.

In Fig. 1.2 the relevant parameters of the eye tracking techniques presented are summarized.

	Search coil	**Electro-Oculography**	**Infra-Red Oculography**	**Video Oculography**
Measurement Typology	Absolute spatial coordinate	Craniotopic coordinates	Head-mounted: craniotopic coordinates; External device: spatial coordinates	Head-mounted: craniotopic coordinates; External device: spatial coordinates
Range of measurement	±90 deg for all 3D space	±70 deg in horizontal plane ±30 deg in vertical plane	±30 deg in horizontal plane ±20 deg in vertical plane	±30 deg in horizontal plane ±20 deg in vertical plane
Temporal Resolution	linked to A/D conversion, 500-1000Hz (depends on software and hardware instrumentation)	linked to A/D conversion, 500-1000Hz (depends on software and hardware instrumentation)	linked to A/D conversion, 500-1000Hz(depends on software and hardware instrumentation)	Depends on camera frame rate: from 30Hz to 1000-1250Hz
Spatial resolution	0.01 deg	1-1.5 deg	0.1 deg	0.1 deg
Discomfort	High	Limited	Limited	Limited
Interference with the subject field of view	None	None	Head-mounted device can interfere with the field of view	Head-mounted device can interfere with the field of view
Tolerance to head movement	- Head has to be in the center of the magnetic field - Additional search coil on the forehead	Not affected by head movement	External device: low tolerance to head movements	-Head movements are tolerated when eye is kept in the field of view of the camera; - Additional sensors allow to re-orientate the camera
Other Notes	- limited recording time and risk of corneal abrasion or lead breakage - lens slippage	- Measurement with closed eyes (during sleeping) - Skin electrodes artifacts - Resting corneo-retinal potential variability	- Not suitable when the subject wears glasses or contact lens	

Figure 1.2: Comparison of different gaze tracking technologies

Magnetic Induction Method. Magnetic scleral search coil technique [8] is the standard research technique providing the highest spatial and temporal resolution, and it can also detect torsional components. This technique is based on the fact that a magnetic field induces a voltage in a coil (search coil) which is attached to the eye. The induced voltage has amplitude proportional to the sine of the angle between the coil axis and the magnetic field direction. The magnetic field is provide by coils mounted at the sides of a cubic frame. Robinson [8] was the first to apply this technique, using a coil secured to the eye by suction. Nowadays the search coil is embedded in a scleral contact lens. The lens is subject to slippage if the lens covers only the cornea. Eye movement is measured in absolute spatial coordinates. Head orientation can also be measured with a search coil mounted on the forehead, and orientation and movement of the eye within the head can be calculated from the orientation of the head and of the eye with respect to the magnetic field [81]. Although the scleral search coil is the most precise eye movement measurement method, it is also the most intrusive method. Insertion of the lens requires care and practice and wearing the lens causes discomfort and risk of corneal abrasion or lead breakage. The requirements to stay in the center of the magnetic field precludes the use of search coils during many natural activities. Thus, this technique is mostly used for research purposes, it is not suitable for clinical routine and moreover it is unfeasible for application on children.

Electro-Oculography. The standard clinical method for recording of eye movements is the electro-oculography. It relies on measurement of electrical potential differences between the cornea and the retina, discovered by DuBois-Reymond in 1849. Skin electrodes are positioned around the eye. The measured potential difference is proportional to the sine of the rotation angle of the eye. For small rotation the proportionality is almost linear; it decreases for higher angles of rotation [41]. The eye movement is measured in craniotopic coordinates and head movement during recording does not affect the measurement. The most important advantage of this methodology is the possibility of recording eye movement with closed eyes, which is relevant requirement during some experimental protocol (e.g. during sleeping phases). The main drawback of this technique are related to the nature of the potential recorded and to the artifacts due to the electrodes properties. As concern the potential, the resting corneoretinic potential (usually of the order of 0.4-1 mV) can be affected by lighting conditions of the environment and by the psycho-physical condition of the subject. The artifacts at the level of the skin electrodes relies on the contact resistance electrode-skin, on the oxidation and polarization of the electrodes. Electro-Oculography is not suitable for our application because electrodes positioned on the face can create discomfort in non-cooperative subjects like children and electrodes artifacts can limit the use of this method during natural activities [30].

Infra-Red Oculography. Infra-red (IR) oculography is based on the recording of the light reflected by the eye when it is lighted with IR light beam. Since IR light is not visible, it does not interfere with the subject vision, moreover the IR detectors are not influenced by environmental lighting

Giuseppina Schiavone

conditions. There are three categories of Infra-red (IR) oculography which use respectively: the corneal reflection, the Purkinje images and the track of the limb. Due to the construction of the eye, when a beam of IR light points to it, four reflections are formed on the eye, called Purkinje images [21]: the first on the front surface of the cornea and it is called corneal reflection, the second image on rear surface of the cornea, the third on the front surface of the lens and the fourth on the rear surface of the lens. By detecting the corneal reflection and the pupil center and by using an appropriate calibration procedure, it is possible to measure the gaze on a planar surface on which calibration points are positioned. Two points of reference on the eye are needed to separate eye movements from head movements. The positional difference between the pupil center and corneal reflection changes with pure eye rotation, but remains relatively constant with minor head movements. The corneal reflection moves in the opposite direction of the eye respect to the pupil center. In other cases both the first and the fourth Purkinje images (Dual-Purkinje images eye trackers are detected. Both reflections move together through exactly the same distance upon eye translation but they move through different distances upon eye rotation. The third method based on photoelectric principle relies on the track of the limb (scleral-iris edge) of the eye by measuring the amount of scattered light. Most photoelectric systems must be mounted close to the eyes, so they may restrict the field of view of the wearer, moreover fast movements of the head can cause slippage of the device on the head leading to mis-alignment of the eye respect to the IR emitter and detector. There exist also external device and a support for keeping the head fixed is needed (i.e. Tobii eye tracker). The major drawbacks of this methodology are the limited linear range, the complicated and time-consuming installation and calibration procedures and poor mechanical stability of the transducer with respect to the eye.

Another consideration, very often underestimated, is related to the potential danger of IR light sources for the retina:

- near IR light, like UV light, causes burns to the retina at high powers.
- the blink reflex does not work with IR: IR light is not registered by the eye, and thus the pupil will not close in order to protect the retina from bright, and you do not realize you are getting burnt
- IR lasers are generally considered to be eye safe for powers below 1mW: legal restrictions and medical advices only concerned with infrared emissions of heat lamps or in the welding process and IEC 825-1 (CENELEC EN60825-1)

Giuseppina Schiavone

- IR Lasers above 5mW are considered definitely dangerous

IR lasers can be considered as IR leds with a much tighter beam spread, thus, although the peak power emission of most of the IR leds used for eye tracking systems is below the dangerous threshold, looking at the led at close range for an extended period of time could cause permanent damage to the retina.

Even if this safety consideration could appear to be very drastic, it is important to be aware of it, especially when observational studies are conducted on very young children in which visual system is not yet completely developed.

Video-Oculography (VOG). Video systems for measuring ocular movements are based on the analysis of images recorded by cameras. This technique quickly improved in terms of performances and reliability thanks to the technological development of digital cameras and computer powerful. VOG provides directly a digital output. Several algorithms are available for the pupil detection in an image frame and pupil centroid coordinates extraction, nevertheless environmental lighting conditions can affect the automatic detection [35] - [59]. Thus, IR light is used together with video recording, so that the pupil appears brighter. This technique is called Pupil Center/-Corneal Reflection (PC/CR) because the IR light produces also the Purkinje images, mentioned before. As in IR Oculography, also VOG can be realized both as wearable device [30] and provides measurements in craniotopic coordinates or external device and provides measurements in spatial coordinates. Head-mounted system (i.e. EyeLink) can be worn without too much discomfort. High resolution and high frame rate CCD and CMOS cameras are used. Reduced dimensions and weight of the actual cameras allow to position them in such a way that they interfere as less as possible with the field of view of the subject [65]. The drawback is that these systems have a low acquisition rate, in general 50-60 Hz, not suitable for recording fast eye movement such as saccadic movements, but suffices for smooth pursuit eye movements. External camera are generally positioned under the screen of a computer, used for calibration and for specific visual stimuli. Head movements are tolerated if the eye is kept in the field of view of the camera. There are devices which include systems of pursuit of the subject and the camera orients automatically so that the eye of the subject is always in its field of view.Existing head-mounted trackers (e.g. [31] - [10] - [91]) are not suitable for children, being too heavy and bulky. A new head-mounted camera, the Wearcam [66], recently developed by the LASA of Ecole Politechnique de Lausanne (within the TACT project too), is specifically designed for children aged between 6 months and 18 months, to be used in a free - play environment. It films

Figure 1.3: Children wearing the WearCam *III* and *III* Ti prototypes [2].

the frontal field of view of the child and a small mirror protruding from the bottom part of the camera reflects the eye portion of the wearer's face (Fig. 1.3).

Wearable VOG technologies appear to be the most suitable for continuous children behavior assessment since: reduced dimensions and weight of the actual cameras allow to position them in such a way that they interfere as less as possible with the field of view of the child, thus can be worn without too much discomfort; high resolution and high frame rate CCD and CMOS cameras are now available and low cost.

1.3.3 Methods for assessing sound localization abilities

Lack of response to auditory stimuli and hearing loses are usually monitored with ABR (Auditory Brain Responses) audiometry [27] and OAE (OtoAcoustic Emissions) tests [33].

Auditory brainstem response, also known as brainstem evoked response (BSER) is an electrical signal evoked from the brainstem of a human or other mammal by the presentation of a sound such as a click. ABR audiometry is a screening test to monitor for hearing loss or deafness, especially notable for its use with newborn infants. ABR is measured by placing elec-

Giuppa Schiavone

trodes (non-invasive and painless) on the head to record the brain's response to sound. Because this test does not rely on behavior, and because of artifacts due to the electrodes positioning and the brain activity it is preferable to administer the test when the child is in a sleepy state. ABR is a method employed to assess the functions of the ears, cranial nerves, and various brain functions of the lower part of the auditory system, prior to the child developing to the point of describing a possible hearing problem.

Otoacoustic emission is a sound which is generated from within the inner ear. Studies have shown that OAEs disappear after the inner ear has been damaged, so OAEs are often used in the laboratory and the clinic as a measure of inner ear health. OAEs are measured by placing a very sensitive microphone in the ear canal to measure the ear's response to sound. They are clinically important because they are the basis of a simple, non-invasive, test for hearing defects in newborn babies and in children who are too young to cooperate in conventional hearing tests.

Both the methodologies are implemented in specialized clinical center and hospitals. Moreover, they are not specifically designed for testing sound localization abilities. Other tests include **behavioral audiometry**, which is a screening test used in infants to observe their behavior in response to certain sounds. It is a purely observational practise which can be affected by the experience of the clinician and his/her ability to detect small, unpredictable changes in the child's behavior, such as fast eye shifting and sudden reflex movements.

Despite the presented state of the art, quantitative methods for assessing and for continuous monitoring of the child's ability to localize sound sources, in particular human voices, in ecological environment do not exist.

Giuppe Schiavone

Chapter 2

The Audio Visuo Vestibular System

In this chapter the hardware configuration adopted for the AVV system is described. Modifications respect to the previous prototype are explained together with the resolution of technical issues. The synchronization procedure designed for the acquisition of multimodal signals is presented. Algorithms used for the local data processing system are described.

Figure 2.1: Child wearing the AVV Cap

2.1 Hardware Design

Among the technologies presented in Chapter 1, *wearable technologies* appear to be very appealing for our purpose.

Being close to the body, wearable devices allow to directly extract relevant information from a child-centered perspective, filtering environmental noise. Wearable systems can consolidate the functionality of multiple devices into a single, integrated system and do not require to modify the environment setting, unlike complex motion tracking systems (i.e. stereo-photogrammetric systems, which require to equip the observational room with several infrared cameras) and sophisticated sound localization systems, requiring multiple points of observation (an array of microphones to be located in specific points of the observational room).

The proposed system has to work like an artificial audio-visuo-vestibular system: mounted on the child's head, it has to be able to localize voices close to him/her, to sense child's head orientation, angular velocity and acceleration and estimate gaze direction by detecting eye orientation. *Ecological and non-obtrusive principles* lead to the use of a colored cap, a widely used garment among children, as wearable support for the AVV system.

Also the *multimodal principle* is achieved. Three different sensors are mounted on the cap:

- a magneto-inertial sensor, positioned on the top of the cap with velcro bands;
- a pair of omni-directional microphones, positioned at two opposite sides of the cap, in correspondence with the ears of the child (binaural configuration);
- a mini-webcam mounted on the visor of the cap, sustained by an ad-hoc light rubber support, with the objective pointing to the face of the child.

The sensors can be easily moved from one cap to another which better fits the cranial circumference of the child (estimated from 35 cm to 49 cm for children from 6 to 24 months). The cap is kept fixed on the head of the child using adjustable elastic bands.

All the sensors on the AVV-Cap send data to a PC via USB, which means that the AVV-Cap is still a wired device. The cables are connected together in correspondence at the back of the cap, before being plugged into the PC. The drawback of the wires is that the AVV-Cap, at this moment, cannot be used in free-play situations, in which the child can move freely, but he/she has to sit on a chair during experimental sessions. Although, first experiments showed that the presence of cables do not affect the child's acceptability of the device, the future prototype will be equipped with wireless communication modules, allowing the child to interact without constraints.

2.1.1 AVV-Cap Components

The three components of the AVV-Cap are described hereafter.

Figure 2.2: AVV-Cap Commercial Components: Magneto-inertial sensor from Xsens Technologies B.V. for orientation tracking; Lavalier microphones for sound localization; web-cam for eye tracking

Magneto-inertial technology [87] for motion capture has been chosen for several reasons: it is sourceless, it relies solely upon gravitational and geomagnetic fields that are ubiquitously present on Earth and does not require additional field sources; it is available in compact packages, limited in dimension and weight; moreover such systems are easy to calibrate and are low cost in respect to other motion tracking systems. The main drawback of this technology is the sensitivity to external magnetic fields (i.e. mobile phones, power stations, etc.), however, it is plausible to assume that in the environments in which the AVV-Cap is designed to be applied (daycare centers) electro-magnetic interferences are limited. The **head tracker** (MTx, Xsens Technologies B.V.) transduces head kinematics in 3D at a frequency of 100 Hz, with a dynamic range of all angles in 3D and angular resolution of $<1°$.

Localization of the speaking voice is achieved by processing the signals from the **two microphones** (MKE 2-ew Gold, Lavalier). The microphones are the input for an external audio card (Tascam US-144) connected to the computer via USB. The audio signal acquisition frequency is set to $44100Hz$, so as to cover an interval of frequencies from 0 to the Nyquist frequency $22050Hz$, in which is contained the frequency range of human auditory perception $(20 - 20000Hz)$. Quantization is set to $16 - bit$ per sample, which provides a dynamic interval of $96dB$, close to human ear dynamic range.

Unexpensiveness, light weight, dimensions, and quality of modern day web-cams allowing them to be used for eye-tracking has driven to choose them for our purpose. The web-cam, that we address as **EyeCam**, is composed of a $1/4in$ CMOS sensor, with a resolution of $640x480$ and a frame rate of 30 frames per seconds. The EyeCam sensor points to the face of the child for detecting eye movements.

Two issues arose relatively to the use of the proposed eye-tracker: *i)* the EyeCam field of view, *ii)* the EyeCam positioning. Hereafter, I will describe the solutions I proposed.

EyeCam field of view

In order to focus the face of the child (target of hight, h, and length, l) at a very small distance (distance target-objective, d) the objective of the camera has been substituted with another with a focal length, f, defined by the following relation:

$$f_{mm} = \frac{d_m x_{mm}}{l_m} \tag{2.1}$$

where x is the CMOS length. Focal length and CMOS length are expressed in mm, target length and distance target-objective are expressed in m.

Considering a distance target-objective of about 0.1 m and a target length of about $0.14m$, for a $1/4in$ CMOS sensor the estimated focal length is equal to $2.5mm$. As a result, a field of view corresponding a face dimension of 12x15 cm is covered (the diagonal Field Of View is 84° for an average of 57° and 71° vertically and horizontally respectively). In Fig.2.3 snapshots of children's faces obtained using the selected mini-objective are shown.

EyeCam positioning

The EyeCam positioning has to interfere as less as possible with the field of view of the child. The solution adopted in the first prototype of the AVV-Cap was to position the EyeCam on the tip of the visor with a metallic hook fixed with screws (Fig.2.4). Nevertheless this configuration loaded too much the tip of the cap which often slid forward on the nose. Thus, I designed with a CAD Design Software (SolidWorks) small bars to be positioned on

Giuppa Schiavone

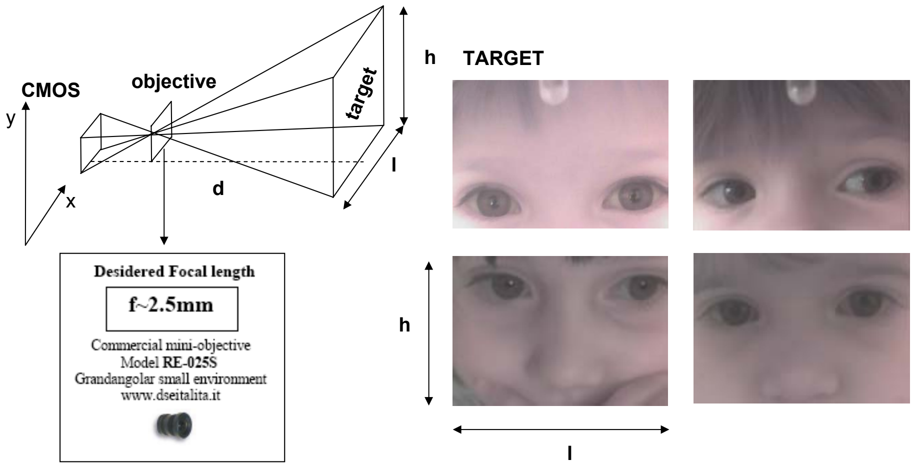

Figure 2.3: Focal length estimation for the mini-objective of the EyeCam, on the left, and children faces detected by the EyeCam, on the left

the visor for distributing the load on a wider surface. The negative of the bars have been printed with a 3D printer, then the sticks are realized by straining polyurethane resin (Prochima SINTAFOAM 1.1), not toxic, rapid cooling resin. Resulting rubber bars are shown in Fig.2.5. Three kind of bars have been designed: in Fig.2.5 sticks A connect the EyeCam to the visor through screws, sticks B pass through holes in the visor (see Fig.2.4), sticks C are optional and can be used to increase the distance target-objective, thus widening the field of view of the EyeCam (see Fig.2.4 longer and shorter configuration). The silicon rubber keeps some flexibility, necessary when it is needed to move the support from one cap to another which better fits with the child's head dimensions.

Giuppa Schiavone

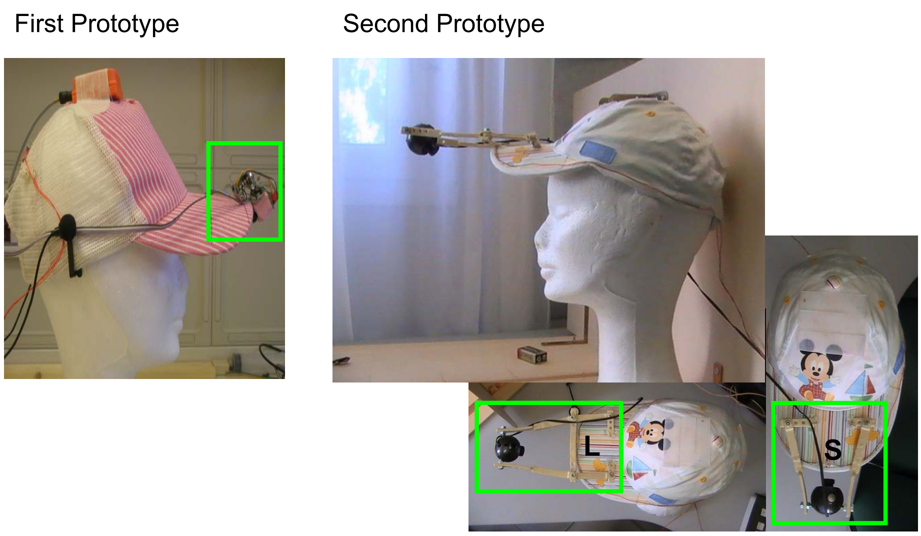

Figure 2.4: EyeCam positioning in the green square: on the left, the first prototype, on the right the second prototype where rubber sticks are used in longer configuration, L, or shorter configuration, S

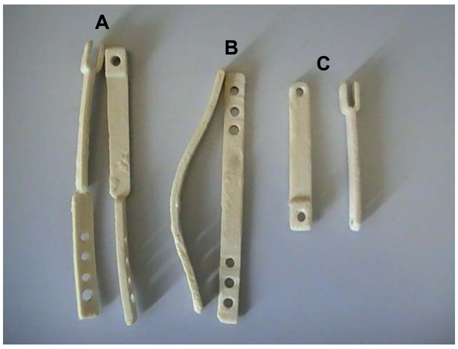

Figure 2.5: Rubber bars for supporting the EyeCam on the visor of the Cap: A connect the EyeCam to the visor; B keep fixed the support on the visor; C optional bars for increasing the distance target-objective

This solution hardly interferes with the child's field of view and enables manual setting of the EyeCam orientation.

2.2 Synchronization system

When different sensors work together and several signals need to be acquired at the same time, synchronization is a step that cannot be skipped, moreover when signals describe complex behavior, including sensori-motor coordination.

Synchronization can be considered a bottleneck of systems in which many events need to be timing coordinated. This is the case of data processing devices which employ modules that are in different clock domains, with each domain synchronized to a clock that is asynchronous with the clocks of other clock domains.

In Fig.2.6 is illustrated the synchronization procedure proposed.

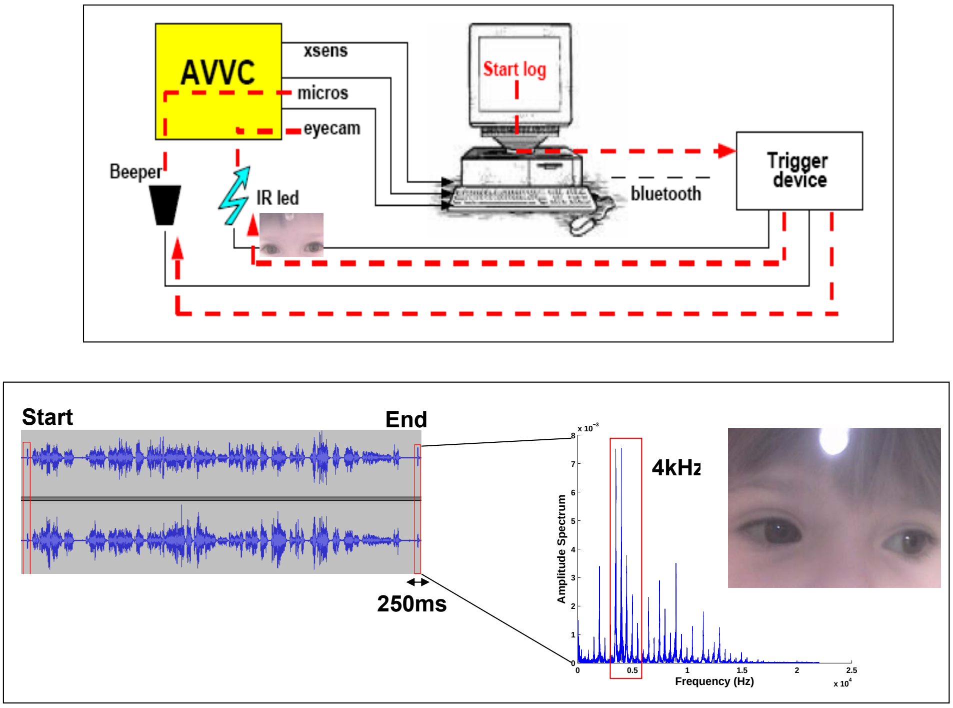

Figure 2.6: Synchronization: upper square, schematic of the synchronization system; bottom square, synchronization signals: beep in the audio signals and flash in the video image

An external electronic device, referred as **trigger device**, has been realized with the function of triggering the stereo signal acquired by the microphones and the video track provided by the EyeCam. The trigger device includes a bluetooth receiver and a PIC microcontroller. The program running on the PIC allows to read characters sent via bluetooth from the computer and produce triggering signals for the synchronization, i.e. at the beginning and at the end of the acquisition it lights an IR led, positioned under the visor of the cap pointing to the EyeCam objective, and it generates a sound, a beep, lasting 250ms and at frequency of 4kHz.

The function of 'master' in the synchronization procedure is played by the acquisition of the magneto-inertial sensor data, managed by a program running in Matlab Environment. As magneto-inertial sensor data starts or ends to be logged, a trigger signal (a character) is sent via bluetooth from the computer to the trigger device, which in its turn synchronizes the EyeCam and the microphone signals with the Infrared flashing and with the beep.

The flash and the beep are then identified in the acquired recordings.

Although, the described procedure is coherent and easy to implement, wired connections between the trigger device and the AVV-Cap can create instability and computer overload can bring a delay in the acquisition of the different channels. Accordingly, there is a need for improved synchronization and other techniques for transferring data between clock domains could be explored in future prototypes (i.e. the use of external clock signal as the reference clock).

2.3 Multimodal Data Processing

AVV-Cap processing is made off-line. In the present work, on-line data processing was not required for children behavioral assessment, although, future development of the device are oriented to a real-time processing system, so that the acquired information can be directly available.

The design of the processing system has to respond to the requirement of *automatic or at least semi-automatic data analysis*, therefore, simple and robust algorithms are required both for low level processing of the signals provided by each sensor and for data integration processing.

The proposed system for data processing is organized as in Fig.2.7. Multimodal signals are provided by the three sensory channels (stereo audio tracks by the microphones located in binaural configuration, video recording by the EyeCam and motion tracking information by the magneto-inertial sensor). After the identification of the synchronization points (beeps in the audio tracks, flashes in the video images), signals are separately processed to extract local features. The mapping process is based on the estimation of significant parameters through calibration curves, which allow to transform local features in features relevant for the behavior assessment.

In this section, the definition and the processing of local features will be described. Mapping processing and calibration procedures are reported in details in Chapter 3.

All the algorithms written for data processing have been implemented in Matlab Development Environment.

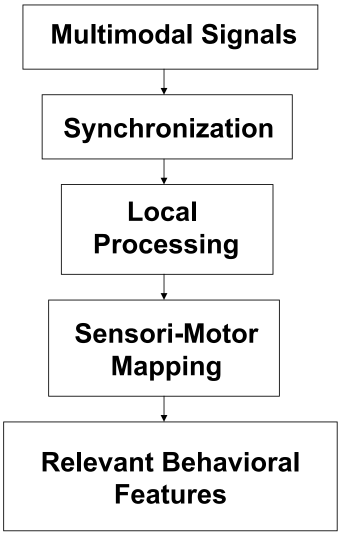

Figure 2.7: Data Processing Diagram

2.3.1 Local Processing

Three are the main local features to be extracted from raw signals:

1. the Time Difference of Arrival (TDOA) of the sound waves at the child's ears

2. the child's head azimuth

3. the coordinates in pixel of the eye position in the video images

The first feature is used for identifying in which direction human voices interacting with the child are located. The second allows to detect child's head orientation in the frontal horizontal plane. The third is used for estimating eye angular position.

TDOA

The TDOA, also addressed as Interaural Time Difference (ITD), is the delay between the time when a sound from a single source reaches one ear and when it reaches the other ear. This delay is due to the different path that the sound wave covers from the point where it is located to the two ears.

In order to determine the delay in the signal captured by the two microphones, a coherence measure has to be defined. Among several techniques (

[17] - [24] - [9]), the most common coherence measure is the simple crosscorrelation [50] between the signals perceived by the two microphones.

Considering windowed frames of N samples with 50% overlap, the cross-correlation function for the a single frame is expressed by:

$$R_{ij}(\tau) = \sum_{n=0}^{N-1} x_i[n]x_j[n-\tau] \tag{2.2}$$

where $x_i[n]$ and $x_j[n]$ are the signals received by microphone i and microphone j, τ is the correlation lag in samples (equally distributed from -1 and 1 ms, which means in samples $-44 < n < 44$).

Computational load required for implementing 2.2 has a complexity of $O(N^2)$. The inverse Fourier transform of the cross-spectrum is computed to reduce such complexity to $O(Nlog_2N)$:

$$R_{ij}(\tau) \approx \sum_{n=0}^{N-1} X_i[k]X_j^*[k]e^{j2k\pi\tau/N} \tag{2.3}$$

where $X_i[k]$ is the discrete Fourier transform of $x_i[n]$ and $X_i[k]X_j^*[k]$ is the cross-power spectrum of $x_i[n]$ and $x_j[n]$.

The drawback is that 2.3 is strictly dependent on the statistical properties of the source signal. Since most signals, including voice, are generally low-pass, the correlation between adjacent samples is high and generates crosscorrelation peaks that can be very wide. The problem of wide cross-correlation peak can be solved by whitening the spectrum of the signal prior to compute the cross-correlation [6]. The resulting whitened cross-correlation, also commonly referred to as Phase Transform (PHAT) technique, is:

$$R_{ij}^{PHAT}(\tau) \approx \sum_{n=0}^{N-1} \frac{X_j^*[k]}{|X_i[k]|\,|X_j[k]|}e^{j2k\pi\tau/N} \tag{2.4}$$

This approach allows to only take the phase of $X_i[k]$ into account, narrowing the wide maxima caused by the correlation between the received signals; it does not require any knowledge about the spectrum of the microphone dependent noises and shows good performance in low-noise, reverberative environments [32] - [95].

Although there exists more robust and accurate algorithms for sound localization, the PHAT algorithm has been chosen because of its simplicity and lower complexity. The TDOA, ΔT_{12} , for each time frame, between the two microphones, can be found by locating the peak in the cross-correlation function:

$$\Delta T_{12} = arg\max_{\tau}(R_{12}^{PHAT}(\tau)) \tag{2.5}$$

Giuseppina Schiavone

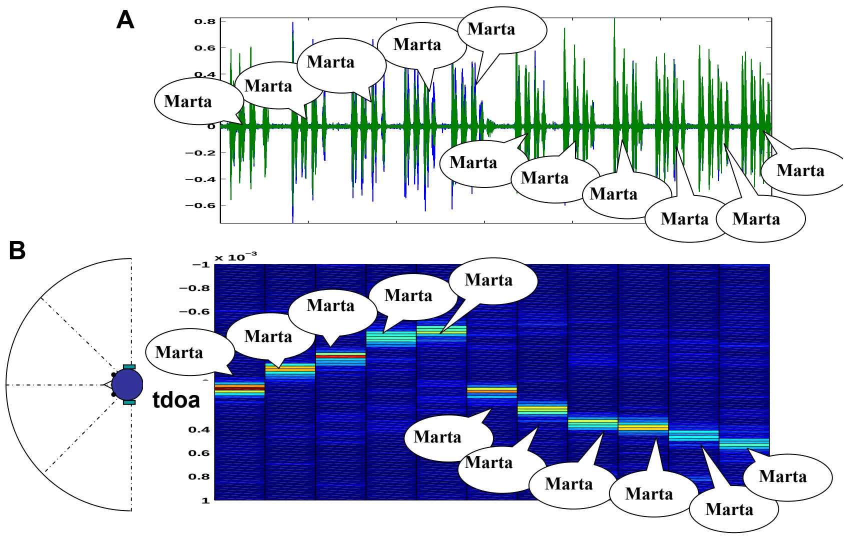

Figure 2.8: A) stereo signal; B) top view cross-correlation

In an example of the cross-correlation function is shown: a sound source moves from one side to the other side of a dummy head where the two microphones are mounted; the TDOAs relative to the maximum values of the cross-correlation function change as a function of the sound source direction.

The resolution in the TDOA estimation is related to the sampling frequency of the signal. As the signal is acquired at 44100 Hz, the uncertainty of the measurement is approximatively equal to $10\mu s$, half of the sampling period [55].

As the TDOA has been obtained, the sound source location in the horizontal plane can be estimated theoretically by the model described in [17]. This model works well when the wavelength of the sound is higher than the dimensions of the head. For higher frequencies, the wavelength of the sound wave is smaller than the dimension of the head and other factors, such as the shadowing of the head, need to be considered.

Moreover the child is free to orient his/her head in the space and the sound sources nearby the child (i.e. caregivers interacting with him) are not fixed. Thus an ad-hoc calibration procedure is needed to correlate an estimated TDOA to a specific angular direction of the sound source.

Head azimuth

The head rotation angle in the horizontal plane (head azimuth, ψ_h) is extracted by performing some computations on the rotation matrix provided by the magneto-inertial sensor.

The sensor calculates the orientation between the sensor coordinate system, **H**, in agreement with the head, and a fixed reference coordinate system, **O**. The fixed reference coordinate system is defined as a right handed Cartesian coordinate system with:

- X positive when pointing in the direction of the nose of the wearer.
- Y according to right handed co-ordinates.
- Z positive when pointing up.

The output provided by the sensor is in form of rotation matrix, R_{OH}, interpreted as the unit-vector components of the sensor coordinate system **H** expressed in **O**. The columns of the matrix R_{OH} are the unit vectors of **H**:

$$R_{OH} = [X_H Y_H Z_H] = \begin{bmatrix} x_H x_O & y_H x_O & z_H x_O \\ x_H y_O & y_H y_O & z_H y_O \\ x_H z_O & y_H z_O & z_H z_O \end{bmatrix} \tag{2.6}$$

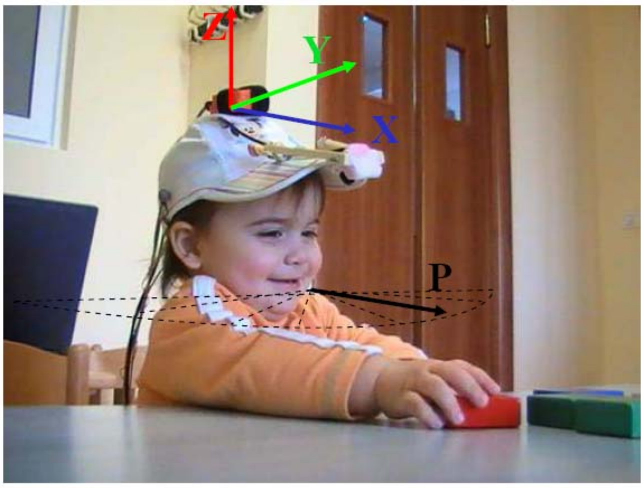

Figure 2.9: Sensor coordinate system and vector **P**, projection of the X_H vector on the xy-plane of the fixed reference co-ordinate system

Giuseppina Schiavone

The azimuth of the head lies on the xy-plane of the fixed reference co-ordinate system and it is defined as the angle between the unit vector X_O of the fixed reference coordinate system and the vector **P**, projection of the X_H vector on the xy-plane of the fixed reference co-ordinate system (Fig. 2.9). The vector **P** is a column vector defined as:

$$R_{OH} = \begin{bmatrix} x_H x_O \\ x_H y_O \\ 0 \end{bmatrix} = \begin{bmatrix} cos(\theta)cos(\psi) \\ cos(\theta)sin(\psi) \\ 0 \end{bmatrix} \tag{2.7}$$

The first two elements of the vector **P** are the first components of the unit vector X_H of the rotation matrix, where θ is the pitch, describing rotations around the axis Y_O, and ψ is the yaw, describing rotations around the axis Z_O.

The azimuth of the head corresponds to the yaw and it is obtained by the trigonometric function $arctan2(y, x)$, where x and y are real arguments and not both equal to zero.

$$\psi = arctan2(y, x) = 2arctan\left(\frac{y}{\sqrt{x^2 + y^2} + x}\right) \tag{2.8}$$

By substituting the arguments x and y with the first two components of the vector **P** , the azimuth is estimated in the range $(-\pi, \pi]$:

$$\psi = 2arctan\left(\frac{sin(\psi)}{1 + cos(\psi)}\right) \tag{2.9}$$

Other kinematics parameters can be estimated, such as the angle of pitch θ, rotations around the axis Y_O describing flexion and extension head movements (tilt forward and backward), and roll ϕ, rotations around the axis X_O describing lateral head oscillations. Considering the RPY rule, according to which the rotation matrix performs the roll first, then the pitch, and finally the yaw, ϕ and θ are defined as:

$$\theta = 2arctan\left(\frac{-R_{31}}{\sqrt{R_{32}^2 + R_{33}^2}}\right) \tag{2.10}$$

$$\phi = 2arctan\left(\frac{R_{32}}{R_{33}}\right) \tag{2.11}$$

where R_{ij} refers to the component of *ith* column and *jth* row.

Giuseppina Schiavone

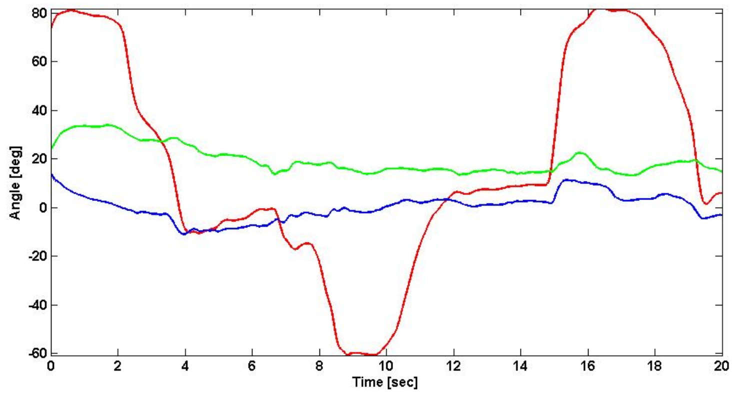

Figure 2.10: Head rotation angle: red, azimuth (yaw); green, pitch; blue, roll

The angular velocity vector, $\omega = [\ \omega_x \quad \omega_y \quad \omega_z\ ]^T$, can also be defined through the rotation matrix, R, and the antisymmetric matrix Ω:

$$\Omega = \frac{dR}{dt}R^T = -R\frac{dR^T}{dt} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{2.12}$$

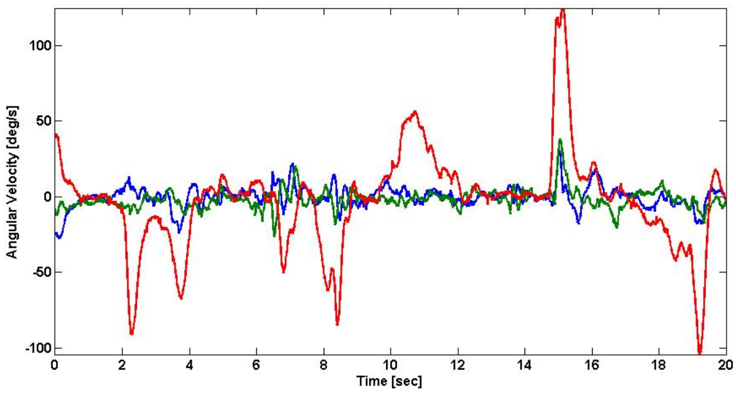

Figure 2.11: Angular velocity vector components: blue, ω_x; green, ω_y; red, ω_z

Fast Fourier Transform can also be applied to detect principal frequency components of the movement.

Eye coordinates in pixel units

Eye movements are detected by identifying the pupils position in frames recorded by the EyeCam. The eye tracker measures eye-in-head movements in craniotopic coordinates, that means that measurements are not affected by head movement because the device is fixed on the head.

Although several algorithms exist for tracking the pupil in an image (such as, histogram based methods, Hough transform methods [42] and template matching methods [93] - [38] - [94], neural networks [12]), an ad-hoc algorithm has been written for the developed eye-tracker.

The main sequence of the algorithm is shown in the flowchart in Fig.2.12.

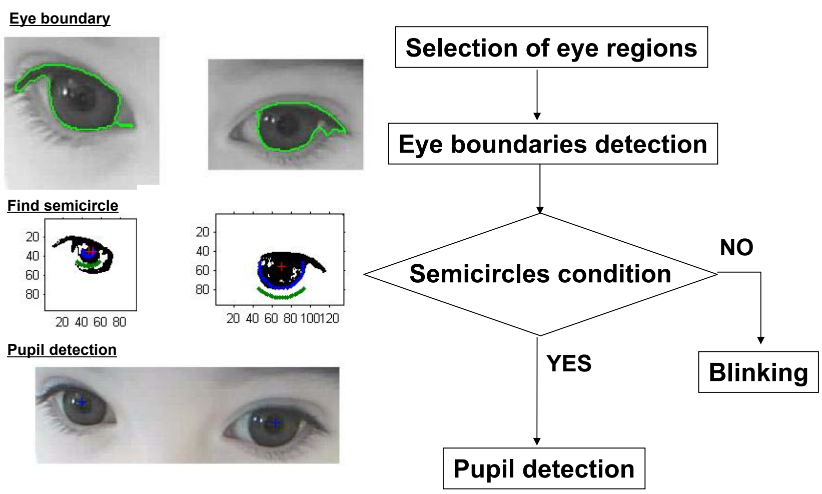

Figure 2.12: Eye tracking algorithm

Regions of the eye are manually selected. The algorithm finds the boundaries of the eye and verifies the semicircles condition on the black and white image.

The semicircles condition is here defined. Two concentric semicircles of radius, R and $R + a$ (with R variable and a constant) move on the image. Pixel by pixel the image is scanned until the algorithm finds that all the pixels along the inner semicircle are black and all the pixels along the outer semicircle are white. When the semicircles condition is not verified the eye is blinking, thus a blinking frame is detected.

When this condition is verified the center of the pupil is found for a determined R. The scanning is repeated incrementing the radius, until the best fit is detected.

In Fig.2.13 the result of the scanning of a set of image frames is shown: horizontal shifting of the left and right eye is estimated. Measurement in pixel units are normalized respect to the width of the eye region.

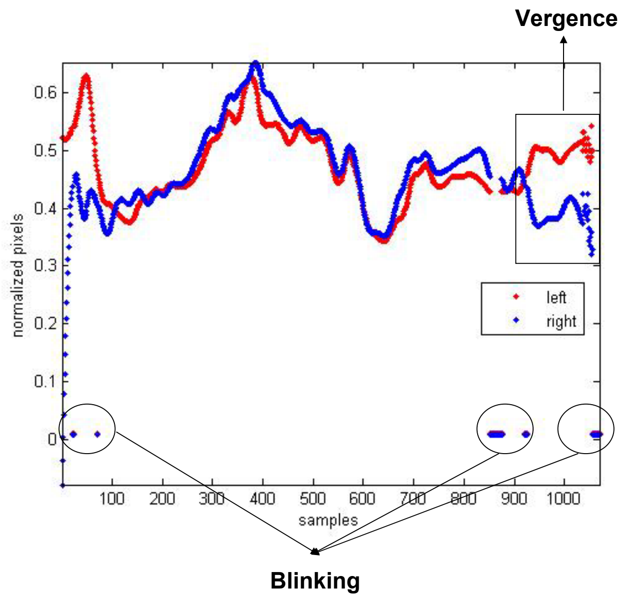

Figure 2.13: Eye horizontal component expresses in normalized pixel: blinking and vergence movements are also detected

Further information on this algorithm and its performance are still under evaluation.

Pupil coordinates expressed in pixels units are transformed in eye angular positions through the calibration procedure described in Chapter 3.

Chapter 3

Ad-Hoc Calibration Procedures for Sensori-Motor Mapping

In this chapter the ad-hoc calibration procedures designed for the AVV system according to the new technological approach are described. Laboratory and in-field for assessing its performances are also reported.

The artificial audio-visuo-vestibular system is provided with two sensori-motor maps: one represents a model which relates estimated TDOAs to specific angular directions of an active sound source, the other defines the function which allows to associate to the horizontal eye coordinate expressed in pixel and angular position for the eye. Sensori-motor maps are obtained through suitable calibration procedures. Since the mapping depends on several factors, such as the positioning of the cap on the child's head, the circumference of the child's head, the distance between the two microphones, the EyeCam orientation respect to the child's face, calibration procedures have to be performed before every experimental session.

Two procedures are required: the *Vestibulo-Auditory Calibration* and the *Vestibulo-Ocular Calibration*. Both procedures integrate sensory information (visual and auditory) to motor information (vestibular).

3.1 Vestibulo-Auditory Calibration

The standard experimental set up for mapping of the TDOAs with angular positions of sound sources, usually, consists of a fixed set of observation points (array of microphones) and fixed set of sound sources, located in known orientations with respect to the observation points.

Giuppa Schiavone

In an ecological environment, which is poorly structured, the typical experimental set up is difficult to reproduce, especially if the head of the child, where the microphones are mounted, is free to move. Thus, an ad-hoc calibration procedure is required for those environments and those experimental conditions.

The proposed Vestibulo-Auditory Calibration is very raw but very effective. It exploits the free movements of the child's head to determine a correlation between the TDOAs and the sound sources directions.

In-the-lab calibration evaluation

In-the-lab calibration has been conducted for comparing the standard procedure with the proposed ecological one. In Fig.3.1, the two experimental settings are schematically shown.

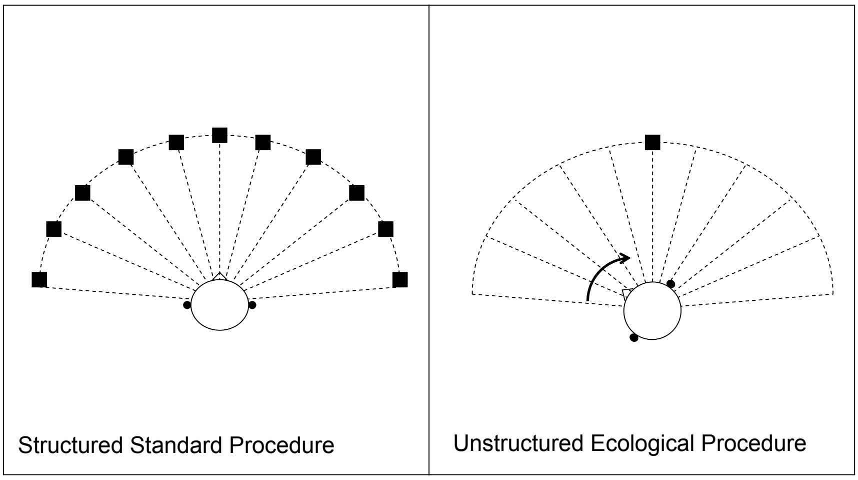

Figure 3.1: Comparison between standard procedure, on the left, and proposed procedure, on the right, for calibrating the AVV-Cap sound localization system: a dummy head wearing two microphones in binaural configuration is in the center of a semicircle; the black squares are the speakers positioned around the semicircle

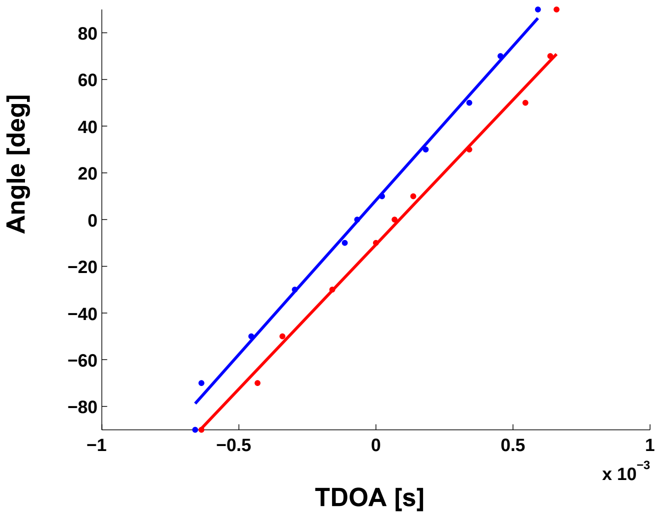

Figure 3.2: TDOA/Angle map obtained in laboratory: blue rots refer to calibration obtained with the standard procedure; red dots refer to calibration obtained with the proposed procedure; blue and red solid lines are the fitting curves

Two microphones are mounted on a dummy head in binaural configuration.

In the standard procedure, addressed as P1, the sound source (speaker) is located in a set of eleven locations ($-90°$, $-70°$, $-50°$, $-30°$, $-10°$, $0°$, $10°$, $30°$, $50°$, $70°$, $90°$) at a distance of $1.2m$ respect to the head, and the head is kept in the primary position, at $0°$.

In the proposed procedure, addressed as P2, the speaker stays in the primary position while the dummy head is oriented in the eleven locations. Both the procedures have been repeated ten times.

While the procedure P1 requires to modify the environment by using several locations for the sound source, the procedure P2 is more suitable for ecological conditions since it exploits the dummy head orientation for TDOA/angle mapping.

Experimental trials proved that TDOAs and angular positions, Ψ, are well correlated with a linear relation:

$$\Psi = G \cdot TDOA + O \tag{3.1}$$

where G is the slope or gain of the curve, expressed in deg/s, and O is the offset, expressed in deg.

All the trials show R-square coefficient values >0.93, confirming that angular position and TDOA are highly correlated with a linear fitting. Fig.3.2 shows a representative TDOA/angle map obtained in a single trial for the two calibration procedures, where blue dots and blue solid line refer to P1, while red dots and red solid line refer to P2.

Difference in offset values are due to misalignment between head orientation and the set of angular positions, and are not considered for the comparison of the two methods. An unpaired two-tailed T-test on the gains of the two procedures shows that the procedures are not statistically discrepant ($p > 0.05$). Thus, confirming that in-field calibration procedure can be performed exploiting child's head movements and without modifying the child's environment.

In-field calibration evaluation [77]

After a period of familiarization with the device, the vestibulo-auditory calibration procedure has been tested on 11 normally developed children between 12 and 24 months of age, at the day-care center *La Primavera del Campus* in Rome.

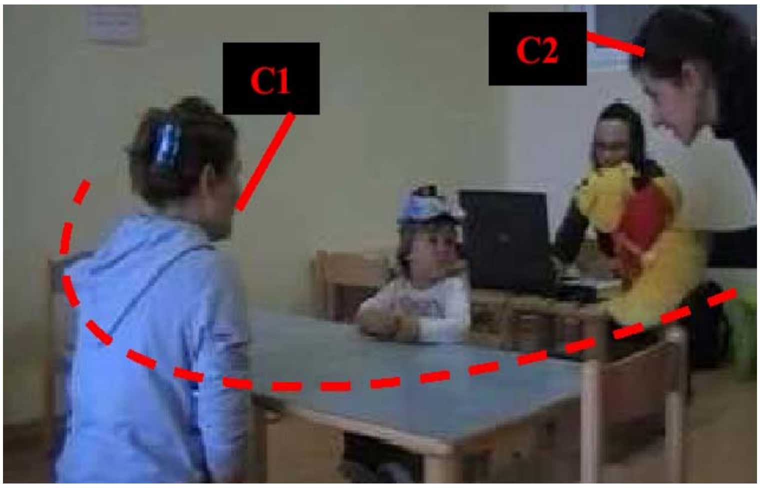

Figure 3.3: Vestibulo-auditory calibration procedure in a room of the day-care center: C1 speaks in front of the child, C2 moves a toy in a semicircle in front of the child.

During the procedure, performed in a room of the day-care center, the child is sitting on a chair at the long side of a desk while wearing the AVVC device (see Fig.3.3). A caregiver (C1) is sitting in front of him/her, on the opposite side of the desk, and exhorts the child to orient the head toward a second caregiver (C2). C2 moves in a semicircle in the frontal space of the child, and captures his/her attention by holding a toy without speaking. C1 represents a fixed sound source in front of the child, while he/she is orienting the head from left to right and viceversa following the toy. The procedure is carried out like a game and children enjoyed it very much.

The calibration curve is determined by correlating the TDOAs estimated by the localization algorithm and the head azimuth estimated from provided by the magneto-inertial sensor.

Output of the vestibulo-auditory calibration procedure is the binaural azimuth, Ψ_b, which is the location of the sound source, in the horizontal plane, respect to the child's head.

$$\Psi_b = G \cdot TDOA + O \tag{3.2}$$

Absolute direction of the speaking caregiver, Ψ_v, is then estimated as the difference between the binaural azimuth, Ψ_b, and the head azimuth, Ψ_h, thus integrating sensory and motor proprioception:

$$\Psi_v = \Psi_b - \Psi_h \tag{3.3}$$

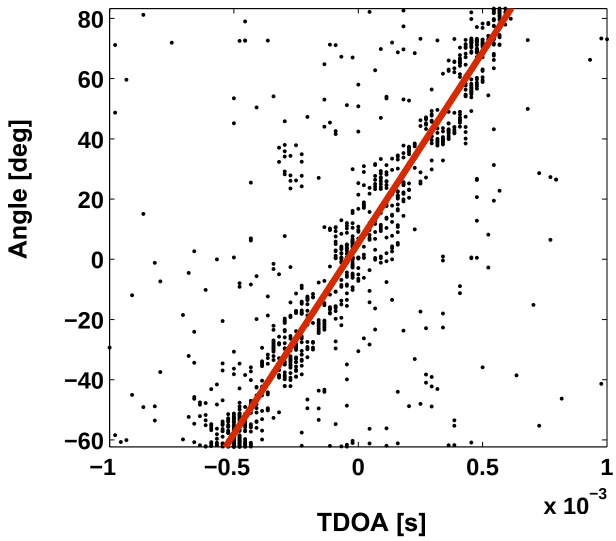

Figure 3.4: Vestibulo-auditory calibration curve for a 24-months-old child: the black dots are the raw data, the red line is the fitted curve (R-square coefficient = 0.95)

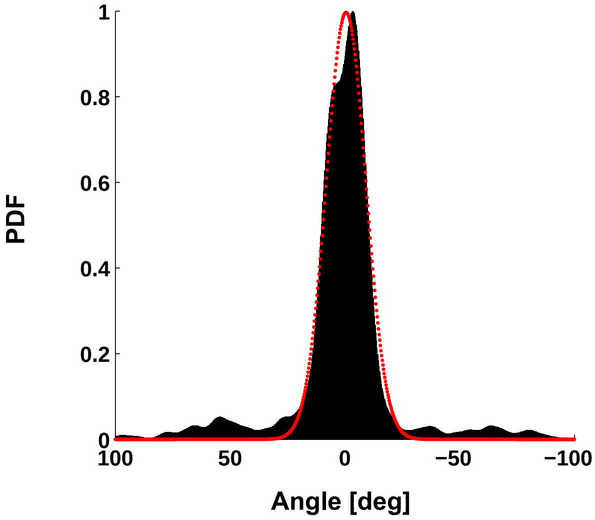

Figure 3.5: Probability Density Function (PDF) of the estimated sound source direction, Ψ_v: the black area represents the raw data, the red line the gaussian fitting of the PDF (mean $= -0.5°$; standard deviation $= 12°$; R-square coefficient $= 0.98$)

The calibration curve relative to a 24-months-old child is shown in Fig.3.4 and the probability density function (PDF) of the estimated sound source direction during the calibration procedure is shown in Fig.3.5. The PDF of the sound source direction, Ψ_v, estimated for each calibration session, has been fitted using a Gaussian model, with 95% confidence bounds. As expected, the probability to find the sound source, the caregiver's voice, is maximum at about 0° (i.e. in front of the child).

The coefficients (mean and standard deviation) of the fitted PDFs, obtained for each subject, together with the R-square coefficients are listed in Table 3.1. R-square coefficients are close to 1 (see Table 3.1), thus confirming that the estimated sound source distribution is well fitted by a Gaussian process.

Giuseppina Schiavone

Table 3.1: PDF parameters of estimated sound direction

Subject	Mean [*deg*]	StD [*deg*]	R^2
1	-0.34	17	0.98
2	0.89	24	0.96
3	0.23	21	0.97
4	-0.07	20	0.98
5	0.02	15	0.99
6	0.37	16	0.98
7	-0.01	20	0.99
8	1.09	18	0.98
9	0.01	15	0.99
10	0.09	15	0.98
11	-0.12	16	0.98

A T-test on the mean values of the PDFs shows that there is not a statistically significant discrepancy among the subjects (p-value=0.16, $\alpha = 0.05$). Also standard deviation values, measure of the width of the bells, are not discrepant (p-value=1, $\alpha = 0.05$, the mean of standard deviation values is 18°).

Compared to the human ability to localize sound source (1°- 5°), the accuracy of the AVVC device is very low. The dispersion of the PDFs is due to several factors: the performance of the algorithm used for sound localization in a noisy and un-controlled environment, such as a day-care centre, the head and trunk movement of the caregiver speaking in front of the child.

These results show that: *i)* the procedure respects the ecological approach, since it is simple and it can be configured as a sort of a game play with the child; *ii)* the calibration is consistent, reliable and there is not statistically difference when comparing the procedure with the standard practice; *iii)* although many factors can affect the AVVC accuracy in localizing sound source, it can be considered satisfactory for experimental conditions.

3.2 Vestibulo-Ocular Calibration

Eye-tracking systems calibration, used for transforming movements of the pupil in pixels to eye positions in degrees, usually, consists of looking at several markers on a screen in order to collect enough data to modify the parameter of an adjustable model, often while keeping the head still. Also,

Guppa Schiavone

this kind of calibration cannot be easily performed when the users are very young children, thus, not cooperative subjects.

The proposed calibration procedure is inspired by the vestibulo-ocular reflex (VOR) which allows generating compensatory eye movements in response to head motion as sensed by the vestibular organs in the inner ear. When the head rotates about any axis (horizontal, vertical, or torsional) distant visual images are stabilized by rotating the eyes about the same axis, but in opposite direction [23]. The gain of the VOR (the ratio of eye angular velocity $\frac{d\Psi_e}{dt}$ to head angular velocity $\frac{d\Psi_h}{dt}$) is typically around -1 when the eyes are focused on a distant target.

During vestibulo-ocular calibration, the subject is asked to rotate the head to the left and to the right while keeping looking at the experimenter who is sitting in front of him/her. Head rotation movements are recorded at a frequency of 100 Hz by the magneto-inertial sensor mounted on the top of the cap.

Given the relation between head and eye angular velocities, the head azimuth correlates with the coordinates of the pupil in the horizontal plane, Px. Linear fitting can be applied to the calibration curve to extract gain, G, and offset, O (see Fig.3.6). This allows expressing the eye orientation, Ψ_e, in degrees rather than in (normalized) pixels:

$$\Psi_e = GPx + O \tag{3.4}$$

Giuseppina Schiavone

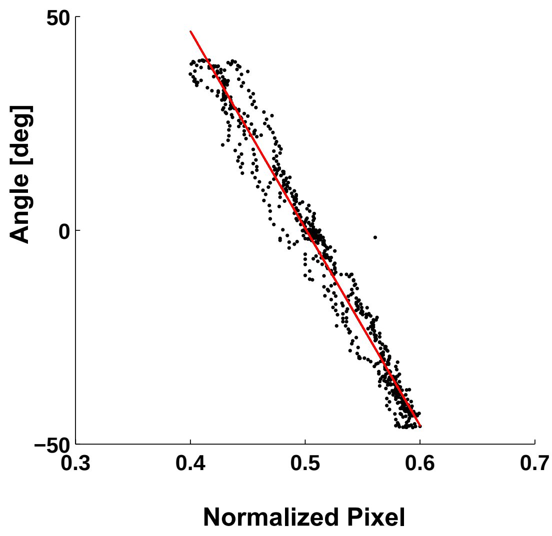

Figure 3.6: Vestibulo-ocular calibration curve: the black dots are the raw data, the red line is the fitted curve (R-square coefficient = 0.97)

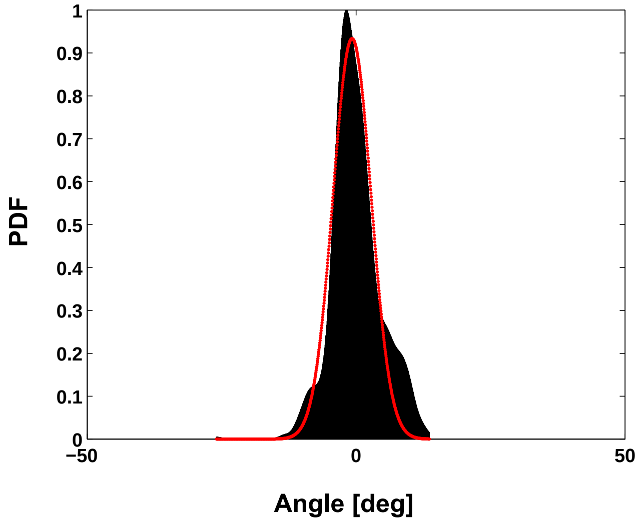

Figure 3.7: Probability Density Function (PDF) of the estimated gaze direction, Ψ_g: the black area represents the raw data, the red line the gaussian fitting of the PDF (mean $= -.7°$; standard deviation $= 5°$; R-square coefficient $= 0.95$)

The gaze direction, Ψ_g, is then estimated as the difference between the eye-in-head orientation, Ψ_e, and the head azimuth, Ψ_h, thus integrating sensory and motor proprioception:

$$\Psi_g = \Psi_e - \Psi_h \tag{3.5}$$

The probability density function (PDF) of the estimated gaze direction during the calibration procedure is shown in Fig.3.7.

In-the-lab calibration evaluation

The calibration procedure has been tested in-lab on 5 subjects. Fig.3.8 shows the calibration curves obtained for each subject, where variation in offset values are due to the different positioning of the cap on each subject's head. R-square coefficient (0.94 ± 0.03) averaged on the subjects confirms that (normalized) pixels and angular positions are highly correlated with a linear fitting.

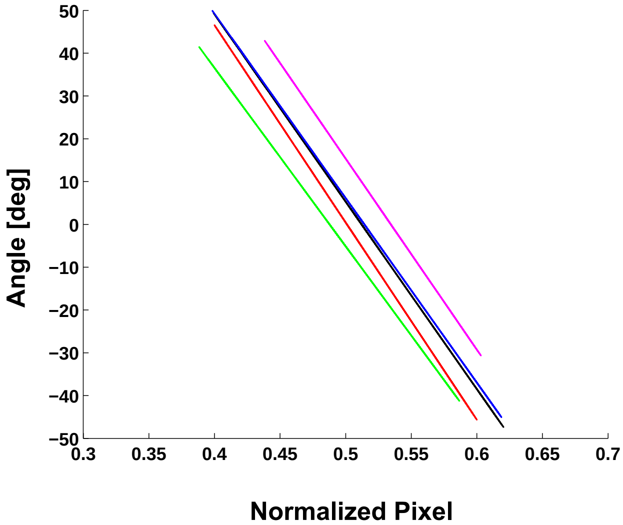

Figure 3.8: Vestibulo-ocular calibration curves: black line refers to subject 1, blue line refers to subject 2, red line refers to subject 3, green line refers to subject 4, magenta line refers to subject 5

Each subject orients the head over an averaged range of $[-43°,42°]$, thus proving that the calibration procedure operates over a range of motion of $\pm 40°$ in the horizontal plane.

A t-test applied to the estimated gaze over time (see Tab.3.2) reveals that the gaze distribution for each subject is not statistically different from a normal distribution with mean 0 ($p > 0.05$). This result confirms that the direction of the gaze is kept in a frontal direction on the experimenter.

Table 3.2: PDF parameters of estimated gaze direction

Subject	Mean $[deg]$	StD $[deg]$	R^2
1	0.3	6	1
2	-0.5	5	1
3	-0.7	5	1
4	-0.4	5	1
5	-0.2	8	1

Gruppe Shiawe

Although this procedure has not been yet tested with children in a day-care center, the obtained results show how, also, pixel/angle mapping can be achieved by exploiting solely head movement. Moreover the vestibulo-ocular calibration is simple and can be easily configured as a game (e.g. we observe that when the caregiver exhorts the child to mime a negation he/she turns his/her head keeping the gaze on the caregiver in front of the child, that is just what it is required in the proposed procedure).

3.3 Summing-up

Fig.3.9 resumes the behavioral relevant features extraction through the sensori-motor mapping.

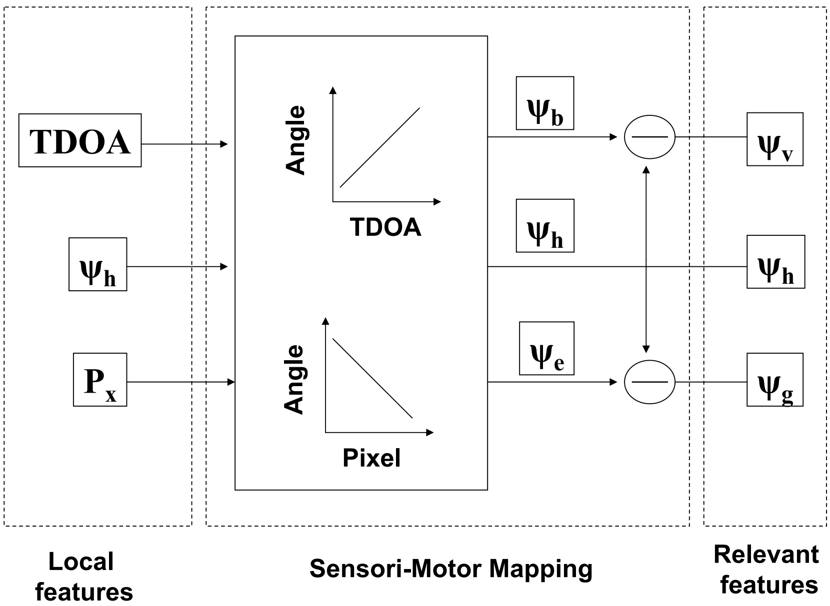

Figure 3.9: Behavioral relevant features extraction process

The output of the processing system provides the principal features which characterize the child's orienting behavior: his/her head orientation in the horizontal plane, his/her gaze and the direction of the sound sources, human voices, interacting with him/her.

Depending on the experimental protocol adopted for investigating child's sensori-motor coordination these features can be integrated to observe specific micro-behaviors, such as, if and for how long the child is looking at a caregiver who is speaking/ at a caregiver who is silent/elsewhere; if there are anticipatory movements or delays in shifting attention towards a social stimulus; which motor strategies the child uses during social interaction, i.e. if he/she orients only the eyes or is able to coordinate eye-head movement; pattern of repetitive behaviors; emotional involvement can also be assessed by measuring the blinking and eye-head kinematics.

Chapter 4

Experiments in day-care center: the social protocol

In this chapter the experimental protocol specifically designed for investigating child orienting behavior and the first results from the experiments carried on children in *La primavera del campus* day-care center are reported. Qualitative evaluation of the performance of the AVV system in detecting relevant behavior compared to the standard video coding technique is presented.

4.1 The Social Protocol

The principal scientific hypothesis is that children with autism spend less time overall looking at people and look more briefly at people and for longer durations at objects [80]-[29].

The experimental scenario, that stimulates the child to perform relevant sensory-motor tasks in social orienting behaviors, has been designed, within TACT project, by G. Stenberg (Dept. of Psychology, University of Uppsala, Sweden).

During the protocol, carried out in a room of a day-care center, the child's chair is placed at the long side of a table. The two experimenters (E_{right} and E_{left}) are sitting one at each end of the table facing each other (see Fig.4.1). Under the table, hidden from the child, are 8 bricks (4 for each experimenter).

The computer, connected to the AVV-Cap, is placed behind the child, out of his/her sight. A video camera is positioned in the room for filming the experimental session.

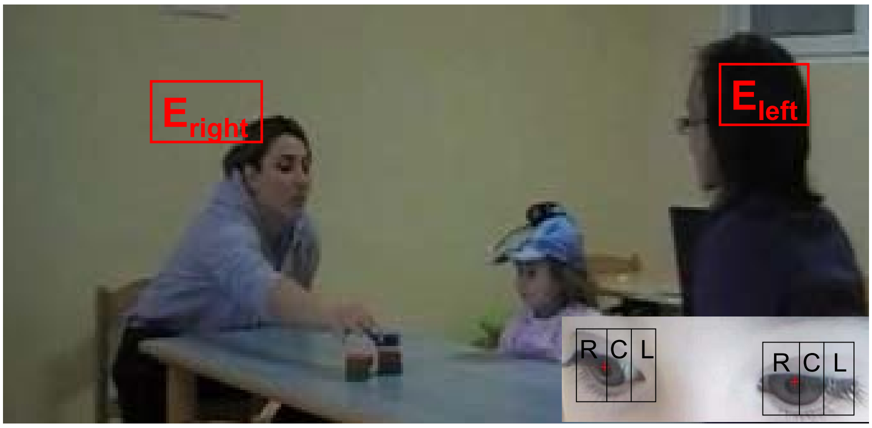

Figure 4.1: Social Protocol Setting

The social protocol consists of four phases:

1. *Talking*: The experimenters are talking to each other alternately, one at the time

2. *Talking and building I*: The experimenters are putting colored bricks in front of the infant alternately, one at the time. Every time the experimenter puts a brick in front of the infant she/he also talks about what he/she is doing.

3. *Talking and building II*: The experimenters are putting colored bricks in front of the infant alternately, one at the time. However, every time an experimenter puts a brick in front of the infant that experimenter remains silent. Instead, the other experimenter talks about what is happening.

4. *Building*: The experimenters continue building with bricks in the same manner, one at the time. However, now they are silent all the time.

The child is expected to perform the following actions: *i)* look at the person who is speaking/acting; *ii)* look at the person who is not speaking/acting;

iii) look at the hand of the person who is putting the colored brick on the table; *iv)* look at the objects on the table. These behaviors are critical: normally developing children are expected to look at the face of other persons while autistic children may look at objects and actions instead.

Normally developed children, between 20 and 36 months old, participated in the experimental sessions after the approval of informative consent by their parents. Vestibulo-auditory calibration procedure was performed for each child before the social protocol. Although the vestibulo-ocular calibration was not yet tested it was possible to estimate only qualitative positions of the eye (i.e. left, center, right, as in Fig.4.1).

In this work, aiming at assessing the performance of the AVV system in respect to standard video coding, only the first two phases of the protocol have been performed, using a total of 6 colored bricks instead of 8. Data collected on one child are reported as an example of the AVV system application.

4.2 Data Analysis

Twofold data analysis has been performed for assessing child's orienting behavior: from one side, videos from the external camera have been codified by using ELAN, professional software developed by the Max Planck Institute (The Netherlands) for the creation of complex annotations on video and audio resources; from the other side AVV-Cap multimodal signals have been processed as described in Chapter 2 and Chapter 3.

Three agents are defined: person to the left, person to the right, child.

While *person to the left* and *person to the right* can perform two main behaviors, talking (**t**) and not talking (**n**), the child's behaviors are classified as *Look to*:

- center (**c**), look at the object;
- left (**l**), look at the face of the experimenter to the left;
- right (**r**), look at the face of the experimenter to the right;
- hand of the experimenter on the left (**lh**),
- hand of the experimenter on the right (**rh**).

A trial is defined every time one of the experimenter talks. Thus, for the first two phases of the protocol, a total of 12 trials have been analyzed.

4.2.1 Video Coding

ELAN software Fig.4.2 allows to manually add annotations on the recorded video. Time intervals in which one of the two experimenters is speaking can be selected and within the same intervals it is possible to identify one of the child's behavior listed above.

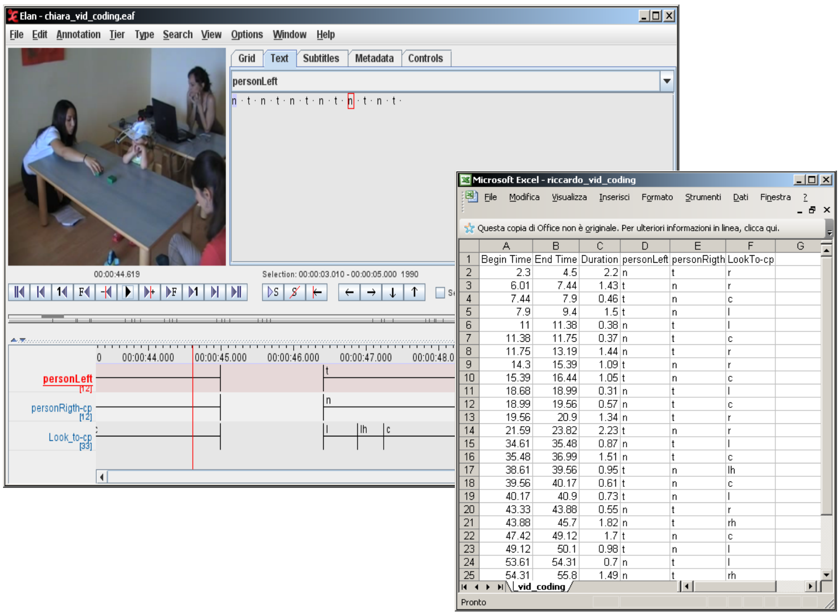

Figure 4.2: ELAN working window and .xls file with annotations exported from ELAN software

Detected behaviors can be saved in tab-delimited text file which can be exported in .xls format (Fig.4.2) for further analysis on the data.

In Fig.4.3 the sequence of behaviors over time is shown for a 26-months-old child.

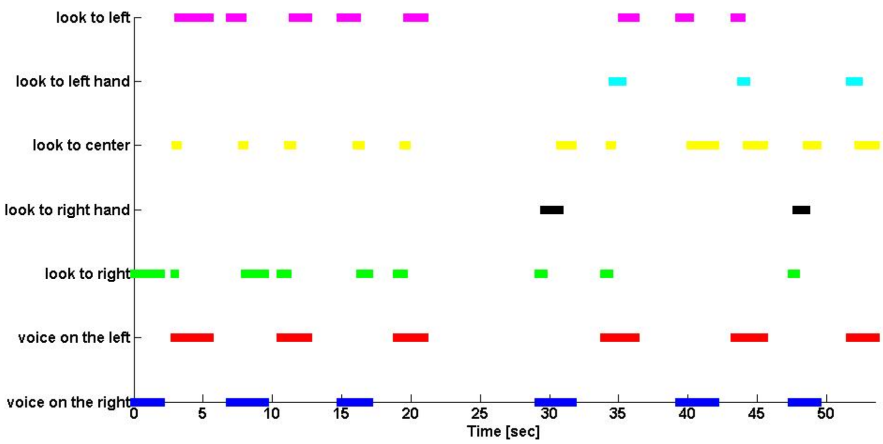

Figure 4.3: Timeline of child's behaviors

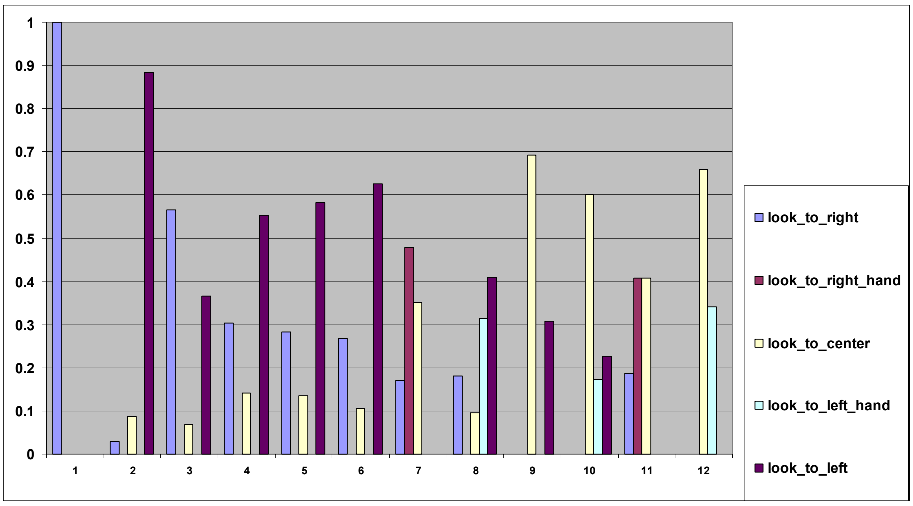

Figure 4.4: Relative duration of the behaviors for each trial

For each trial, i, the relative time, $t_{i,j}$, spent to look at a specific direction, j, is defined as:

$$t_{i,j} = \frac{\Delta t_{j,i}}{\Delta t_i} \tag{4.1}$$

where $i = 1, 2, ...N(N = 12)$, $j = (c, l, r, lh, rh)$, Δt_i is the trial duration, $\Delta t_{j,i}$ is the duration of the behavior within the trial.

Experimenter on the right of the child speaks in the trials 1,3,5,7,9,11, experimenter on the left speaks in the trials 2,4,6,8,10,12.

In the first six trials, in Fig.4.4, composing the first phase of the protocol, the child spends more time looking at the experimenter who speaks. In the second phase of the protocol the child spends more time looking at the object and at the movements of the hand of the experimenter while he/she is putting the bricks on the table, rather then to the experimenter who is speaking. This behavior could be interpreted as a selective attention to the action rather then to the voice, as if auditory information become less important for the child.

4.2.2 AVV-Cap coding

As described in Chapter 2 and Chapter 3, signals provided by the AVV system are processed to extract relevant behavioral features. In Fig.4.5 the experimental scenario is reconstructed. The multimodal processing system allows to automatically distinguish if somebody is speaking or is silent (Fig.4.5 A), estimates the child's head azimuth (Fig.4.5 B), localizes the experimenters' voice direction (Fig.4.5 C), detects child's eye position and eye blinking (Fig.4.5 D).

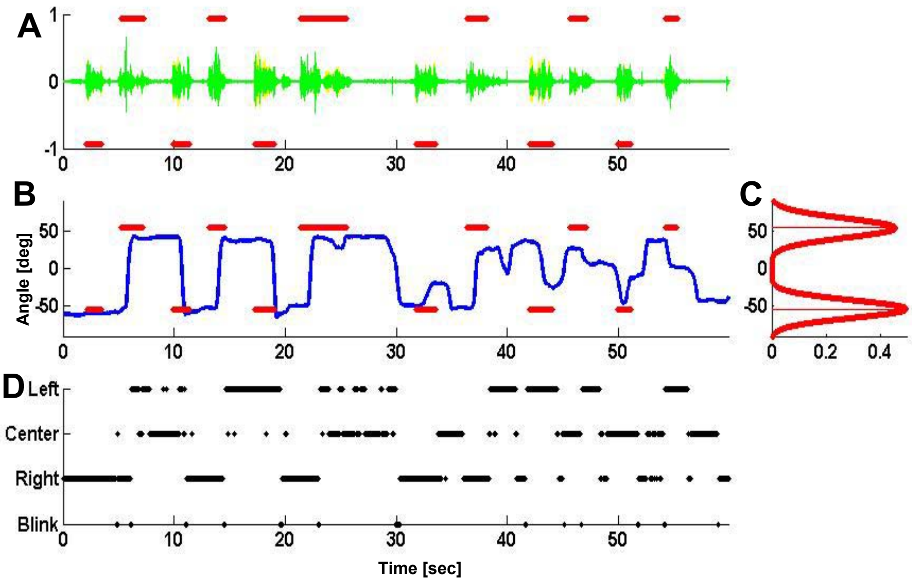

Figure 4.5: Reconstructed experimental scenario. A, red lines, talking intervals in the stereo tracks. B, blue line child's head azimuth, red lines, estimated voice directions. C, PDF of the estimated voice directions. D, eye positions and blinking

In order to obtained a timeline of the child's behavior, like the one realized with the video coding technique (Fig.4.3), head azimuth angles, ψ_h, have been classified in five regions of the horizontal plane (see Fig. 4.6): the child's head is oriented to center when $-15° <= \psi_h <= 15°$, to his/her right when $\psi_h <= -35°$, to his/her left when $\psi_h >= 35°$; the child is looking to the hand of the experimenter on his/her right when $-35° < \psi_h < -15°$, on his/her left when $15° < \psi_h < 35°$.

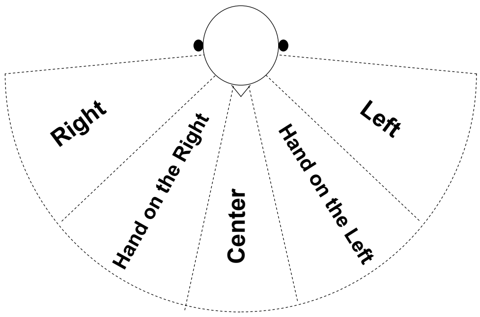

Figure 4.6: Frontal horizontal plane regions for orienting directions classification

Considering only the vestibular information and the estimated location of the sound sources, the behavior timeline reconstructed with the AVV system is shown in Fig.4.7.

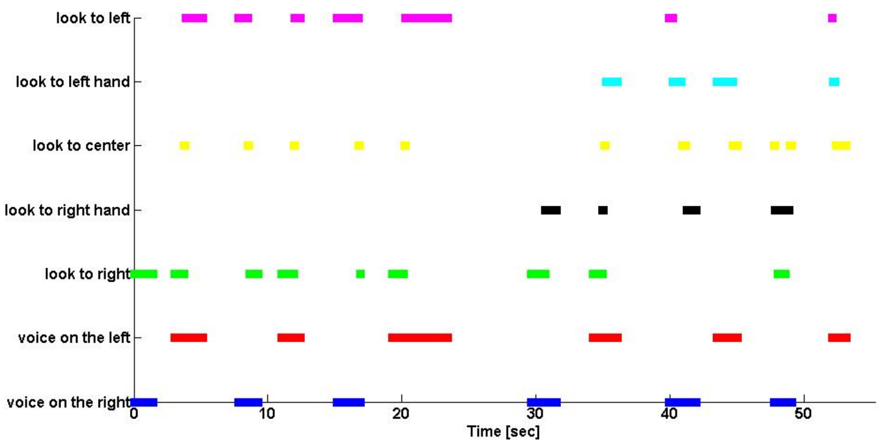

Figure 4.7: Timeline of child's behaviors

Information eye directions is added by following the rules in Fig.4.8: i.e. in phase I of the protocol, if head azimuth is in a central position and the eye is oriented to the left then the gaze is directed to the left; in phase II, if the head is oriented to the hand on the left and the eye is in a central position, then the child is looking to the left hand movement.

		Phase I					Phase II			
		Eye								
		b	**l**	**c**	**r**		**b**	**l**	**c**	**r**
azimuth	**c**	b	l	c	r		b	lh	c	rh
	l	b	l	l	c		b	l	l	lh
	r	b	c	r	r		b	rh	r	r
	rh						b	c	rh	r
	lh						b	l	lh	c

Figure 4.8: Rules for integrating head and eye direction in the definition of the gaze: b, blinking; c, center; r, right; l, left; rh, hand of the experimenter on the right; lh, hand of the experimenter on the left

The completed timeline of the child's behaviors can then be obtained, as shown in Fig.4.9, where also the child's eye blinking is represented.

The relative time, $t_{i,j}$, spent to look at a specific direction, can be also estimated as in 4.1. Same considerations, done for results in Fig.4.4, can be done looking at Fig.4.10: the first phase of the protocol, the child spends more time looking at the experimenter who speaks; in the second phase of the protocol the child spends more time looking at the movements of the hand of the experimenter while he/she is putting the bricks on the table, rather then to the experimenter who is speaking. The data provided by the AVV-cap can give a more detailed description of the child's orienting behavior

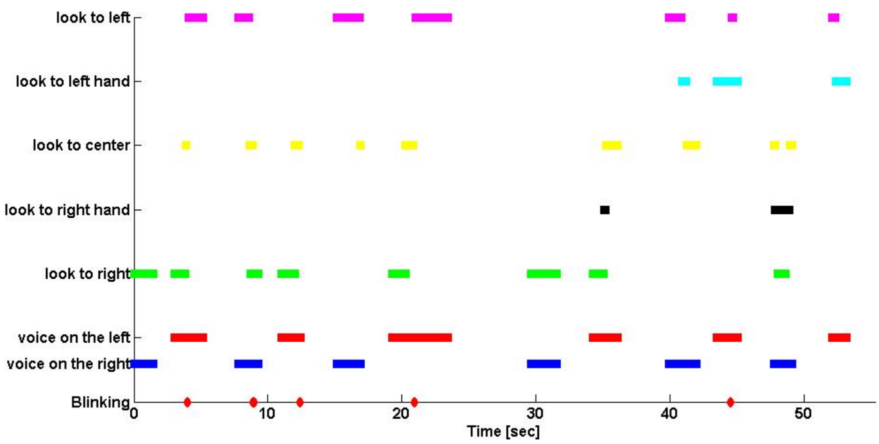

Figure 4.9: Complete Timeline of child's behaviors

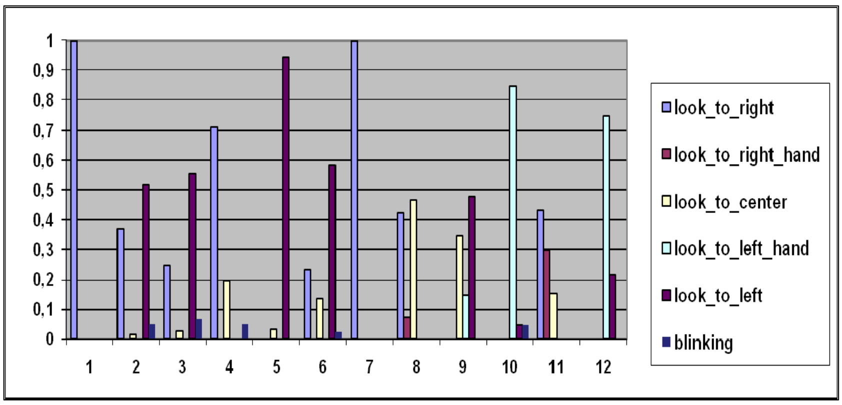

Figure 4.10: Relative duration of the behaviors for each trial

4.2.3 Comparison: AVV system vs video coding

At the present, qualitative considerations can be drawn on the comparison between data provided by the AVV-Cap and data coming out with a standard video coding.

Comparing Fig.4.3 and Fig.4.9, there is not a statistically difference among the intervals of talking (trials) detected with the two procedures. In Table.4.1 the duration of each trial measured with the AVV-cap and with the video coding procedure is reported. The distribution of the difference between the two estimates does not differ from a normal distribution with mean 0 ($p = 0.1912 > 0.05$).

Table 4.1: Trials duration: AVV-Cap vs Video Coding

trial	duration [sec] (AVV-cap)	duration [sec] (video coding)	difference
1	1.4	1.9	0.5
2	2.1	2.5	0.1
3	1.5	2.5	1
4	1.4	1.9	0.5
5	1.8	2	0.2
6	4.2	1.9	2.3
7	1.9	2.4	0.5
8	1.8	2.2	0.4
9	2	2.6	0.6
10	1.5	2.1	0.6
11	1.3	1.8	0.5
12	1.1	1.8	0.7

Table 4.2: Frequency of occurrence of Child's behaviors: AVV-Cap vs Video Coding

child's behavior	frequency (AVV-cap)	frequency (video coding)
look to l	7	8
look to lh	3	3
look to c	9	11
look to rh	3	2
look to rh	8	9

Also the frequency of occurrence (see Table. 4.2) of the child's behaviors described in 4.2 does not differ statistically when behaviors are estimated

Guppe Shiavo e

with the AVV-cap and with ELAN software. Nevertheless, some considerations need to be done.

The child's behavior in the first phase of the protocol represented by the AVV system is almost specular to the representation obtained with the video coding; as regard to the second phase of the protocol it seems that some behaviors are missing in the behavior timeline provided by the AVV system. This last result can be explained by the contribution of eye movements to the gaze, as defined in the rules in Fig.4.8. From one side, it can happen that the experimenter who is coding the video takes into account only the child's head orientation and neglects the eye orientation, when it is not well recognizable from the video of the external camera. From the other side, rules in Fig.4.8 are a simplification which not always represent well the behavior.

Other remarks can be done:

1. While the annotation procedure can be subjected to the ability of the person who is screening the video of detecting particular behaviors, the signals processed by the AVV system are un-effected by subjective evaluation.

2. Video coding can be a long and time consuming procedure, especially when more then one parameter at a time need to be considered (i.e. duration of talking, head movement together with eye movement). The AVV processing system can process multimodal data with semi-automatic functions and rules.

3. Although the AVV-Cap provides results very close to those obtained with the video coding procedure, the AVV-Cap is thought to be a tool complementary to it and to be able to provide other relevant information, not directly quantifiable with naked-eye: such as eye-tracking, frequency of blinking, head kinematics.

In the future, error metrics for quantitatively assessing the difference between the two methodologies will be defined.

4.2.4 Child's head stability during fixation towards a social stimulus

Within the described social protocol other interesting parameters can be used for child's behavior assessment, in particular head kinematics parameters. Monitoring the control of head stability during social interaction can provide information on child's attentional skills and child's emotional involvement.

Postural control, together with facial expressions, eye contact and gesture, belongs to the class of nonverbal behaviors. Lack of postural control may interfere with the development of social interaction and communication in the first half of first years of the infant: head and trunk control are fundamental

Gppa Shiavone

for establishing eye-contact with the mother [39]-[89]. Many theories have emphasized the importance of the social contact between mother and infant in promoting later cognitive and social-emotional development. A study conducted by Wijnroks et al. [90]- [85] showed that postural control could predict later cognitive development and inattention of infants who were born prematurely.

In attentional tasks, postural stability is a priority to stabilize head and gaze position. Good control and good coordination of eye-head system is a prerequisite also for learning and goal-oriented behavior in infancy, such as visual exploration and reaching. Through exploration, infants learn about the object's properties and characteristics, and about the effects they have on objects, all of which contribute directly to infants cognitive development.Postural dysfunctions, in this case, may interfere with prolonged and detailed exploration and manipulation of objects, which in turn, give the infant less opportunity to learn about the properties of objects, hence knowledge of the object world.

Head and body cues can also communicate accurate information about affect [64], about whether a child understands a caregiver's message and if the child is engaged in the interaction. One might speculate that, when infants frequently over-stretch while exploring objects, they will have difficulties learning to focus and sustain their attention [74].

For example when we say *Yes* we tilt forward the head or when we are concentrated on somebody speaking with us we unconsciously try to keep our head as fixed as possible.

Very few is known on this subject especially in children. Parameters describing head tilting and shaking during fixation towards a social stimulus can be useful to support the diagnosis of attention disorders in very young children.

An example of the analysis which I am running is here described. Considering the data of the child reported above, other then head azimuth also tilt and lateral oscillation of the head can be measured with the AVV-cap and angular velocity can be estimated, as described in Chapter 2.

When head angular velocity around the vertical axis, ω_z, is kept within an interval of $\pm 10°/s$, head azimuth, ψ_h, can be considered fixed and pointing to one direction (to experimenter on the left or on the right or to the center). For $\psi_h \geq 35°$ and $\psi_h \leq -35°$ the head azimuth is oriented toward the two experimenters. For $- < 35° \psi_h < 35°$ the head azimuth is oriented toward a non social stimulus, the objects on the table.

Guseppina Schiavone

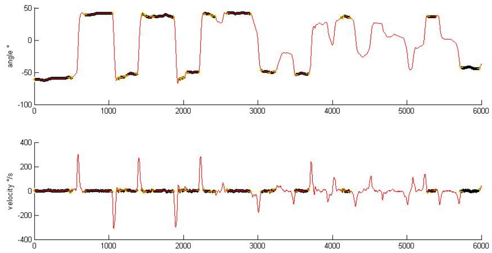

Figure 4.11: Intervals of fixation to social stimuli: red solid line, head azimuth and head angular velocity in the horizontal plane; black solid lines, fixation intervals; yellow crosses, beginnings and ends of the fixation intervals

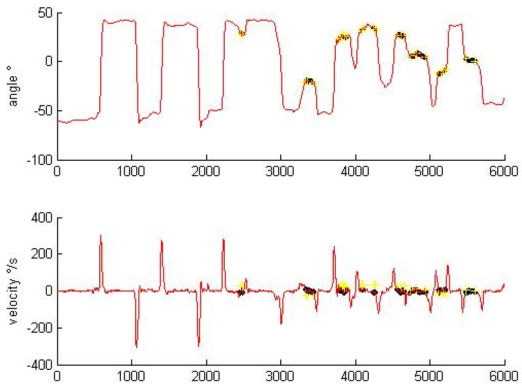

Figure 4.12: Intervals of fixation to non social stimuli: red solid line, head azimuth and head angular velocity in the horizontal plane; black solid lines, fixation intervals; yellow crosses, beginnings and ends of the fixation intervals

Only intervals of duration higher than $200ms$ (corresponding to 20 samples, at a sampling frequency of $100Hz$) have been considered for both social (Fig. 4.11) and non social stimulus (Fig. 4.12). This interval of time is set as the minimum duration of fixation in which perception takes place [34] - [37].

22 intervals of fixation have been detected for social stimuli, 15 intervals for non social stimuli. For each interval mean, $\bar{x}$ and standard error, S_E, have been determined considering a 95% confidence interval:

Guseppina Schiavone

$$S_E = \frac{\sigma}{\sqrt{n}} \tag{4.2}$$

where n is the number of sample of each interval ($n \geq 20$), σ is the standard deviation.

If the data are assumed to be normally distributed within the interval, the upper and lower 95% confidence limits can be expressed as:

$$Upper95\%Limit = \bar{x} + (S_E * 1.96) \tag{4.3}$$

$$Lower95\%Limit = \bar{x} - (S_E * 1.96) \tag{4.4}$$

In Fig. 4.13 and Fig. 4.14 error bars and means of pitch and roll within every interval are reported. Negative values of pitch describe head tilt forward, positive values describe tilt backward; lateral tilt to left is described by positive roll values, to right by negative roll values.

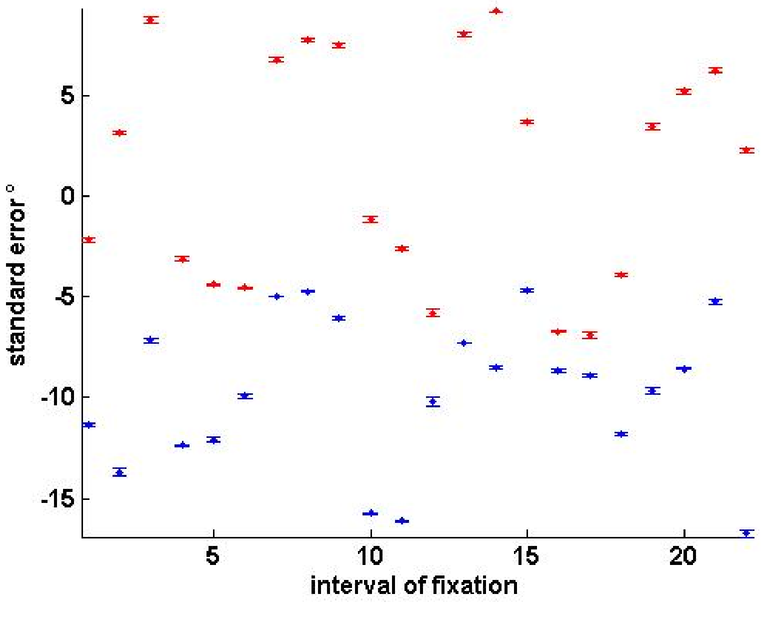

Figure 4.13: Mean and standard error of pitch, blue, and roll, in each interval of fixation to social stimuli

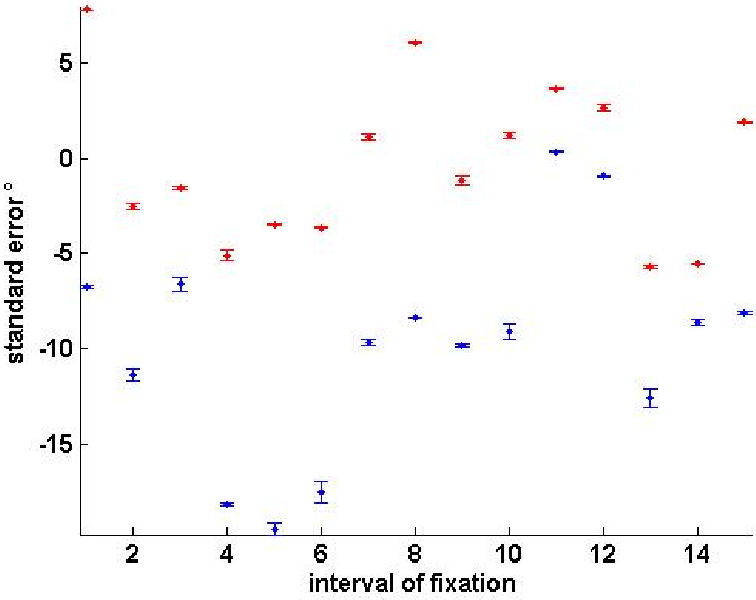

Figure 4.14: Mean and standard error of pitch, blue, and roll, in each interval of fixation to non social stimuli

Preliminary statistical observations can be drawn on a single subject: applying a two-tailed unpaired t-test to the standard error obtained for pitch and roll in both orienting conditions (towards social stimuli and towards non social stimuli), there is not a statistically difference between variations of pitch and variations of roll when the child is fixating a social stimulus ($\alpha = 0.05$, $p - value = 0.3$); the same result can be observed for fixations towards a non social stimulus ($\alpha = 0.05$, $p - value = 0.08$), although, in this condition pitch values vary slightly more then roll values. Comparing pitch and roll variations between the two conditions: while roll variations do not statistically differ for social and non social fixation intervals ($\alpha = 0.05$, $p - value = 0.5$), pitch variations in social intervals statistically differ from pitch variations in non social intervals ($\alpha = 0.05$, $p - value = 0.004$), where pitch standard error is higher. This last result can be explained considering that the objects on the table (non social stimuli) are positioned in a lower vertical position respect to the caregiver's face (social stimuli) and the child's head, thus higher pitch rotations are required to fixate the non social stimulus.

Even if very few can be inferred from the behavior of a single subject, at first sight one an say that there is not difference during attentional task

Guppe Shawe

towards social and non social stimuli.

The measurement of these parameters for a wide number of normally developed children has a twofold advantage: from one side it can be used as a standard to which compare behaviors of child at risk of attention disorders; from the other side it provides information on the postural control in sitting and head control during social interaction which are a necessary prerequisite for a correct development of communicative and attentional processes, and for supporting exploratory and learning processes.

Conclusions

In this chapter the main outcomes of this research work will be summarized and I will outline future research directions, stressing the multiple application fields of the designed device.

This work has given an overview of the major goals underlying the design and the use of the AVV-Cap as an unobtrusive, wearable technology for assessing very young children sensory-motor coordination during social situations. The focus on children from 12 to 36 months comes from evidences in literature that there are clearly observable behaviors which are important predictors of autistic disorder in pre-verbal children.

The behavior representation obtained with the AVV-Cap is a complementary tool to the standard video scoring used to analyze children's behavior. It has the advantage of being objective, it does not rely on the subjective qualitative judgement of the experimenter.

Moreover, the multimodal approach allows to evaluate the child's behavior taking into consideration both sensory information (hearing and vision) and motor information (eye-head coordination), standing out during the social interaction. Micro-behaviors can be easily inferred from the behavior representation provided by the AVV-Cap. Although the device accuracy is low, compared to traditional observational instruments, and new solutions, both hardware and software, needed for improving its performance, I proved it is suitable for application in un-controlled, ecological environments.

Results of such studies are promising not only from a diagnostic point of view. A wearable device, which is well accepted by children, can operate as an extended child-robot interface.

Giuppe Schiavone

What is going on now?

Because of its performance and its suitability for monitoring children behaviors during their first three years of life in ecological environments, the AVV-Cap is thought to be a powerful candidate tool for supporting other experiments within a new European Project, IM-CLeVeR (Intrinsically Motivated Cumulative Learning Versatile Robots) [1].

IM-CLeVeR aims to develop a new methodology for designing robots controllers that can: (1) cumulatively learn new efficient skills through autonomous development based on intrinsic motivations, and (2) reuse such skills for accomplishing multiple, complex, and externally-assigned tasks. During skill-acquisition, the robots will behave like children at play which acquire skills autonomously on the basis of 'intrinsic motivations'. During skill-exploitation, the robots will exhibit fast learning capabilities and a high versatility in solving tasks defined by external users due to their capacity of flexibly re-using, composing and readapting previously acquired skills.

The robots controllers design will be modeled on the results of empirical experiments run with monkeys, children, and human adults, aiming at investigating three fundamental scientific and technological issues: (1) the mechanisms of abstraction of sensory information; (2) the mechanisms underlying intrinsic motivations;(3) hierarchical recursive architectures which permit cumulative learning.

Within this research framework empirical investigations on children will be conducted by the Laboratory of Developmental Neuroscience and the Laboratory of Biomedical Robotics and Biomicrosystems of Università Campus Bio-medico with the support of previous instrumented toys and wearable devices developed within Tact project.

The application of the AVV-Cap will play an important role in experimental protocols designed for investigating learning processes driven by curiosity and subsequent cumulative learning process (see Report D3.1 [4]). In particular, the eye-tracker and the orientation tracker mounted on the AVV-Cap could provide information on the contribution of vergence movements in prehension of objects [60] (see pilot study in Fig.4.15) and on the eye-hand coordination [73] in very young children (see pilot study in Fig.4.16).

Figure 4.15: Pilot study on the contribution of the development of vergence movements on prehension

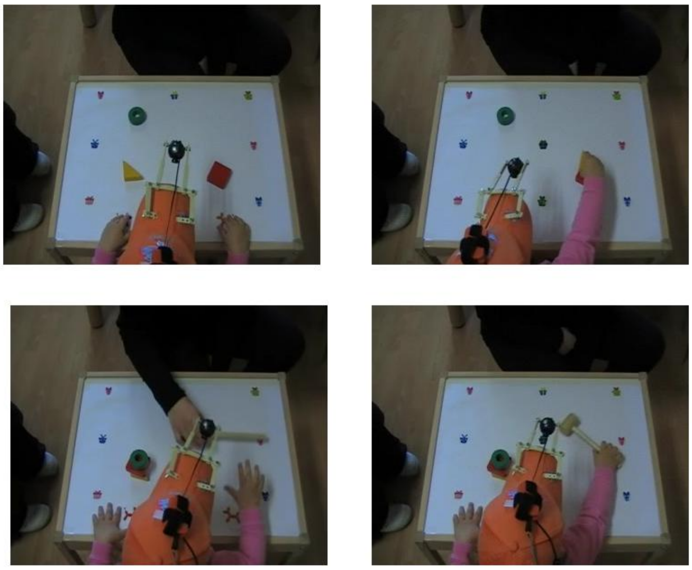

Figure 4.16: Pilot study on eye-hand coordination in peripersonal space

Results obtained from these studies could be useful for describing the development of the peripersonal space in infancy. Peripersonal space refers to the space surrounding the body, which can be reached by the limbs [56]-[16]. Within this space the child explores the world and learns about objects properties and functions.

Future Research Direction: from early diagnosis to robot-mediated therapy

Further work on the AVV-Cap will aim at developing a wireless communication module and a real-time signal processing system. Wireless connections will enable the child to move in a room without constrains, thus allowing to design new protocols for assessing social behavior. Real-time operating systems will provide the advantage of retrieving relevant information in a short time, thus allowing other systems, i.e. robotic platforms, to use them and improve the interaction.

In the last thirty years several robots, able to interact with children for educational and therapeutic purposes, have been developed. Particular attention is oriented to robotic application to rehabilitate children with autism who generally show communicative and other social skills deficits (see [25], [69], [70], [26], [88]).

Interactional dynamics defined by synchronized and coordinated responses between the robot and the child have strong influences on regulation and naturalness in interaction [7]. Human-robot interaction, modeled on human-human interaction [47], has impact on the ability of the robot of attention seeking behavior to initiate interactions and of attention keeping behavior to sustain interactions [71] and more in general on its believability, as a life-like appearance system [45].

In the majority of the trials presented in literature, the robot interactional dynamics are triggered by the experimenter who (taking the role of the robot) interpreted the interactions and the meaning of the children's behavior, and then selected the appropriated responses for the robot (as in [71] - [53] - [11]). Simple action-reaction movements can be performed autonomously by mobile robots, where heat sensors and infrared sensors allow to detect the presence of the child and the contact with him/her [76]. Motion tracking system mounted on the robot, as in [15], can analyze gross arm movements of the child that in turn trigger the robot to imitate the child. In some cases the robot, provided with built-in behaviors, performs movement on its own, to catch the attention of the child or can behave on the basis of attention maps, created by detecting the locations of a human face, toys and moving

Gyuppa Shiavo

objects in the environment, as in [52].

In the absence of the remote control by the experimenter, the robot has to record and reorganize the interaction by using only its proprioceptive data, recognize facial expressions and emotions of the child, detect his/her body movements and gaze, sense touch and discriminate voices. In a natural familiar environment, different from a well-controlled laboratory situation, factors like changes of lighting conditions, displacements of objects together with the complex and unpredictable child's behavior can affect the robot 'interpretation' of the interaction.

The combination of passive wearable technology with social robots could improve the child-robot interactional dynamics, in particular in situations in which the child has to face the robot for the interaction (as in Robota doll [15], Infanoid [44], Keepon [54], ESRA [11] and Kaspar [72]). Moreover, this integrated approach could provide quantitative evaluation of child-robot interaction, still immature and based on subjective assessment [46]-[63].

The use of passive technology was already explored in [67] where instrumented gloves, accelerometers embedded in armbands and in a hat, pressure sensors in the shoes, were used only to remotely control the movements of the robot. Other recent works refer to application not suitable for children and for their natural environments: in [51] wearable sensors and stationary sensors positioned in a room are used to improve the behavior of an interactive guide robot; in [47] and in [82] body movement analysis of human-robot interaction is assessed by using a motion tracking system with wearable optical markers, in the first work, and by using a combination of accelerometers and gyroscopes, in the second work.

In the same line of this integrated approach, future development of the AVV system can contribute to enhance robotic research in child-robot interaction since: it can provide systematic quantitative evaluations of interaction dynamics between child and robot; temporal and structural coordination of the robot interacting with the child can be improved using the information on the child's behavior provided by the AVV-Cap.

In a future scenario, autonomous social robots will gather and integrate both proprioceptive information (signals provided by sensors mounted on the robot) and information coming from wearable devices (signals processed from a child-centered perspective), such as the AVV-Cap. The knowledge of the child's state (e.g. his/her orientation in the space, his/her attention direction, and his/her emotional state) will guide the robot to interact with him/her in a coordinated and synchronized way, enabling the robot to be more socially responsive and closing the loop of child-robot interaction.

List of Publications

Abstracts

2008

- **G. Schiavone**, D. Campolo, E.Guglielmelli, F. Keller, Wearable technology for monitoringorienting behavior in children, in Proceedings of the First International Symposium on Perception and Action in Early Development (PAED), Rome, 11 December, 2008

- **G.Schiavone**, D.Campolo, F.Keller, E.Guglielmelli, Multimodal analysis of orienting behaviortowards speaking sound sources based on binaural cues and head tracking, Poster for Conference of Neuroscience, Washington DC, 15-19 November, 2008

- F.Taffoni, **G.Schiavone**, D.Campolo, F.Keller, E.Guglielmelli, Early diagnosis of autism: new perspective and devices, Poster for Primo Convegno Nazionale di Bioingegneria, Pisa, 3-5July 2008

2009

- **G. Schiavone**, D. Campolo, F. Keller, E. Guglielmelli, Multimodal device for assessing children orienting behavior in ecological environments, 4th International Conference on Spatial Cognition (ICSC), Rome, 14-18 September, 2009

- I. Gaudiello, D. Caligiore, **G. Schiavone**, A. Salerno, F.Sergi, L. Zollo, E. Guglielmelli, D. Parisi, G. Baldassarre, R. Nicoletti, A. M. Borghi, Effect on space representation of using a tool and a button, 4th International Conference on Spatial Cognition (ICSC), Rome, 14-18 September, 2009

Peer-Reviewed Conference Papers

2007

- L. Picardi, B. Noris, O. Barbey, A Billard, **G. Schiavone**, F. Keller, Cl. V. Hofsten, WearCam: A head mounted wireless camera for monitoring gaze attention and for the diagnosis of developmental disorders in young children, In Proceedings of the 16th IEEE International Simposium on Robot and Human Interactive Communication (RO-MAN), JeJu Island, Korea, 26-29 August,2007.

2008

- D. Campolo, F. Taffoni, **G. Schiavone**, C. Laschi, F. Keller, E. Guglielmelli, A Novel Technological Approach Towards the Early Diagnosis of Neurodevelopmental Disorders, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Vancouver, Canada, August 20-24, 2008

- D. Campolo, F. Taffoni, **G. Schiavone**, C. Laschi, F. Keller, E. Guglielmelli, Towards the early diagnosis of neurodevelopmental disorders: a novel technological approach, International Conference on Technology and Applications in Biomedicine (ITAB), pp. 531 - 534, Tokyo, Japan, 30-31 May, 2008.

Journals

2009

- **G. Schiavone**, D. Formica, F. Taffoni, D. Campolo, E. Guglielmelli, F.Keller, Multimodal ecological technology: from child's social behavior assessment to child-robot interaction improvement, conditionally accepted to Journal of Social Robotics

Book Chapters

2009

- D. Campolo, F. Taffoni, **G. Schiavone**, D. Formica, E. Guglielmelli, F. Keller, Neuro-Developmental Engineering: towards early diagnosis of neuro-developmental disorders, accepted to Recent Advances in Biomedical Engineering

Other PhD activities

2007

- Research activity in European NEST/Adventure Project (TACT); topic: Design of new wearable, multimodal, ecological, unobtrusive device and approach for early diagnosis of neuro-developmental disorders; contribution in editing deliverables; participation to project meetings.

- **28 May - 3 August**: Internship in LASA Laboratory, Ecole Polytechinique Federale de Lausanne, Switzerland; topic: Research on Binaural Sound Localization Algorithm, CASA (Computational Auditory Scene Analysis), BICA (Binaural Independent Component Analysis).

- **5-12 September**: Attendance to the *3rd Summer School European University in Surgical Robotics*, in Montpellier.

- **14-15 April 2007**: Attendance to the *IEEE International Conference on Robotics and Automation* (ICRA07), in Rome; Attendance to the (IEEE RAS) Workshop on technical challenges for dependable robotics, in Rome; Attendance to the (IEEE RAS) Workshop on perspective in rehabilitation robotics, in Rome.

2008

- Research activity in European NEST/Adventure Project, TACT; topic: Design of new wearable, multimodal, ecological, unobtrusive device and approach for early diagnosis of neuro-developmental disorders; contribution in editing deliverables; participation to project meetings.

- **June-August**: Student Visitor in *Fondazione Peppino Scoppa* (Angri, SA, Italy); topic: Pilot Study on Application of an Audio-Visuo-Vestibular Device in COR (Conditioned Orienting Reflex) Audiometry for objectifying Audiometric Tests.

- **15-19 September** : Attendance to the *XXVII Scuola Annuale di Bioingegneria* in Brixen (Italy); topic: Sistemi indossabili intelligenti per la salute e la protezione.

- **15-19 November**: Attendance to the *Conference of Neuroscience*in Washington DC.

2009

- **Since May**: Research activity in European Project IM-CleVeR (Intrinsically Motivated Cumulative Learning Versatile Robots) Project; topic: design and development of mechatronic objects for behavioural experiments with children, design of experimental protocols for studying curiosity-driven learning and cumulative learning in children; contribution in editing deliverables; participation to project meetings.

- **March-May** : Teaching experience: Lessons on Gyroscope, Magnetometer and Accelerometer Sensors for the course of Ing. Sergio Silvestri and Ing. Domenico Campolo, 'Laboratory of Bioengineering and Mechanical Measurements', for the faculty of Biomedical Engineering in Università Campus Bio-Medico, Rome, Italy.

- **September 14-19** Attendance to the *4th International Conference on Spatial Cognition* in Rome, Italy.

Bibliography

[1] http://im-clever.noze.it/.

[2] http://lasa.epfl.ch/research/devices/wearcam/autism/index.php.

[3] http://tact.unicampus.it/.

[4] Im-clever deliverable 3.1.

[5] Campolo D , Taffoni F , Formica D , Schiavone G , Keller F and Guglielmelli E. Technology for ecological assessment of spatial cognition in infants. *Submitted to Journal of NeuroEngineering and Rehabilitation*, 2009.

[6] Omologo M and Svaizer P. *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pages II273–II276, 1994.

[7] Kendon A. Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32:100–125, 1970.

[8] Robinson D A. A method of measuring eye movements using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Electron. BME*, 10:137–145, 1963.

[9] Aarabi P, Zaky S. Iterative spatial probability based sound localization. *in Proceedings of the 4th World Multiconference on Circuits, Systems, Computers, and Communications,Athens, Greece,*, July 2000.

[10] Allison R S, Eiyenman M, Cheung B S K. *IEEE transaction on Biomedical Engineering*, 43(11):1073–1082, 1996.

[11] Scassellati B. Quantitative metrics of social response for autism diagnosis. *Proc. RO-MAN*, 2005.

[12] Baluja S, Pomerleau D . Non-intrusive gaze tracking using artificial neural networks. *Technical Report CMU-CS-94-102. Carnegie Mellon University*, 1994.

[13] Baron-Cohen S, Ring H A, Wheelwright S, Bullmore E T, Brammer M J, Simmons A, Williams S C R. Social intelligence in the normal and autistic brain: an fmri study. *European Journal of Neuroscience*, 11(6):1891–1898, 1999.

[14] Baron-Cohen S, Wheelwright S, Hill J, Raste Y, and Plumb I. The reading the mind in the eyes test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 2001.

[15] Billard A, Robins B, Dautenhahn K and Nadel J. Building robota, a mini-humanoid robot for the rehabilitation of children with autism. *the RESNA Assistive Technology Journal*, 19, 2006.

[16] Brain W R. Visual disorientation with special reference to lesions of the right hemisphere. *Brain*, 64, 1941.

[17] Brandstein M S, Silverman H. A robust method for speech signal time-delay estimation in reverberant rooms. *In Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, Munich, Germany*, April 1997.

[18] Campolo D, Laschi C, Keller F, Guglielmelli E. A mechatronic platform for early diagnosis of neurodevelopmental disorders. *RSJ Advanced Robotics Journal*, 21(10):1131–1150, 2007.

[19] Caronna EB, Milunsky JM, Tager Flusberg H. Autism spectrum disorders: clinical and research frontiers. *Archives of Diseases Childhood*, 93(6), 2008.

[20] Ceponiene R, Lepisto T, Shestakova A, Vanhala R, Alku P, Naatanen R and Yaguchi K. Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc Natl Acad Sci, USA*, 100(9):5567–5572, Apr 2003.

[21] Cornsweet TN and Crane HD. Accurate two-dimensional eye tracker using first and fourth purkinje images. *J. Opt. Sc. Am.*, 63:921–928, 1973.

[22] Courchesne E, Townsend J, Akshoomoff N A, Saitoh O, Yeung-Courchesne R, Lincoln A J, James H E, Haas R H, Schreibman L, Lau L. Impairment in shifting attention in autistic and cerebellar patients. *Behav Neurosci.*, 108(5):848–865, 1994.

[23] Crawford JD, Vilis T. Axes of eye rotation and listing's law during rotations of the head. *Journal of Neurophysiology*, 65(3):407–423, 1991.

[24] Datum MS, Palmieri F, Moise A. An artificial neural network for sound localization using binaural cues. *Journal of the Acoustical Society of America*, 100(1), 1996.

[25] Dautenhahn K. Robots as social actors: Aurora and the case of autism. *In Proc. CT 99, The Third International Cognitive Technology Conference*, pages 359–374, 1999.

[26] Dautenhahn K, Werry I. Towards interactive robots in autism therapy. *Pragmatics and Cognition*, 12, 2004.

[27] Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S. Health a critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment. *Health Technology Assessment*, 1, 1997.

[28] Dawson G, Meltzoff A N, Osterling J, Rinaldi J, Brown E. Children with autism fail to orient to naturally occurring social stimuli. *J. Autism Dev. Disord*, 28:479–485, 1998.

[29] Dawson G, Toth K, Abbott R, Osterling J, Munson J, Estes A, and Liaw J. Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*, 40(2), 2004.

[30] DiScenna A, Das V, Zivotofsky A, Seidman S, Leigh R J. Evaluation of a video tracking device for measurement of horizontal and vertical eye rotations during locomotion. *Journal of Neuroscience Methods*, 58:89–94, 1995.

[31] Dongheng Li, Babcock J, Parkhurst DJ. *Proceedings of the 2006 symposium on Eye tracking research and application*, pages 95–100, 2006.

[32] Donohue K D, Hannemann J and Dietz H G. Performance for phase transform for detecting sound sources in reverberantand noisy environments. *Signal Processing*, July 2007.

[33] Kemp DT. Stimulated acoustic emissions from within the human auditory system. *Journal of the Acoustical Society of America*, 64, 1978.

[34] Javal E. Essai sur la physiologie de la lecture. *Annales d'ocullistique*, 80, 1878.

[35] Eizenman M, Frecker R C, and Hallet P E. Precise non-contacting measurement of eye movements using corneal reflex. *Vis. Res.*, 24:167–174, 1984.

[36] Mullen EM. *Mullen Scales of Early Learning: AGS Edition. Circle Pines, MN: AGS.Mullen.* 1995.

[37] Erdmann B and Dodge R. Psychologische untersuchung über das lesen auf experimenteller grundlage. *Niemeyer: Halle*, 1898.

[38] Feng GC and Yuen PC. Multi-cues eye detection on gray intensity image. *Pattern Recognition*, 34:1033–1046, 2001.

[39] Fogel A, Dedo J Y, McEwen I. Effect of postural position on the duration of gaze at mother during face - to - face interaction in three-six-month-old infants. *Infant Behavior and Development*, 15, 1992.

[40] Frith U, Soares I. Research into earliest detectable signs of autism: what the parents say. *Communication*, 27(3):17–18, 1993.

[41] Byford GH. Non-linear relations between the corneo-retinal potential and horizontal eye movements. *J. Physiol. (London)*, 168:14P–15P, 1963.

[42] Ghow G, Li X. Towards a system for automatic facial feature detection. *Pattern Recognition*, 26:1739–1755, 1993.

[43] Baranek GT. Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9-12 months of age. *Journal of Autism and Developmental Disorders*, 29:213–224, 1999.

[44] Kozima H. Infanoid: A babybot that explores the social environment. *in Dautenhahn, K. et al. (eds), Socially Intelligent Agent, Kluwer Academic Publishers*, 2002.

[45] Bates J. The role of emotion in believable agents. *in Comm. of the ACM*, 37(7):122–125, 1994.

[46] Kanda T, Ishiguro H, and Ishida T. Psychological analysis on human-robot interaction. *Int. Conf. on Robotics and Automation*, pages 4166–4173, 2001.

[47] Kanda T, Ishiguro H, Imai M, and Ono T. Body movement analysis of human-robot interaction. *In International Joint Conference on Artificial Intelligence*, pages 177–182, 2003.

[48] Kemp B, Janssen AJMW, van der Kamp B . Body position can be monitored in 3d using miniature accelerometers and earth-magnetic field sensors. *Electroencephalography and Clinical Neurophysiology*, 109:484–488, 1998.

[49] Klin A, Lang J, Cicchetti D V, and Volkmar F R. Interrater reliability of clinical diagnosis and dsm-iv criteria for autistic disorder: Results of the dsm-iv autism field trial. *Journal of Autism and Developmental Disorders*, 30(2):163–167, 2000.

[50] Knapp C H and Carter G. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoustics, Speech and Signal Processing*, 24(4):320–327, 1976.

[51] Koide Y, Kanda T, Sumi Y, Kogure K, Ishiguro H. An approach to integrating an interactive guide robot with ubiquitous sensors. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[52] Kozima H and Nakagawa C. Interactive robots as facilitators of children's social development. *In: A. Lazinica (eds.) Mobile Robots: Towards New Applications, Vienna: Advanced Robotic Systems*, pages 269–286, 2007.

[53] Kozima H and Nakagawa C. Longitudinal child-robot interaction at preschool. *In AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*, pages 27–32, 2007.

[54] Kozima H, Nakagawa C, Yasuda Y, and Kosugi D. A toy-like robot in the playroom for children with developmental disorder. *in Proceedings of the International Conference on Development and Learning*, 2004.

[55] Battista L. Thesis: Caratterizzazione sperimentale di un sensore d'angolo di suono prodotto da una sorgente monofrequenziale, università campus bio-medico di roma. 2009.

[56] Ladavas E, Di Pellegrino G, Farne A, and Zeloni G. Neuropsychological evidence of an integrated visuotactile representation of peripersonal space in humans. *Journal of Cognitive Neuroscience*, 10, 1998.

Giuppa Schiavone

[57] Lepisto T, Kujala T, Vanhala R, Alku P, Huotilainen M, and Naatanen R. The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Res.*, pages 1066–1147, 2005.

[58] Kipp M. Anvil-a generic annotation tool for multimodal dialogue. *In Proc. Eurospeech.*, 2001.

[59] Landau U M. Estimation of a circular arc centre and its radius. *Comput. Vis Graph. Image Process.*, 38:317–326, 1987.

[60] Mon-Williams M, Dijkerman HC. The use of vergence information in the programming of prehension. *Experimental Brain Research*, 128:578–582, 1999.

[61] Mosconi M W, Cody-Hazlett H, Poe M D, Gerig G, Gimpel-Smith R, Piven J. Longitudinal study of amygdala volume and joint attention in 2- to 4-year-old children with autism. *Arch Gen Psychiatry*, 66(5):509–516, 2009.

[62] Newschaffer CJ, Croen LA, Daniels J et al. The epidemiology of autism spectrum disorders. *Annu Rev Public Health*, 28, 2007.

[63] Ogata T and Sugano S. Emotional communication between humans and the autonomous robot which has the emotion model. *IEEE Int. Conf. on Robotics and Automation*, pages 3177–3182, 1999.

[64] Ekman P. Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology*, 2(10):726–735, 1965.

[65] Pelz J, Canosa R, Babcock J, Kucharczyk D, Silver A, and Konno D. Portable eyetracking: Astudy of eye movements. *in Proceedings of SPIE, Human Vision and Electronic Imaging, San Jose, CA, USA*, pages 566–582, 2000.

[66] Picardi L, Noris B, Schiavone G, Keller F, Von Hofsten C and Billard A G. *Poceedings of the 16th International Symposium on Robot and Human Interactive Communication*, 2007.

[67] Plaisant C, Druin A, Lathan C, Dakhane K, Edwards K, Vice J M and Montemayor J. A storytelling robot for pediatric rehabilitation. *in Proceedings of ASSETS*, 2000.

[68] Reidsma D, Jovanovic N, and Hofs D. Designing annotation tools based on properties annotation problems. *In Measuring Behavior, 5th Int. Conf. On Methods and Techniques in Behavioral Research*, 2005.

Gruppe Shiavone

[69] Robins B, Dautenhahn K, Boekhorst R, Billard A. Robotic aissistants in therapy and education of children with autism. *Univ Access in the Information Society*, 4(2), 2005.

[70] Robins B, Dautenhahn K, Dickerson P, Stribling P. Robotmediated joint attention in children with autism. *Interaction Studies*, 5, 2004.

[71] Robins B, Dautenhahn K, Nehaniv C L, Mirza N A, Francois D, Olsson L. Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. *Proc. IEEE RO-MAN*, pages 716–722, 2005.

[72] Robins B, Dickerson P, and Dautenhahn K. Robots as embodied beings - interactionally sensitive body movements in interactions among autistic children and a robot. *14th IEEE International Workshop on Robot and Human Interactive Communication*, 2005.

[73] Rotman G, Troje N F, Johansson R S, and Flanagan J R. Eye movements when observing predictable and unpredictable actions. *J Neurophysiol*, 96, 2006.

[74] Ruff H A, Rothbart M K. Attention in early development. *Oxford: Oxford University Press.*, 1996.

[75] Rutherford M D, Baron-Cohen S and Wheelwright S. Reading the mind in the voice: A study with normal adults and adults with asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, 32(3), 2002.

[76] Salter T, Michaud F, Dautenhahn K, Letourneau D, Caron S. Recognizing interaction from a robot's perspective. *In Robot and Human Interactive Communication IEEE International Workshop on*, pages 178–183, 2005.

[77] Schiavone G, Campolo D, Keller F, Guglielmelli E. Calibration of a multimodal head-mounted device for ecological assessment of social orienting behavior in children. *In Proceedings of IEEE/RSJ International Conference on Intelligent RObots and Systems, Hyatt Regency, St. Louis, USA*, 2009.

[78] Sigman M D, Kasari C, Kwon J H and Yirmiya N. Responses to the negative emotions of others by autistic, mentally retarded, and normal children. *Child Development*, 63:796–807, 1992.

Gueppa Schiavone

[79] Sparrow S S, Balla D, and Cicchetti D, Vineland. *Adaptive Behavior Scales, Expanded Edition. Circle Pines, MN: American Guidance Service.* 1984.

[80] Swettenham J, Baron-Cohen S, Charman T, Cox A, Baird G, Drew A et al. The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry*, 39, 1998.

[81] Haslwanter T. Mathematics of three-dimensional eye rotations. *Vision Res.*, (42):1053–1061, 1995.

[82] Tao Liu Inoue Y, Shibata K. A wearable sensor system for human motion analysis and humanoid robot control. *IEEE International Conference on Robotics and Biomimetics*, pages 43–48, 2006.

[83] Teitelbaum P, Teitelbaum O, Nye J, Fryman J, and Maurer RG. Movement analysis in infancy may be useful for early diagnosis of autism. *In Proceedings of the National Academy of Sciences, USA*, 95:13982–13987, 1988.

[84] Trevarthen C and Daniel S. Disorganised rhythm and synchrony: Early signs of autism and rett syndrome. *Brain and Development*, 27:25–34, 2005.

[85] Van Beek Y, Hopkins B, Hoeksma J, and Samson J F . Prematurity, posture and the development of looking behaviour during early communication. *Journal of Child Psychology and Psychiatry*, 35, 1994.

[86] Volkmar F R, Chawarska K, Klin A. Autism in infancy and early childhood. *In A. Kazdin (Ed.), Annual Review of Psychology*, 53:315–36, 2005.

[87] Welch G, Foxlin E. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 22:24–38, 2002.

[88] Werry I, Dautenhahn K, Ogden B, Harwin W. Can social interaction skills be taught by a social agent. the role of a robotic mediator in autism therapy. *in Beynon M, Nehaniv CL, Dautenhahn K (eds.), Cognitive Technology: Instruments*, 2001.

[89] Wijnroks L, Kalverboer A F. Early mother child interaction and quality of attachment in preterm infants. *In W. Koops, J. B. Hoeksma, and*

Gpppa Shiavone

D. van den Boom, Eds., Development of interaction and attachment: Traditional and non-traditional approaches. North Holland Publishers., 1997.

[90] Wijnroks L, van Veldhoven N . Individual differences in postural control and cognitive development in preterm infants. *Infant Behavior and Development*, 26, 2003.

[91] Winfield D, Dongheng Li, Babcock J, Parkhurst D J. Towards an open-hardware open-software toolkit for robust low-cost eye tracking in hci applications. *Iowa State University Human Computer Interaction Technical Report ISU-HCI*, April 2005.

[92] Young L, Sheena D. Survey of eye movement recoding methods. *Behavior research methods and instrumentation*, 7(5):397–429, 1975.

[93] Yuasa M, Yamaguchi O, Fukui K. Precise pupil contour detection based on minimizing the energy of pattern and edge. *IEICE Trans. Information and Systems*, E87(D):105–112, 2004.

[94] Yuille AL, Hallinan PW, and Cohen DS. Feature extraction from faces using deformable templates. *Int. J. Computer Vision.*, 8:99–111, 1992.

[95] Zhang C, Florencio D, Zhang Z. Why does phat work well in low noise, reverberative environments? *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, 2008.

Giuseppina Schiavone