Int J Soc Robot
DOI 10.1007/s12369-010-0080-9

Multimodal Ecological Technology: From Child's Social Behavior Assessment to Child-Robot Interaction Improvement

**Giuseppina Schiavone · Domenico Formica ·
Fabrizio Taffoni · Domenico Campolo ·
Eugenio Guglielmelli · Flavio Keller**

Accepted: 17 September 2010

Abstract The development of sensorimotor coordination in infancy is fundamental for regulating interactional dynamics with peers and adults. In this work we present a multimodal device to systematically assess children's orienting behavior in social situations. Technological choices are emphasized with respect to ecological requirements. Also ad-hoc calibration procedures are presented which are suitable to unstructured environments. Preliminary tests carried out at a local daycare with 12–36 months old typically developing infants prove the in-field usability of the proposed technology. Considerations on the future development of the device underscore the meaningful contribution that such platform can offer to child-robot interaction research.

Keywords Child-robot interactions · Autism · Ecological assessment · Multimodal signals · In-field calibration

1 Introduction

Social orienting refers to the ability of young children to spontaneously orient themselves to naturally occurring social stimuli in their environment. Failure to orient to social stimuli during early development may contribute to the later-emerging social, communicative and cognitive impairments [16].

The assessment of social behavior in infancy during daily-life activity is usually based on observational methodologies or on questionnaires distributed to parents and to day care teachers. Recently, approaches using expensive systems of multiple cameras placed in rooms of the house or in day care play rooms have been proposed. Issues related to the use of video surveillance systems are the huge amount of data to process off-line, and the high sensitivity of such systems to changes of lighting conditions. Moreover, recording with external cameras can present occlusion problems and it is often impossible to detect child's micro-behaviors, such as gaze shifting, or complex eye-hand coordination during social interaction. More sophisticated technologies for fine measurement of the baby's responsiveness to social stimuli require structured environments, typical of laboratory settings or clinical centers, and they are not suitable for investigations in ecological environments, such as the home and day care centers.

This work was partly supported by a grant from the European Union, TACT (Thought in Action), FP6-NEST/ADVENTURE program, contract no. 015636 and by the Academic Research Fund (AcRF) Tier1 (RG 40/09), Ministry of Education, Singapore.

G. Schiavone (✉)
Advanced Concepts Team, European Space Research and Technology Center, European Space Agency,
2200 AG Noordwijk, Netherlands
e-mail: giuseppina.schiavone@esa.int

D. Formica · F. Taffoni · E. Guglielmelli
Biomedical Robotics and Biomicrosystems Laboratory,
Università Campus Bio-Medico, 00128 Rome, Italy

D. Formica
e-mail: d.formica@unicampus.it

F. Taffoni
e-mail: f.taffoni@unicampus.it

E. Guglielmelli
e-mail: e.guglielmelli@unicampus.it

D. Campolo
School of Mechanical and Aerospace Engineering, Nanyang Technological University, 639798 Singapore, Singapore
e-mail: d.campolo@ntu.edu.sg

F. Keller
Developmental Neuroscience Laboratory, Università Campus Bio-Medico, 00128 Rome, Italy
e-mail: f.keller@unicampus.it

In this work we explore a novel approach based on wearable technology for monitoring and assessing social orienting behaviors in typically developing children from 12 to 36 months of age. The main requirement for the proposed technology is to be suitable for ecological environments, in order to monitor the child during daily-life, avoiding stress and uncomfortable situations, and to provide quantitative and objective evaluations on the child's early development of social and sensori-motor abilities.

The quantitative evaluations of early human orienting behavior could have also a great impact, not only on the assessment of child development, but also, on the design and implementation of interactional dynamics between humans, in particular children, and robots.

In the last thirty years several robots able to interact with children for educational and therapeutic purposes have been developed (see [12, 13, 39, 41, 54]).

Interactional dynamics defined by synchronized and coordinated responses between the robot and the child have strong influences on regulation and naturalness in interaction [4, 11, 20, 22, 40]. 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 [28, 40, 46]). 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 [44]. Motion tracking system mounted on the robot, as in [5], can analyze gross arm movements of the child that in turn trigger the robot to imitate the child. In some cases the robot, provided of 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 objects in the environment, as in [29].

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 face 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 technology that we propose will enable to overcome these limitations.

In this work we present a multi-modal head-mounted device, the Audio-Visuo-Vestibular (AVV)-Cap, designed for assessing social orienting behavior in typically developing infants from 12 to 36 years old. The AVV-Cap worn by the child would act as an extended interface of the robot, gathering and sending information on child behavior (e.g. direction of child's attention) to the robot, thus allowing to regulate the interaction as naturally as possible.

The developed technology has, thus, twofold challenges: from one side it aims at providing un-obtrusive instruments for monitoring and quantitatively assessing young children social orienting behaviors during daily-life; from the other side it will offer a novel prospective for designing and regulating child-robot interactional dynamics.

The paper is structured as follows. The principles of design of the AVV system, its components, its functioning and in-field calibration procedures are described in Sect. 2. Preliminary results from application of the AVV system in a day-care center with typically developing children are presented in Sect. 3. We discuss how these results and future developments of the AVV system could contribute to improve the design of autonomous social robots and close the loop in the child-robot interaction in Sect. 4.

2 Principles of Design

The design of the AVV system is based on three main principles [7]:

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 orienting behavior in social situations, that means, it has to be able to detect gaze, child's head kinematics, and localize sound stimuli, in particular human voices, with respect to the child.

Other constraints on the design of the AVV system are associated to the processing system. Current tools [23, 37] for coding videos and screening of recorded data result tedious and time consuming. Huge amount of data can arise from the collection of multimodal information. The aim of the AVV system is to provide a device able to acquire multimodal data and process them with automatic or at least semi-automatic modality. Therefore, simple and quiet robust algorithms are required both for low level processing of the signals provided by each sensor and for data integration processing.

The listed principles of design impose limits on the system accuracy, which, therefore, does not seek to be accurate as other state-of-the-art techniques for behavior assessment.

Although this limitation, the proposed method may enable researchers to gain new insights on the contents of social orienting behavior from the child's point of view.

2.1 Available Technology for Behavior Assessment

Current technologies for behavior assessment are more often unimodal, meaning that only one feature, among movement analysis, gaze tracking and sound localization ability, is monitored at time.

The most popular technology for gait and posture analysis are stereophotogrammetric systems, such as Motion Capture System from Vicon. They are based on optical devices and markers attached on the body. These systems have a high accuracy and they are useful to capture fast motion data for analysis. However, they have the drawback of being expensive and often require highly structured environments. Moreover, a lot of data processing work is necessary. Other technologies, mostly used for tracking orientations, are based on magnetic and inertial sensing. In the first case, an electronic compass allows to estimate heading, it solely relies on the geomagnetic field, i.e. it does not require any artificial source; measurements can be altered by ferromagnetic influence of surrounding objects. In the second case, miniaturized accelerometers and gyroscopes sense, respectively, acceleration and angular velocity. Magneto-inertial technologies appear to be very appealing since: they are highly unobstrusive due to the availability of miniaturized off-the shelf devices; can provide high accuracy in orientation tracking; are sourceless; do not require any structuring of the environment; are low-cost.

Several technologies, then, exist for recording eye movements. Magnetic scleral search coil [38] is the standard research technique providing the highest spatial and temporal resolution, and it can also detect torsional components, but it is limited to a clinical setting due to discomfort, limited recording time 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. The standard clinical method for recording of eye movements is the electro-oculography, which allows also measurements with closed eyes but it suffers from low resolution, drift, noise, poor vertical measurements, and motion and EMG artefacts that limit its use during locomotion and other natural activities [17]. The most popular methods are Infrared (IR) Oculography and Video oculography (VOG). IR oculography uses infrared lighting to illuminate the pupil and then extracts the eye orientation by triangulation of the IR spotlight reflections or other geometrical properties. 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. In VOG, the tracking of eye movements is achieved by the digital processing of video images of the eyes. The position of the pupil and of characteristic iris landmarks is detected by image processing algorithms and used to calculate the rotation angles of the eye. There exists external video trackers systems, such as, the Tobii eye tracker, which constrain considerably the experimental setup while being very sensitive to head movements. Head-mounted video trackers also exist [3, 18, 55], they do not suffer of head movements but usually they are not suitable for children, being too heavy and bulky. A new head-mounted camera, the Wearcam [35], 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 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. 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.

Lack of response to auditory stimuli and hearing loses are monitored with ABR (Auditory Brain Responses) audiometry [14] and OAE (otoacoustic emissions) tests [21], both the methodologies are implemented in specialized clinical center and hospitals and therefore they are not directly available to continuously monitor the development of hearing functions. Other quantitative methods for assessing child's ability to localize sound sources in ecological environment do not exist.

Despite the presented technologies, the AVVC device offers a multimodal approach to systematic assessment of early social attention skills.

2.2 Hardware Solution

Wearable technologies appear to be very appealing for our purpose since: being close to the body, allow to directly extract relevant information from a child-centered perspective, filtering environmental noise; can consolidate the functionality of multiple devices into a single, integrated system; do not modify the environment setting, unlike complex motion tracking system (i.e. stereophotogrammetric systems, which require to equip the observational room of several infrared cameras) and sophisticated sound localization systems, requiring multiple points of observation (array of microphones to be located in specific points of the observational room).

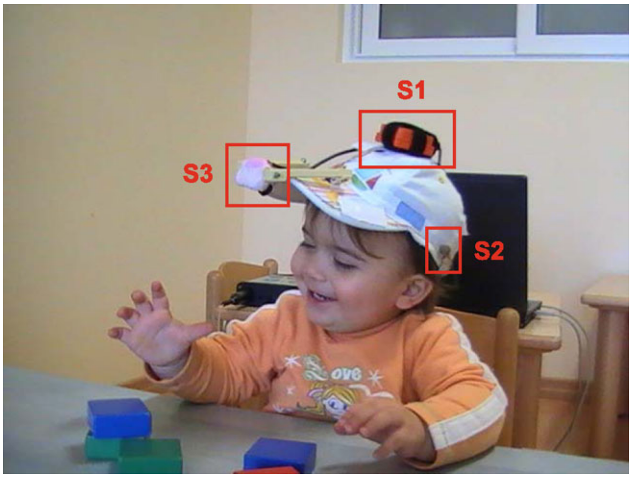

Fig. 1 Audio-Visuo-Vestibular Cap: S1, magneto-inertial sensor; S2, pair of microphones in binaural configuration; S3, mini-webcam

Thus, a colored cap, widely used garment among children, has been used as wearable support for the AVV system.

The most important senses involved in orienting behavior are vestibular (orienting of head position and sensory organs response-ability to the environment), hearing (how the ears and the brain work together to intake and organize auditory information), and sight (how the eyes and the brain work together to intake and organize visual information).

Thus three different sensors are mounted on the cap (Fig. 1): a magneto-inertial sensor (S1), positioned on the top of the cap with velcro bands; a pair of omni-directional microphones (S2), positioned at two opposite side of the cap, in correspondence of the ears of the child (binaural configuration); a mini-webcam (S3) 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.

Magneto-inertial technology [52] 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 low cost 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 AVVC is designed to be applied (daycare centers) electro-magnetic interferences are limited. The head tracker (MTx, Xsens Technologies B.V.) transduces head kinematics in 3 dimension at a frequency of 100 Hz, with a dynamic range of all angles in 3D and angular resolution of $<1^\circ$.

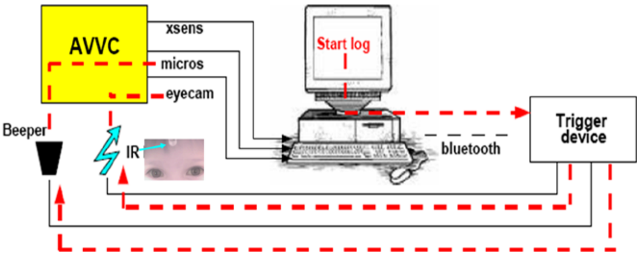

Fig. 2 Synchronization system: the trigger device receives, via bluetooth, a trigger signal from the PC and lights an Infrared led, under the visor of the cap, and generate a sound wave

Localization of the speaking voice is achieved by processing the signals from the two microphones (MKE 2-ew Gold, Lavalier). The audio signal acquisition frequency is set to 44100 Hz, so as to cover an interval of frequencies from 0 to the Nyquist frequency 22050 Hz, in which is contained the frequency range of human auditory perception (20–20000 Hz). Quantization is set to 16-bit per sample, which provides a dynamic interval of 96 dB, 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 us to choose them for our purpose. The web-cam, that we address as *eye-cam*, is composed of a $1/4$ in CMOS sensor, with a resolution of 640×480 and a frame rate of 30 frames per seconds. The rubber support which sustains the eye-cam is designed to hold the camera at a distance of about 10 cm from the face of the child, interfering as less as possible with his field of view, and enabling manual setting of the orientation of the eye-cam. The eye-cam sensor points to the face of the child and it has been provided with a mini-objective of 2.5 mm focal length (Model RE-025S) in order to cover a field of view (the diagonal Field Of View is 84° for an average of 57° and 71° vertically and horizontally respectively), corresponding a face dimension of 12×15 cm.

All the sensors on the AVVC send data to a PC via USB, which means that the AVVC 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 AVVC, 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 the presence of cables do not affect the child's acceptability of the device, as preliminary trials have demonstrated, we count to develop wireless communication modules in the future prototype, which allow the child to interact without constraints.

The synchronization system, for the multimodal data acquisition, is shown in Fig. 2 and consists of an external device, referred as trigger device, which lights an Infrared led, positioned under the visor of the cap, and generates a sound

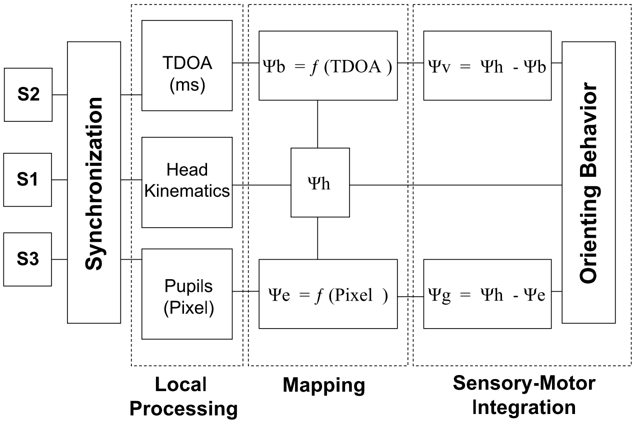

Fig. 3 Processing phases: signals detected from magneto-inertial sensor (S1), microphones (S2) and from the eye-cam (S3) are processed to estimate head kinematics (i.e. head azimuth, Ψ_h), TDOA (Time Difference Of Arrival) and pixels relative to the eyes coordinates; through the mapping process TDOA and pixels are transformed, respectively, in binaural azimuth (Ψ_b) and eye-in-head azimuth (Ψ_e); sensory-motor integration combines head azimuth with binaural azimuth, providing the location of voices (Ψ_v) close to the child, and with the eye-in-head azimuth, providing an estimation of the gaze (Ψ_g)

of duration 250 ms and frequency 4 kHz, at the beginning and at the end of the acquisition. As magneto-inertial sensor data starts or ends to be logged a trigger signal is sent via bluetooth to the trigger device that synchronizes the eye-cam and the microphone signals with the Infrared light and with produced sound.

2.3 Multimodal Data Processing

The AVVC works like an artificial audio-visuo-vestibular system. Although in the present work only the horizontal component of the orienting behavior has been considered, the AVVC worn by the child is able to localize voices close to him/her, to sense child's head orientation, angular velocity and acceleration and to estimate gaze direction by detecting eye orientation.

In the scheme in Fig. 3 the phases of the processing are shown. After the identification of the synchronization points, signals detected by the three sensors are separately processed to extract local features. Through a mapping process and a sensory motor integration process, local features are transformed in relevant behavioral features.

AVVC processing is made off-line. Although, in the present work, on-line data processing was not required for children behavioral assessment, we intend to provide the AVVC with real-time signal processing, so that the acquired information can be directly used for future application in child-robot interaction.

Local Processing In the first phase of the AVVC processing system, each sense modality is processed independently, thus to obtain a peculiar feature from each signal: Time Difference Of Arrival (TDOA) is estimated from the microphones signal, head kinematics from the magneto-inertial sensor and pixels relative to the eye coordinates from eye-cam recordings.

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. The TDOA is the binaural cue on which most of the state-of-the-art sound localization systems relies [2]. Among several techniques [1, 6, 10, 51] used for the estimation of TDOA, τ, the Phase Transform (PHAT) generalized cross-correlation technique has been selected [24]:

$$\tau = \arg\max_{\beta} \int_{-\infty}^{\infty} W(\omega) X_1(\omega)\overline{X_2(\omega)} e^{j\omega\beta}\, d\omega \qquad (1)$$

where, $X_1(\omega)$ and $X_2(\omega)$ are the Fourier Transform of the signal at each microphone, $X_1(\omega)\overline{X_2(\omega)}$ is the cross-power spectrum, and $W(\omega)$ is a normalizion factor defined as:

$$W(\omega) = \frac{1}{|X_1(\omega)\overline{X_2(\omega)}|} \qquad (2)$$

This approach allows to only take the phase of $X_i(\omega)$ into account, narrowing the wide maxima caused by the correlation between the received signals. Although PHAT algorithm is a suboptimal solution, because each frequency bin of the spectrum contributes the same amount to the final correlation, and this makes the system less robust to noise, PHAT technique has emerged as the standard in sound localization. It does not require any knowledge about the spectrum of the microphone dependent noises (prior knowledge of the noise spectra is usually difficult to obtain, especially in real, reverberant environments), it results simple to be implemented, and shows good performances in reverberative environments [32].

TDOA/angle mapping is required to determine the angular position of a sound source with respect to the head once the TDOA is estimated. The mapping is achieved with the calibration procedure, described in details in the next paragraph.

Information provided by the magneto-inertial sensor is used to estimate head kinematics: the inertial sensors can provide esteem of the head tilt with respect to gravity as well as angular velocity of the head; the magnetic sensors, used as a compass, can determine the amount of head rotation on the horizontal plane, also referred as head azimuth, Ψ_h.

Eye movements are detected by identifying the pupils position in frames recorded by the eye-cam. 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. Pupil coordinates expressed in pixels units need to be transformed in eye angular positions and this is achieved by the calibration procedure.

Mapping and Sensory-Motor Integration Two calibration procedures are required to calibrate the AVVC device: the first one is the *vestibulo-auditory calibration* which relates the estimated TDOA with the binaural azimuth (Ψ_b), that is the angular position of the sound source in the horizontal plane; the second one is the *vestibulo-ocular calibration*, which allows to derive an angle of orientation of the eyes in the horizontal plane, the eye azimuth (Ψ_e), from the pixel coordinates of the pupil. Both the vestibulo-auditory calibration and the vestibulo-ocular calibration are processes of the mapping phase (Fig. 3).

- *Vestibulo-Auditory Calibration Procedure*

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. 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. The proposed vestibulo-auditory calibration is properly designed for those environments. 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 has been conducted for comparing the standard procedure with the proposed ecological one. Two microphones are mounted on a dummy head. In the first procedure, P1, the sound source (speaker with experimenter voice) was located in a set of eleven locations (−90°, −70°, −50°, −30°, −10°, 0°, 10°, 30°, 50°, 70°, 90°) at a distance of 1.2 m respect to the head, and the head was kept in the primary position, at 0°. In the second procedure, P2, the sound source stayed in the primary position while the dummy head was 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.

TDOAs (τ) and angular positions, Ψ, are correlated with a linear relation:

$$\Psi = G_b \cdot \tau + O_b \tag{3}$$

where G_b is the slope or gain of the curve, expressed in °/s, and O_b is the offset, expressed in °. All the trials show R-square coefficient values >0.93, confirming that angular position and TDOA are highly correlated with a linear fitting. Figure 4 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.

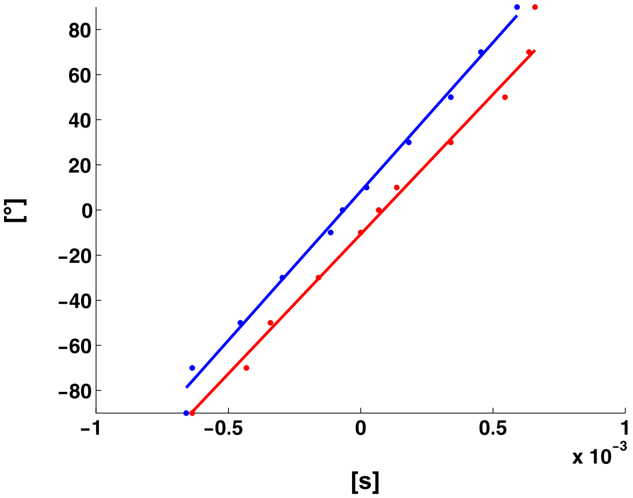

Fig. 4 (Color online) TDOA/Angle map obtained in laboratory: *blue dots* refer to calibration obtained when the dummy head was kept in the primary position and the sound source moved in different positions; *red dots* refer to calibration obtained when the dummy head was oriented in different directions and the sound source stayed in the primary position; *blue* and *red solid lines* are the fitting curves

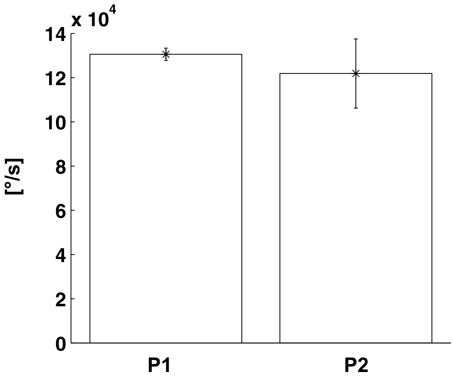

Fig. 5 Mean and Standard deviations of the gain of the curves, estimated on ten trials for each calibration procedure (P1 and P2)

Comparison between the procedures is achieved by considering the slope of the curves. Mean and standard deviation of the gains for P1 and P2 are shown in Fig. 5. 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.

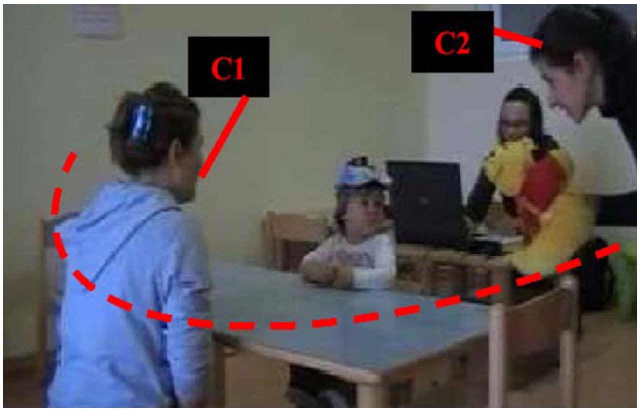

Fig. 6 Vestibulo-auditory calibration procedure in a room of the day-care center: C1, caregiver speaking in front of the child, C2, caregiver moving a toy in a semicircle in front of the child.

During in-field calibration procedure, conducted in a room of a day-care center, the child is sitting on a chair at the long side of a desk while wearing the AVVC device (Fig. 6). 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 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 (obtained as in (3)), which is the location of the sound source, in the horizontal plane, respect to the child's head.

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:

$$\Psi_v = \Psi_b - \Psi_h \tag{4}$$

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. The calibration curve relative to a 24-months-old child is shown in Fig. 7 and the probability density function (PDF) of the estimated sound source direction during the calibration procedure is shown in Fig. 8. The PDF of the sound source, computed with a normal kernel smoother, has been fitted using a Gaussian model, with 95% confidence bounds. The gain of the calibration curves obtained with children does not differs statistically from the values obtained in laboratory conditions, differently from the offset of the curves which depends on the positioning of the cap on the child's head.

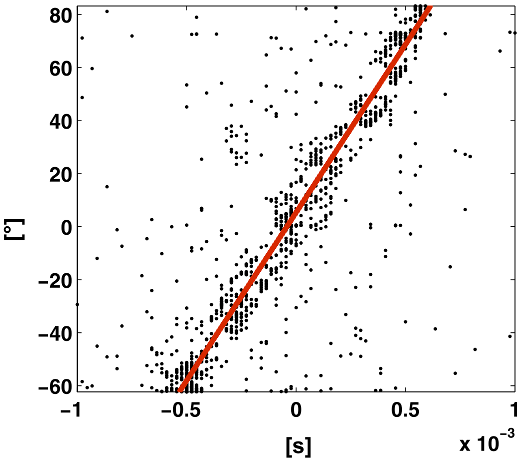

Fig. 7 (Color online) Vestibulo-auditory calibration curve for a 24-months-old child: the *black dots* are the raw data, the *red line* is the fitted curve (R-square coefficient = 0.95)

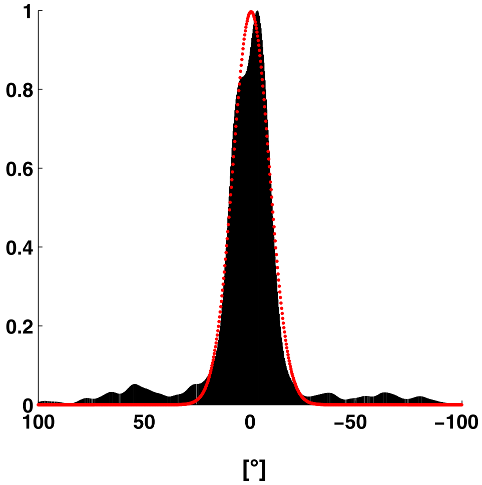

Fig. 8 (Color online) Probability Density Function (PDF) of the estimated sound source direction, Ψ_v, during the calibration procedure in Fig. 7: the *black area* represents the raw data, the *red line* the gaussian fitting of the PDF (mean = −0.5°; standard deviation = 12°; R-square coefficient = 0.98)

Results, reported in detailed in [47], 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 and reliable; (iii) although many factors (such as the noisy and un-controlled environment, the head and trunk movement of the caregiver speaking in front of the child) can affect the AVVC accuracy in localizing sound source, it can be considered satisfactory for experimental conditions in which two speakers are located in angular positions higher than 18° respect to the child.

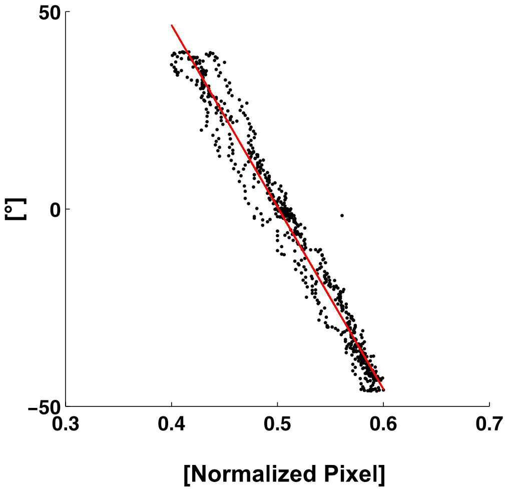

Fig. 9 (Color online) Vestibulo-ocular calibration curve (in-lab evaluation): the *black dots* are the raw data, the *red line* is the fitted curve (R-square coefficient = 0.97)

– *Vestibulo-Ocular Calibration Procedure*

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

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 [9]. 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, P_x. Linear fitting can be applied to the calibration curve to extract gain, G_e, and offset, O_e (see Fig. 9). This allows expressing the eye orientation, Ψ_e, in degrees rather than in (normalized) pixels:

$$\Psi_e = G_e P_x + O_e \tag{5}$$

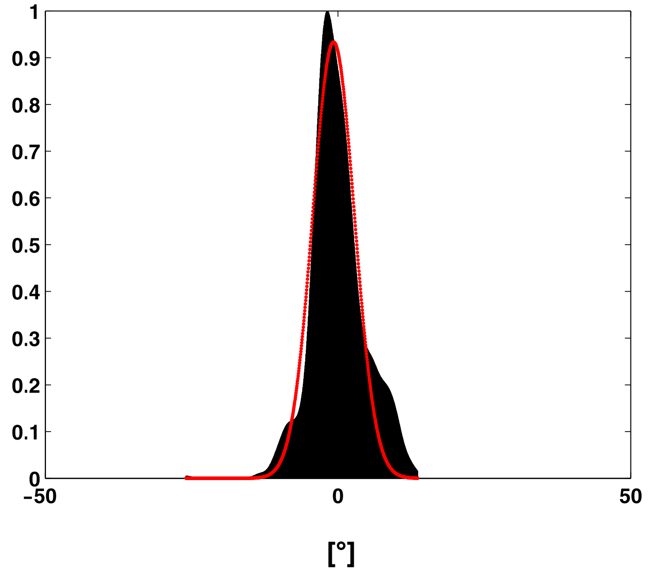

Fig. 10 (Color online) Probability Density Function (PDF) of the estimated gaze direction, Ψ_g, during the calibration procedure in Fig. 9: the *black area* represents the raw data, the *red line* the gaussian fitting of the PDF (mean = −0.7°; standard deviation = 5°; R-square coefficient = 0.95)

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:

$$\Psi_g = \Psi_e - \Psi_h \tag{6}$$

The probability density function of the estimated gaze direction, has been computed with a normal kernel smoother and fitted using a Gaussian model, with 95% confidence bounds (Fig. 10).

The calibration procedure has been tested in-lab on 5 subjects, between 25 and 27 years old. Figure 11 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.

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 ±40° in the horizontal plane. A t-test applied to the estimated gaze over time (see Table 1) 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. The standard deviation mean is around 5°, which is quite high value if compared to accuracy of state-of-the-art eye-tracking methods. However, because of the working conditions (i.e. ecological environments where the cooperation of

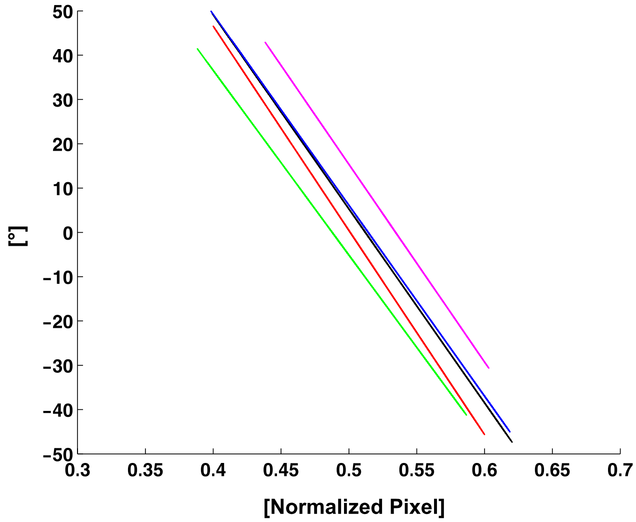

Fig. 11 (Color online) Vestibulo-ocular calibration curves: *black line* refers to subject 1, *blue line* refers to subject 2, *red line* refers to subject 3, *green line* refers to subject 4, *magenta line* refers to subject 5

Table 1 Probability density function mean and standard deviation (STD) of the gaze estimated for each subject, expressed in degree; p-value of the t-test performed on the estimated gaze

Subject	Mean	STD	p-value
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

the user cannot be expected) we do not seek to be provide an highly accurate system.

The proposed procedure is based on a human physiological reflex which matures and develops in infants very early, within the first two months of life [34] and this allows to obtain a pixel/angle mapping by exploiting solely head movement.

Further evaluations need to be performed to estimate the sensitivity of the calibration when the procedure is not perfectly executed. Although only preliminary tests have been made on adults, the vestibulo-ocular calibration can be easily configured as a game (e.g. we observed 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), therefore the next step will be to test it with children.

Orienting Behavior The last phase of the AVVC data processing system consists in the reconstruction of the child orienting behavior, integrating all the relevant features: location of the speaking voice, head orientation of the child and his/her gaze. Other information can be also available, such as velocity and acceleration of head orientation, velocity of gaze shifting, and blinking. Several micro-behaviors can be quantitative assessed: 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.

Fig. 12 Experimental Scenario. One caregiver to the left (C_{left}), another to the right (C_{right}) of the child, speaking and positioning colored blocks on the table in front of the child. On the *right corner*, a snapshot of the child's eyes with three possible positions for the eye (left, L, center, C, right, R), empirically defined

3 AVVC's Application in a Day-Care Center

The AVVC has been used in the day-care center *La Primavera del Campus* (Rome) for assessing children's orienting behaviors in a social protocol.

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

3.1 Social Protocol

As shown in Fig. 12, the child, wearing the AVVC, sits on a chair, and two caregivers sit next to him/her, each one at the opposite side of a desk (C_{left} and C_{right}). In a first phase (speaking phase) each caregiver, alternately, explains to the child what he is going to do. In a second phase (speaking and acting phase) each caregiver puts an object (three colored blocks, positioned one at a time) on the table in front of the child, far enough from the child avoiding he/she can reach it, describing the action (e.g. "I'm putting a red block").

The aim of the protocol is to investigate shifts in attention of the child from the object to the caregiver and viceversa, in presence or absence of speech. The child is expected to

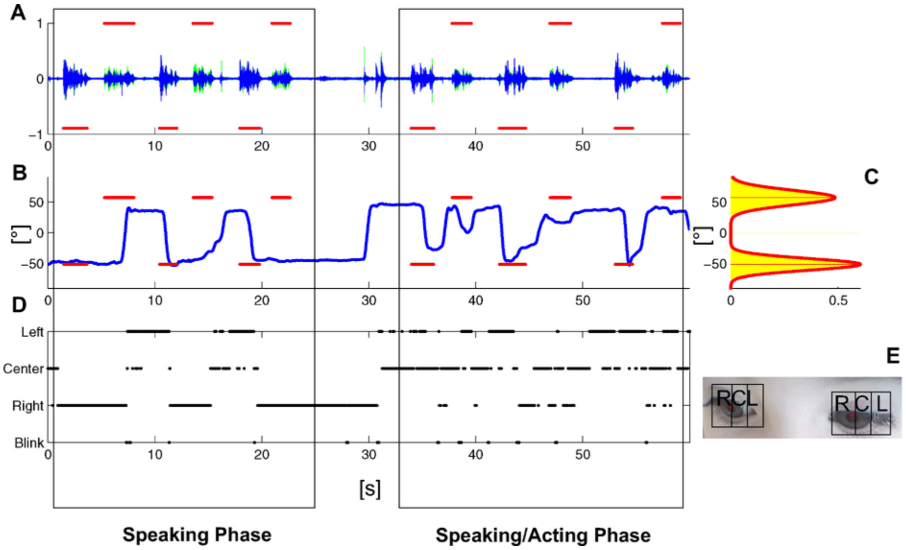

Fig. 13 (Color online) Behavior representation over time. (**A**): stereo signal from microphones, *red lines*, sound source activation; (**B**): *blue line*, head azimuth (negative angles correspond to rotation to the right, positive angle to rotation to the left), *red lines* sound sources location; (**C**): probability density distribution of the estimated sound sources location; (**D**): eye orientation (left, center, right) and blinking; (**E**): snapshot with the three eye positions (left, center, right)

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 objects on the table.

Typically developing children, between 20 and 36 months old, participated to the experimental sessions after the approval of informative consent by their parents. Vestibulo-auditory calibration procedure was performed for each children 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. 12).

In this work data of one child of 33 months old are shown in Fig. 13 as an example of the information provided by the AVVC. The child's behavior can be reconstructed and quantitatively assessed with a semiautomatic procedure. The AVVC processing system allows to: identify when somebody is speaking (Fig. 13A); estimate where the person who is speaking is located (Fig. 13B and 13C); detect where the child's head is oriented (Fig. 13B) during the two phases of the protocol; detect if the child's eyes are in a central position or oriented to the left or to the right (Fig. 13D and 13E).

These studies are still ongoing, but preliminary evaluation of the results have showed how the device respects the ecological approach: children can wear the AVVC without discomfort, they enjoy both the calibration procedure and the social protocol, thus spontaneous and natural behaviors can be observed.

4 Conclusion and Future Directions

This paper has given an overview of the major goals underlying the design and the use of the AVVC as a unobtrusive, wearable technology for assessing very young children sensory-motor coordination during social situations. Focus on children from 12 to 36 months comes from evidences in literature that there are clearly observable behaviors which are important predictors of neuro-developmental in pre-verbal children [50].

The behavior representation obtained with the AVVC results a complementary tool to the standard video scoring used to analyze children's behavior. It has the advantage of being objective, do not rely on the subjective qualitative judgement of the experimenter. Moreover, the multi-modal 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. Although in this work only the horizontal component of the orienting behavior has been considered, we intent to extend the analysis to 3D orienting behaviors, thus that others micro-behaviors can be inferred from the AVVC data.

Compared to traditional laboratory observational instruments, the proposed device is not highly accurate but its performance are satisfactory for the application in uncontrolled, ecological environments.

Results of such studies show promise for assessing social orienting behaviors of typically developing children.

Further work on the AVVC will aim at developing a wireless communication module and a real-time signal processing system.

Wireless connections will enable the child to move around in the room without constrains, thus allowing to implement new protocols for assessing social behavior. Moreover, the absence of cables will ensure a better suitability for future applications on infants affected by neuro-developmental disorders, who show evident impairments in social orienting. Lack of social orienting to other's faces [48]; difficulty of the child in shifting attention to speech sounds, or to being called by his/her name

[8, 30, 43]; unusual patterns of eye-contact and gaze directionality [5] are, in fact, early characteristic symptoms of neuro-developmental disorders such Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD) [15]. Tests with a wireless version of the AVVC on children with autism will definitively prove/disprove the effectiveness of the device as a complementary instrument to standard diagnostic measures, such as the ADOS-G (the Autism Diagnostic Observation Schedule-Generic) [31].

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.

The combination of 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 [5], Infanoid [27], Keepon [26], ESRA [46] and Kaspar [42]). Moreover, this integrated approach could provide quantitative evaluation of child-robot interaction [19, 33, 45, 53].

The use of wearable technology was already explored in [36] 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 [25] wearable sensors and stationary sensors positioned in a room are used to improve the behavior of an interactive guide robot; in [20] and in [49] 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 AVVC.

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 AVVC. 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.

Acknowledgements The authors would like to thank the personnel of the daycare *La Primavera del Campus* for helping with the preliminary tests and Gunilla Stenberg for the design of the experimental scenario.

References

1. Aarabi P, Zaky S (2000) Iterative spatial probability based sound localization. In: Proceedings of the 4th world multiconference on circuits, systems, computers, and communications, Athens, Greece, July 2000
2. Aarabi P (1998) Multi-sense artificial awareness. MASc Thesis, Department of Electrical and Computer Engineering, University of Toronto, Ontario, Canada
3. Allison RS, Eiyenman M, Cheung BSK (1996) Combined head and eye. Tracking system for dynamic testing of the vestibular system. IEEE Trans Biomed Eng 43(11):1073–1082
4. Bates J (1994) The role of emotion in believable agents. Commun ACM 37(7):122–125
5. Billard A, Robins B, Dautenhahn K, Nadel J (2006) Building robota, a mini-humanoid robot for the rehabilitation of children with autism. RESNA Assist Technol J 19:37–49
6. Brandstein MS, Silverman H (1997) 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
7. Campolo D, Laschi C, Keller F, Guglielmelli E (2007) A mechatronic platform for early diagnosis of neurodevelopmental disorders. RSJ Adv Robot J 21(10):1131–1150
8. Ceponiene R, Lepistö T, Shestakova A, Vanhala R, Alku P, Näätänen R, Yaguchi K (2003) Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. Proc Nat Acad Sci USA 100(9):5567–5572
9. Crawford JD, Vilis T (1991) Axes of eye rotation and Listing's law during rotations of the head. J Neurophysiol 65(3):407–423
10. Datum MS, Palmieri F, Moise A (1996) An artificial neural network for sound localization using binaural cues. J Acoust Soc Am 100(1):372–383
11. Dautenhahn K, Werry I (2000) Issues of robot-human interaction dynamics in the rehabilitation of children with autism
12. Dautenhahn K (1999) Robots as social actors: Aurora and the case of autism. In: Proceedings CT 99, the third international cognitive technology conference, August, San Francisco, pp 359–374
13. Dautenhahn K, Werry I (2004) Towards interactive robots in autism therapy. Pragmat Cogn 12:1–35
14. Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S (1997) Health. A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment. Health Technol Assess 1:1–176
15. Dawson G, Meltzoff AN, Osterling J, Rinaldi J, Brown E (1998) Children with autism fail to orient to naturally occurring social stimuli. J Autism Dev Disord 28:479–485
16. Dawson G, Toth K, Abbott R, Osterling J, Munson J, Estes A, Liaw J (2004) Early social attention impairments in autism: social orienting, joint attention, and attention to distress. Dev Psychol 40(2):271–283
17. DiScenna A, Das V, Zivotofsky A, Seidman S, Leigh RJ (1995) Evaluation of a video tracking device for measurement of horizontal and vertical eye rotations during locomotion. J Neurosci Methods 58(1–2):89–94
18. Dongheng Li, Babcock J, Parkhurst DJ (2006) OpenEyes: a low-cost head-mounted eye-tracking solution. In: Proceedings of the 2006 symposium on eye tracking research and application, San Diego, California, pp 95–100
19. Kanda T, Ishiguro H, Ishida T (2001) Psychological analysis on human-robot interaction. In: Int conf on robotics and automation (ICRA 2001), pp 4166–4173

20. Kanda T, Ishiguro H, Imai M, Ono T (2003) Body movement analysis of human-robot interaction. In: International joint conference on artificial intelligence, IJCAI, pp 177–182
21. Kemp DT (1978) Stimulated acoustic emissions from within the human auditory system. J Acoust Soc Am 64:1386–1391
22. Kendon A (1970) Movement coordination in social interaction: some examples described. Acta Psychol 32:100–125
23. Kipp M (2001) Anvil-a generic annotation tool for multimodal dialogue. In: Proceedings of the 7th European conference on speech communication and technology, Eurospeech, September 2001, pp 1367–1370
24. Knapp CH, Carter G (1976) The generalized correlation method for estimation of time delay. IEEE Trans Acoust Speech Signal Process 24(4):320–327
25. Koide Y, Kanda T, Sumi Y, Kogure K, Ishiguro H (2004) An approach to integrating an interactive guide robot with ubiquitous sensors. In: IEEE/RSJ international conference on intelligent robots and systems, IROS, 28 September–2 October 2004, Sendai, Japan
26. Kozima H, Nakagawa C, Yasuda Y, Kosugi D (2004) A toy-like robot in the playroom for children with developmental disorder. In: Proceedings of the international conference on development and learning, San Diego, CA
27. Kozima H (2002) Infanoid: a babybot that explores the social environment. In: Dautenhahn K et al (eds) Socially intelligent agent. Kluwer Academic, Dordrecht, pp 157–164
28. Kozima H, Nakagawa C (2007) Longitudinal child-robot interaction at preschool. In: AAAI spring symposium on multidisciplinary collaboration for socially assistive robotics, March 2007, pp 27–32
29. Kozima H, Nakagawa C (2007) Interactive robots as facilitators of children's social development. In: Lazinica A (ed) Mobile robots: towards new applications, Vienna: advanced robotic systems, pp 269–286
30. Lepisto T, Kujala T, Vanhala R, Alku P, Huotilainen M, Näätänen R (2005) The discrimination of and orienting to speech and non-speech sounds in children with autism. Brain Res 1066:147
31. Lord C, Rutter ML, Goode S, Heemsbergen J (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord 19:185–212
32. Mungamuru B, Aarabi P (2004) Enhanced sound localization. IEEE Trans Syst Man Cybern B: Cybern 34(3):1526–1540
33. Ogata T, Sugano S (1999) Emotional communication between humans and the autonomous robot which has the emotion model. In: IEEE int conf on robotics and automation (ICRA'99), pp 3177–3182
34. Ornitz EM, Kaplan AR, Westlake JR (1985) Developmet of the vestibule-ocular reflex from infancy to adulthood. Acta Otolaryngol 100:180–193
35. Picardi L, Noris B, Schiavone G, Keller F, Von Hofsten C, Billard AG (2007) In: RO-MAN'07: proceedings of the 16th international symposium on robot and human interactive communication
36. Plaisant C, Druin A, Lathan C, Dakhane K, Edwards K, Vice JM, Montemayor J (2000) A storytelling robot for pediatric rehabilitation. In: Proceedings of ASSETS'2000, Washington DC, Nov 2000. ACM, New York
37. Reidsma D, Jovanovic N, Hofs D (2005) Designing annotation tools based on properties annotation problems. In: Measuring behavior 2005, 5th int conf on methods and techniques in behavioral research, 30 August–2 September 2005
38. Robinson DA (1963) A method of measuring eye movements using a scleral search coil in a magnetic field. IEEE Trans Biomed Electron BME 10:137–145
39. Robins B, Dautenhahn K, Dickerson P, Stribling P (2004) Robot mediated joint attention in children with autism. Interact Stud 5:161–198
40. Robins B, Dautenhahn K, Nehaniv CL, Mirza NA, Francois D, Olsson L (2005) Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: lessons learnt from an exploratory study. In: Proc. IEEE RO-MAN'05. IEEE Press, New York, pp 716–722
41. Robins B, Dautenhahn K, te Boekhorst R, Billard A (2005) Robotic assistants in therapy and education of children with autism. Univ Access Inform Soc 4(2):105–120
42. Robins B, Dickerson P, Dautenhahn K (2005) Robots as embodied beings—interactionally sensitive body movements in interactions among autistic children and a robot. In: 14th IEEE international workshop on robot and human interactive communication, ROMAN05, Nashville, USA
43. Rutherford MD, Baron-Cohen S, Wheelwright S (2002) Reading the mind in the voice: a study with normal adults and adults with asperger syndrome and high functioning autism. J Autism Dev Disord 32(3):189–194
44. Salter T, Michaud F, Dautenhahn K, Letourneau D, Caron S (2005) Recognizing interaction from a robot's perspective. In: IEEE international workshop on robot and human interactive communication, ROMAN, 2005, pp 178–183
45. Salter T, Michaud F, Larouche H (2010) How wild is wild? A taxonomy to characterize the 'wildness' of child-robot interaction. Int J Soc Robot. Published online: 11 August 2010
46. Scassellati B (2005) Quantitative metrics of social response for autism diagnosis. In: Proc. ROMAN
47. Schiavone G, Campolo D, Keller F, Guglielmelli E (2009) Calibration of a multimodal head-mounted device for ecological assessment of social orienting behavior in children. In: The 2009 IEEE/RSJ international conference on intelligent robots and systems, Hyatt Regency, St Louis, USA, Oct 11–15
48. Sigman MD, Kasari C, Kwon JH, Yirmiya N (1992) Responses to the negative emotions of others by autistic, mentally retarded, and normal children. Child Dev 63:796–807
49. Liu T, Inoue Y, Shibata K (2006) A wearable sensor system for human motion analysis and humanoid robot control. In: IEEE international conference on robotics and biomimetics, ROBIO'06, 17–20 Dec 2006, pp 43–48
50. Volkmar FR, Chawarska K, Klin A (2005) Autism in infancy and early childhood. Ann Rev Psychol 56:315–336
51. Watanabe H, Suzuki M, Nagai N, Miki N (1989) A method for maximum likelihood bearing estimation without nonlinear maximization. Trans Inst Electron Inf Commun Eng J72A(8):303–308
52. Welch G, Foxlin E (2002) Motion tracking: no silver bullet, but a respectable arsenal, motion tracking survey. IEEE Comput Graph Appl 22(6):24–38
53. Welch KC, Lahiri U, Warren Z, Sarkar N (2010) An approach to the design of socially acceptable robots for children with autism spectrum disorders. Int J Soc Robot. Published online: 07 July 2010
54. Werry I, Dautenhahn K, Ogden B, Harwin W (2001) 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
55. Winfield D, Li D, Babcock J, Parkhurst DJ (2005) 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

Giuseppina Schiavone received her B.Sc. (2004), M.Sc. (2006) and the Ph.D. (2010) degrees in Biomedical Engineering from Università Campus Bio-Medico, Rome, Italy. Currently, she is Young Graduate Trainee (YGT) in the area of neuroscience at Advanced Concepts Team, European Space Research and Technology Center, European Space Agency (ESA), Noordwijk, Netherlands. Her research interests:

computational neuroscience (models of selective visual attention based on spiking neural networks); social neuroscience and social robotics; applied behavior analysis (wearable technologies and sensor fusion algorithms applied to neuro-developmental engineering).

Domenico Formica received his B.Sc. (2002) and M.Sc. (2004) and Ph.D. degrees in biomedical engineering (2008) from the Università Campus Bio-Medico, Rome, Italy. Since May 2008, he has been a Postdoctoral Fellow with the Laboratory of Biomedical Robotics and Biomicrosystems, Università Campus Bio-Medico. His current research interests include the areas of mechatronic technologies for the study of human motor control, with particular attention to neuro-development, quantitative algorithms for clinical assessment of patients with neuromuscular disorders, and novel robotic devices for rehabilitation motor therapy after neurological injury.

Fabrizio Taffoni received his B.Sc. (2003), M.Sc. (2005) and the Ph.D. (2009) degrees in Biomedical Engineering from Università Campus Bio-Medico, Rome, Italy. Since March 2009, he has been a Postdoctoral Fellow with the Laboratory of Biomedical Robotics and Biomicrosystems, Università Campus Bio-Medico. Currently, he is involved in the design and development of mechatronic platforms for human and animal (capuchin monkeys) behavioral analysis. His research interests are in fields of robotics and neuroscience.

Domenico Campolo received his B.Sc. degree (1998) from University of Pisa, Italy; his M.Sc. (1999) and his Ph.D. (2002) degrees in Micro-Engineering from Scuola Superiore Sant'Anna, Pisa. In 2000–2003, he was at UC Berkeley (USA) as a Visiting Scholar and, after 2002, as a Postdoc working on the MFI (Micromechanical Flying Insect) project. From 2003 to 2009, he was a Postdoc and, later, a Research Assistant Professor at the Laboratory of Biomedical Robotics and Biomicrosystems of Università Campus Bio-Medico in Rome, Italy. Since 2009, he is Assistant Professor at the School of Mechanical and Aerospace Engineering, Nanyang Technological University (NTU) in Singapore. His research interests include mechatronic technologies with application to phenomics and neuro-developmental engineering; Development of Robotic platforms for Neuroscience; human-robot interaction. Biomimetic micro-robotics, with application to the biomedical field (nonlinear control issues in flapping-wing flight; design, fabrication and development of biologically inspired actuators and sensors).

Eugenio Guglielmelli received his M.Sc. (1991) in Electronics Engineering, Ph.D. (1995) in Biomedical Robotics from the University of Pisa, Italy. He is currently Full Professor of Bioengineering at Università Campus Bio-Medico (Rome, Italy) where he serves as the Head of the Laboratory of Biomedical Robotics and Biomicrosystems. His main research interests are in the fields of human-centered robotics, biomechatronic robot design, rehabilitation and assistive robotics, and neurorobotics. He serves as Associate Editor of the IEEE Transactions on Robotics, the IEEE Robotics and Automation Magazine, and the CRC International Journal of Applied Bionics and Biomechanics.

Flavio Keller received his M.D. (1982) from the University of Zurich. He received a Postdoctoral training in neuropharmacology and developmental neurobiology at the University of Zurich, and in molecular neurobiology with Eric R. Kandel at Columbia University. Full professor of Physiology and Neurophysiology both at the Medical and the Bioengineering faculty of Università Campus Bio-Medico, Rome, and founder and Head of the Laboratory of Developmental Neuroscience. He is also an Associate Researcher at the Perception-Movement-Action Research Centre (PMARC) of the University of Edinburgh. He is author of numerous publications in major scientific journals, including Science, Neuron and other neuroscience journals.