Recognition of Affective State for Austist from Stereotyped Gestures

Marcos Yuzuru O. Camada[1], Diego Stéfano[2], Jés J. F. Cerqueira[2], Antonio Marcus N. Lima[3], André Gustavo S. Conceição[2] and Augusto C. P. L. da Costa[2]

[1]*IFbaiano, Rua Baro de Camaçari, 118, Centro, 48110-000, Catu, Bahia, Brazil*

[2]*Electrical Engineering Department, Federal University of Bahia, Rua Arstides Novis, 02, Federação, 40210-630, Salvador, Bahia, Brazil*

[3]*Electrical Engineering Department, Federal University of Campina Grande, Aprigio Veloso, 882, Bodocongo, 58109-100, Campina Grande, Paraíba, Brazil*

Keywords: HRI, HMM, Fuzzy Inference System, Autism, Stereotyped Gesture, Assistive Robotic.

Abstract: Autists may exhibit difficulty in interaction (social and communication) with others and also stereotyped gestures. Thus, autists have difficulty to recognize and to express emotions. Human-Robot Interaction (**HRI**) researches have contributed with robotic devices able to be mediator among autist, therapists and parents. The stereotyped behaviors of these individuals are due to their defense mechanism from of their hypersensitivity. The affective state of a person can be quantify from poses and gestures. This paper proposes a system is able to infer the defense level of autists from their stereotyped gestures. This system is part of the socially assistive robot project called **HiBot**. The proposed system consist of two cognitive subsystems: Hidden Markov Models (**HMM**), in order to determine the stereotyped gesture, and Fuzzy Inference System (**FIS**), to infer activation level of these gestures. The results of these simulations show this approach is able to infer the defense level for an task or the presence of the robot.

1 INTRODUCTION

Autism Spectrum Disorder (**ASD**) belongs to the group of pervasive developmental disorders which is characterized by deficits in social interaction, communication, and stereotyped (or unusual) behaviors (Levy et al., 2009). The autist has difficulty to express and to recognize social cues, as emotion through facial and body expression and gaze eyes. The major treatments for **ASD** rely on psychiatric medications, therapies and behavioral analysis (or both).

Both software (Parsons et al., 2004) and robotic devices (Goodrich et al., 2012) have been developed to aid the treatment of autism. The design of such robotic devices is naturally demand for multidisciplinary teams, because they may involve different fields of health, engineering and computing.

The use of robots as social partners for autistic children has already been proposed (Dautenhahn, 2003; Goodrich et al., 2012) within the field of Human Robot Interface (**HRI**). Theses devices can behave as mediators among autists, therapist, parents. The affective state recognition of a person is essential for a social partner robot.

A human can express himself through verbal and non-verbal, such as face, body and voice (Zeng et al., 2009). Researches have focused on face expression, but studies have also shown body cues are as powerful as facial cues in conveying and recognizing of emotions. The quantification of the human affective state from the poses and gestures (Camurri et al., 2003) has been proposed as a way to recognize emotions. In addition, (Kuhn, 1999) assumes **stereotyped gestures** are defense mechanisms of autists due to their hypersensitivity.

For these reasons, we propose a system to affective state recognition from the **stereotyped gestures** of autists in this paper. The gestures are recognized using Hidden Markov Models (**HMM**). A Fuzzy Inference System (**FIS**) is used to infer the affective state (defense level) of the autist from the gesture recognized and kinetic of joint groups.

This paper is organized as follows. In Section 2, we define affective state, body expression, and the relationship between autism. Classification and inference tools are described in Section 3. Details of the proposed system architecture is presented in the Section 4. Experiments of proposed model and their results are discussed in Section 5. Finally, a general discussions about the contributions of this paper and

Camada, M., Stéfano, D., Cerqueira, J., Lima, A., Conceição, A. and Costa, A.
Recognition of Affective State for Austist from Stereotyped Gestures.
DOI: 10.5220/0005983201970204
In *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2016) - Volume 1*, pages 197-204
ISBN: 978-989-758-198-4

future works are present in Section 6.

2 BACKGROUND

2.1 Affective State

Although the human emotional state is present only in mind, and some unconscious signals from body allow us to infer the mood. Particular models are essential to define the human affective state. There are two major approaches (Russell, 2003): (i) the *discrete* approach considers the human experiences can be expressed by a small set of emotions (*e.g.* happiness, sadness, fear, anger, surprise and tenderness) and these emotions are basic and experimented independently from each other; and (ii) *affective dimensions* approach, also called *Core Affective* by (Russell, 2003), assumes that emotions are appropriately represented in an emotional plan of Valence/Arousal.

2.2 Body Expressions

Studies on body language have advanced, though they are still few if compared with researches on facial expressions or voice. Two properties about emotional quality from body expression are considered (Wallbott, 1998): (i) *static configuration (posture)*, and (ii) *dynamic or movement configuration (gesture)*. However, most of body cues may indicate only activation level of the person. Thus, these cues just work to differentiate emotions. The energy (power) of movements is of these cues. The highest values related to hot anger, elated joy and terror emotions, the lowest values corresponded to sadness and boredom.

A way to get relevant emotional features from the full-body movements is through the Quantity of Motion (QoM) (Camurri et al., 2003). QoM can revel activation level, for example, during dance performance showed that movements of the limbs associated with anger and joy are significantly high values of QoM.

Now, let $v_l(f)$ denote the module of velocity of each limb l at time frame f as

$$v_l(f) = \sqrt{\dot{x}_l(f) + \dot{y}_l(f) + \dot{z}_l(f)}, \tag{1}$$

where $\dot{x}_l(f)$, $\dot{y}_l(f)$ and $\dot{z}_l(f)$ are cartesian velocities.

The body kinematic energy, $E_{tot}(f)$, can be an approximated by sum of the kinematic energy of each limb as

$$E_{tot}(f) = \frac{1}{2}\sum_{l=1}^{n} m_l v_l(f)^2, \tag{2}$$

where m_l is the approximated of the limb mass based on biometrics anthropometric tables (Dempster and Gaughran, 1967).

2.3 Autism Spectrum Disorders

Autism Spectrum Disorder (**ASD**) is a neuropsychiatric disorder characterized by severe damage in the socialization and communication processes. Generally, autist may also have a unusual pattern or stereotyped behaviors (Levy et al., 2009). Novel researches have been indicated that several factors are associated with the **autism**. Some of these known factors are genetic, neurological anomalies and psychosocial risks (Levy et al., 2009).

(Kuhn, 1999) assumes that the stereotypic behaviors in autists are defense mechanism due to their hypersensitivity. Some **stereotyped gestures** that can be noted are: (i) **Body Rocking** (repetitive movement to forward and backward of the upper torso); (ii) **Top Spinning** (walk in a circle); (iii) **Hand Flapping** (swing motion of the hands up and down); (iv) **Head Banging** (hitting head on the floor or wall). The **Head Banging** was not considered in this papers specially because the trajectory of their movement is similar to **Body Rocking**.

3 CLASSIFICATION AND INFERENCE TOOLS

In this paper, we propose the use of two cognitive tools for recognizing the **stereotyped gesture** and inference of autist defense level: **HMM** (Subsection 3.1) and **FIS** (Subsection 3.2), respectively. Figure 1 (B) shows this proposed model.

3.1 Hidden Markov Models

Hidden Markov Models (**HMM**) are doubly stochastic models, because they have an underlying Markov chain and to transit their stochastic states symbols need to be emitted. This emitting process is itself a stochastic process, once it has a probability distribution over the states and to following the timing of the transitions. Since the symbol output probability distribution of a continuous **HMM** is given by a mixture of Gaussians, a **HMM** can be expressed as $\lambda = (A, c, \mu, U)$, where: A is the matrix of transition probabilities; c is a set of coefficients (weights for each Gaussian in the mixture of Gaussians); μ represents the averages of each Gaussian in the mixture and U also represents the covariance matrices of the Gaussians.

The **HMM** can be applied for supervised learning pattern recognition tasks. The training process of the **HMM** consists of the presentation of sequences of outputs (training sequences) from a particular system.

A training algorithm adjusts the **HMM**'s parameters in such a way that when a new observation sequence from the system being modelled is given as input to the **HMM**, the probability that the model was generated will be presented in output. This discussion leads to the three basic problems of the **HMM** (Rabiner, 1989; Fink, 2008):

Problem 1 - To find the probability that the **HMM** λ generated a given sequence of observation symbols $\mathbf{O} = O_1, O_2, ..., O_T$, where T is the length of the sequence ($P(\mathbf{O}|\lambda)$);

Problem 2 - To find the underlying optimal state sequence $\mathbf{Q} = q_1, q_2, ..., q_T$ of λ that was needed to generate $O_1, O_2, ..., O_T$ ($P(Q|\mathbf{O}, \lambda)$);

Problem 3 - To adapt the model parameters in order to maximize $P(\mathbf{O}|\lambda)$.

3.2 Fuzzy Inference System

Fuzzy Inference Systems (**FIS**) are widely used for problems what the variables of the real worlds are complex or unclear. These systems are knowledge-based (or rule-based) and such knowledge can be obtained from human experts. It can be defined by fuzzy rules of $IF - THEN$ type. Each $IF - THEN$ rule is a statement in which some words are represented by continuous membership functions (Wang, 1997). The value of the membership function informs the degree of membership into a fuzzy set.

(Wang, 1997) indicates three types of fuzzy systems. They differ basically about how they deal with inputs and outputs variables of the system:

Pure Fuzzy System - It is a generic model with inputs and outputs based on words of natural language;

TSK (Takagi-Sugeno-Kang) Fuzzy System - It has input variables combine words of natural language and Real values, but output variables are only real values;

Mamdani - Both Fuzzifier and Defuzzifier associates translate real input to output variables into natural language.

Fuzzyfier and Defuzzifier are based on membership functions. The membership functions translate variable values from one universe to another. There are different membership functions, such as Singleton, Gaussian, S-Shape and Z-Shape.

A Singleton membership, μ_I, represents a set I which is associated to a crisp number α_I, such that,

$$\mu_I(x) = \begin{cases} 1, & \text{if } x = \alpha_I \\ 0, & otherwise. \end{cases} \tag{3}$$

A Gaussian membership, μ_G, represents a set G, where G is a gaussian curve, and it is defined by average $\overline{X}$ and standard deviation σ,

$$\mu_G(x) = e^{-\frac{(x-\overline{X})^2}{2\sigma^2}}. \tag{4}$$

A S-shape membership, μ_S, represents a set S, where S is a "S" curve, and it is defined two parameters α_S and β_S such that,

$$\mu_S(x) = \begin{cases} 0, & x \leq \alpha_S \\ 2\left(\frac{x-\alpha_S}{\beta_S-\alpha_S}\right)^2, & \alpha_S \leq x \leq \frac{\alpha_S+\beta_S}{2} \\ 1-2\left(\frac{x-\beta_S}{\beta_S-\alpha_S}\right)^2, & \frac{\alpha_S+\beta_S}{2} \leq x \leq \beta_S \\ 1, & x \geq \beta_S \end{cases}. \tag{5}$$

A Z-shape membership, μ_Z, represents a set Z, where Z is a "Z" curve , and it is defined two parameters α_Z and β_Z such that,

$$\mu_Z(x) = \begin{cases} 1, & x \leq \alpha_Z \\ 1-2\left(\frac{x-\alpha}{\beta_Z-\alpha_Z}\right)^2, & \alpha_Z \leq x \leq \frac{\alpha_Z+\beta_Z}{2} \\ 2\left(\frac{x-\beta}{\beta_Z-\alpha_Z}\right)^2, & \frac{\alpha_Z+\beta_Z}{2} \leq x \leq \beta_Z \\ 0, & x \geq \beta_Z \end{cases}. \tag{6}$$

4 SYSTEM OVERVIEW

This paper is part of the robotics project called **HiBot** (see Figure 1 (A)). The **HiBot** has been developing in the laboratory of robotics at the Electrical Engineering Department, Federal University of Bahia. It aims to be a platform for experiments on Human Robot Interaction (**HRI**).

The **HiBot** has a sets of affective actuators and sensors. These sets are arranged by modules. The affective actuators aim to promote interaction with through social protocols (face and voice expressions). That way, we define two modules: (i) Voice Synthesis Module (VSM) and (ii) Facial Expression Module (FEM). The affective sensors aim to get affective cues conveyed by different modais, such as face, body, voice and electroencephalography (EEG). Thus, **Hibot** has four affective sensory modules: (i) Facial Expression Recognition Module (FERM); (ii) Body Expression Recognition Module (BERM); (iii) Voice Expression Recognition Module (VERM); and (iv) EEG Recognition Module (EEGRM). We focus on the BERM in this paper.

The BERM gets **stereotyped gestures** from autists to recognize his affective stare (defense level).

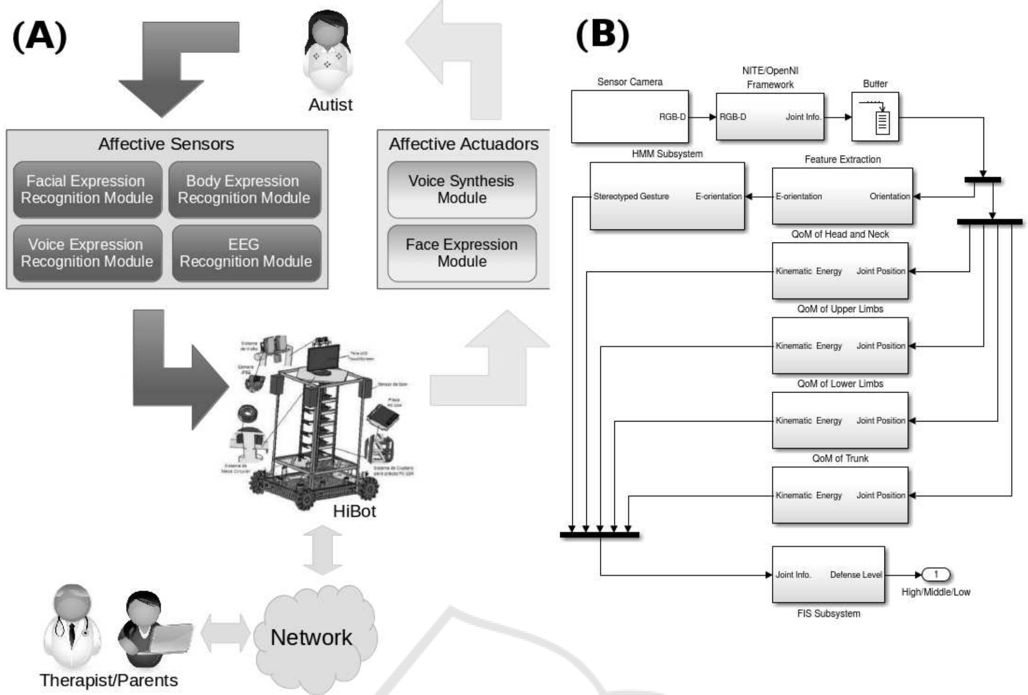

Figure 1: (A) General model of HiBot project and (B) Body Expression Recognition Module architecture.

The architecture of this module is shown in of Figure 1(B). The camera sensor is a Kinect®. This device has a set of sensors (RGB and IR cameras, accelerometer, and microphone array) and also a motorized tilt (Microsoft, 2013). We used the RGB and IR cameras in this paper. The IR camera provides depth information of environment and objects. The IR Emitter projects on the environment (and objects) several infrared lasers. The IR Sensor captures the IR lasers projected. Due to this, Kinect® is able to infer the distance (depth) between objects and IR sensor.

OpenNI/NITE frameworks are used for development of 3D sensing middleware and applications. Currently, they are maintained by Structure Sensor (Structure, 2013). These software extract information about position and orientation of the joint of an attending person. The joint considered in this paper are: head, neck, shoulders, elbows, hands (wrists), trunk, hips, knees and foots (ankles).

The joint data (orientation and position) from 40 frames are stored in a *Buffer*. Joint orientation data are processed by feature extraction algorithm (Subsection 4.1). Likewise, Quantity of Motion (Equation 2) is applied on the joint position data.

The feature extraction algorithm results are used by HMM Subystem (Subsection 4.2). Thus, the inputs of FIS Subsystem are the stereotyped gesture recognized by HMM Subsystem and the QoMs of each joint group. FIS Subsystem must infer the defense (stress) level of target autist.

4.1 Feature Extraction

The joint orientation data are computed by a algorithm of feature extractions. This procedure was applied in order to both reduce the dimensionality (from 4 to 3) of the acquired input data and obtain a meaningful representation of this data. Results from this algorithm are the input to **HMM** subsystem.The following subsections describe the step-by-step this algorithm.

4.1.1 Merging

The first step of feature extractions algorithm consist in merging the four quaternion streams into only one stream. This was achieved by averaging the components. So, let s^i, for $1 \leq i \leq 4$, denote the i-th quaternion stream. The resulting signal is given by

$$\bar{s} = \frac{\sum_{i=1}^{4} s^i}{4} \qquad (7)$$

4.1.2 Frequency Spectrum

After the merging step, the frequency spectrum of the signal $\tilde{s}$, evaluated from the Fast Fourier Transform (FFT) algorithm, is appended to it, to generate the signal $s = [\tilde{s} \quad FFT(\tilde{s})]$.

4.1.3 Short-Time Analysis

The Short-Time Analysis is performed with the signal s divided in segments having $M = M_1 + M_2$ samples centered at the $\hat{n}$-th sample, given by $s_{\hat{n}}(m) = s(\hat{n} + m)$, with $-M_1 \leq m \leq M_2$. This segment is further multiplied (element-wise) with a Hamming window function given by

$$w(n) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{M-1}), & \text{if } 0 \leq n \leq M-1 \\ 0, & \textit{otherwise} \end{cases}, \quad (8)$$

being the segments obtained with an overlapping of 30%.

The energy of each segment $s_{\hat{n}}$ is then calculated as

$$E_{\hat{n}} = \sum_{j=-M_1}^{M_2} s_{\hat{n}}(j), \quad (9)$$

that is the first component of the observation vectors. Second and third components are respectively the first and second derivatives of this energy with respect to $\hat{n}$.

4.2 HMM Subsystem Setup

In this work, the **HMMs** is applied to recognize gestures from sequences of joint orientation acquired with sensor camera Kinect® and stored in *Buffer* of size 40. The features are extracted (see Subsection 4.1) from these sequences to generate arrays of feature vectors. These arrays, in turn, will represent the sequence of observation symbols **O**. Each gesture is associated with a **HMM**.

The training procedure is given by the solution to the third problem. Let $\lambda_i = (A_i, c_i, \mu_i, U_i)$ denote the **HMM** associated to the i-th gesture, with a given initial condition, the training procedure should adapt these parameters using enough (typically several) observation symbols sequences $\mathbf{O}^i_{train}$ from the i-th gesture such that the likelihood that the resulting model is given by

$$\bar{\lambda}_i = (\bar{A}_i, \bar{c}_i, \bar{\mu}_i, \bar{U}_i). \quad (10)$$

The mechanism to evaluate the aforementioned likelihood is given by the solution to the first problem. The HMM training uses 200 samples by **stereotyped gesture** (100 for each activation group).

Each gesture has an **HMM** where **Body Rocking** and **Hand Flapping** have 3 states, and **Top Spinning** has 4 states. The number of Gaussian mixtures was the same in all three case is 3.

4.3 FIS Subsystem Setup

FIS Subsystem infers the state of defense from the **stereotyped gesture** recognized by **HMM Subsystem**. Model of **FIS** subsystem has 5 inputs: (i) Stereotyped Gesture, (ii) QoM Head/Neck, (iii) QoM Upper Limbs, (iv) QoM Lower Limbs and (v) QoM Trunk. The output of this model is *Defense Level*.

The first input of the fuzzifies is the **stereotyped gesture** recognized by **HMM** subsystem. Thus, this fuzzifier has 3 linguistic variables: (i) **Body Rocking** (*BR*), (ii) **Hand Flapping** (*HF*) and (iii) **Top Spinning** (*TS*). These linguistic variables are defined by *Singleton* membership function (Equation 3). The parameters α_I for these linguistic variables are respectively 1, 2 and 3. Figure 2 (A) shows these linguistic variables and its values.

The processing of QoM (see Equation (2)) is executed in the following joint groups: *Head/Neck*, *Upper Limbs*,*Lower Limbs* and *Trunk*. Thus, the inputs *QoM Head/Neck*, *QoM Upper Limbs*, *QoM Lower Limbs* and *QoM Trunk* maps QoM of joint groups.

Figure 2(B)-(E) show 4 inputs of **FIS** Subsystem with three linguistic variables: *Low*, *Middle* and *High*. Linguistic variable *Low* is defined by *Z-Shape* membership function (see Equation (6)). The values of parameters α_Z and β_Z are as follows,

$$[\alpha_Z, \beta_Z] = [\bar{X}_{Low}, Max_{Low}], \quad (11)$$

where $\bar{X}_{Low}$, Max_{Low} are average and maximum values of low subgroup related to the training samples.

Linguistic variable *Middle* is represented by *Gaussian* membership function. The values of its parameters σ and $\bar{X}$ are defined as,

$$[\sigma, \bar{X}] = [\sigma_{LH}, \bar{X}_{LH}], \quad (12)$$

where σ_{LH}, $\bar{X}_{LH}$ are standard deviation and average related to high and low subgroups of the training samples.

Finally, linguistic variable *High* is defined by *S-Shape* membership function (Equation (5)). The values of the parameters α_S and β_S are as follows,

$$[\alpha_S, \beta_S] = [Min_{High}, \bar{X}_{High}], \quad (13)$$

where Min_{High} and $\bar{X}_{High}$ are respectively minimum and average values of high subgroup related to the training samples.

FIS Subsystem is based on **Mamdani**. That way, we define 15 weighted rules. Table 1 shows these rules and their respective weights.

Body Rocking and **Hand Flapping** are defined respectively by *head/neck* and *upper limbs*. **Top Spinning** is defined by *lower limbs* and *trunk*. Besides that, a weight is assigned to each rule. The last

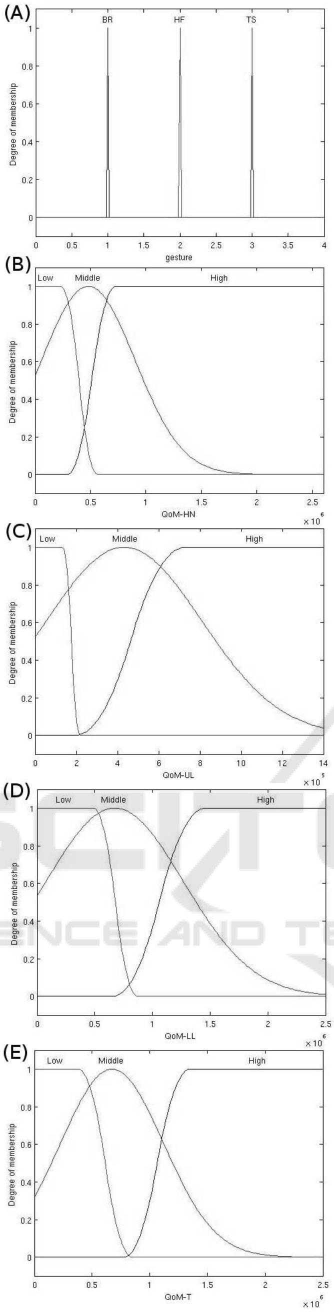

Figure 2: Input of FIS Subsystem: (A) Gesture, (B) QoM of Head and Neck, (C) QoM of Upper Limbs, (D) QoM of Lower Limbs, and (E) QoM of Trunk.

column (**W.**) in Table 1 define the values of weight of rules. The weight is a real value that can be 0.25, 0.50 or 1.00.

The value 1.00 is assigned to the weights of the rules related to **Body Rocking** and **Hand Flapping** gestures. The weight values for rules of the gesture **Top Spinning** depends on the difference of activation level between joint groups *upper limb* and *trunk*. It is assigned to the maximum, medium ad minimum differentiation, 0.25, 0.50 and 1.00, respectively.

Table 1: Defining the rules and their weights (W.) with Defense Level (D. Level) for stereotyped gestures (Ge.): Body Rocking (BR), Hand Flapping (HF) and Top Spinning (TS). Linguistic variables HIGH (HI.), MIDDLE (MI.) and LOW (LO.) are defined according to each QoM (Q.) of joint groups Head/Neck (H/N), Upper Limbs (UL), Lower Limbs (LL) and Trunk.

Ge.	Q. H/N	Q. UL	Q. LL	Q. Trunk	D. Level	W.
BR	HI.	any	any	any	HI.	1.00
BR	MI.	any	any	any	MI.	1.00
BR	LO.	any	any	any	LO.	1.00
HF	any	HI.	any	any	HI.	1.00
HF	any	MI.	any	any	MI.	1.00
HF	any	LO.	any	any	LO.	1.00
TS	any	any	HI.	HI.	HI.	1.00
TS	any	any	HI.	MI.	HI.	0.50
TS	any	any	HI.	LO.	MI.	0.25
TS	any	any	MI.	HI.	HI.	0.50
TS	any	any	MI.	MI.	MI.	1.00
TS	any	any	MI.	LO.	LO.	0.50
TS	any	any	LO.	HI.	MI.	0.25
TS	any	any	LO.	MI.	LO.	0.50
TS	any	any	LO.	LO.	LO.	0.50

The output of the **FIS** Subsystem has three Gaussian membership function: *Low*, *Middle* and *High* (see Figure 3). The defuzzifier uses Centroid method, aggregation *Maximum* and implication *Minimum*. These membership functions are equally distributed on the universe of values. The output of defuzzifier represents the defense level of a person with **autism**.

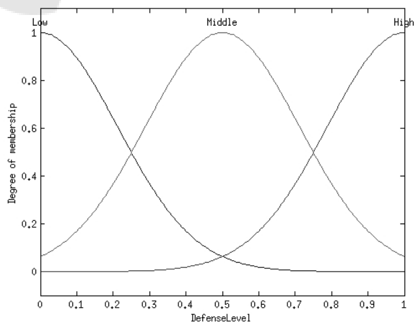

Figure 3: Defuzzifier (defense level) with 3 Gaussian membership function: Low, Middle or High.

5 EXPERIMENTS

Simulations of the BERM allow us to analyzing its behavior and also expected results. MATLAB® and **HMM/FIS** toolboxes (Murphy, 1998) were used to simulate the BERM. MATLAB® is a high-level language and interactive environment well known by the community of scientists and engineers.

5.1 Methodology

For each gesture, we defined two simulation scenarios: *high* and *low* activation. In this way, an actor performed repeatedly each scenarios of the **stereotyped gestures**. These gestures were recorded using Kinect® device and OpenNI/NiTE frameworks. Thus, the RGB-D image frames were stored together with position and orientation metadada of each joint.

Figure 4 shows RGB-D images of **stereotyped gestures**: **Body Rocking** (A.1 and A2), **Hand Flapping** (B.1 and B.2) and **Top Spinning** (C.1 and C.2).

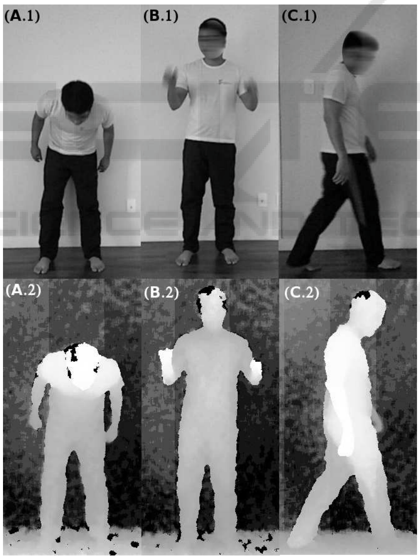

Figure 4: RGB-D images of stereotyped gestures performed by an actor.

After that, the samples were manually extract. Each sample has data about joints (position and orientation) of 40 image frames. The scenarios of each **stereotyped gesture** have 150 samples of which: 100 samples were used to training and other 50 were used to simulation.

The parameter values of QoM membership function high (Equation (13)), middle (Equation (12)) and low (Equation (11)) were defined from the training samples of HMM.

Results of these simulations present the defense level for each gesture. In this way, we analyze whether these results coincided with expected defense level. These results are discussed in the following section.

5.2 Results and Discussion

In order to represent the statistics of the simulations results, we use the confusion matrix. Simulation results show the **HMM** Subsystem recognized all stereotyped gestures **Hand Flapping** and **Top Spinning**. Although the results for **Body Rocking** lower than the other gestures, their performance was 86% hit (see Table 2). The efficiency of the **HMM** is due to two reasons: (i) **stereotyped gestures** are well-defined and distinct from themselves. (ii) the **HMM** Subsystem should not differentiate among subgroups of gestures (high or low activation).

Table 2: Confusion matrix of recognition stereotyped gesture by HMM Subsystem.

	Body Rocking	Hand Flapping	Top Spinning
Body Rocking	86%	0%	14%
Hand Flapping	0%	100%	0%
Top Spinning	0%	0%	100%

Tables 3, 4, 5 show the performance of **FIS** Subsystem for each **stereotyped gesture**. We consider defense level is high for values above or equal to 0.5. Therefore, the defense level is low for values below 0.5.

Table 3 shows the defense level for gesture **Body Rocking** presents better adjustments values for high activation (98%) than for low activation (82%).

Table 3: Confusion matrix of activation level for Body Rocking.

	High	Low
High	98%	2%
Low	18%	82%

However, Table 4 shows gesture **Hand Flapping** presents better adjustments values of defense level for low activation (100%) than for high activation (96%).

Table 4: Confusion matrix of activation level for Hand Flapping.

	High	Low
High	96%	4%
Low	0%	100%

The defense level for the gesture **Top Spinning** showed positive performance for the two activation levels (see Table 5).

Table 5: Confusion matrix of activation level for Top Spinning.

	High	Low
High	100%	0%
Low	0%	100%

The results of simulations in Tables 2, 3, 4 and 5 show relevant results for the proposed model. However, the proposed model may present lower performance with a autist in real world than with an actor. The idiosyncrasy of each person may influence the gesture recognition process and inference of defense level. In addition, the ASD (Autism Spectrum Disorder) presents different behavioral aspects which may vary according to the severity. Thus, it necessary to specify the target autistic spectrum.

Although the confusion matrix does not show the variation in trend of defense level, this is a major requirement in the process of interaction between the robot and autistic. Thus, it is possible to analyze the interactive process is effective or not.

6 CONCLUSION

This paper proposed a system model to infer the defense level of autist from the **stereotyped gestures** (**body rocking**, **hand flapping** and **top spinning**). These gestures were performed by an actor. The cognitive model consists of **HMM** and **FIS** Subsystems.

The simulation results demonstrate this approach is adequate and promising to recognize the defense level from **stereotyped gestures**. **HMM** Subsystem classifies these gestures correctly. **FIS** Subsystem is able to correctly infer for most simulations, showing better results for **Top Spinning**.

The **BERM** will be used in the **HiBot** to recognize the affective state of the autist, more precisely during interaction with others sensors.

The next steps after this paper are:

1. Creating and using a database with genuine autistic gestures (not actors);
2. Specifying the target autistic spectrum;
3. Integrating this module *Body Expression Recognition Module* (BERM) with the other modules of **HiBot**.

REFERENCES

Camurri, A., Lagerlof, I., and Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human Computer Studies*, 59(1–2):213–225.

Dautenhahn, K. (2003). Roles and functions of robots in human society- implications from research in autism therapy. *Robotica*, 21(4):443–452.

Dempster, W. and Gaughran, G. (1967). Properties of body segments based on size and weight. *American Journal of Anatomy*, 120(7414):33–54.

Fink, G. (2008). *Markov Model for Pattern Recognition: From Theory to Applications*. Berlin: Springer.

Goodrich, M., Colton, M., Brintonand, B., Fujiki, M., Atherton, J., and Robinson, L. (2012). Incorporating a robot into an autism therapy team. *IEEE Intelligent Systems*, 27(2):52–59.

Kuhn, J. (1999). Stereotypic behavior as a defense mechanism in autism. *Harvard Brain Special*, 6(1):11–15.

Levy, S., Mandell, D., and Schultz, R. (2009). Autism. *The Lancet*, 374(9701):1627–1638.

Microsoft (2013). Kinect for windows sensor components and specifications. https://msdn.microsoft.com/en-us/library/jj131033.aspx. Last accessed on Oct 21, 2013.

Murphy, K. (1998). Hidden markov model (hmm) toolbox for matlab. http://www.cs.ubc.ca/ murphyk/Software/HMM/hmm.html. Last accessed on Jan 10, 2015.

Parsons, S., Mitchell, P., and Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 34(4):449–466.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172.

Structure (2013). Support openni. http://structure.io/openni. Last accessed on Nov 21, 2013.

Wallbott, H. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896.

Wang, L. (1997). *A Course in Fuzzy Systems and Control*. PTR, Prentice Hall.

Zeng, Z., Pantic, M., Roisman, G., and Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58.