An Architecture with a Mobile Phone Interface for the Interaction of a Human with a Humanoid Robot Expressing Emotions and Personality

Antonio Chella[1], Rosario Sorbello[1], Giovanni Pilato[2], Giorgio Vassallo[1], Giuseppe Balistreri[1], and Marcello Giardina[1]

[1] DICGIM Università degli Studi di Palermo, RoboticsLab, Viale delle Scienze, 90128 Palermo, Italy
[2] ICAR, CNR Viale delle Scienze, 90128, Palermo, Italy

Abstract. In this paper is illustrated the cognitive architecture of a humanoid robot based on the proposed paradigm of Latent Semantic Analysis (LSA). This paradigm is a step towards the simulation of an emotional behavior of a robot interacting with humans. The LSA approach allows the creation and the use of a data driven high-dimensional conceptual space. We developed an architecture based on three main areas: Sub-conceptual, Emotional and Behavioral. The first area analyzes perceptual data coming from the sensors. The second area builds the sub-symbolic representation of emotions in a conceptual space of emotional states. The last area triggers a latent semantic behavior which is related to the humanoid emotional state. The robot shows its overall behavior also taking into account its "personality". We implemented the system on a Aldebaran NAO humanoid robot and we tested the emotional interaction with humans through the use of a mobile phone as an interface.

Keywords: Humanoid Robot, Emotions, Personality, Latent Semantic Analysis.

1 Introduction

The concept to have a humanoid robot expressing emotions appear mind-boggling to the human beings; but it is possible to try to reproduce some emotional behaviors inside the robots in order to improve the interaction with people. In the past few years, many efforts have been made experimenting human robot social interaction [10], [3] through the emotions [1], [2], [4], [5]. Lun et al. [6] describe an emotional model and affective space of a humanoid robot and establish their studies on the psycho-dynamics psychological energy and affective energy conservation law. ARKIN et al. [7], [8], [9] show a robotic framework called TAME for human-robot interaction that tries to connect together in the same system affective phenomena as attitudes, emotions, moods and trait. Miwa et al. [1] have illustrated a mechanism for humanoid robot WE-4RII to express human-like emotions in a natural way. Research by Breazeal [2] has shown a humanoid

R. Pirrone and F. Sorbello (Eds.): AI*IA 2011, LNAI 6934, pp. 117–126, 2011.

robots emotion model as an important basis in social interaction with humans. The interaction of an humanoid robots with humans and the environment should not be mechanical and deterministic in order to avoid a predictable and trivial behavior. To reach this goal we have modeled a Cognitive Architecture that drives the behavior of a humanoid robot taking into account its own personality", the context in which it is immersed, the present stimulus and the most recent tasks executed by the robot. This cognitive architecture of a humanoid robot is based on the paradigm of Latent Semantic Analysis [11], [12]. This paradigm is a step forward towards the simulation of an emotional behavior of a robot interacting with humans [13], [14], [15]. The approach presented integrates traditional knowledge representation and associative capabilities provided by geometric and sub symbolic information modeling. The paper is structured as follows: In section 2, we will describe the creation of the emotional space of the robot. Section 3 reports on the cognitive architecture of the humanoid robot expressing emotions and personality. Finally section 4, shows the emotional interaction system with humans through the use of a mobile phone as an interface.

2 The Emotional Space Creation

The architecture of the robot is based on the creation of a probabilistic emotional conceptual space automatically induced from data. The LSA technique has been widely explained and details can be found in [16], [14], [17]; for completeness we briefly recall it here.

The approach is based on the application of the Truncated Singular Value Decomposition (TSVD) technique, preceded and followed by specific pre- and post-processing phases that allow to give a probabilistic interpretation of the induced vector space.

The methodology is inspired to the Latent Semantic Analysis (LSA) approach, which is a technique based on the vector space paradigm, used to extract and represent the meaning of words through statistical computations applied to a large corpus of texts. The paradigm defines a mapping between words and documents belonging to the corpus into a continuous vector space S, where each word, and each document is associated to a vector in S [18], [19]. Even if Truncated SVD (TSVD) has been traditionally applied in the text classification and information retrieval fields [20], it can be successfully applied to any kind of dyadic domain. Let us consider a corpus made of N text chunks, each one expressing an emotion, and let M be the number of words taken into consideration in the whole data-set. The goal is to realize a mapping between the M words and the N text chunks verbally describing emotions into a continuous vector space S, where each word, as well as each emotion is associated to a emotional *knoxel* in S. Like knoxels in conceptual spaces, from the mathematical point of view, an emotional knoxel is a vector in the probabilistic emotional space; from the conceptual point of view, it is the epistemologically basic element at the level of analysis considered. Let B be the $M \times N$ matrix whose $(i,j)-th$ entry is the square root of the sample probability of the $i-th$ word belonging to the $j-th$ text chunk. The Singular Value Decomposition of the matrix B is performed, so that B is decomposed in the product of three matrices: a column-orthonormal $M \times N$ matrix U, a column-orthonormal $N \times N$ matrix V and a $N \times N$ diagonal matrix , whose elements are called singular values of B.

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\boldsymbol{T}}$$

Let us suppose that B's singular values are ranked in decreasing order. Let R be a positive integer with $R < N$, and let U_R be the $M \times R$ matrix obtained from U by suppressing the last N-R columns, $_R$ the matrix obtained from by suppressing the last N-R rows and the last N-R columns and V_R be the $N \times R$ matrix obtained from V by suppressing the last N-R columns. Then

$$\mathbf{B}_R = \mathbf{U}_R\boldsymbol{\Sigma_R}\mathbf{V}_{\boldsymbol{R}}^{\boldsymbol{T}}$$

is a $M \times N$ matrix of rank R. The matrix B_R is then post-processed in order to obtain a new matrix Ψ in this manner:

$$[\psi]_{ij} = \begin{cases} 0 & \text{if } [b_R]_{ij} < 0 \\ \frac{[b_R]_{i,j}}{\sqrt{\sum [b_R]_{ij}^2 < 0}} & \text{otherwise} \end{cases} \tag{1}$$

It can be shown that the illustrated procedure can be seen as a statistical estimation process, being Ψ the probability amplitude estimated starting from the sample probability amplitude B.

3 The Cognitive Architecture

The architecture of the system presented is inspired to the approach illustrated in [21], [22], and it is organized, as shown in figure 1, in three main areas:

- The Sub-conceptual Area processes the perceptual data that come from the sensors.
- The Emotional Area is constituted by a "conceptual" space where emotions, behaviors, personality of the robot, together with the perceptual data are mapped as emotional knoxels. The coding of the personality distorts the perception of reality of the robot.
- The Behavioral Area reasons about the environment, and chooses the most adequate behaviors in order to react to the external stimuli emphatically. In this process, the linguistic area takes into account both the personality of the robot and the behaviors adopted in the recent past.

3.1 The Sub-conceptual Area

The sub-conceptual area is aimed at receiving stimuli from the different modalities of the robot and controlling its actuators at a low level also. Two main modules compose this area:

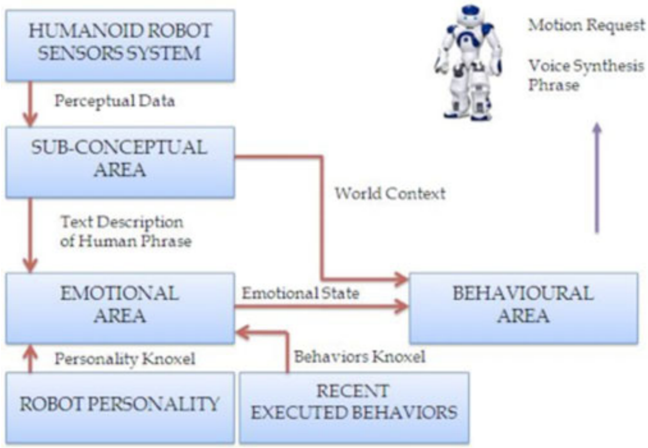

Fig. 1. The Emotional Humanoid Robot Architecture

– The *MotionModule* controls the actuators of the robot in order to perform the movement requested by the Behavioral Area.
– The *PerceptualModule* processes the raw data coming from robot sensors in order to obtain information about the environment, and the external stimuli.

The perceptual information sensed by the processing of the raw sensor data is associated to their English description: as an example, if the system recognizes a "red hammer", it will be associated to the sentence "I can see a red hammer". The use of a verbal description of the information retrieved from the environment allows its easy mapping into the emotional area. These natural language descriptions associated to each modality will be the input of the conceptual area.

3.2 The Emotional Area

The emotional area enables the robot able to find emotional analogies between the current status and the previous knowledge stored in the robot using the semantic space of emotional states. The associative area is built up in order to reflect and encode not only emotions and objects that provoke emotions, but also the personality of the robot. This is true in a twofold manner:

– documents used to induce the associative space characterize its dimensions and concepts organization beneath the space
– personality is also encoded as a set of knoxels that play the role of attractors of perceived objects beneath the space

A corpus of documents dealing with emotional states has been chosen in order to infer the space which plays therefore the role of a "probabilistic emotional space". Emotional states have been coded as "emotional knoxels" in this space using verbal description of situations capable to evoke feelings and reactions.

Environmental incoming stimuli are encoded using natural language and then are mapped in the space, to find analogies with the emotions encoded in it. A large amount of documents has been selected from several publicly available on-line sources. The excerpts have been organized in homogeneous paragraphs both for text length and emotion. The set of documents used has been obtained through an accurate selection of excerpts associated to feelings. We have selected the following emotional expressions: sadness, fear, anger, joy, surprise, love and a neutral state.

A matrix has been organized where the 6 emotional states and the neutral state have also been coded according to the procedure illustrated in the previous section that leads to the construction of a probabilistic emotional space.

A corpus of 1000 documents, equally distributed among the seven states, including other documents characterizing the personality of the robot, has been built. This set of documents represents the affective knowledge base of the robot. Each document has been processed in order to remove all words in literature named "stopwords" that do not carry semantic information like articles, prepositions and so on.

According to the technique outlined in section two, a 87×1000 terms documents matrix (B) has been created where M=80+7 is the number of words plus the emotional states and N=1000 is the number of excerpts. The generic entry $b_{i,j}$ of the matrix is the square root of the sample probability of the $i-th$ word belonging to the $j-th$ document. The TSVD technique, with K=150, has been applied to B in order to obtain its best probability amplitude approximation Ψ. This process leads to the construction of a K=150 dimensional conceptual space of emotions S. The axes of S represent the "fundamental" emotional concepts automatically induced by TSVD procedure arising from the data. In the obtained space S, a subset of n_i documents for each emotional state corresponding to one of the six "basic emotion" E_i has been projected in S using the folding-in technique. According to this technique each excerpt is coded as the sum of the vectors representing the terms composing it. As a result, the $j-th$ excerpt belonging to the subset corresponding to the emotional state E_i is represented in S by an associated vector $em_j^{(i)}$ and the emotional state E_i is represented by the set of vectors $\{\mathbf{em}_j^{(i)} : j = 1, 2, Kn_i\}$

The personality of the robot is encoded as a set of knoxels derived from documents describing the fundamental personality characteristics of the robot. As an example, if the robot has a "shy" personality, a set of documents dealing with shyness, bashfulness, diffidence, sheepishness, reserve, discretion, introversion, reticence, timidity, and so on are used to construct a cloud of "personality knoxels" that represent attraction points for the external stimuli that cause emotions in the robot. Therefore the $i-th$ excerpt belonging to the subset corresponding to the personality characteristic is mapped as a vector p_i in S. The inputs from the sense channels are coded in natural language words or sentences describing them and projected in the conceptual space using the folding-in technique. These vectors, representative of the inputs from the channels, are merged together as a weighted sum in a single vector $stim(t)$ that synthesizes the inputs stimuli from environment at instant t:

The input stimuli are therefore biased through the computation of the contribution of personality attractors in the space as the weighted sum of the knoxels p_i representing the personality of the robot, and therefore are called "personality knoxels". Each personality knoxel can be weighted with coefficients w_i (with $0 \leq w_i \leq 1$) in order to fine tune the personality influence upon the robot's behaviors.

$$\hat{stim}(t) = \frac{stim(t) + \sum w_i p_i}{||stim(t) + \sum w_i p_i||}$$

This procedure represents the common process that arises in human beings, when reality is "filtered" and interpreted by personality. The emotional semantic similarity between the vector $\hat{stim}(t)$ and the knoxels that code the six emotions in S, plus the "neutral" state, can be evaluated using the cosine similarity measure between each $em_j^{(i)}$ and $\hat{stim}(t)$:

$$sim(\hat{stim}(t), em_j^{(i)}) = \frac{\hat{stim}(t) \cdot em_j^{(i)}}{||\hat{stim}(t)|| \cdot ||em_j^{(i)}||}$$

A higher value of $sim(\hat{stim}(t), em_j^{(i)})$ corresponds to a higher value of similarity between the emotion evoked from the input and the emotion E_i associated with the vector $em_j^{(i)}$. The semantic similarity measure is calculated between $\hat{stim}(t)$ and each $em_j^{(i)}$. The vector $em_j^{(i)}$, which maximizes the quantity expressed in the formula, will be the inferred emotional state E_i. This process will activate the emotional stimulus "i" with a given intensity given by:

$$I_i(j) = sim(\hat{stim}(t), em_j^{(i)})$$

3.3 The Behavioral Area

The purpose of the Behavioral Area (figure 2) is to manage and execute behaviors coherently with the emotional state inferred by the Emotional Area, the personality of the robot, the environmental status, the stimulus given by the user and the recent past behavior adopted by the robot. A behavior is described as a sequence of primitive actions sent directly to the robot actuators. Each emotional state E_i is related with different behaviors $b_k^{(E_i)}$ in order to give the robot a non-monotonous, non-deterministic, and non-boring response. The choice of the behavior is a function $g()$ of the emotion aroused in the robot as a function of the environment, the stimulus perceived and the personality of the robot, and a function of the recent behaviors adopted in the past by the robot. Among the behaviors associated to the emotional state E_i, the behavior $b_k^{(E_i)}$ is selected through the evaluation of a score associated to each one of them.

$$Bbi(k, t) = \mu_k r - \frac{\lambda_k}{t}$$

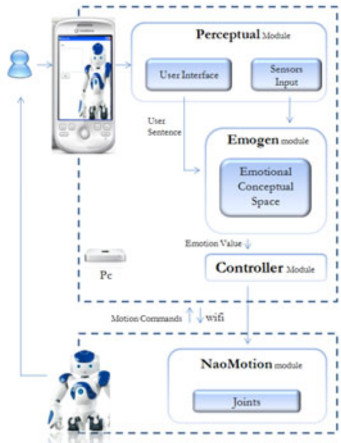

Fig. 2. An high-level description of the architecture

where r is a random value ranging from 0 to 1, $t(t > 0)$ is the time elapsed by the instant at which the $k-th$ behavior $b_k^{(E_i)}$ associated to the emotional state E_i has been executed and the instant at which this assessment is made; μ_k and λ_k are the weights assigned to the random value and to time elapsed respectively. The response with the highest weight is chosen and executed. Since the emotional stimulus is also weighted, thanks to the I "intensity" parameter. Thus, the reaction will be executed with the same intensity: movements of the parts of the body will be quicker, faster or slower. Summing up, the resulting behavior is therefore a composite function:

$$b_k^{(E_i)} = F(g(emotion, personality, stimulus), Bb(k,t), I_i(j))$$

If the emotional state is classified as "neutral", a standard behavior (lie down, sleep, and so on) is randomly selected.

4 Expected Results

The Architecture of Emotional Humanoid Robot NAO, as shown in the figure 3, was modeled using the cognitive architecture described in the previous section. Complex Humanoid Behaviors have been developed to express each of these seven emotions: sadness, fear, anger, joy, surprise, love, and a neutral. A human being can interact by talking to the robot through its voice recognition system. The Architecture of the robot elaborates the user sentence considered it as present stimulus and the image information of the camera system and generates the correct behavior taking into account also its own "personality", the context in which it is immersed, and the most recent behaviors executed by the robot. As a preliminary test bed, we considered a human being talking to the

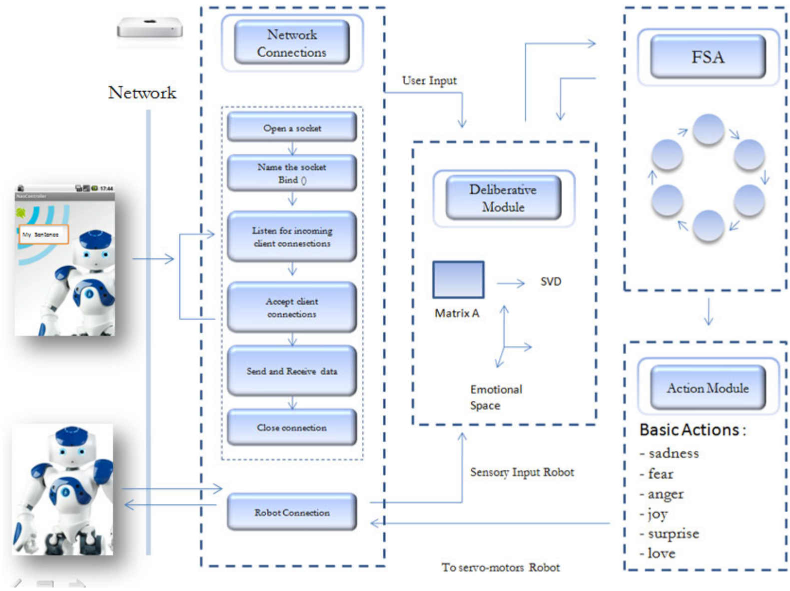

Fig. 3. The Architecture of Emotional Humanoid Robot NAO

Table 1. Samples of results of interaction: in bold it is highlighted the empathic behavior triggered by the user sentence. PE stands for the "Prevalent Emotion".

Sentence	**Sadness**	**Anger**	**Fear**	**Joy**	**Surprise**	**Love**	**PE**
I found myself confuse by sudden feelings of astonishment	-0,09	-0.09	-0.12	+0.08	**+0.50**	+0.08	Surprise
I don't know if my feeling for you is just pleasure or bond	-0.15	-0.33	-0.23	+0.50	+0.28	**+0.51**	Love
I want to share with you the delight of ecstasy for your help	-0.30	-0.25	-0.13	**+0.50**	+0.08	+0.47	Joy
I notice the sense of trepidation and anxiety in your eyes	+0.12	+0.32	**-0.50**	-0.13	-0.12	-0.11	Fear
I am irate and I come to you because the work you did	+0.15	**+1.00**	+0.65	-0.50	-0.17	-0.43	Anger
I am under depression for the trouble that I encountered	**+0.33**	+0.05	0.08	-0.20	-0.06	-0.15	Sadness

humanoid robot through a mobile phone. If the human being, for example, sends the following sentence to the robot as a stimulus,: “I found myself confused by sudden feelings of astonishment”, the dominant emotional state activated in the LSA space of the robot with an “open” personality is “SURPRISE”, and the robot executes a “SURPRISE” behavior.

The experiments, summarized in table 1, have been conducted in order to test the emotional capabilities of the robot by reproducing various emotive situations. A set of 100 sentences has been evaluated in order to understand if the behavior of the robot was accurate. While the 81% of the set was considered almost correct, in 19% of the trials the behavior of the robot was judged as being below the performance expected.

5 Conclusion and Future Works

The results shown in this paper demonstrate the possibility for a humanoid robot to generate emotional behaviors recognizable by humans. In the future we want to focus our attention on increasing the emotional interaction with humans.

References

1. Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Roccella, S., Carrozza, M.C., Dario, P., Takanishi, A.: Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII Integration of Humanoid Robot Hand RCH-1. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai International Center, Sendai, Japan, September 28-October 2, vol. 3, pp. 2203–2208 (2004)
2. Breazeal, C.: Emotion and Sociable Humanoid Robots. International Journal Human-Computer Studies 59, 119–155 (2003)
3. Bruce, A., Nourbakhsh, I., Simmons, R.: “The Role of Expressiveness and Attention in Human-Robot Interaction” AAAI Technical Report FS-01-02 (2001)
4. Liu, Z., Pan, Z.G.: An Emotion Model of 3D Virtual Characters in Intelligent Virtual Environment. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 629–636. Springer, Heidelberg (2005)
5. Monceaux, J., Becker, J., Boudier, C., Mazel, A.: Demonstration: First Steps in Emotional Expression of the Humanoid Robot Nao. In: International Conference on Multimodal Interfaces, Cambridge, Massachusetts, USA, pp. 235–236 (2009)
6. Xie, L., Wang, Z.-L., Wang, W., Yu, G.-C.: Emotional gait generation for a humanoid robot. International Journal of Automation and Computing. Inst. of A.Chinese Ac. of Sc. (2010)
7. Moshkina, L., Arkin, R.C.: Beyond Humanoid Emotions: Incorporating Traits, Attitudes, and Moods. In: Kobe, J.P. (ed.) Proc. 2009 IEEE Workshop on Current Challenges and Future Perspectives of Emotional Humanoid Robotics (May 2009)
8. Arkin, R.C., Fujita, M., Takagi, T., Hasegawa, R.: An Ethological and Emotional Basis for Human-Robot Interaction. Robotics and Autonomous Systems 42, 191–201 (2003)
9. Arkin, R., Fujita, M., Takagi, T., Hasegawa, R.: Ethological Modeling and Architecture for an Entertaiment Robot. In: IEEE Int. Conf. on Robotics & Automation, Seoul, pp. 453–458 (2001)

10. Chella, A., Barone, R.E., Pilato, G., Sorbello, R.: Workshop An Emotional Storyteller Robot. In: AAAI 2008 Spring Symposium on Emotion, Personality and Social Behavior, March 26-28. Stanford University, Stanford (2008)
11. Prendinger, H., Ullrich, S., Nakasone, A., Ishizuka, M.: MPML3D: Scripting Agents for the 3D Internet. IEEE Transactions on Visualization and Computer Graphics 17(5), 655–668 (2011)
12. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: SentiFul: A Lexicon for Sentiment Analysis. IEEE Transactions on Affective Computing 2(1), 22–36 (2011)
13. Anzalone, S.M., Cinquegrani, F., Sorbello, R., Chella, A.: An Emotional Humanoid Partner. In: Linguistic and Cognitive Approaches To Dialog Agents (LaCATODA 2010) At AISB 2010 Convention, Leicester, UK (April 2010)
14. Chella, A., Pilato, G., Sorbello, R., Vassallo, G., Cinquegrani, F., Anzalone, S.M.: An Emphatic Humanoid Robot with Emotional Latent Semantic Behavior. In: Carpin, S., Noda, I., Pagello, E., Reggiani, M., von Stryk, O. (eds.) SIMPAR 2008. LNCS (LNAI), vol. 5325, pp. 234–245. Springer, Heidelberg (2008)
15. Menegatti, E., Silvestri, G., Pagello, E., Greggio, N., Cisternino, A., Mazzanti, F., Sorbello, R., Chella, A.: 3D Models of Humanoid Soccer Robot in USARSim and Robotics Studio Simulators. International Journal of Humanoids Robotics (2008)
16. Agostaro, F., Augello, A., Pilato, G., Vassallo, G., Gaglio, S.: A Conversational Agent Based on a Conceptual Interpretation of a Data Driven Semantic Space. In: Bandini, S., Manzoni, S. (eds.) AI*IA 2005. LNCS (LNAI), vol. 3673, pp. 381–392. Springer, Heidelberg (2005)
17. Pilato, G., Vella, F., Vassallo, G., La Cascia, M.: A Conceptual Probabilistic Model for the Induction of Image Semantics. In: Proc. of the Fourth IEEE International Conference on Semantic Computing (ICSC 2010), September 22-24. Carnegie Mellon University, Pittsburgh (2010)
18. Landauer, T.K., Foltz, P.W., Laham, D.: Introduction to Latent Semantic Analysis. Discourse Processes 25, 259–284 (1998)
19. Colon, E., Sahli, H., Baudoin, Y.: CoRoBa, a Multi Mobile Robot Control and Simulation Framework. Int. Journal of Advanced Robotic Systems (2006)
20. Thagard, P., Shelley, C.P.: Emotional analogies and analogical inference. In: Gentner, D., Holyoak, K.H., Kokinov, B.K. (eds.) The Analogical Mind: Perspectives from Cognitive Science, pp. 335–362. MIT Press, Cambridge (2001)
21. Chella, A., Frixione, M., Gaglio, S.: An Architecture for Autonomous Agents Exploiting Conceptual Representations. Robotics and Autonomous Systems 25, 231–240 (1998)
22. Chella, A., Frixione, M., Gaglio, S.: A cognitive architecture for robot self-consciousness. Artificial Intelligence in Medicine 44(2), 147–154 (2008)