A Connexionist Model for Emotions in Digital Agents

Jean-Claude Heudin
Devinci Media Lab, Pôle Universitaire Léonard de Vinci, Paris – La Défense, France

Keywords: Emotion, Affect, Mood, Neural Network, Artificial Creature, Digital Agent.

Abstract: This paper introduces a bio-inspired model of affects for digital agents. This model provides three distinct layers: emotions as short-term affect, moods as medium-term affect, and personality as a long-term affect. It describes an implementation based on a connexionist architecture using a dedicated neural network designed for the "Living Mona Lisa" research project.

1 INTRODUCTION

In this paper we introduce a bio-inspired model of affects for digital agents. Modeling emotions is a recurrent problem in the design of artificial creatures, including virtual characters and robots. When we ask to anyone what is the difference between a machine and a human, emotion is always the answer before any other aspect.

One of our long term projects is to design a software architecture for believable conversational agents. We have conducted experimentations in the past showing that a "multi-personality" approach – called "schizophrenic" – leads to believable and complex characters (Heudin, 2011). However, we have also concluded that an emotion engine is required to balance between these different "personalities" resulting in coherent and pertinent behaviors.

There have been many artificial emotion models proposed in the past. One of the most complete, mixing short, medium and long-term aspects of emotional behaviors, was designed by Gebhard with ALMA (Gebhard, 2005). We have also proposed a similar approach with the first version of EVA (Heudin, 2004). Both approaches were based of the PAD model (Pleasure, Arousal and Dominance) proposed by Mehrabian (Mehrabian, 1996).

In this paper, we first describe a new model for implementing a bio-inspired model of affects in digital agents. This model is a layered neural network architecture implementing three levels of affects: short-term emotions, mi-term moods and long-term personality.

In the second part of the paper, we describe the implementation of this model in the "Living Mona Lisa" research project. The aim of this project is to design an interactive installation displaying an animated, high-resolution, full-scale reproduction of the famous painting of Leonardo da Vinci. This project is conducted in the spirit of the "Living Art" approach by a multidisciplinary team including researchers, artists and students from the *Institute of Internet and Multimedia* and *Strate School of Design* in Paris. Living Art is a burgeoning field that uses Artificial Intelligence to create interactive works of art, bridging digital technologies and more traditional art forms (Aziosmanoff, 2015).

The paper concludes by showing qualitative results and discussing the future steps in our research.

2 THE EMOTION MODEL

2.1 A Layered Model of Affects

We propose here a new layered model of affects based on three main interacting forms of affects:

Emotion reflects a short-term affect, usually bound to a specific event, action or object, which is the cause of this emotion. After its elicitation emotions usually decay and disappear from the individual's focus (Becker, 2001).

Mood reflects a medium-term affect, which is generally not related with a concrete event, action or object. Moods are longer lasting stable affective states, which have a great influence on human's cognitive functions (Morris, 1989).

Heudin, J-C.
A Connexionist Model for Emotions in Digital Agents.
DOI: 10.5220/0005669202720279
In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016) - Volume 2*, pages 272-279
ISBN: 978-989-758-172-4

Personality reflects long-term affect. It shows individual differences in mental characteristics (McCrae, 1992).

2.1.1 Personality

This layer is based on the "Big Five" model of personality (McCrae, 1992). It contains five main variables with value varying from 0.0 (minimum intensity) to 1.0 (maximum intensity). These values specify the general affective behavior by the traits of openness, conscientiousness, extraversion, agreeableness and neuroticism.

Openness is a general appreciation for art, emotion, adventure, unusual ideas, imagination, curiosity, and variety of experience. The trait distinguishes imaginative people from down-to-earth, conventional people. People who are open to experience are intellectually curious, appreciative of art, and sensitive to beauty. They tend to be, compared to closed people, more creative and more aware of their feelings. They are more likely to hold unconventional beliefs. People with low scores on openness tend to have more conventional, traditional interests. They prefer the plain, straightforward, and obvious over the complex, ambiguous, and subtle. They may regard the arts and sciences with suspicion, regarding these endeavors as abstruse or of no practical use. Closed people prefer familiarity over novelty. They are conservative and resistant to change.

Conscientiousness is a tendency to show self-discipline, act dutifully, and aim for achievement. The trait shows a preference for planned rather than spontaneous behavior. It influences the way in which we control, regulate, and direct our impulses. The benefits of high conscientiousness are obvious. Conscientious individuals avoid trouble and achieve high levels of success through purposeful planning and persistence. They are also positively regarded by others as intelligent and reliable. On the negative side, they can be compulsive perfectionists and workaholics.

Extraversion is characterized by positive emotions and the tendency to seek out stimulation and the company of others. The trait is marked by pronounced engagement with the external world. Extraverts enjoy being with people, and are often perceived as full of energy. They tend to be enthusiastic, action-oriented. In groups they like to talk, assert themselves, and draw attention to themselves. Introverts lack the exuberance, energy, and activity levels of extraverts. They tend to be quiet, low-key, deliberate, and less involved in the social world. Their lack of social involvement should not be interpreted as shyness or depression. Introverts simply need less stimulation than extraverts and more time alone.

Agreeableness is a tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. The trait reflects individual differences in concern with for social harmony. They are generally considerate, friendly, generous, helpful, and willing to compromise their interests with others. Agreeable people also have an optimistic view of human nature. They believe people are basically honest, decent, and trustworthy. Disagreeable individuals place self-interest above getting along with others. They are generally unconcerned with others' well-being, and are less likely to extend themselves for other people. Sometimes their skepticism about others motives causes them to be suspicious, unfriendly, and uncooperative.

Neuroticism is the tendency to experience negative emotions, such as anger, anxiety, or depression. Those who score high in neuroticism are emotionally reactive and vulnerable to stress. They are more likely to interpret ordinary situations as threatening, and minor frustrations as hopelessly difficult. Their negative emotional reactions tend to persist for unusually long periods of time, which means they are often in a bad mood. At the other end of the scale, individuals are less easily upset and are less emotionally reactive. They tend to be calm, emotionally stable, and free from persistent negative feelings. Freedom from negative feelings does not mean that low scorers experience a lot of positive feelings. Frequency of positive emotions is a component of the Extraversion domain.

2.1.2 Moods

We choose a bio-inspired approach which tries to mimic the effects of three important monoamine neurotransmitters involved in the Limbic system. They are endogenous chemicals that transmit signals across synapses from neurons to other neurons. The three virtual neurotransmitters are:

Dopamine (D) is related to experiences of pleasure and the reward-learning process. It is a special neurotransmitter because it is considered to be both excitatory and inhibitory.

Norepinephrine (N) helps moderate the mood by controlling stress and anxiety. It is an excitatory neurotransmitter that is responsible for stimulatory processes.

Serotonin (S) is associated with memory and learning. An imbalance in serotonin levels results in an increase in anger, anxiety, depression and panic. It is an inhibitory neurotransmitter.

Let say that the values for these three virtual neurotransmitters are between 0.0 (minimum value) to 1.0 (maximum value). They form a three dimensional mood space called the "Lövheim Cube" of emotion (Lövheim, 2012). In this model, the three monoamine neurotransmitters form the axes of a 3D coordinate system, and the eight basic emotions, labeled according to the Affect Theory of Silvan Tomkins (Tomkins, 1991) are placed in the eight corners (cf. figure 1).

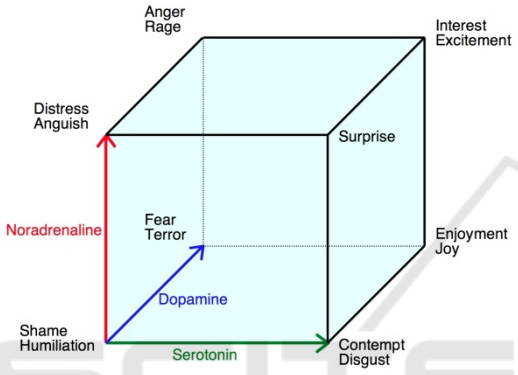

Figure 1: The Lövheim Cube.

This model attempts to organize affects into discrete categories and connect each one with its typical response. For example, the affect "joy" is observed through the display of smiling. There are eight basic affects (2 positives and 6 negatives) listed with a low/high intensity label for each affect and accompanied by its biological expression:

- **Enjoyment/Joy**: smiling lips wide and out.
- **Interest/Excitement**: eyebrows down, eyes tracking, eyes looking, closer listening.
- **Surprise/Startle**: eyebrows up, eyes blinking.
- **Anger/Rage**: frowning, a clenched jaw, a red face.
- **Contempt/Disgust**: the lower lip raised and protruded head forward and down.
- **Distress/Anguish**: crying, rhythmic sobbing, arched eyebrows, mouth lowered.
- **Fear/Terror**: a frozen stare, a pale face, coldness, sweat, erect hair.
- **Shame/Humiliation**: eyes lowered, head down and averted, blushing.

2.1.3 Emotions

Emotions are very short term affects with relatively high intensities. They are triggered by inducing events, which suddenly increase one or more neurotransmitters:

- D is both excitatory and inhibitory and mainly involved in pleasure/rewards.
- N is excitatory and increase active vs. passive feelings.
- S is inhibitory and increase positive vs. negative feelings.

After a short time, neurotransmitter values decrease due to a natural decay function. Most of the time, the system tends to return toward an attractor, which is a point in the system's phase space. This attractor is the transposition in the Lövheim Cube of the personality traits. This is not the neutral mood, which is by definition in the center of the 3D space:

$$D = N = S = 0.5$$

2.1.4 Primordial Emotions

Craig and Denton include pain in a class of feelings they name, respectively, "homeostatic" (Craig, 2003) or "primordial" emotions (Derek, 2006). These are feelings such as hunger, thirst and fatigue, evoked by internal body states, communicated to the central nervous system by interoceptors, which motivate behavior aimed at maintaining the internal milieu at its ideal state. They distinguish these feelings from the "classical emotions" such as joy, fear and anger, which are elicited by environmental stimuli. In our model, we choose to implement two basic primary emotions:

Energy is the machine transposition of internal feelings such as hunger, thirst and fatigue.

Pain measures the amplitude of unpleasant feelings often caused by intense, noxious or damaging stimuli.

2.2 Implementation

In this section, we describe an approach for implementing the previous layered model of affects using a connexionist neural-based architecture. The resulting implementation is called an Emotion Engine.

2.2.1 Anna

We choose to implement our own connexionist javascript-based framework called ANNA:

Algorithmic Neural Network Architecture. This architecture can be described as a deep highly non-linear neural network.

More precisely an application can include an arbitrary number of interconnected *networks*, each of them having its own interconnection pattern between an arbitrary number of *layers*. Each layer is composed of a set of *neurons*.

Each neuron has an arbitrary number of *weighted inputs*, a *single output*, and an *operator* function that computes the output given the inputs. This function can be a classical and homogenous activation function or any heterogeneous non-linear programmed function. In other words, each neuron can be programmed as a dedicated cell. Each input's weight can be learned using a machine learning algorithm, or statically programmed, or dynamically tuned by another network.

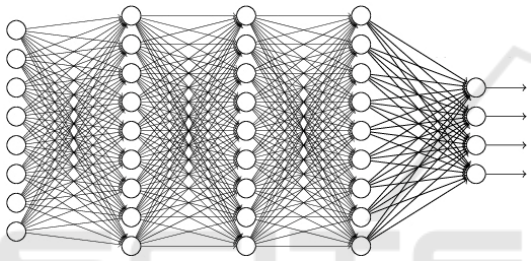

Figure 2: A typical ANNA deep feed-forward network with a full connection pattern between layers.

This general-purpose architecture enables to design any types of feed-forward, recurrent or heterogeneous complex sets of networks.

2.2.2 Emotion Engine

The next figure gives a simplified diagram of the Emotion Engine. Each square represents a small network composed of one or more layers. There are seven networks:

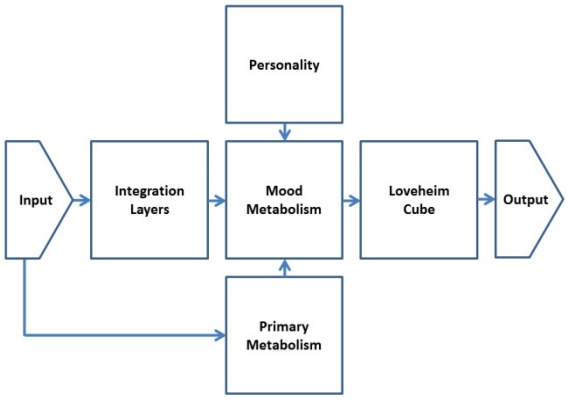

Figure 3: The Emotion Engine architecture.

Input: connect and convert external signals to the internal representation.

Integration: computes the three DNS virtual neurotransmitters spike values.

Personality: implements the personality layer based on the big five traits.

Primary Metabolism: implements the Energy and Pain system.

Mood Metabolism: computes the current DNS mood values with a decay function.

Lövheim Cube: convert the DNS mood values into the eight main affects using a distance calculation.

Output: select the emerging mood and computes its level.

Most of the networks use a feed-forward layers or a simple layer. All neurons have dedicated programmed operator functions with the exception of the Integration network, which uses a classical nonlinear weighted sum and a local supervised back-propagation learning scheme.

As an example, the Lövheim Cube implements a Euclidian distance function between the current DNS mood values and each main affect, that is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

The Emotion Engine is updated by propagating the inputs using a cyclic trigger called "lifepulse". The frequency of this signal ranges from a few milliseconds to a few seconds depending on the application.

3 THE LIVING MONA LISA

The first research prototype implementing the model of affects is the "Living Mona Lisa" installation. The aim of this project is to create an interactive and animated reproduction of the famous painting from Leonard da Vinci in the framework of the Living Art approach (Aziosmanoff, 2015).

3.1 Architecture Overview

The Living Mona Lisa architecture is based on three major and straightforward building blocks: the Sensory Module, the Artificial Intelligence Module and the Display Module. These three building blocks are connected together and form with the user(s) an interacting closed loop (cf. figure 4).

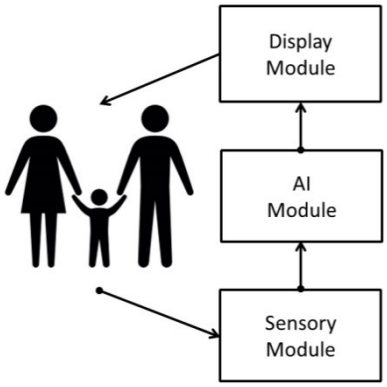

Figure 4: The Living Mona Lisa architecture. The sensory module captures the behaviors of spectators, the AI module computes Mona Lisa's emotional state, and the display module updates Mona Lisa's emotional expression.

3.2 Sensors

The sensory module is responsible for sensing the environment and sending pertinent information to the AI module. Typical information includes the presence of one or more persons, their position, their moves, facial expressions, recognition of some keywords, noise, etc.

The sensory module is implemented using a Microsoft Kinect 2 sensor system allowing the detection of up to six people with advanced facial tracking (Microsoft, 2014).

3.3 Emotion Engine

The AI module is the central part of the architecture, implementing the Emotion Engine described in this paper (cf. figure 5). The Primary Metabolism, that is Energy and Pain, was not implemented in the first version of the prototype.

The inputs of the Emotion Engine are connected to a set of 13 variables coming from the Sensory Module. The "lifepulse" update rate is set to 20 milliseconds. The "spike" value for each DNS signal is 0.01.

Mona Lisa's personality traits are Openness = 0.7; Conscientiousness = 0.8; Extraversion = 0.4; Agreeableness = 0.6; Neuroticism = 0.1. The Metabolism returns to this point in the DNS space at the decay rate of 0.01 per cycle.

The outputs are the following:

- The selected emotional expression as a string: "Neutral", "Shame", "Distress", "Surprise", "Disgust", "Fear", "Anger", "Interest", "Enjoyment".
- The intensity of this emotion: a float value in the range 0.0 to 1.0.
- The behavior of the eyes regarding the position of the main user: "Follow", "Avoid", "Fixed", "Indefinite". The face can move with an angle in the range of -50° to +50° left-right and up-down according to the normal in the center of the face.
- A "blink" trigger for the eyes.

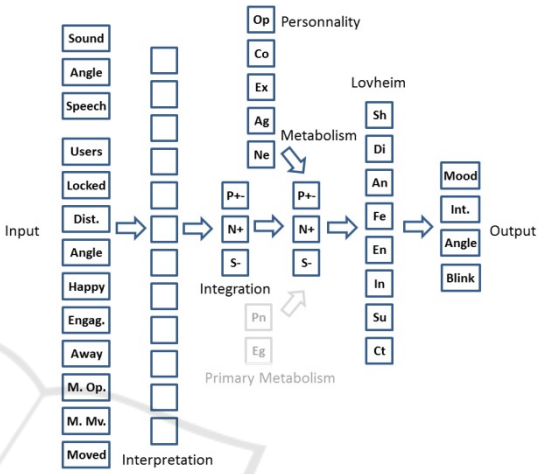

Figure 5: Diagram of the Emotion Engine's layers.

3.4 Display

The display module embodies the autonomous Mona Lisa character.

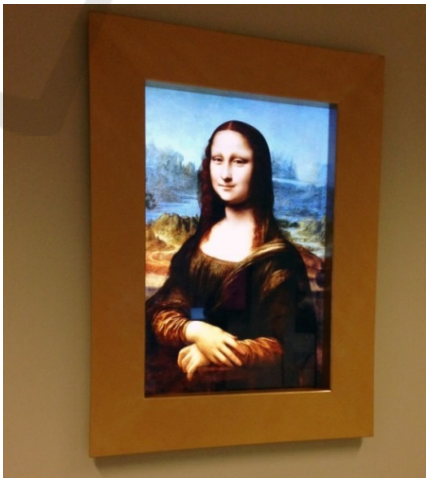

Figure 6: The Living Mona Lisa installation.

The display module is a reproduction of the painting at its original scale (77 cm x 53 cm not including the frame). The hardware uses an Ultra High Definition 55' LCD screen in portrait mode. The image is

composed of a 3D model made of over 500.000 polygons with advanced texturing techniques (cf. figure 7). We use the Unity Pro 3D rendering engine for real-time animation of this 3D model (Unity, 2015).

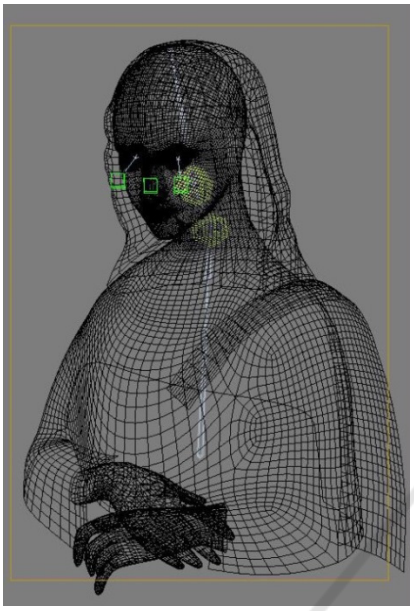

Figure 7: The Living Mona Lisa 3D model.

4 RESULTS

The next figures give some example views of the final rendering showing different facial expressions.

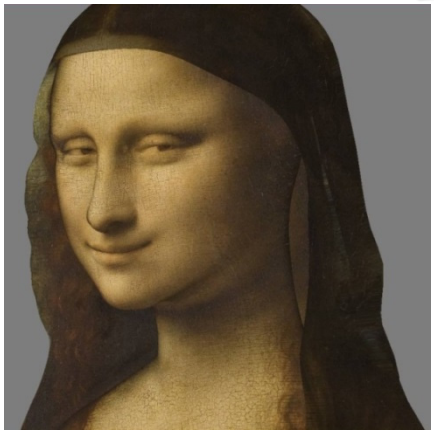

Figure 8: A close-up to the Living Mona Lisa with a new emotional expression compared to the genuine painting.

Figure 9/10: Living Mona Lisa showing its original expression (top) and an "interest" expression combined with face move and user follow behavior (down).

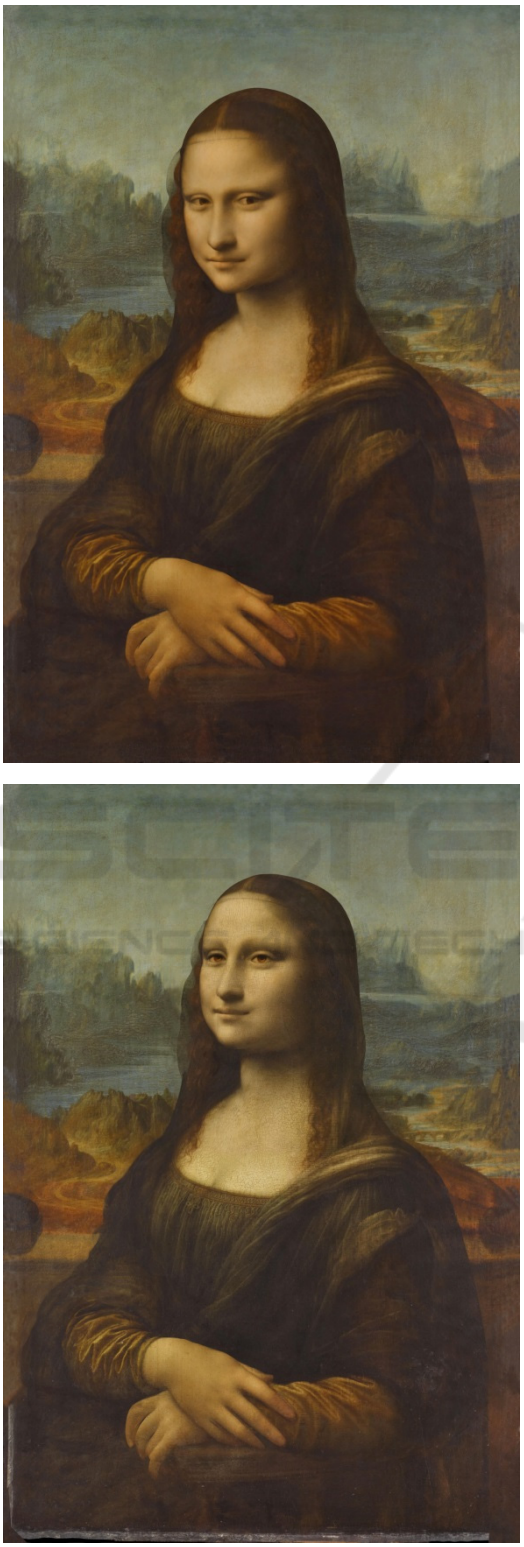

Figure 11/12: Living Mona Lisa showing an "anger" expression (top) and a "joy" expression combined with face move and "indefinite" eyes behavior (down).

The Living Mona Lisa prototype was showed to the public during *Futur en Seine* exhibition in Paris organized by *Cap Digital* during June 2015. First results were encouraging and show that the system was successfully functioning in the respect of the genuine masterpiece and known history of the real Mona Lisa character. All people interacting with the system were impressed by the "presence" of Mona Lisa.

However, there are some limitations that we must solve in the next version of the prototype. In the current version, the emotional behaviors are limited to the face of the character. Thus, we want to extend the expression to the entire body. For example the hands could have slightly moves according to emotions.

In the current version, the Emotion Engine selects one of the main emotions, one of the edges of the Lövheim Cube, according to the distance with the current DNS coordinates. A more realistic approach could be to express more subtle combinations of moods: for example one can be at the same time sad and surprised.

5 CONCLUSIONS

In this paper we introduced a layered model of affects for digital agents and an implementation as an Emotion Engine using a connexionist bio-inspired neuron-based architecture. The model provides three distinct affects: emotions as short-term affect, moods as medium-term affect, and personality as a long-term affect.

The model was implemented in an interactive installation called "Living Mona Lisa".

In parallel with the design of a new version of this "non-verbal" prototype, we also work for implementing the model in a "multi-personality" conversational agent in order to obtain a better balance between the different behaviors according to the context of the verbal interaction with a user.

ACKNOWLEDGEMENTS

The Living Mona Lisa project is partly funded by the French *Région Île-de-France*. We want to thank *Le Louvre Museum* and *La Réunion des Musées Nationaux* (RMN) for their contribution.

For their participation at every phase of the Living Mona Lisa project, I would like to thank Florent Aziosmanoff (author) and Dominique

Sciamma. Also, I would like to thank the Living Mona Lisa team at the Institute of Internet and Multimedia (IIM): Marc Bellan, Fabrice Houlné, Emanuel Perotti, Frédéric Rolland-Porché and all the students who have contributed to this project.

REFERENCES

Aziosmanoff, F., 2015. *Living Art Fondations*, CNRS Editions, Paris.

Becker, P., 2001. Structural and Relational Analyses of Emotion and Personality Traits. Zeitschrift für Differentielle und Diagnostische, *Psychologie*, 22, 3, 155-172.

Craig, A.D., 2003. Interoception: The Sense of the Physiological Condition of the Body. *Current Opinion in Neurobiology*, 13(4), 500-505.

Derek, A.D., 2006. *The Primordial Emotions: The Dawning of Consciousness*. Oxford University Press.

Gebhard, P., 2005. ALMA – A Layered Model of Affect. *Proceedings of the Fourth Int. Joint Conference on Autonomous Agents and Multiagent Systems*, 29-36.

Heudin, J.-C., 2004. Evolutionary Virtual Agent. *Proceedings of the Intelligent Agent Technology Int. Conference* (IAT 2004), IEEE/WIC/ACM, 93-98.

Heudin, J.-C., 2011. A Schizophrenic Approach for Intelligent Conversational Agent, *3rd Int. ICAART Conference on Agents and Artificial Intelligence*.

Lövheim, H., 2012. A New Three-dimensional Model for Emotions and Monoamine Neurotransmitters, *Med Hypotheses*, 78, 341-348.

McCrae, R.R. and John, O.P., 1992. An Introduction to the Five Factor Model and its Implications. *Journal of Personality*, vol. 60, 171-215.

Mehrabian, A., 1996. Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament, *Current Psychology*, vol. 14, 261-292.

Morris, W. N., 1989. *Mood: The Frame of Mind*. New York: Springer-Verlag.

Microsoft, 2015. *Kinect 2 Manual*.

Tomkins, S.S., 1962-1991. *Affect Imagery Consciousness*, vol. I-IV, London: Tavistock and New York: Springer.

Unity, 2015. www.unity3d.com.