Natural, affect aware interfaces: Gesture and Body Expressivity Aspects

Eirini Kalatha
Cultural Technology and Communication Department
University of the Aegean, Mytilene, Greece
ctm12017@ct.aegean.gr

George Caridakis[1 2]
Intelligent Systems, Content and Interaction Laboratory
National Technical University of Athens, Greece
gcari@image.ntua.gr

Abstract— **Recently, a growing number of researchers focus their work on a fledgling field related to the identification, recording, interpreting, processing and simulation of emotion and affective states. Affective computing, along with the wide range of related application areas, is expanded increasingly due to the emerging demand for natural, intelligent, adaptive and personalized multimodal interaction context. This paper focuses on the gesture and bodily expressivity analysis, investigating and reporting the entire spectrum of related aspects such as issues involved in affectively enhanced corpora (stimuli and input streams), respective expressivity formalization approaches as well as machine learning and pattern recognition aspects (feature extraction, multimodal fusion and classification). Finally, the synthesis counterpart of affectively enhanced behavior is examined and future directions of the research area are identified.**

Keywords— intelligent and adaptive interfaces, personalised and natural interaction, affective computing; gesture and body expressivity

I. Introduction

Current setting within many aspects of computing requires user analysis that goes beyond typical static, predefined, conventional user feedback and interaction. As computing goes social, multimodal and natural, adaptive and intelligent some aspects of the approach towards the user resemble the early years of Human Computer Interaction (HCI) where the user was sidelined and the overall design and implementation effort was focused on the machine counterpart. Accordingly, in the lightning fast evolution of computing today HCI seems to fall behind in terms of keeping pace with advancements. The requirements are set from research areas such as Intelligent, Personalized and Adaptive Interfaces, User and Group (Social) modeling and Context-Aware, Natural and Multimodal Interaction and Affective Computing faces the challenge of meeting them.

As part of the Affective Computing research area, non verbal user behavior analysis is attracting a continuously growing attention in bibliography. Gesture and full body, within the Natural Interaction context, is considered one of the underrepresented modalities researched in this area when compared to speech and facial expressions. Even Physiology and Electroencephalography signals have been more researched perhaps due to their strong correlation to Medical disciplines. Current paper surveys work on body movement qualitative analysis and formulization and focuses on expressivity aspects of this non verbal modality. It reports on the entire spectrum of aspects and intermediate steps that qualitative, expressivity analysis of body motion and gesturing require. On the one hand, related work on higher level issues are discussed from the modeling and formalization point of view whereas more technical aspects are also presented. Corpora are examined in terms affective enhancement and stimuli and recorded modalities and streams. Moreover, respective expressivity formalization approaches and their correlation with broader emotion modeling are discussed. Approaching the problem from a Digital Signal Processing, Machine Learning and Pattern Recognition perspective respective aspects are also discussed.

Regarding the origins of affective computing, even if they date back to early philosophical enquiries concerning emotion, the more modern branch of affective computing is defined by Rosalind Picard [38]. In this paper the nature of motivation, feelings, and emotions is sought, the detection of various emotional states is researched and the requirements for effective interaction applied to a variety of conditions between humans and computers are refereed [42].

II. Background Information

A. Emotion

The meaning of emotions has for many years remained uncommented due to the fact that the meaning of the term “emotion” is neither unique nor accurately defined. Emotions are classified into two main categories: “primary” and “secondary” emotions. The term “primary emotions” characterizes the emotions which are experienced by the social mammals and they are represented as an immediate consequence of some kind of event related to particular types of expression and changes; such as facial expressions, behavioral tendencies, and physiological standards. “Secondary emotions” constitute ‘wealthy’ emotional states, which emerge from cognitive procedures including estimation of perceptible or imaginary conditions.

Many of the researchers have examined a wide range of emotions but few of them have managed to represent the emotion as a usable result in the field of graphic and computer vision. The two most remarkable and popular representations are Whissel’s Wheel of Emotions and Plutchick’s Wheel of Emotions. Whissel’s Wheel of Emotions is a simple representation which is able to “describe” a wide range of emotions. It is based on valence and activation level. The term valence is referred to the impact of various emotional states or other individuals to a specific individual. This impact is characterized either as positive or negative and it is a common characteristic of various emotional states due to the fact that all

[1] George Caridakis is also affiliated with the Cultural Technology and Communication Department, University of the Aegean

of them have an impact. The activation level is based on the observation that the experiencing of each emotional state has as consequence the tendency of expressing specific reactions. This observation led to the exertion to characterize the expression on the basis of the facial activation. Plutchick's Wheel of Emotions is a three-dimensional model of emotions that describes the relations among emotion concepts, which are analogous to the colors on a color wheel.

In retrospect, according to some recent findings it has been proved that emotion consists an ubiquitous feature of any HCI and the systems' designers have to take it into account in order to design usable and intelligent systems [27].

B. *Theory of Bodily Motion*

The topic of whether body movements and body postures are indicative of specific emotions has been a controversial issue, due to the fact that some studies have found evidence for specific body movements accompanying specific emotions whether others indicate that movement behavior may be only indicative of the intensity of emotion, but not of its quality. For the time being the bodily expression of felt emotion is associated with emotion-specific changes in gait parameters and kinematics [13].A wide range of disciplines has been dealt with bodily motion such as psychology, kinesiology, choreography, computer graphics and many others. To be more specific, in psychology, studies related to body motion with moving light displays (MLD) and gestures have been conducted. In kinesiology, models of the human body have been developed in order to explain how it functions mechanically and how one might increase its movement efficiency.

Moreover, in computer graphics the researchers were dealt with the synthesis of bodily motion whereas in choreography with the modulation of bodily movement's high-level descriptions for the notation of dance, ballet, and theatre. Regarding the most common notations the Labanotation (LN), the Ekshol–Wachmann, and the effort–shape notation have been included. In the context of this paper it is appropriate to describe the LN theory. First and foremost, the Labanotation model has been introduced by Rudolf von Laban in 1928, who was a dance artist and theorist. This system uses abstract symbols in order to describe movement and provides a well-structured language with rich vocabulary and clear semantics, based on Laban Movement Analysis (LMA) [17] which serves both for designing dance documentation software and modeling human computer interaction based on movement and gestures [31], [40], [17].

Furthermore, LN systems' usage is widespread in the scientific community as far as analysis related to dance, theater, sports and gymnastics is concerned. The main feature of this particular system is the fact that it captures both the directional and the qualitative aspects of movement [17]. As mentioned above it is a widespread due to the fact a great number of studies such as [17], [23], [24], [30], [31], [40], [34] and [45].

III. CORPORA

A. *Stimuli*

The main stimuli used in related studies in order to trigger the manifestation of gestures and body movements are dance, music, film and virtual reality (i.e. avatars). As far as dance stimulus is concerned many applications exist that utilize it. For instance in [10], an implementation of Laban annotation scheme for dance is presented. The model acts on the effort and shape components of the LMA and the way a gesture looks depending on values such as the strength of the gesture and its tempo. Moreover, in [4] expressive gestures in human full-body movement, namely in dance performances are classified.

Regarding the stimulus of music some of the applications that utilize it are [6], [14] and [15]. Specifically, in [6] a system allowing users to express themselves through their full-body movement and gesture and to control in real-time the generation of an audio-visual feedback is described. The systems analyses in real-time the user's full-body movement and gesture, extracts expressive motion features and maps the values of the expressive motion features onto real-time control of acoustic parameters for rendering a music performance. In [14] has been showed that the movement of the head provides important information for identifying the emotional intention in marimba performances and in [15] that the head movement is important for observers to discriminate between pianist's performances played with different expressive intentions.

As mentioned above even a film or a part of it can be used as a stimulus. For example in [18] cross-cultural differences were compared in the facial expressions of Japanese and American subjects while viewing both neutral films and films intended to cause stress [29]. One of the conclusions drawn by the survey was the fact that that the Japanese viewers were more unwilling to show their real expressions. In addition in people's physiological reactions as they view films in a controlled cinema-like environment are recorded.

Virtual reality consists another stimulus which is widely used in related studies in order to trigger the manifestation of gestures and body movements. For instance in [39] a virtual rap dancer that invites users to join him in a dancing activity is designed. To be more specific, users' dancing movements are tracked by a video camera and guide the virtual rap dancer in his own dance movements and gestures [33].

B. *Sensors and inputs*

Sensors constitute basic components that give substance to the concept of affective computing; owing to the fact that their primary function is to detect a signal or stimulus and produce a measurable output. The sensors related to affective computing from the gesture and body expressivity perspective are divided into four main categories which are Motion Capture (MC) system, stereo camera, image sensors and accelerometer.

Specifically, MC constitutes the process of recording the movements or postures in the real word. It is used widespread in a wide range of fields such as in military, entertainment, sports, medical applications and for computer vision and robotics validation . MC systems are divided into two subcategories; optical and non-optical systems and MC technology is characterized by several advantages. For instance, it is a rapid and real-time technique, cheap and requires minimal device calibration. Some of the disadvantages of MC are the fact that the movement that does not follow the laws of physics cannot be captured, its sensitivity to metal and the limited range (10-15 m^2) that it has. Many applications exist that use MC systems. For instance in [17] the researchers use motion capture or video technology in order to record a particular performance of specific dancers including personal style-even mistakes- rather than the choreography itself. Moreover, in [13] a type of MC is used, named optoelectronic stereophotogrammetry. To be more precise, this method is based on passive retro-reflective markers which are placed on specific anatomical landmarks on the body. Each marker's position is tracked using high-speed video cameras and the each marker's centroid position is calculated for each instant in time [13]. In addition, in [1] a motion capture system or a video camera enable to capture a human's affective posture. The main purpose of this study is to map a gestural pattern implementing a CALM network. In [29] a MC system is used in order to collect 3D affective postures of 13 human subjects and its data are used in order to build affectively expressive avatars.

The stereo camera approach consists a method of distilling a noisy video signal into a coherent data set that a computer can begin to process into actionable symbolic objects, or abstractions. It is one of the approaches that used in the broader fields of computer vision and machine vision. To be more specific, in stereo camera approach, two cameras with a known physical relationship are correlated via software. By finding mappings of common pixel values, and calculating how far apart these common areas reside in pixel space, a rough depth map can be created.

The stereoscopic image processing technique is used in applications such as robotic control and sensing and crowd dynamics monitoring. Xbox Kinect sensor constitutes a variant of this technique due to the fact that it is able to create a depth map of an image by using an infrared camera. One of the studies which uses stereo camera technique and it has been studied in the context of this essay is in [47] where a stereo camera located on top of a wall size display allows to continuously track the 3D position of the head and hand of a user in a large interaction space.

An image sensor is a device that converts an optical image into an electronic signal and it is used mostly in digital cameras, camera modules and other imaging devices. It is one of the most common sensors and it is used in a great number of studies. For instance in [29], static posture images of affectively expressive avatars are used in order to conduct recognition experiments with subjects from three cultures.

Accelerometer inertial sensors consists one of the most common approach which is capable of a vast range of sensing. It enables to measure acceleration forces in one, two, or three orthogonal axes and it is typically used in one of three modes below:

- as an inertial measurement of velocity and position,
- as a sensor of inclination, tilt, or orientation in 2 or 3 dimensions or
- as a vibration or impact sensor.

The acceleration forces may be static or dynamic - caused by moving or vibrating the accelerometer. One of the studies using this type of sensors is [4], where analysis and classification of expressive gesture in human full-body movement and in particular in dance performances are presented.

C. *Feauture extraction and expressivity formalisation*

Processing of the signals which are described above must take place in order to draw precise conclusions related to emotional computing. Fourier transformation consists one of the most common signal processing method that is used in several applications, as for example in [27] and in [44]. The basic traits of this method are the fact that exams signals in three dimensions: width, time, frequency and contributes to the development of effective techniques, which deal with the noise introduced by the channel to the broadcast signal.

Cohen's kappa coefficient is a statistical measure of agreement for categorical items. In [19] was first mentioned the term of kappa like statistic. Generally, it is characterized as a robust measure due to the fact that κ takes into account the agreement occurring by chance. Specifically, Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories.

Video processing consists a special case of signal processing, where input and output signals are video files or video streams. In [22] the existence of a number of early efforts to detect non basic affective states, such as attentiveness, fatigue and pain from face video is referred.

There exist techniques from the image and signal processing domain which can be applied to define motion sequences, such as motion multiresolution filtering. In particular multiresolution filtering describes a range of digital filter-bank techniques which typically pass a signal through a cascade of lowpass filters to produce a set of short-time bandpass or low-pass signal components.

In [48] the authors investigate the expressivity formalization of natural interaction, namely 3D body movement, by performing Principal Component Analysis on different approaches. The approaches that are validated are silhouette, motion volumes and joint rotations and the introduced Fading Silhouette Motion Volumes (FMSV) is the most efficient in terms of natural interaction expressivity formalization.

The most comprehensive approach modeling expressiveness of the body is the dimensions of expressiveness due to the fact that they cover the entire spectrum of the expressive parameters which are associated with emotion. According to related bibliography six parameters of expressiveness are defined. Each of these parameters varies according to the expressive medium in which it applies. In the case of gestures, the expressiveness is determined at the different phases of each gesture and by the way two gestures are combined. Six expressivity parameters that are quite popular are:

- Overall activation: consists the movement's 'quantity' during a conversation
- Spatial extent: characterizes the expanding or condensing of the space that is used in front of the user (space gestures)
- Temporal: indicates the speed of hand's movement during a gesture and distinguishes fast from slow gestures
- Fluidity: distinguishes smooth / elegant than sudden / abrupt gesture and shows the continuity between movements
- Power/Energy: refers to the dynamic qualities of the movement
- Repetitivity: the tendency toward a rhythmic repetition of a certain movement. [5]

D. Machine learning approaches

User's emotions, expressions or physiological responses are sorted out through machine learning approaches owing to the fact that this procedure consists the main aim of affective computing applications. Some of the most common machine learning methodologies are neural networks (NNs), Support Vector Machines (SVMs) classification, Hidden Markov Models (HMMs) and decision trees. First and foremost, NNs constitute machine learning methods, based on biological standards, structures and using processes that imitate those of the human brain. NNs are very popular in the field of emotional computing because of their basic traits. To be more precise, they are able to learn by example, are capable of pattern recognition, are considered as distributed memory and memory correlation and finally, they present high tolerance to noisy learning data. There are many applications in which this method was used as a classification method. For instance, in [28] a NN is utilized in order to learn how the emotion state should be influenced by the occurrence of stimuli. Moreover, in [46] a NN has been used in order to recognize facial expressions and in [12] it has been used in order to classify emotional expressions such as joy, sorrow, disgust, surprise, terror, neutrality and angriness.

In [22] it is mentioned that the multilayer perceptron (MLP) back-propagation artificial neural network consists the simplest feedforward neural network that uses back propagation to train. Various layers of neurons, each one of them completely connected to the next one, comprise an MLP. Also, in [26] it has been shown that an MLP, which is used in order to identify different human emotions (sadness, joy, anger, and fear) from gestures, was one of the most effective classifiers for their study.

A widespread learning algorithm, which is used to predict a model's output values, based on input values that are presumed to be independent, are decision trees. Some of the most notable advantages of this method are the fact that it is inexpensive to construct and easy to interpret, it provides an extremely fast method for classifying unknown records and its accuracy is comparable to that of other classification techniques, for many simple data sets. It is one of the most common classification methods which is used in a great amount of applications. In [36]a decision tree learning algorithm has been used in order to identify the different morphological and dynamic characteristics of the amused, polite, and embarrassed smiles of the corpus. C4.5 classifier constitutes a specific decision tree which is widespread used. In particular, it is a generator of supervised symbolic classifiers based on the concept of entropy. In [8], an implementation of the C4.5 decision tree was adopted during the researchers' investigation related to the dynamic variations of gestures used by a pianist to communicate emotional expressive intentions while playing. Furthermore, in [26] a decision tree classifier based on the C4.5 algorithm constitutes one of the five different classifiers were used in the learning experiments.

SVM classification consists a relatively modern method of supervised machine learning. SVMs learn to distinguish between two categories, in their simplest form, whereas with diminutive algorithmic differences, they can separate examples from several categories. Some of the most notable advantages of this method are its capacity for quality of generalization and ease of training and the fact that it can model complex, real-world problems and tends to perform well even on multivariate data sets. This method is widespread in applications related to affective computing. For example, in [37] SVM has been used for pose recognition. To be more specific, in order to train the SVM classifiers, the researchers have applied images synthesized using motion capture data and animation software as well as real images captured by two video cameras.

HMMs constitute generative models, which are the most commonly used techniques for the recognition of affective states and for the detection of the temporal segments. A great number of studies, such as [24], have detected the temporal segments of facial expressions by using HMMs. In addition, in [32] is presented continuous recognition of the emotional content of movies using a HMM which classifies dimensional attributes into discrete levels. In particular, each video frame is classified into one of seven discrete categories. Furthermore, HMM is used in [37] for dynamic gesture recognition.

Classic machine learning methods such as HMM and neural network, which are described above, can be considered as special cases of Bayesian networks (BNs). Bayesian network consists a probabilistic graphical model (GM). Substantially, it is a high-level representation of a probability distribution over a set of variables that are used for building a model of the problem domain. The most remarkable trait of BN is the fact that its structure can be used as a compact representation for many naturally occurring and complex problem domains.

In addition, BN constitutes a widespread technique that is one of the state of the art techniques for emotion recognition. For instance, in [7] the researchers trained and tested a model with a Bayesian classifier, using a multimodal corpus with eight acted emotions and ten subjects of five different nationalities. To be more specific, a separate Bayesian classifier was used for each modality i.e. face, gestures and speech. Moreover, in [26], which is described above, one of the five different classifiers which were used in machine learning experiments was a naïve bayes classifier.

The nearest-neighbor method is perhaps the simplest of all machine learning algorithms algorithms. Precisely, the k-nearest neighbor algorithm (k-NN) is a non-parametric method for classifying objects and a type of instance-based learning. There exist many studies where this algorithm is used as a machine learning approach, as for instance in [8]. In this study a nearest neighbor method based on DTW distance (1NN-DTW) is used as the classifier for the system that is presented. As the researchers mentioned a new gesture is assigned to the gesture's emotion in the training set to a minimum distance, that is its nearest neighbor.

Furthermore, in [41] k-NN and Bayesian are used as classification techniques in order to recognize isolated sign. To be more specific, in this paper various spatio-temporal feature-extraction techniques with applications to online and offline recognitions of isolated Arabic Sign Language gestures are mentioned.

IV. SYNTHESIS COUNTERPART

Nowadays, agents, especially embodied conversational agents (ECA) have received large amounts of interest owing to the fact that a great amount of researchers are dealt with the potential for embodied agents to enhance interactions with computers. Agents are autonomous software modules with perception and social ability to perform goal-directed knowledge processing over time, on behalf of humans or other agents in software and physical environments and they are divided in intelligent or cognitive agents and in emotive agents. Specifically, intelligent or cognitive agents have cognitive abilities such as anticipation, understanding, learning and communication in natural language whereas some of the most remarkable emotive agents' traits are the fact that they have cognitive knowledge processing abilities, they can have personalities and can detect personalities of others, they are sensitive to emotional inputs and they can understand emotions of self, as well as of others and they can use the understood emotions for further knowledge processing and actions.

On the other hand the Embodied Conversational Agent is anthropomorphic virtual simulation that communicates multimodally with the user through voice, facial expression, gaze, gesture, and body movement and consists one of the most widespread agent. The emotion's expressivity and the personality's exhibition in a consistent manner contribute to the increase of the believability and the life-likeness of an agent.

As far as the applications related to embodied agents are concerned, a wide range of fields have dealt with them. For instance, experiments in health [2], learning [3], commerce, therapy [21], video games [25], and military systems [20] have been conducted. Some of the applications related to virtual agent are [16], [43] and [35]. To be more specific in [16] a generic, modular and interactive architecture for embodied conversational agent, called Greta, is presented. In [33] a system that takes video data as input, extracts movement characteristics, and finally synthesizes the animation of a virtual agent is referred. For each gesture performed by a human, the agent responds with a gesture that exhibits the same characteristics of movement. Moreover, in [43] the AVLaughterCycle application is presented. Specifically, this application endows a virtual agent with the capability of joining its conversational partner's laughs, by displaying a laugh response related to an input laughter. In [35] an embodied agent that acts as a dancer and invites human partners to dance with her is presented.

In retrospect, while agents with affective behavior are becoming even more popular as virtual interaction partners, their effect on human interaction partners is not so renowned, therefore many challenges related to affective computing in conjunction with virtual agents emerge.

V. CONCLUSIONS

Current paper surveys work on body movement qualitative analysis and formalization and focuses on expressivity aspects of this non verbal modality. It reports on the entire spectrum of aspects and intermediate steps that qualitative, expressivity analysis of body motion and gesturing require. On the one hand, higher level issues and related work are discussed from the modeling and formalization point of view whereas more technical aspects are also presented. Corpora are examined in terms affective enhancement and stimuli and recorded modalities and streams. Moreover, respective expressivity formalization approaches and their correlation with broader emotion modeling are discussed. Approaching the problem from a Digital Signal Processing, Machine Learning and Pattern Recognition perspective respective aspects are also discussed.

VI. ACKNOWLEDGMENTS

Current research work has been partially funded by the Greek Ministry of Education, Education and Lifelong Learning program, project code MIS 380328 "Cross-disciplinary research in Affective Computing for physiology and activity recognition in assistive environments - STHENOS".

REFERENCES

[1] Bianchi-Berthouze, N., & Kleinsmith, A. (2003). A categorical approach to affective gesture recognition. *Connection science , 15* (4), pp. 259-269.

[2] Bickmore, T. and Picard, R. (2005) "Establishing and Maintaining Long-Term Human-Computer Relationships" ACM Transactions on Computer Human Interaction (ToCHI), 59(1): 21-30.

[3] Burleson, W., & Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. Intelligent Systems, IEEE, 22(4), 62-69.

[4] Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies , 59* (1), pp. 213-225.

[5] Caridakis, G., & Karpouzis, K. (2011). Full body expressivity analysis in 3D Natural Interaction: a comparative study. *Affective Interaction in Natural Environments workshop, ICMI 2011 International Conference on Multimodal Interaction*. Spain.

[6] Castellano, G., Bresin, R., Camurri, A., & Volpe, G. (2007). User-centered control of audio and visual expressive feedback by full-body movements. Berlin: Springer.

[7] Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech. In *Affect and Emotion in HCI* (pp. 92-103). Springer Berlin Heidelberg.

[8] Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., & Scherer, K. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception , 26* (2), pp. 103-119.

[9] Castellano, G., Villalba, S. D., & Camurri, A. (2007). Recognising Human Emotions from Body Movement and Gesture Dynamics. In *Affective computing and intelligent interaction* (pp. 71-82). Berlin: Springer.

[10] Chi, D., Costa, M., Zhao, L., & Badler, N. (2000). he EMOTE model for effort and shape. *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (pp. 173-182). ACM Press/Addison-Wesley Publishing Co.

[11] Cohen, I., Sebe, N., Garg, A., Chen, L., & Huang, T. S. (2003). Facial Expression Recognition from Video Sequences: Temporal and Static Modeling. *Computer Vision and Image Understanding , 91*, pp. 160-187.

[12] Courgeon, M., Martin, J.-C., & Jacquemin, C. (2008). MARC: a Multimodal Affective and Reactive Character. *Proceedings of the 1st Workshop on AFFective Interaction in Natural Environments.*

[13] Crane, E., & Gross, M. (2007). Motion Capture and Emotion: Affect Detection in Whole Body Movement . In *Affective computing and intelligent interaction* (pp. 95-101). Berlin: Springer.

[14] Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception , 24* (5), pp. 433-454.

[15] Davidson, J. W. (1994). What type of information is conveyed. *Journal of Human Movement Studies , 6*, pp. 279-301.

[16] de Sevin, E., Niewiadomski, R., Bevacqua, E., Pez, A. M., Mancini, M., & Pelachaud, C. (2010). Greta, une plateforme d'agent conversationnel expressif et interactif. *Technique et Science Informatiques , 29* (7).

[17] El Raheb, K., & Ioannidis, Y. (2012). A labanotation based ontology for representing dance movement. In *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication* (pp. 106-117). Berlin: Springer.

[18] Friesen, W. V. (1973). Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules. PhD Thesis.

[19] Galton, F. (1892). Finger Prints Macmillan.

[20] Gratch, J., & Marsella, S. (2004). Adomain-independent framework for modeling emotion. Cognitive Systems Research, 5(4), 269-306.

[21] Grolleman, J., van Dijk, B., Nijholt, A., & van Emst, A. (2006). Break the habit! Designing an e-therapy intervention using a virtual coach in aid of smoking cessation. In Persuasive Technology (pp. 133-141). Springer Berlin Heidelberg.

[22] Gunes, H., & Piccardi, M. (2009). Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics , 39* (1).

[23] Hatol, J. (2006). *MOVEMENTXML: A representation of semantics of human movement based.* A thesis submitted for the Master Degree in the School of Interactive Arts and Technology, Simon Fraser University.

[24] Hutchinson-Guest, A. (2005). Labanotation: The System of Analyzing and Recording Movement. New York: 4th edn. Routledge.

[25] Isbister, K. (2006). Better game characters by design: A psychological approach. Taylor & Francis US.

[26] Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., & Driessen, P. F. (2005). Gesture-Based Affective Computing on Motion Capture Data. *Affective Computing and Intelligent Interaction* , pp. 1-7.

[27] Karg, M., Kuhnlenz, K., & Buss, M. (2010). Recognition of affect based on gait patterns. *Systems, Man, and Cybernetics, Part B: Cybernetics , 40* (4).

[28] Kesteren, J. A. (2001). *A supervised machine-learning approach to artificial emotions.* University of Twente, Department of Computer Science.

[29] Kleinsmith, A., De Silva, P. R., & Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers , 18* (6), pp. 1371-1389.

[30] Kojima, K., Hachimura, K., & Nakamura, M. (2002). LabanEditor: Graphical Editor for Dance Notation. *Proceedings 11th IEEE International Workshop on Robot and Human Interactive Communication.*

[31] Loke, L., Larssen, A. T., & Robertson, T. (2005). Labanotation for Design of Movement-Based Interaction. *Proceedings of Second Australasian, Conference on Interactive Entertainment.* Australia.

[32] Malandrakis, N., Potamianos, A., Evangelopoulos, G., & Zlatintsi, A. (2011). A supervised approach to movie emotion tracking. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*

[33] Mancini, M., & Castellano, G. (2007). Real-time analysis and synthesis of emotional gesture expressivity. *Proc. of the Doctoral Consortium of Intl. Conf. on Affective Computing and Intelligent Interaction.*

[34] Nakamura, K., & Hachimura, K. (2006). An XML representation of Labanotation, LabanXML, and its implementation on the notation editor Laban Editor. *Review of the National Center for Digitization (Online Journal) , 9*, pp. 47-51.

[35] Nijholt, A., Reidsma, D., & Welbergen, H. (2006). Establishing Rapport with a Virtual Dancer.

[36] Ochs, M., Niewiadomski, N., Brunet, P., & Pelachaud, C. (2012). Smiling virtual agent in social context. *13* (2), pp. 519-532.

[37] Peng, B., Qian, G., Ma, Y., & Li, B. (2011). Multifactor feature extraction for human movement recognition. *Computer Vision and Image Understanding , 115*, pp. 375-389.

[38] Piccard, R. (1997). Affective Computing. *MIT press.*

[39] Reidsma, D., Nijholt, A., Poppe, R., Rienks, R., & Hondorp, H. (2006). Virtual rap dancer: invitation to dance. *CHI'06 extended abstracts on Human factors in computing systems* , pp. 263-266.

[40] Rett, J., Dias, H., & Ahuactzin, J. (2010). Bayesian reasoning for Laban Movement Analysis used in Human Machine Interaction. *Reasoning-based Intelligent Systems , 1*, pp. 13-35.

[41] Shanableh, Tamer, Assaleh, K., & AI-Rousan, M. (2007). Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language. *Systems, Man, and Cybernetics, Part B: Cybernetics , 37* (3), pp. 641-650.

[42] Sloman, A. (1999). Review of: Affective Computing. *AI magazine* .

[43] Urbain, J., Niewiadomski, R., Bevacqua, E., Dutoit, T., Moinet, A., Pelachaud, C., et al. (2010). AVLaughterCycle. *Journal on Multimodal User Interfaces , 4* (1), pp. 47-58.

[44] Weinland, D., Ronfard, R., & Boyer, E. (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding , 104* (2), pp. 249-257.

[45] Wilke, L., Calvert, T., Ryman, R., & Fox, I. (2005). From dance notation to human animation: The LabanDancer project. *Computer Animation and Virtual Worlds , 16* (3-4), pp. 201-211.

[46] Yasunari, Y., Sung-Ill, K., Takako, K., & Tetsuro, K. (2000). Effect of Sensor Fusion fo Recognition of Emotional States Using Voice, Face Image and Thermal Image of Face. Proceedings. 9th IEEE International Workshop on. IEEE.

[47] Viallet, J. E., & Carbini, S. (2007). A simple hand posture recognition for efficient single handed pointing and selection. *Proceedings of GW2007-7th International Workshop on Gesture in Human-Computer Interaction and Simulation.*

[48] G. Caridakis, K.Moutselos, I. Maglogiannis, "Natural Interaction expressivity modeling and analysis", 6th International Conference on PErvasive Technologies Related to Assistive Environments, Rhodes, Greece, May 29-31, 2013