Engineering Applications of Artificial Intelligence ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Associating gesture expressivity with affective representations

Lori Malatesta [a], Stylianos Asteriadis [b], George Caridakis [a,d], Asimina Vasalou [c], Kostas Karpouzis [a]

[a] Image, Video and Multimedia Lab, National Technical University of Athens, Greece
[b] Department of Knowledge Engineering, University of Maastricht, Netherlands
[c] London Knowledge Lab, UCL, United Kingdom
[d] Department of Cultural Technology, University of the Aegean, Greece

ARTICLE INFO

Keywords:
Affective computing
Gesture expressivity parameters
Association of representations
Modelling expressivity
Expressivity judgment study
Neuro-fuzzy network

ABSTRACT

Affective computing researchers adopt a variety of methods in analysing or synthesizing aspects of human behaviour. The choice of method depends on which behavioural cues are considered salient or straightforward to capture and comprehend, as well as the overall context of the interaction. Thus, each approach focuses on modelling certain information and results to dedicated representations. However, analysis or synthesis is usually done by following label-based representations, which usually have a direct mapping to a feature vector. The goal of the presented work is to introduce an interim representational mechanism that associates low-level gesture expressivity parameters with a high-level dimensional representation of affect. More specifically, it introduces a novel methodology for associating easily extracted, low-level gesture data to the affective dimensions of activation and evaluation. For this purpose, a user perception test was carried out in order to properly annotate a dataset, by asking participants to assess each gesture in terms of the perceived activation (active/passive) and evaluation (positive/negative) levels. In affective behaviour modelling, the contribution of the proposed association methodology is twofold: On one hand, when analysing affective behaviour, it can enable the fusion of expressivity parameters alongside with any other modalities coded in higher-level affective representations, leading, in this way, to scalable multimodal analysis. On the other hand, it can enforce the process of synthesizing composite human behaviour (e.g. facial expression, gestures and body posture) since it allows for the translation of dimensional values of affect into synthesized expressive gestures.

1. Introduction

Human behaviour has been often studied both for the purposes of inferring internal affective aspects from observations, and for fostering the quality and believability of synthesized actions of virtual characters. Whether studying low level behavioural cues – and, thus, focusing on aspects like gestures, body, posture and facial expressions – or emphasizing high level phenomena such as affective states and dispositions, often the case is that researchers will adopt different computational models or representations in order to model emotions and related phenomena. Such variation is not necessarily problematic since it enables the investigation of subtle differences amongst approaches. Nevertheless, when it comes to building an affect aware system (either targeting analysis, or synthesis components) such variations raise interesting questions regarding the correspondence between entities in each representation scheme.

Affective cues can be broadly categorised across two broad families, verbal and non-verbal communication channels. In the presented work, we focus on the analysis of non-verbal behavioural channels. A substantive body of research has focused on non-verbal behaviour, within the fields of psychology, cognitive science and human computer interaction, stressing the importance of *qualitative* expressive characteristics of body motion, posture, gestures, facial expressions (Ioannou et al., 2007), eye gaze (with eye gaze still necessitating specialized hardware (Bengoechea et al., 2013; Jennett et al., 2008) and overall human action recognition during an interaction session (Wallbott, 1998; Pelachaud, 2008; Knapp and Hall, 2013). Qualitative affective cues contain significant information about the user's non verbal behaviour and communication. In this context computationally formulated qualitative expressivity features (e.g. fluidity of a gesture performance) correspond to the intermediate layer between extracted quantitative features (e.g. coordinates of the hand position) and the conveyed emotion in the form of the adopted affective representation approach.

Behaviour expressiveness is an integral part of the communication process since it can provide information about the current

E-mail addresses: lori@image.ntua.gr (L. Malatesta), stelios.asteriadis@maastrichtuniversity.nl (S. Asteriadis), gcari@image.ntua.gr, gcari@aegean.gr (G. Caridakis), A.Vasalou@ucl.ac.uk (A. Vasalou), kkarpou@image.ntua.gr (K. Karpouzis).

http://dx.doi.org/10.1016/j.engappai.2016.01.010
0952-1976/

emotional state and the profile of the interlocutor, as well as his performance.

Although the research field of human behaviour analysis has primarily focused on human to human interaction, there has been growing recognition of the importance of accounting for Human-Computer Interaction (HCI). Most examples of studies incorporating gesture expressivity in the HCI context (Vinciarelli et al., 2012; Caramiaux et al., 2015), however, have tended to focus on the expressively-enhanced *synthesis* of gestures by virtual agents and ECAs (Caridakis et al., 2007; Cassell et al., 2004; Martin et al., 2006; Kipp et al., 2007; Pelachaud, 2008; Hartmann et al., 2005) which, many times, follow animation patterns that depend on low-level features (e.g. tracking) and only partly depend on semantic interpretation of human's emotional or cognitive state.

The present work focuses on expressivity in gesturing. The approach adopted is holistic in the sense that the gestures studied are not recognized or broken down to their components. Emphasis is given to the expressive content of a closed set of singular gestures with clear semantic meaning (such as waving goodbye or clapping). The contribution of the research work presented in this article lies in the association of existing results on automatically calculated expressivity parameters (Caridakis et al., 2006), with dimensional representations of affect. This is done by incorporating a properly, data-driven trained neuro-fuzzy network. The proposed association allows for the inclusion of expressivity parameters in the fusion process with other modalities that commonly use dimensional representations of affect.

2. Related work

In computational behaviour analysis, according to a survey paper by Kleinsmith and Bianchi-Berthouze (2013), research on non-verbal affect recognition has mostly focused on facial expressions starting with the FACS coding system developed by Ekman and Friesen (1977) and moving to more recent computational approaches (Zhao et al., 2003; Pantic and Rothkrantz, 2000). The research shift towards bodily expressions has only started recently. According to the same survey, specific features of bodily expressivity have been identified to contribute to the recognition of specific affective states. In the case of upper body expressivity and gestures, there exist several manual annotation approaches on gesture analysis (Foster, 2004; Ferré et al., 2007; Kipp and Martin, 2009), while research on the *automatic analysis* of gesture expressivity is ongoing (Varni et al., 2010; Caridakis et al., 2006; Sanghvi et al., 2011; Pantic et al., 2007; Griffin et al., 2013; Glowinski et al., 2011, Kleinsmith and Bianchi-Berthouze (2007), Kleinsmith et al. (2011)), rendering human action analysis asymmetrically less studied with regards to its synthesis counterpart. The main reason behind this is that robust software or dedicated motion capture hardware (Pfeiffer et al., 2013) are needed in order to support analytic methods such as hand trajectory extraction which returns an abundance of data of high detail and richness, especially when it comes to these observations taking place in spontaneous, natural interaction contexts (Cowie et al., 2008). The attribution of affective labels on such data is not a straight forward task. Gesture expressivity – similarly to other types of bodily expressivity – can be interpreted in various ways, leaving a lot of room for subjective assessment. In order to establish ground truth for emotion expression, it is common practice to rely on the judgment of observer coders (Kleinsmith and Bianchi-Berthouze, 2013).

Another trend that has attracted attention in non-verbal behaviour studies is the role of multimodality (Caridakis et al., 2010): signals coming from different emotional channels (Zeng et al., 2009) inform a system's computational intelligence module regarding the emotional or cognitive state of the user. Synergy of multiple modalities (Kapoor et al., 2007) is expected to overcome problems related to reliability, noise and personalization. Other typical examples are reported in Castellano et al. (2009) and Sanghvi et al. (2011) where a Bayesian network uses information coming from posture and gaze, in order to detect engagement with a robot companion (Van Breemen et al., 2005) that is able to pose various expressions. Salem et al. (2012) also investigated the gesture and posture expressivity aspects from a Human Robot Interaction perspective.

Expressivity of body movement (Laban and Lawrence, 1974) is a qualitative cue that is, or at least should be, incorporated in the design process of such applications. In the words of Alex Pentland (1996): "The problem, in my opinion, is that our current computers are both deaf and blind: they experience the world only by way of a keyboard and a mouse… I believe computers must be able to see and hear what we do before they can prove truly helpful". Moving a step further, we might add, that they should also interpret appropriately what they see and hear.

Behaviour expressiveness is an integral part of the communication process since it can provide information about the person's current emotional state, the profile of the interlocutor and metrics of his/her performance. Many researchers have studied characteristics of human movement and coded them in binary categories such as slow/fast, restricted/wide, weak/strong, small/big, unpleasant/pleasant in order to properly model expressivity. *Expressivity dimensions* are considered as the most complete approach to body expressivity modelling, since they cover the entire spectrum of expressivity parameters related to emotion and affect (Karpouzis et al., 2007; Sykes, 2003).

3. Motivation

Non-verbal behaviour has been frequently broken down to its communicative functions (start/end conversation, emphasize, depict object etc) and the behaviours that manifest these functions (nod, body posture, gaze aversion etc.) (Kopp et al., 2006; Vilhjálmsson et al., 2007). One communicative function can be expressed through one or more behaviours and, vice versa, one single behaviour can express one or more functions. In our case we have chosen as behaviours a closed set of expressive gestures with a non-ambiguous semantic meaning. Our goal is to investigate and attempt to quantify how the same gestures, with the same functions, can convey different affective messages through their expressivity features.

An important aspect when studying gesture expressivity is that of subjectivity, mostly in terms of perceiving the conveyed emotion when a gesture is performed. In the case of facial expressions, several sets of universally recognisable emotions exist, with Ekman's being the most prominent (Ekman and Friesen, 1977). However, in the case of gesturing, the cultural background of the interactants plays an important role both when performing a gesture, as well as interpreting it in the receiving end. Kita (2009) (Schroder, 2004) elaborates on the culture-specific conventions for form-meaning associations in emblem gestures (e.g., the thumbs-up sign), and on how cognitive and cultural differences shape iconic and deictic gestures expressing spatial or temporal concepts. As a result, a perceiver-based annotation scheme is needed so as to obtain labels and ratings which can be used as 'ground truth' for any machine learning approach.

This work extends research by Caridakis et al. (2006) on gesture expressivity parameters. These parameters are the result of a qualitative approach to modelling non-verbal upper body expressivity based on computer vision algorithms. Castellano et al. (2009) and, later, Glowinski (Glowinski et al., 2011) have also studied abstract representations of gesture expressivity and their relation to emotion expression and perception, however they rely on machine learning and data processing to arrive at relations between parameters, values and emotion perception, without taking into account the inherent subjectivity in the observed emotion classes. An overview of the

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

parameters' definition and their computational formalization is provided in Section 4.3.

The proposed methodology is a step towards investigating the relationship between gesture expressivity and higher level representations of affect, keeping an expressivity parameters' view point. A correlation between gestures and the dimension of activation has already been established in an influential study by Wallbott (Wallbott and Scherer, 1986). Previous work on gesture expressivity parameters, carried out by Caridakis et al. (2006) is used as a starting point in order to investigate whether the information captured in these parameters is sufficient to make judgments on the affective dimension of evaluation (i.e., the pleasantness/unpleasantness a gesture encompasses). An additional goal is to render expressivity parameters more versatile by introducing an association methodology of these parameters with the dimensions of activation and evaluation. According to a survey on multimodal computer interaction by Jaimes and Sebe (2007), a common meaningful representation framework is required for all modalities in order to achieve a late (semantic/decision) fusion scheme of different modalities. By choosing these dimensions of activation and evaluation as the common representation scheme, the incorporation of expressivity parameters in such a fusion architecture will be greatly facilitated.

The structure of the rest of the paper is as follows: Section 4 sets the theoretical grounds of the semantic bonds between gestures and expressivity parameters and gives a detailed definition of the models adapted for our analysis and representation. Section 5 presents the user perception study conducted, its design and results, as well as the reliability measures taken in order to use it as ground truth. Section 6 evaluates statistical relations between expressivity parameters and maps them to the Activation/Evaluation dimensions through fuzzy reasoning.

4. Gestures in human-computer interaction

4.1. Expressivity in gestures

Various ways of grouping and classifying hand movement have been put forward depending on its function and linguisticity (Kendon, 1988; McNeill, 1992). In most cases gestures are considered to complement spoken language. Nevertheless, hand movements, voluntarily or not, tend to convey additional information, besides speech, regarding the internal mental processes of the speaker including emotional aspects of expression. Our work isolates gestures from speech and aims at focusing less on the function and more on the expressive content of hand movement. A hand movement is classified as semiotic when it communicates meaningful information and results from shared cultural experience. We chose a set of semiotic gesture classes such as clapping, waving, raising hand, etc. According to the approach adopted, their iconic, metaphoric, deictic or beat function (according to McNeill's classification (McNeill, 1992)) is not investigated. The focus is rather on their expressive content, since each chosen gesture class is performed with varying expressivity in order to capture different nuances in a systematic manner. We study the way such differences occur and are perceived both by humans and machines.

4.2. Expressivity parameters and dimensions

Expressivity parameters are a typical example of an intermediate level of representation of affective information. They do not correspond to low level tracking features, nor do they directly relate to higher level representations such as emotion labels or dimensions. They lie in an intermediate level capturing qualitative information on expressivity. In order to empower their applicability and allow for their combined usage with other tracked modalities, we identified the need to come up with an association methodology that will link expressivity parameters to a more versatile representation such as the chosen target representation of dimensional emotion representation (Russell, 1980; Whissell, 1989).

Fontaine et al. (2007) conducted a cross-cultural study on emotional experience and concluded that there exists no single universal solution on the number of necessary dimensions for capturing and representing affective information. Moreover, these figures vary depending on the studied behaviours. Castellano et al. (2012) discuss the perceptual aspect of social agents' expressive behaviour while Kret et al. (2013) approach the issue using physiology and gaze input streams. In order to distinguish complex expressions of emotion, we might be forced to introduce more dimensions beyond the typical Activation-Evaluation dyad. For example, according to their findings, in order to distinguish emotional expressions of surprise, they stress the necessity of the less commonly used dimension of novelty (or unpredictability). Authors in Glowinski et al. (2011) encode human motion using the notion of Sample Entropy (Hong and Newell, 2008), in order to account for the presence of emotion during interactions, not as an occasional occurrence but as a factor constantly influencing behaviour. In our case, we purposely chose to restrict our target representation to the dimensions of Activation and Evaluation.

In contrast to categorical and appraisal based approaches the dimensional emotion representation approached are gradual and represent aspects of emotion concepts (e.g. good/aroused/powerful) as dimension of an emotional space (Schroder, 2009). The Activation and Evaluation space is a representation derived from psychology research and represents emotional states in terms of two dimensions: the activation dimension measures how dynamic the emotional state is whereas the evaluation dimension is a global measure of the positive or negative feeling associated with the state. Alternative terms for the two dimensions include arousal (activation) and valence or pleasure (evaluation) levels and is also related to the PAD (Pleasure, Arousal, Dominance) affective model.

There are two reasons behind this choice: Firstly, expressivity parameters (a detailed definition is provided in the following section) are designed to capture qualitative information and thus attempting a mapping to more than two dimensions increases the risk of arbitrary decisions. Secondly, in order to achieve the aforementioned association and introduce a representation bridging mechanism between expressivity parameters and dimensions, we rely on capturing human perception of expressivity. The selection of only two dimensions is necessary in order not to overload the recruited raters.

4.3. Gesture expressivity parameters

Many researchers have studied characteristics of human movement and coded them in binary categories such as slow/fast, restricted/wide, weak/strong, small/big, unpleasant/pleasant in order to properly model expressivity. We utilize the representation scheme of the expressivity dimensions described in Hartmann et al. (2005), extended in 3D depth in Caridakis et al. (2013), as the most complete approach to expressivity modelling, since it covers the entire spectrum of expressivity parameters related to emotion and affect. Derived from the field of expressivity synthesis five parameters have been defined:

- Overall activation.
- Spatial extent.
- Temporal.
- Fluidity.
- Power.

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

In order to provide a more strict definition of these gesture expressivity parameters, let us consider a gesture G as a sequence, of T frames, consisting of 2D image coordinates of the left and right hand respectively. The coordinates of hands are relative to the $(x_{li}^G, y_{li}^G)(x_{ri}^G, y_{ri}^G)$ position of the head which is defined as the centre of the bounding box of the region of the head as provided by a face detection module (Viola and Jones, 2001) and normalized with reference to the diagonal of this box which is considered indicative of the size of the head. These transformations are required in order to ensure that the coordinates are invariant to the position and the distance of the user with regards to the camera, parameters that are not known a priori. Thus, a gesture is formally defined as:

$$G = [((x_{l1}^G, y_{l1}^G), (x_{r1}^G, y_{r1}^G)), ((x_{l2}^G, y_{l2}^G), (x_{r2}^G, y_{r2}^G)), \cdots, ((x_{lT}^G, y_{lT}^G), (x_{rT}^G, y_{rT}^G))] \quad (1)$$

For simplicity reasons $(x_{li}^G; y_{li}^G)$ will be referred to as L_i^G from this point forward. Additionally, the quantity of motion D_i for one hand during the time period between frame i and frame $i+1$ is defined as the norm of the vector defined by the coordinates of the hand in the respective frames:

$$D_i = \left|\overrightarrow{(x_i, y_i)(x_{i+1}, y_{i+1})}\right| \quad (2)$$

Overall activation is considered as the quantity of movement during a dialogic discourse and is formally defined as the sum instantaneous quantities of motion:

$$OA_G = \sum_{i=1}^{T-1} D_{li}^G + D_{ri}^G \quad (3)$$

Spatial extent is expressed as the expansion or the condensation of the used space in front of the user (gesturing space). Let e_i be the norm of the vector defined by $(x_{li}, y_{li}), (x_{ri}, y_{ri})$ 2D points corresponding to the left and right hand during time. The spatial extent expressivity parameter corresponds to the maximum value of this feature during the stroke phase of the gesture:

$$SE_G = \max e_i, i \in [1, T], e_i = \left|\overrightarrow{(x_{ri}, y_{ri})(x_{li}, y_{li})}\right| \quad (4)$$

The temporal expressivity parameter denotes the speed of hand movement during a gesture and dissociates fast from slow gestures. Given that quantity Di denotes instantaneous hand speed during time, the temporal expressivity parameter is defined as the arithmetic mean of this quantity and *OA* since, as defined earlier, corresponds to the discrete integral, temporal expressivity is given by Eq. (5):

$$TE_G = \frac{OA}{T} \quad (5)$$

On the other hand, the energy expressivity parameter refers to the movement of the hands during the stroke phase of the gesture. Gestures are constituted of three phases: preparation, stroke and withdrawal. The message is primarily conveyed during the stroke phase, while the phases of preparation and withdrawal occur while the hands move from and to their neutral position respectively. The formalization of the energy expressivity feature according to this definition however is far from trivial since the automatic detection of the gesture phases is quite a challenging task. Alternatively, we opted to associate this parameter qualitatively with the first derivative of the norm of D, which refers to the acceleration of hands during a gesture:

$$PO = |D|' \quad (6)$$

Fluidity differentiates smooth/elegant from sudden/abrupt gestures. This concept attempts to denote the continuity between hand movements and is suitable for modelling modifications in the acceleration of the upper limbs. Under this prism, we formally define gesture's fluidity as the variation of the energy expressivity parameter as described in the previous paragraph:

$$FL = var(PO) \quad (7)$$

The reader is prompted to note that the quantity *FL* is reversely proportional to the notion of fluidity. Thus, a gesture with high value of the *FL* expressive parameter demonstrates low fluidity and consequently is categorized as a sudden/abrupt gesture. Inversing the definition of fluidity is not a trivial process since the upper and lower bound of the measure are not a priori known.

4.4. Feature extraction

In order to extract expressivity features from a video sequence of a gesture, it is necessary to detect and track the movement of the actor's hands and face. In order to do so, several approaches have been reviewed. Amongst them only video based methods were considered, since motion capture or other intrusive techniques may interfere with the person's emotional state which is a crucial concern in this kind of analysis, while depth sensor devices would require more dedicated hardware. The major factors taken under consideration are computational cost and robustness, resulting in an accurate, near real-time, skin detection and tracking module.

The overall process is described in detail in Caridakis et al. (2006). Briefly, it consists of the creation of moving skin masks and tracking their centroid throughout the subsequent frames of the video depicting a gesture. A real time colour model of the human skin is constructed by sampling the upper area of a box containing the headas provided by the Viola-Jones head detection module (Viola and Jones, 2001). This sampling box corresponds to the forehead of the user and is defined wrt to the resulting face bounding box (e.g. forehead box width equals ¾ of the bounding box width). Such an adaptive approach tackles illumination issues which often impede the process of modelling and detecting human skin. Additionally, it enhances robustness since the head detection module rarely outputs false positives. Skin-like moving candidate regions are subject to appropriate morphological operations and correspondence is based on size, position and direction heuristic criteria in a multiple-criteria, reward/penalty schema. Object correspondence between two frames is performed by a heuristic algorithm based on skin region size (pixel count), distance with reference to the previous classified position of the region, flow alignment and spatial constraints. In case of occlusions (hand object merging and splitting), a new matching of the left-most candidate object to the user's right hand and the right-most object to the left hand is established.

5. User perception study on expressivity

One of the thorniest issues in the field of affective computing is the fact that ground truth is difficult to come by, annotations are limited to one or two raters, it can be subjective, and inter-rater agreement is often very low and non-reliable, especially in naturalistic data. This often motivates researchers to focus on acted corpora, where a common analysis approach is to rely on the defined labels as ground truth. However, even in acted data, where expressed emotions are known, when it comes to interpreting the expressivity of a gesture, there is a lot of room for subjectivity. Moreover, it is not a rare case that the representation chosen to code the affective content influences the quality of the data collected. In order to overcome these obstacles in the case of expressivity parameters, within the frame of the proposed work, a two-phase user perception study has been designed and conducted. The goal was to collect annotation data on the available videos by human raters and investigate ways of comparing these perception ratings against the employed computer vision

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

components' results. The two-phase expressivity perception study consisted of a pilot phase and a full scale study. It was designed based on the available videos and their analysis in terms of expressivity parameters from Caridakis et al. (2006). Following this strategy, it has been made possible to establish an analytical framework, using fuzzy logic, able to map processed visual features to reliable affective annotation from a large population (more than 100 raters employed in phase II). The result of this analysis was that a system able to imitate human perception of motion-related emotional content was developed, based, not on discrete, subjective interpretations of the actors, but on external viewers' perception (Fig. 1).

5.1. Data Collection

Seven volunteer-actors (with no professional acting experience) were asked to perform the following seven gesture classes in front of a digital video camera. Each gesture class was performed and recorded more than once by each volunteer-actor, according to the expressivity categories appearing in Table 1. Each category corresponds to a quadrant on Whissell's wheel (Whissell, 1989). A quadrant is determined by pairs of plus and/or minus that define the half-axis of activation (vertical axis) and evaluation (horizontal axis). A neutral-in terms of expressivity-performance corresponds to a circle with (0,0) as its center and a relatively small radius. Gestures were performed only with the emotional colouring that made sense (with regards to their semantic meaning) and, thus, a subset of the four quadrants; e.g., the "oh my God" gesture cannot be neutral. A total of 123 video sequences were recorded in a controlled laboratory setting. Each sequence is between two and five seconds duration. The data collected belong to the broader category of acted emotional expressivity data. Fig. 2 holds snapshots from two different raise hand gestures.

Due to occlusions and limited weaknesses of the tracker, the computer vision component did not succeed in returning expressivity parameter values for the full set of 123 videos. Instead, values for a subset of 67 videos were collected.

In the first phase of the experiment (Section 5.2), a small scale forced-choice perception pilot test with a small number of participants and a restricted set of videos was conducted in order to acquire a quantitative feel of people's understanding of expressivity parameters. In the second phase (Section 5.3), using the same web interface, a full scale perception test was run with all videos and 100 participants.

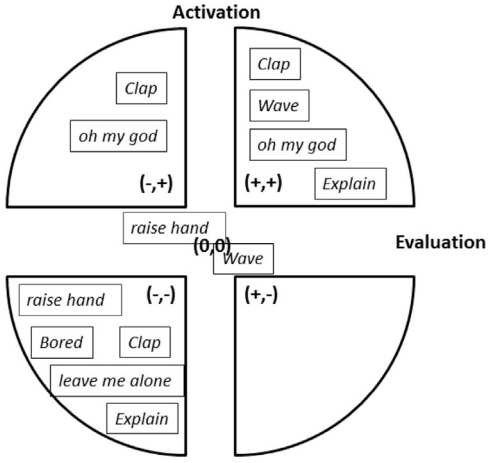

Fig. 1. Acted gestures.

Table 1
Correlation coefficients of users' expressivity parameters against machine extracted values.

	Ov.Act.	Sp.Ext	Speed	Fluid.	Power
Avg/16	0.38	0.78	0.53	−0.20	0.57
Median/16	0.38	0.76	0.50	−0.21	0.61
Avg/12	0.36	0.74	0.50	−0.22	0.52
Median/12	0.38	0.76	0.48	−0.18	0.55

5.2. Phase I: perception of expressivity parameters

Twenty people participated in the first phase of the experiment and watched a subset of 16 videos in custom-made forced-choice web interface. The selection of this restricted subset was based on the results of the computer vision component. Only videos with valid results on all expressivity parameters returned through the computer vision component were used. Participants were recruited in person and their contribution was submitted remotely with no supervision. The group was formed by ten male and ten female raters, aging from 25 to 35, all of the same nationality and of similar higher education background (holders of a master's degree or PhD candidates). They were purposely chosen so as not to be familiar with concepts of affective computing and in particular of expressivity parameters. These parameters were first explained in the introductory page, where the participants' consent was requested. Only data from participants that completed the experiment were taken into account. Each video was presented on a separate page along with five sliders corresponding to each expressivity parameter. Videos ranged from 4 to 8 seconds duration.

The original videos were pre-processed for the requirements of the perception study. More specifically, the face of the person in each video sequence was blurred in order to avoid facial expression effects in the judgment of the human raters. The sound was also muted in order to account for similar confounding effects of voice pitch as well as the semantic content of utterances. Participants were asked to view each video as many times as they desired, in order to use provided sliders and rate the perceived behaviour on the five expressivity scales. Videos were randomised differently for each participant.

Feedback from the participants was recorded through a short interview after completing the online experiment. They reported difficulties in grasping the concepts of the parameters and that they "got the hang" of rating properly only after a couple of videos. It is worth mentioning that they found the expressivity parameters overlapping in meaning and had difficulties discerning "subtle nuances".

One could argue that in acted emotional expressions there are actually three aspects on which one could perform analysis and study their correlation or derive useful conclusions concerning either emotional expression evolution, dynamics, etc. or the user's personality or even divergence from the expected or instructed emotional display.

These three aspects could be the following:

1. *Instructed emotion*: Either in terms of direct guidance or in terms of the induced emotion, in either case, an explicit predetermined emotional state represented as a specific emotional category or a region in some emotional space (Whissell's wheel/PAD values)
2. *The automatic analysis output* in the form of features related to affect or like an emotional label or any other emotion representation entity. Although this aspect is not always indicative of the quality of automatic affective analysis, it is the fused outcome of both the feature extraction and the classification

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

Fig. 2. Intermediate frames from the raise hand gesture.

capabilities of the machine learning or any other method employed to perform the automatic analysis.

3. *The expressed emotion* in terms of how it was perceived by either expert annotators or regular viewers. Although the volunteer-actors may have been instructed to perform an emotional display according to some scenario, this does not entail that they succeeded in conveying the emotion or that the emotion was successfully perceived.

Thus, while the previous two aspects are absolutely defined, inter-rater agreement (or rather disagreement) and other factors governing human annotations establish the third aspect as fuzzy, though indispensable for a complete affective analysis study. This is the aspect we focus on in the perception study, with regard to expressivity parameter perception.

We followed a similar approach to Caridakis et al. (2007) who calculated the correlation of the automatic analysis output with the instructed behaviour quadrants. A strong correlation was found only between power and overall activation with the activation dimension. In this phase of the experiment, we investigated the correlation of the automatic analysis results with the participants' ratings for each expressivity parameter (Table 1). Initially, all 16 sets of ratings for each participant were taken into account. We also calculated correlation coefficients when only taking into account ratings for the last 12 videos that each user viewed to control for a novelty effect with the rating scheme. In both cases, we correlated the average of users' scores for each expressivity parameter of each video (rows marked as 'Avg' in Table 1 for 16 and 12 videos respectively), as well as their median scores (rows marked as 'Median' in Table 1 for 16 and 12 videos respectively) against machine extracted values.

There was no significant variance between the different data views investigated. We measured significant correlation only in the case of the spatial extent feature. Power and speed also correlated relatively well. The low values for overall activation and the negative correlation for the case of fluidity support the view from the qualitative findings that users did not conceive these parameters correctly. These values might also be attributed to the videos' particularly short duration.

The chosen subset of 16 videos included only four out of seven volunteer-actors that participated in the video filming. For each of the four actors (featuring in four videos each) we analysed the correlation of their scores separately. From this approach it is worth mentioning that one student was perceived by raters with less power than the corresponding instruction although the automatic analysis did recognise the expected values. This was expected since the subjects were not professional actors and individual differences are bound to change the way expressivity is perceived.

We also investigated the correlation of human rated power with human rated speed which returned a very high value (0,95) which indicates what the users already reported, that they had difficulty in discerning the differences between these two expressivity features.

5.3. Phase II: rating of expressivity perception using dimensions

In work by Caridakis et al. on the same corpora (Caridakis et al., 2007), the authors identified an association between overall activation and the dimension of activation while using the instructed behaviour quadrants as their ground truth. A core research challenge of the current study was to investigate if conclusions on the evaluation dimension can be drawn from bodily expressivity alone without any cues from facial expressions and vocal features. Thus, this second phase of the experiment was a forced choice design during which participants rated their perceived expressivity of stimuli on two affective dimensions.

We asked 103 participants (56 male and 47 female from 25 to 45 years of age) to use a web interface similar to that of study 1 where a random selection of 40 out of the total 67 videos was displayed in random order to each user. Sliding bars were used to collect user expressivity ratings on the dimensions of activation and evaluation. Similarly to the first phase, an explanation of the affective dimensions was provided in the introductory page, were the participants consent was requested. Each video was presented on a separate page along with two sliders corresponding to each dimension. Again, participants were asked to view each video as many times as they desired in order to use provided sliders and rate the perceived behaviour. For each participant, we started taking into account their ratings after the fourth video sequence, ignoring the first three to control for novelty effects.

Feedback from the participants was recorded through a short interview after completing the online experiment. The rating process using dimensions was easy to grasp and intuitive. Users were satisfied with their ratings and seemed to get familiar with the task at hand early on in the process.

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

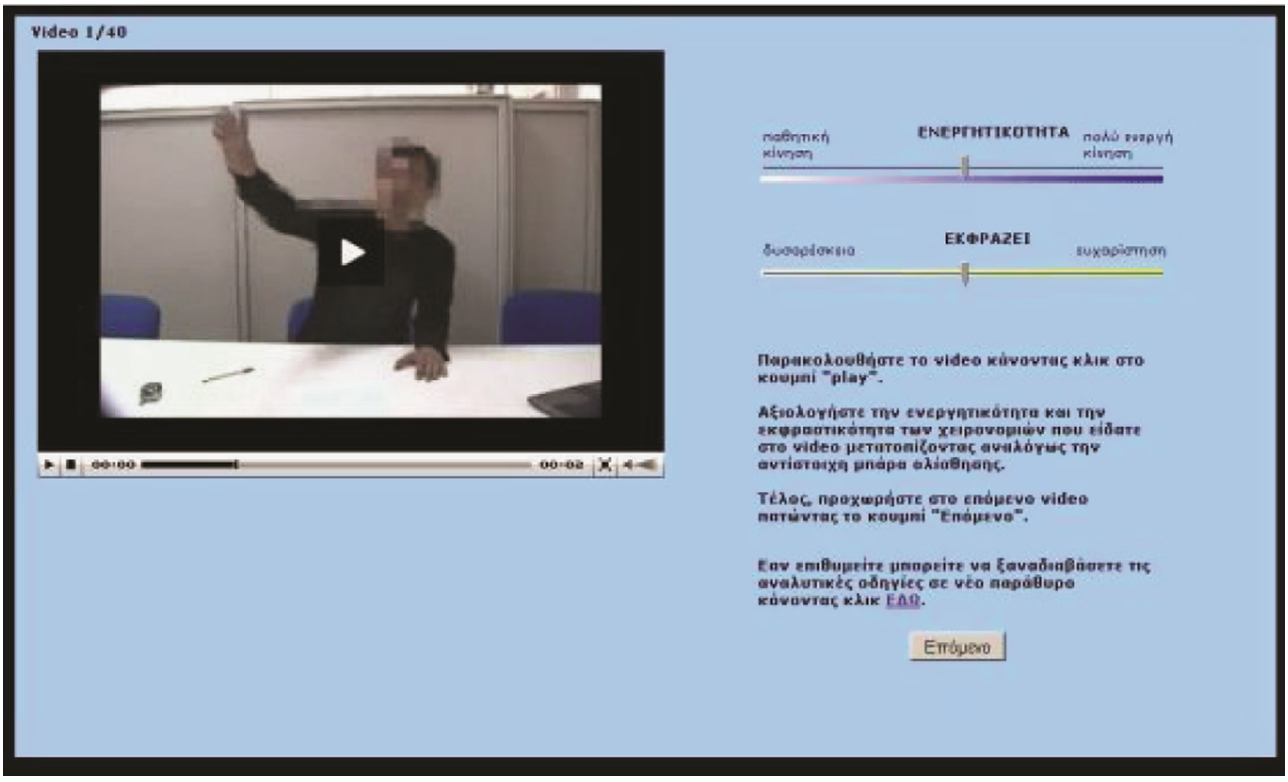

Fig. 3. Web interface snapshot of expressivity perception test using dimensions.

Prior to investigating whether the system-generated expressivity parameters could be associated with the evaluation and activation dimension values collected by the experiment, we had to establish that the mean ratings reported by participants for these dimensions are reliable.

In order to estimate inter-rater consistency, we calculated the Intraclass Correlation Coefficient (ICC) (Shrout and Fleiss, 1979). As ICC approaches the value of 1, less variance will be explained by the effect of the participant, as a given video will tend to yield a similar set of ratings. Using the statistical analysis toolbox SPSS, participants were considered as items and the 67 videos were the cases. There was no way to know which videos each participant rated. For this reason, a one-way random model was chosen that treated the participant as a random factor.

For the activation dimension, the average measure ICC was . 941. For the evaluation dimension, the average measure ICC was . 958. These results suggest that the mean ratings for each video are reliable and can be used in our subsequent analyses.

6. Associating expressivity parameters onto activation – evaluation dimensions

Having established a satisfactory inter-rater agreement, we moved on to investigate the association of calculated expressivity parameter values for the 67 videos, with values -obtained through annotation- for the dimensions of activation and evaluation.

Looking into ways of associating the two different representations of expressivity on this set of data we first needed to investigate which expressivity parameters seemed to affect the corresponding dimension values. In other words we needed a robust way to evaluate the appropriateness of each of the parameters in estimating the dimensions of activation and evaluation.

6.1. Fisher's test

In order to evaluate the appropriateness of each of the expressivity parameters for estimating dimensions, we used Fisher's exact test (Fisher, 1954), as was done in (Asteriadis et al., 2012), where the focus was to infer relations between expressivity features (2D, 3D body analysis, gestures and face expressivity) and affective cues. To this aim, we quantized the values of Activation and Evaluation to the closest integers (0 and 1), thus splitting the dataset in two groups for each dimension. A 3-bin histogram of low, medium and high values for each of the expressivity parameters was calculated for each of the two groups, one for low-high activation and one for low-high evaluation. The resulting histograms for the low and high values of each dimension separately, were compared against each other.

Fisher's exact test for histograms comparison was preferred over other methods (such as the chi-square method), because it is suitable for small scale data. Indeed, in the current dataset, it is often the case that there are only a few instances with low or high values at the corresponding histogram bins (for example, the temporal parameter did not have a lot of instances in the third bin in the case of high activation judgments). Fisher's exact test is ideal in depicting such differentiations in cases of small samples.

The statistical test indicates the rejection of an expressivity parameter if its histogram values, for each dimension, are not significantly different ($p > 0.05$). In our case, we were led to rejecting the Overall Activation parameter, as a non-useful parameter at estimating the Activation dimension. This is qualitatively explained if one takes into account the fact that, by definition, overall activation is especially sensitive during the whole video process. Thus, while raters intuitively focused on the depicted gesture itself, the automatic parameter extraction takes into account the total number of the frames in a gesture, considering information not related to the gesture under consideration (apex and offset phases of the gesture) (Fig. 3).

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

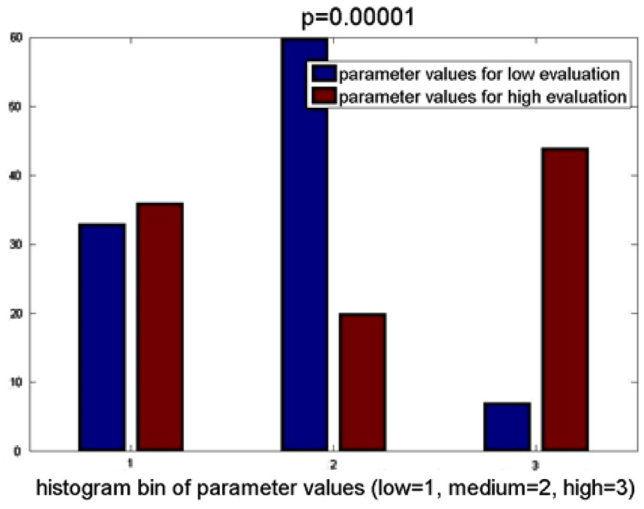

Fig. 4. Histogram bins of maximum spatial extent parameter values (low=1, medium=2, high=3). Axis x corresponds to the parameter values (quantized in low, medium, high values) and axis y counts the number of the corresponding instances for low/high evaluation.

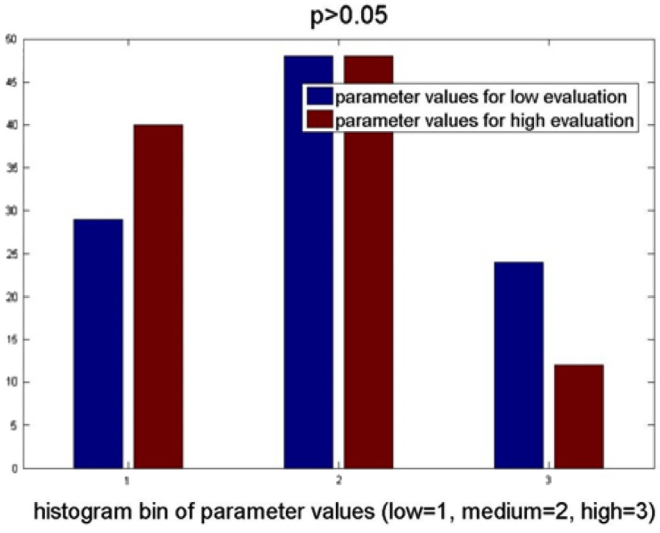

Fig. 5. Histogram bins of power parameter values (low=1, medium=2, high=3). Axis x corresponds to the parameter values (quantized in low, medium, high values) and axis y counts the number of the corresponding instances for low/high evaluation.

Similarly, in the case of the Evaluation dimension, the Power parameter was discarded. The qualitative explanation for this is the fact that the same "amount" of Power may express either pleasure or displeasure in a gesture. Figs. 4–7 show representative examples of expressivity parameters' distributions for both dimensions, where, for each expressivity parameter and dimension, two histograms are compared against each other: one corresponding to the distribution of the expressivity parameter across low activation/evaluation values and one across high activation/evaluation values

6.2. Neuro-fuzzy system

For mapping expressivity parameters to dimensions, a Sugeno-type fuzzy[1] (Takagi and Sugeno, 1985) inference system was built for estimating Activation, while a different model was used for the dimension of Evaluation. Such types of systems perform well in

[1] Matlab fuzzy logic toolbox.

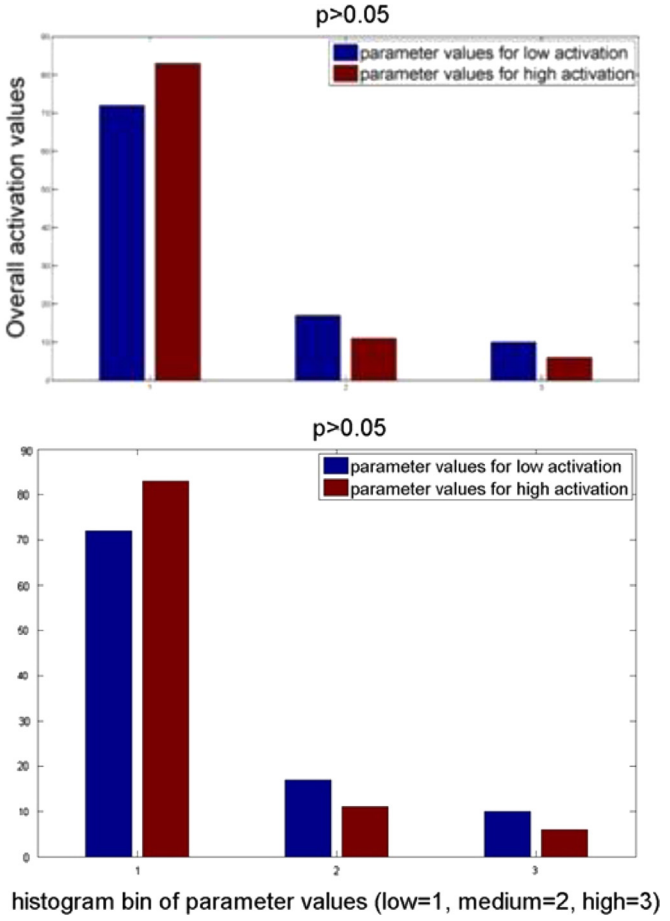

Fig. 6. Histogram bins of overall activation parameter values (low=1, medium=2, high=3). Axis x corresponds to the parameter values (quantized in low, medium, high values) and axis y counts the number of the corresponding instances for low/high activation.

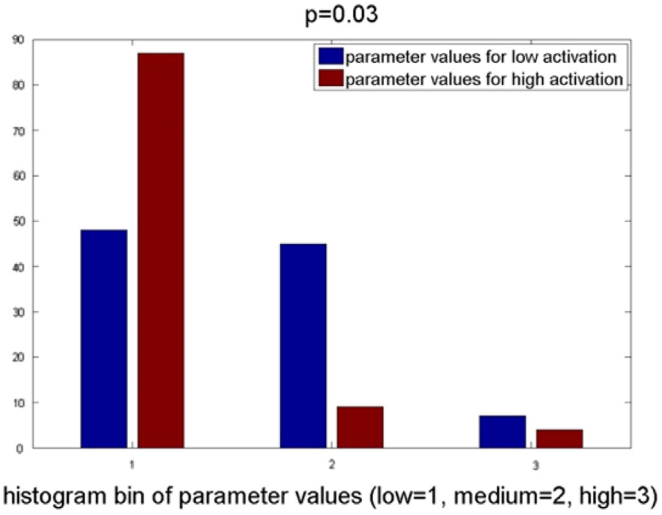

Fig. 7. Histogram bins of fluidity parameter values (low=1, medium=2, high=3). Axis x corresponds to the parameter values (quantized in low, medium, high values) and axis y counts the number of the corresponding instances for low/high activation.

approximation and generalization problems. They are used in various applications ranging from simple neuro-fuzzy models (Jang, 1993), to multilayer classifiers (Mitra and Pal, 2002; Nauck and Kruse, 1997; Cho and Kim, 2002). The underpinning rationale of fuzzy systems is that behavioural states cannot belong to certain classes, but they take fuzzy

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

Table 2
Neuro-Fuzzy System decision accuracy for two different sets of expressivity parameters for estimating Activation.

	Absolute mean error ± std
Spatial Extent, Temporal, Fluidity, Power	0.12 ± 0.10
Spatial Extent, Temporal, Fluidity, Power, Overall Activation	0.123 ± 0.10

Table 3
Neuro-Fuzzy System decision accuracy for two different sets of expressivity parameters for estimating Evaluation.

	Absolute Mean Error ± std
Spatial Extent, Temporal, Fluidity, Overall Activation	0.21 ± 0.17
Spatial Extent, Temporal, Fluidity, Power, Overall Activation	0.24 ± 0.17

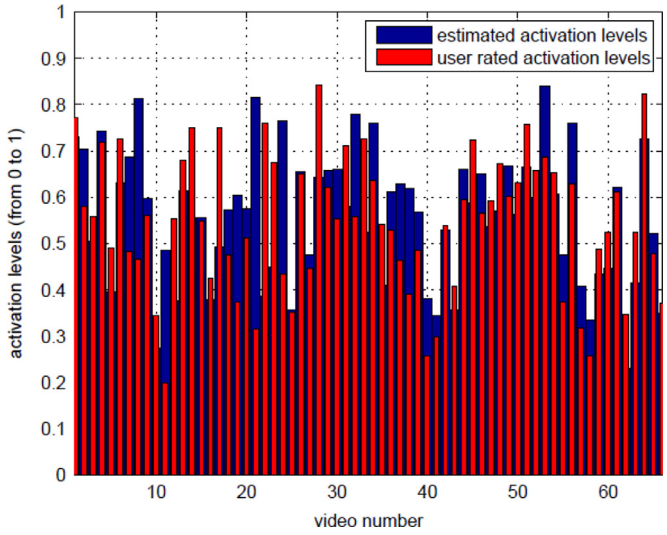

Fig. 8. Activation values: neuro-fuzzy network predictions (blue) and user ratings (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

values. Furthermore, given the exploratory nature of the research, we were uncertain of the mapping between parameters and dimensions, and, so we could not investigate any linear or nonlinear model beforehand. Building neuro-fuzzy systems and letting a training algorithm decide the weights and the whole structure of our model was thus an appropriate approach. In particular, the feature vectors used as inputs to the neuro-fuzzy systems consisted of expressivity parameters (normalized from 0 to 1). In the case of estimating levels of activation, different configurations of feature vectors were tested and, based on the analysis provided above, the input parameter vector that gave the best results were those of *Spatial Extent, Temporal, Fluidity, Power*. In a similar manner, for estimating levels of Evaluation, the most correlated parameters were those of *Spatial Extent, Temporal, Fluidity, Overall Activation* (see next Section for a discussion on the results).

Fuzzy systems consist of rules (e.g. high Overall Activation values, combined with high Temporal, low Fluidity and low Power values leads to low Activation) and map input parameters to semantic fuzzy sets (modelled through membership functions that describe the extent to which a parameter belongs to 'high' or 'low' values). An inference mechanism estimates the extent to which a rule is triggered at a specific instance and combines the outputs of each rule for providing an overall estimate of the output (here, Activation and Evaluation). In this work, since no prior knowledge of the data structure

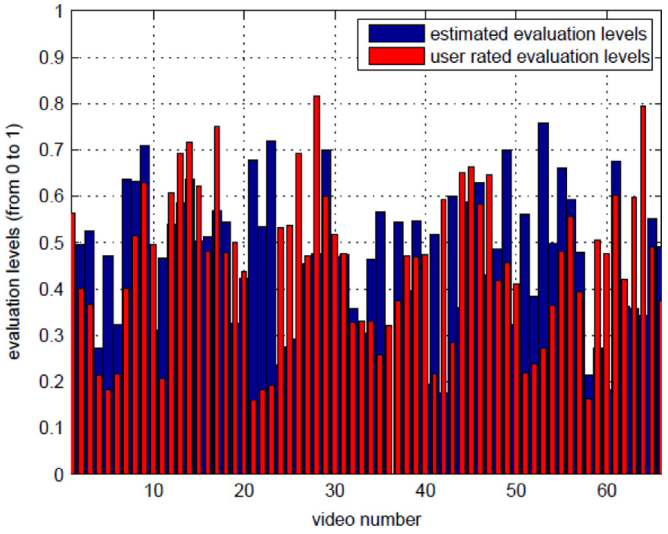

Fig. 9. Evaluation values: neuro-fuzzy network predictions (blue) and user ratings (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and their behaviour was known, we used data clustering for inferring the optimal number of rules (Chiu, 1994), as well as least squares and back-propagation gradient-descent for inferring membership function centres and widths (Jang, 1993). In particular, prior to training, our data were clustered using the sub-cluster algorithm described in Chiu (1994). This algorithm, instead of using a grid partition of the data, clusters them and, thus, leads to fuzzy systems deprived of the curse of dimensionality. For clustering, many radius values for the cluster centres were tried and the one that gave the best trade-off between complexity and accuracy was 0.333 for all normalized inputs and outputs both in the case of activation and evaluation experiments. The number of clusters created by the algorithm determines the optimum number of the fuzzy rules.

7. Results

Tables 2 and 3, as well as Figs. 8 and 9 summarize the results of the overall accuracy of our system in estimating each dimension based on the expressivity features extracted by the hands trajectories. We conducted experiments using the features selected after Fisher's tests but, also, we verified that (Tables 2 and 3), including the features discarded by the test, would not improve the whole accuracy (or possibly, it would deteriorate results due to noise). The absolute errors, as well as the standard deviations, corresponding to the performance of each fuzzy system in relation to the values given by the raters, verify the validity of our choice for neuro-fuzzy inference logic, as well as our prior intuition that hand gesture-dependent expressivity features play a key role at estimating affective dimensions.

Training was done using a leave-one-video-out protocol (input (expressivity) parameters were all normalized with the maximum values of the parameters of the training data), while Gaussian membership functions for the fuzzy rule sets were considered. For training, the remaining video-sequences were used and the process was repeated until all videos were used as validation data. Furthermore, for each validation video sequence, we used the average of all ratings per dimension, as target dimension level, excluding those whose distance from the average was three times the standard deviation. This helped avoid completely unexpected ratings that could be due to rater's tiredness or other exterior factors.

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

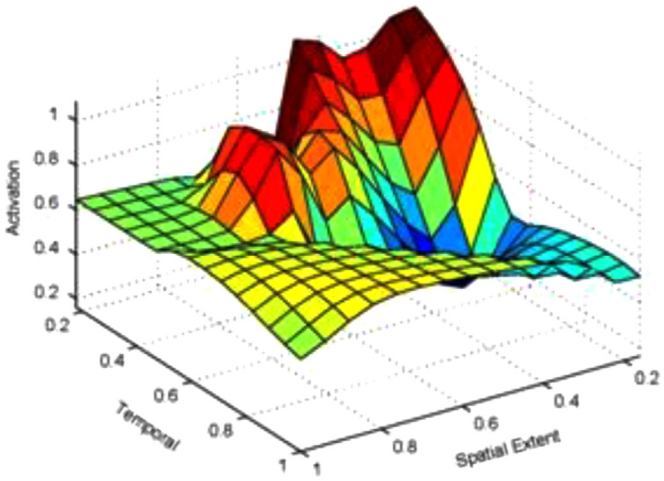

Fig. 10. Correlation between temporal/spatial extent and the dimension of Activation.

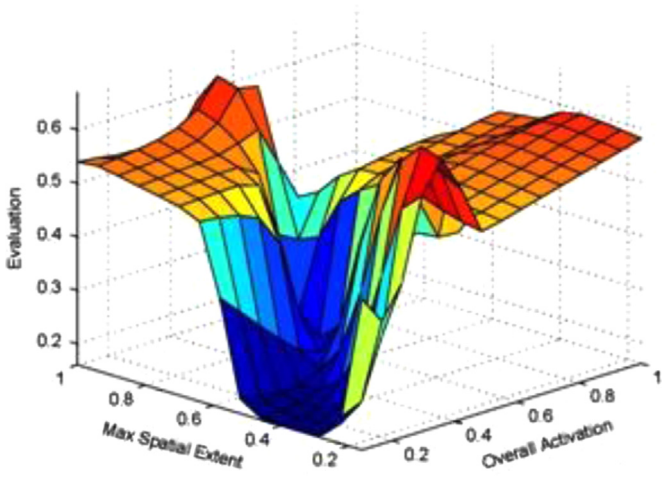

Fig. 12. Correlation between spatial extent/overall activation and the dimension of Evaluation.

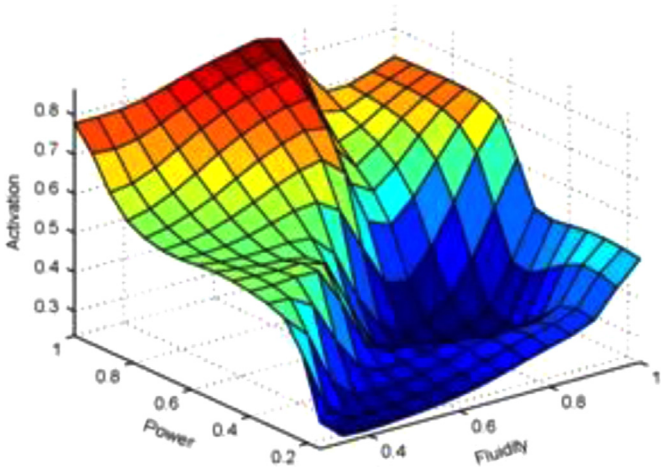

Fig. 11. Correlation between power/fluidity and the dimension of Activation.

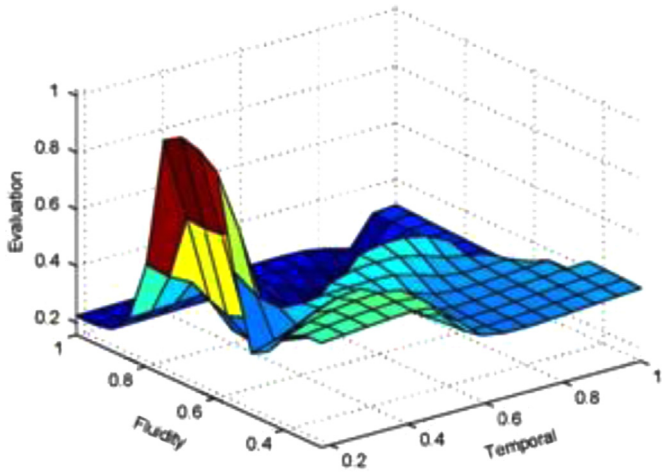

Fig. 13. Correlation between Temporal extent/fluidity and the dimension of Evaluation.

7.1. Association methodology-association rules

A visual inspection of the fuzzy system rules can be seen in Figs. 10–13. From these figures, we can conclude the following:

(1) Association with the Activation dimension
Small temporal and spatial extent, are related to high values in activation (see Fig. 10). Also, in Fig. 11, we can see that, for low values of power, no matter how high the values of fluidity are, the estimates of activation are low. On the contrary, when power and fluidity take higher values, there is an increase in activation. This can be intuitively explained since, fluidity, due to the way it is defined, takes high values for sudden movements.

(2) Association with Evaluation dimension
From Fig. 12, the correlation between the dimension of evaluation with spatial extent and overall activation can be seen. Small spatial extent is related to low levels of evaluation (pleasure), while the same is valid also for low values of the overall activation parameter. Fig. 13 shows the correspondence of evaluation with fluidity and temporal extent. It can be seen that, as temporal extent increases, gesture interpretation tends to correspond to lower values of pleasure. It is also observed that, for a particular "medium" level of fluidity, evaluation takes high values.

8. Conclusions

The purpose of this paper is to associate expressivity parameters with dimensional representations of affect. The approach presented was based on acted gestures and their expressiveness. Especially in terms of the dimension of evaluation, such an association with gesture expressivity had yet to been shown.

The presented results are promising since the activation/evaluation values predicted by the neuro-fuzzy network showed little deviation from the ground truth collected in the user perception study.

The introduced methodology provided an interim representation mechanism with a set of association rules between expressivity parameters and affective dimensions. These rules can be used for the analysis of other corpus data. They can also function as a stepping stone towards the analysis of naturalistic and spontaneous non-verbal expressions. The application of these rules in novel contexts is challenging and can lead to further improvement of the methodology and its applicability. Future work will thus focus on applying these rules, along with the automatic analysis of expressivity parameters to novel, either acted or non-acted corpora where ground truth is harder to come by.

The proposed association methodology between expressivity parameters and activation/evaluation dimensions is a valuable

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

tool that will allow for the inclusion of expressivity parameters in the fusion process of multiple modalities coded in dimensional representations and is sought to be a big step towards bridging different representations schemes. With the advent of new, non-intru sive motion capture devices (e.g. depth sensors), three-dimensional motion information, new human-machine interfaces emerge. Endowing machines with the ability to easily interpret humans' emotional state, can give a significant boost to today's classical interfaces, while it can open new fields of human-centric research.

References

Asteriadis, S., Caridakis, G., Malatesta, L., Karpouzis, K., 2012. Natural Interaction Multimodal Analysis: Expressivity Analysis towards Adaptive and Personalized Interfaces. In: Proceedings of the 2012 Seventh International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), IEEE, December, pp. 131–136.

Bengoechea, J., Cerrolaza, J., Villanueva, A., Cabeza, R., 2013. Evaluation of accurate eye corner detection methods for gaze estimation. In: Proceedings of the 17th European Conference on Eye movements, Lund, Sweden, August.

Caramiaux, B., Donnarumma, M., Tanaka, A., 2015. Understanding gesture expressivity through muscle sensing. ACM Trans. Computer-Hum. Interact. (TOCHI) 21 (6), 31.

Caridakis, G., Karpouzis, K., Wallace, M., Kessous, L., Amir, N., 2010. Multimodal user's affective state analysis in naturalistic interaction. J. Multimodal User Interfaces 3 (1), 49–66.

Caridakis, G., Moutselos, K., Maglogiannis, I. 2013. Natural Interaction expressivity modeling and analysis. In: Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments, ACM, May, p. 40.

Caridakis, G., Raouzaiou, A., Bevacqua E., Mancini, M., Karpouzis, K., Malatesta, L. Pelachaud C., 2007. "Virtual agent multimodal mimicry of humans", Language Resources and Evaluation, 41 (3–4), Special issue on Multimodal Corpora, Springer, pp. 367-388.

Caridakis, G., Raouzaiou, A., Karpouzis, K., Kollias, S., 24–26 May., 2006. Synthesizing Gesture expressivity based on real sequences. In: Workshop on Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models. LREC 2006 Conference, Genoa, Italy.

Cassell, J., Vilhjálmsson, H., Bickmore, T., 2004. BEAT: the behaviour expression animation toolkit. In: Prendinger, Ishizuka (Eds.), Life-Like Characters: Tools, Affective Functions, and Applications Embodied Conversational Agents. Springer, New York, NY.

Castellano, G., Mancini, M., Peters, C., McOwan, P.W., 2012. Expressive copying behavior for social agents: a perceptual analysis. IEEE Transactions Syst. Man Cybern. Part A: Syst. Hum. 42 (3), 776–783.

Castellano, G., Pereira, A., Leite,I., Paiva, A., McOwan, P.W., 2009. Detecting user engagement with a robot companion using task and social interaction-based features. In: Proceedings of the 2009 International Conference on Multimodal Interfaces. ACM, NewYork, NY, USA , pp. 119–126.

Chiu, S., 1994. Fuzzy model identification based on cluster estimation. J. Intell. Fuzzy Syst. 2, 267–278.

Cho, S., Kim, J., 2002. Combining multiple neural networks by fuzzy integral for robust classification. IEEE Trans. Syst. Man Cybern. 25, 380–384.

Cowie, R., Douglas-Cowie, E., Karpouzis, K., Caridakis, G., Wallace, M., Kollias, S., 2008. Recognition of emotional states in natural human-computer interaction. In: Journal of Multimodal User Interfaces, Springer Berlin Heidelberg, pp. 119–153.

Ekman, P., Friesen, W., 1977. Manual for the Facial Action Coding System. Consulting Psychology Press, Palo Alto.

Ferré, G., Bertrand, R., Blache, P., Espesser, R. and Rauzy, S., 2007. Intensive gestures in French and their multimodal correlates. In: Proceedings of Interspeech, Antwerp, Belgium, Cederom.

Fisher, R., 1954. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Fontaine, J., Scherer, K., Roesch, E., Ellsworth, P., 2007. The world of emotions is not two-dimensional. Psychol. Sci. 18, 1050–1057 (Blackwell Publishing).

Foster, M., 2004. Corpus-based planning of deictic gestures in COMIC. Nat. Lang. Gener., 198–204 (Springer).

Glowinski, D., Mancini, M., 2011. Towards Real-Time Affect Detection Based on Sample Entropy Analysis of Expressive Gesture. In: D' Mello, S., et al. (Eds.), ACII 2011, Part I, LNCS 6974. Springer-Verlag, Berlin Heidelberg, pp. 527–537.

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K., 2011. Toward a minimal representation of affective gestures. IEEE Trans. Affect. Comput. 2 (2), 106–118.

Griffin, H. J., Aung, M. S., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., & Bianchi-Berthouze, N., 2013. Laughter type recognition from whole body motion. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, pp. 349–355.

Hartmann, B., Mancini, M., Buisine, S., Pelachaud, C., 2005. Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: Proceedings of the fourth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1095–1096.

Hong, S.L., Newell, K.M., 2008. Entropy Conservation in the Control of Human Action. Nonlinear Dyn. Psychol. Life Sci. 12 (2), 163.

Ioannou, S., Caridakis, G., Karpouzis, K., Kollias, S., 2007. Robust feature detection for facial expression recognition. EURASIP J. Image Video Process. 2.

Jaimes, A., Sebe, N., 2007. Multimodal human-computer interaction: a survey. Comput. Vis. Image Underst. 108, 116–134 (Elsevier).

Jang, J., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23, 665–685.

Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A., 2008. Measuring and defining the experience of immersion in games. Int. J. Hum. – Comput. Stud. 66, 641–661.

Kapoor, A., Burleson, W., Picard, R.W., 2007. Automatic prediction of frustration. Int. J. Hum. – Comput. Stud. 65, 724–736.

Karpouzis, K. Caridakis, G. Kessous, L., Amir, N., Raouzaiou, A., Malatesta, L. Kollias, S., 2007. Modeling naturalistic affective states via facial, vocal, and bodily expressions recognition. Artifical Intelligence for Human Computing, Springer, pp. 91–112.

Kendon, A., 1988. How gestures can become like words. In: Potyatos, F. (Ed.), Crosscultural Perspectives in Nonverbal Communication. Hogrefe, Toronto, Canada, pp. 131–141.

Kipp, M., Neff, M., Kipp, K., Albrecht, I., 2007. Towards Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. Intell. Virtual Agents, 15–28 (Springer).

Kipp, M., Martin, J., 2009. Gesture and Emotion: Can basic gestural form features discriminate emotions? In: Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction.

Kita, S., 2009. Cross-cultural variation of speech-accompanying gesture: A review. Language and Cognitive Processes 24 (2), 145–167.

Kleinsmith, A., Bianchi-Berthouze, N., 2007. Recognizing Affective Dimensions from Body Posture, 4738. Springer, Berlin, p. 48 (Lecture Notes in Computer Science).

Kleinsmith, A., Bianchi-Berthouze, N., 2013. Affective body expression perception and recognition: a survey. IEEE Trans. Affect. Comput. (IEEE)

Kleinsmith, A., Bianchi-Berthouze, N., Steed, A. 2011. Automatic recognition of non-acted affective postures Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, IEEE, vol. 41, pp. 1027–1038.

Knapp, M., Hall, J., 2013. Nonverbal Communication in Human Interaction. Wadsworth Pub Co., Belmont.

Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., Vilhjálmsson, H., 2006. Towards a common framework for multimodal generation: The behaviour markup language. Intell. Virtual Agents, 205–217.

Kret, M.E., Stekelenburg, J.J., Roelofs, K., De Gelder, B., 2013. Perception of face and body expressions using electromyography, pupillometry and gaze measures. Front. Psychol. 4.

Laban, R., Lawrence, F., 1974. Effort: Economy in Body Movement. Plays, Boston.

Martin, J.C., Niewiadomski, R., Devillers, L., Buisine, S., Pelachaud, C., 2006. Multimodal complex emotions: gesture expressivity and blended facial expressions. Int. J. Humanoid Robot. 3, 269–292.

McNeill, D., 1992. Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago, USA.

Mitra, S., Pal, S., 2002. Fuzzy multi-layer perceptron, inferencing and rule generation. IEEE Trans. Neural Netw. 6, 51–63 (IEEE).

Nauck, D., Kruse, R., 1997. A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy Sets Syst. 89, 277–288 (Elsevier).

Pantic, M., Rothkrantz, L.J.M., 2000. Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 22 (12), 1424–1445.

Pantic M., Pentland A., Nijholt A., Huang T.S., 2007. Human Computing and Machine Understanding of Human Behaviour: A Survey. In: Artificial Intelligence for Human Computing, (Eds.), Springer, Lecture Notes in Artificial Intelligence, vol. 4451, pp. 47–71.

Pelachaud, C., 2008. "Studies on gesture expressivity for a virtual agent", Speech Communication, Special issue in honor of Björn Granstrom and Rolf Carlson. vol. 51, pp. 630–639.

Pentland, A., 1996. Smart Rooms Scientific American, vol. 274, pp. 54–62.

Pfeiffer, T., Hofmann, F., Hahn, F., Rieser, H., Röpke, I. 2013. "Gesture semantics reconstruction based on motion capturing and complex event processing: a circular shape example". In: Eskenazi, M. Strube, M., Di Eugenio, B., Williams, J.D., (Eds), Proceedings of the Special Interest Group on Discourse and Dialog (SIGDIAL) 2013 Conference. Association for Computational Linguistics, pp. 270–279.

Russell, J.A., 1980. A circumplex model of affect. J. Person. Soc. Psychol. 39, 1161–1178.

Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., Joublin, F., 2012. Generation and evaluation of communicative robot gesture. Int. J. Soc. Robot. 4 (2), 201–217.

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A., 2011. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: Proceedings of the 6th international conference on Human robot interaction. HRI'11, ACM, NewYork, NY, USA, pp. 305–312..

Schroder, Marc, 2004. "Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions. Affective dialogue systems. Springer Berlin Heidelberg, pp. 209–220.

Shrout, P., Fleiss, J.L., 1979. Interclass correlations: use in assessing rater reliability. Psychol. Bull. 7 (86), 420–428.

Sykes, J., 2003. Affective gaming: measuring emotion through the gamepad. In: CHI 2003: New Horizons, ACM Press, pp.732–733.

Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modelling and control. IEEE Trans. Syst. Man Cybern. 15 (1), 116–132 (Institute of Electrical and Electronics Engineers).

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010

Van Breemen, A.J.N., Yan, X., Meerbeek, B., 2005. icat: an Animated User-interface Robot with Personality. In: AAMAS. pp. 143–144.

Varni, G., Volpe, G., Camurri, A., 2010. A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. IEEE Trans. Multimed. 12 (6), 576–590.

Vilhjálmsson, H., Cantelmo, N., Cassell, J., E. Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C. others, 2007. The Behaviour Markup Language: Recent Developments and Challenges Intelligent Virtual Agents, pp. 99–11.

Vinciarelli, A., et al., 2012. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. IEEE Trans. Affect. Comput. 3.1, 69–87.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. IEEE Comput. Vis. Pattern Recognit. 1, 511–518.

Wallbott, H., 1998. Bodily expression of emotion. Eur. J. Soc. Psychol. 28, 879–896.

Wallbott, H., Scherer, K., 1986. How universal and specific is emotional experience? Evidence from 27 countries on five continents. Soc. Sci. Inform. MSH 25, 763.

Whissell, C., 1989. The Dictionary of Affect in Language, Theory, Research and Experience: The Measurement of Emotions, vol. 113. Academic Press, San Diego.

Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S., 2009. A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31 (1), 39–58.

Zhao, W., Chellappa, R., Rosenfeld, A., 2003. Face recognition: a literature survey. ACM Comput. Surv. 35, 399–458.

Please cite this article as: Malatesta, L., et al., Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.01.010