Draft. Final version in: Proc. 5th Biannual Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013)

Are Discrete Emotions Useful in Human-Robot Interaction? Feedback from Motion Capture Analysis

Matthew Lewis, Lola Cañamero
Embodied Emotion, Cognition and (Inter-)Action Lab, School of Computer Science & STRI,
University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, UK
Email: matt-l@semiprime.com, L.Canamero@herts.ac.uk

***Abstract*—We have conducted a study analyzing motion capture data of bodily expressions of human emotions towards the goal of building a social expressive robot that interacts with and supports hospitalized children. Although modeling emotional expression (and recognition) in (by) robots in terms of discrete categories presents advantages such as ease and clarity of interpretation, our results show that this approach also poses a number of problems. The main issues relate to the loss of subtle expressions and feelings, individual features, context, and social interaction elements that are present in real life.**

I. Introduction

Discrete (and particularly basic) emotions have at the same time been the object of much controversy and a powerful driving force for research into the psychology of human emotions. While other approaches, such as dimensional, component processes and dynamical system models have gained increased support, the discrete approach is still very widely used. In affective computing and robotics research, discrete emotions occupy a similar position: broadly adopted mid and late 1990s, they are still the main paradigm in human-robot interaction studies, particularly regarding automatic recognition and expression, even though researchers (e.g., [1], [2], [3]) are increasing aware of their limitations. Regarding specifically bodily expression, much research in psychology, virtual agents and robotics is devoted to the investigation of postural and movement characteristics of basic and other discrete affective states, primarily with the purpose of analysis and (automatic) recognition of human emotions (see [4] for a comprehensive analysis and [1] for a review), and more recently for synthesis in animated characters and robots [5], [6], [7]. Researchers working on bodily expression for the purpose of synthesis generally draw on data and methods stemming from recognition studies. However, recognition studies are focused on finding population trends and averages, which are suitable for finding archetypal examples of different classes of emotions. However, as pointed out by Bänziger and Scherer [2] and discussed below in this paper, the data collected for perceptual studies are not necessarily suitable for synthesis for human-robot interaction. This is especially true in long-term interaction, where partners would be expected to adapt to each other's individual modes of expression, and where a limited number of expressions in a partner would be perceived as unnatural.

In line with the above, we carried out an empirical investigation, starting from data collection, examining the suitability of a discrete emotion approach in our interactions—that of a humanoid robot companion for children and young adolescents, as part of the European project ALIZ-E (www.aliz-e.org). Our aim was not to be systematic, but to analyze examples of individual cases, since a robot should be able to interact with, and be adapted to, any and each individual, and individuals express emotions in different and unique ways. Therefore we were not not looking for "good" exemplars of different emotions. For this reason, we did not collect data from a large number of individuals in order to calculate an average, or from which we could choose our "best" expressions of an emotion. Instead, we collected data from just two individuals in two areas where discrete emotions play very different roles—dance and theater. For the same reason, we did not consider to what extent these expressions would be considered typical examples of different classes of emotion (what in HRI and related domains are known as "validation studies").

For the sake of grounding on and comparison with the literature, we based it on the seminal study by Camurri et al. [3], which collected and analyzed video data from dancers portraying four emotions—Anger, Fear, Grief, and Joy. To that set we added Pride, as it is very important from both developmental and educational perspectives in our target user population [8]. While we found broad correlations with our reference [3] and similar studies, our results also revealed a number of limitations and issues in the application of the categorial approach to human-robot interaction.

The remainder of the paper is organized as follows: In Section II we describe our data collection and analysis methods, in Section III we describe our results, and in Sections IV and V we consider how useful these results might be in the context of human-robot interaction.

II. Methods

A. Data Collection

Data from two performers were collected using different induction methods based on each performer's normal practice:

1) A professional dancer (male) performing short dance sequences. Scenario outlines designed to evoke affective responses were provided to the dancer who chose music for the scenario. Following his normal practice, emotions were not mentioned during the session; scenarios were referred to as "scenario one" etc. He chose two fixed step sequences in advance that were used for all scenarios, along with free step sequences choreographed for each scenario. A number of takes were made until the dancer felt he had given a successful performance.

 http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6681414

2) A professional actor (female) coached by a director (male) who, following their standard practice, provided scenarios for improvisation designed to evoke specific named emotions: sadness, fear, pride, anger and fear. In some scenarios another actor (female, motion not captured) interacted with the source actor. Work on each emotion continued for around 10 minutes, until the director was satisfied.

The use of a dancer allowed us to compare our results with the foundational study [3]; it also allowed us to examine a method of expression that is built from movement, and hence where we might expect movement to carry the clearest affective signal. In addition, the ALIZ-E project incorporates a dance activity during the child-robot interactions and hence the ability to communicate and read an internal emotional state during this interaction was considered valuable. An actor provided a contrasting set of expressive motion data; acted emotion portrayals of daily life situations were also chosen as they provide a more standard form of expression, are expected to give rise to more prototypical expressions, and are widely used in studies of emotion, both static (postures) and dynamic (motion) expressions, see [2] for a discussion.

We used the Xsens MVN inertial motion capture system[1] to capture full-body 3D motion. 3D skeleton motion capture was chosen for the ease with which the captured motion could be mapped onto humanoid virtual agents for motion synthesis, as well as the ability to isolate the motion of individual body parts. Camera-based systems such as Microsoft's Kinect offer similar capabilities at a consumer level, but the inertial sensor-based system allowed motion information to be collected from limbs which might be hidden from a camera. Results should be transferable to any system which models human motion using a skeletal model. The sampling frequency was 120Hz. The proprietary Xsens software calculated values such as joint angles from the raw data.

Performances were not validated with human observers since we were interested in using the data as naturally expressed by different individuals, without any selective pruning of difficult or ambiguous cases. Moreover, we used only one actor and one dancer, rather than many, as we are interested in developing a robot that can interact with any real-life individual with their own unique ways of expressing themselves, not with a hypothetical human created from statistical averages. See Section IV for further discussion of this.

B. Metrics for Expressive Motion Analysis

The study by Camurri et al. [3], as well as the numerous studies stemming from it, used video processing techniques to study expressive performances of dance and music. Drawing on Laban's effort parameters for movement analysis, they defined metrics such as *Quantity of Motion* (QoM) and a *Contraction Index* (CI) based on the extracted silhouette of the performer and which were calculated using EyesWeb[2]. We adopted these metrics as they have been widely used for the analysis of expressive motion from videos, becoming a sort of standard. However, given the 3D nature of our motion capture data, we adapted the definitions originally formulated for 2D video data.

C. Defining QoM and CI for 3D motion capture data

We (re)defined *QoM* and *Expansion Index* (EI, intuitively the converse of CI) metrics for 3D data. In order to evaluate our definitions we considered desirable properties for motion analysis metrics:

P-1. Independence from the size of the person.
P-2. Independence from the orientation of the person.
P-3. Independence from the sampling frequency.
P-4. Independence from the motion capture system.
P-5. Statistically positive association with the corresponding metric from video processing.

P-2 does not hold for the video processing QoM or CI, since both depend on the position of the camera. P-3 does not hold for the video processing QoM since it considers the change in silhouette between frames, although the n parameter can be adjusted to take into account more or fewer preceding frames.

1) Quantity of Motion: We considered three potential definitions of QoM for 3D data:

QoM-1. The sum of the speeds of each skeleton segment, divided by the number of segments and by the "height" of the performer.[3]
QoM-2. As QoM-1, but with the origin fixed at the pelvis segment, to give motion within the personal space.
QoM-3. The sum of the absolute changes of the joint angles between samples, multiplied by the sample rate. For each joint the single angle is that given by Euler's Rotation Theorem.

QoM-1 and QoM-2 are not independent of the motion capture system used since the addition of extra segments, e.g., more spinal joints, while increasing the completeness of the captured motion, weights the calculated QoM more towards those regions of the body where the segments have been added. An unevenly weighted metric (rather than a simple average, as used here) could compensate for this.

QoM-3 is largely independent of extra *internal* joints being recorded and of the motion capture system used. Therefore, we adopted this definition for the main part of our analysis.

2) Expansion Index: We considered two definitions:

EI-1. The surface area of the convex hull of all the joint positions provided by the Xsens software, plus calculated top of head, and right and left fingertip points. This was divided by the square of the height of the head-neck joint for a standing figure to give a unitless index.
EI-2. The length of a five-sided perimeter connecting the head-neck, wrist, and foot-toe joints. In frames where using an elbow gave a longer perimeter this was substituted for the corresponding wrist. This was divided by the head-neck joint height to give a unitless index.

[1] www.xsens.com/en/general/mvn
[2] www.eyesweb.org
[3] This was actually the height of the head-neck joint, since Xsens did not supply the height to the top of the head.

 http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6681414

Draft. Final version in: Proc. 5th Biannual Humaine Association
Conference on Affective Computing and Intelligent Interaction (ACII 2013)

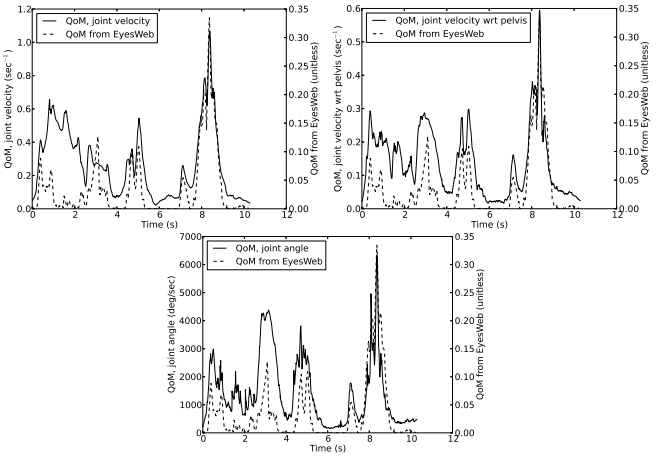

Fig. 1. Comparison of video processing QoM with: QoM-1 (top left), QoM-2 (top right) and QoM-3 (bottom).

EI-2 and EI-1 are comparable for "typical" body poses. However, EI-2 is simpler and hence more easily calculated in real-time for interaction purposes. While EI-1 is computationally more complex, it is intuitively more correct and also covers unusual poses. EI-1 is thus more appropriate for our current analysis purposes, where real-time is not an issue.

D. Comparison of video processing and 3D metrics

To confirm that our definitions were analogues of the existing metrics, we selected one of our dance performances with a variety of movements and generated an animation using Autodesk 3ds Max™. We then used EyesWeb to calculate QoM and CI from this video, and compared them to the values calculated from the motion capture data—see figure 1 for QoM. The metrics were seen to follow each other. Finding such correspondence for all our proposed metrics suggests that QoM and EI are robust concepts, i.e., although some differences do occur, they are broadly insensitive to changes in the way in which they are calculated. In the remainder of the paper, we use QoM-3 (joint angles), and EI-1 (convex hull).

III. Analysis and Results

A. Acting performances

For each emotion, we selected sequences of motion capture data in which the performance wasn't interrupted (e.g., by an instruction from the director) and where there was no change in "scenario" (e.g. from standing to sitting, or from one story to another). We then calculated QoM and EI for each sample. Since it would not be useful to compare these metrics in different "contexts"—the QoM when walking will typically be higher than when standing whatever the mood—for the purposes of analysis we divided the sequences into three groups: standing, walking, and interacting with the other actor. The results are shown in figures 2 and 3. Note that the standing and interacting selections for the sadness mood are the same: the actor was standing while being comforted.

B. Dance sequences

For the dance sequences, we calculated QoM and EI over the selected performances. The results are shown in figures 4 and 5.

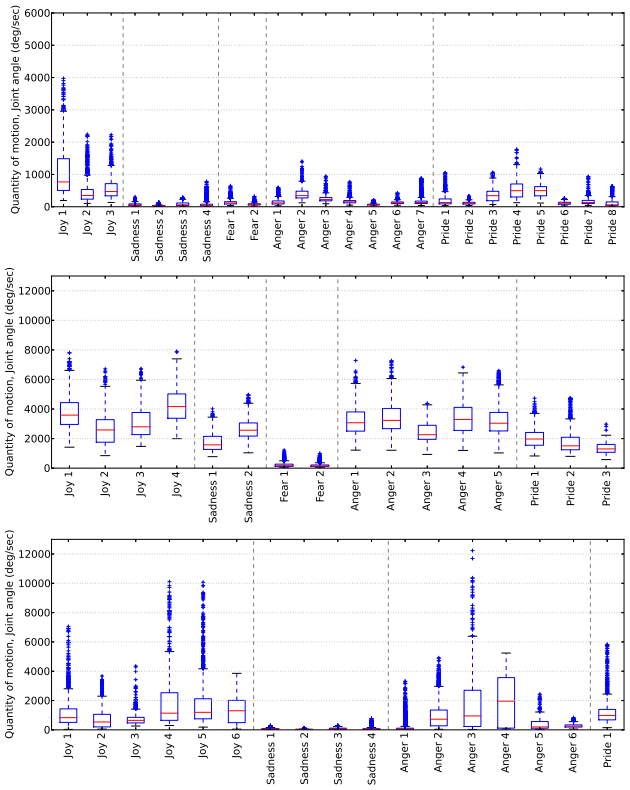

Fig. 2. QoMs for acted sequences: standing (top, y-axis expanded), walking (center) and interacting (bottom).

C. General observations

As our metrics have been constructed, and verified, to follow existing 2D metrics we would expect our results to follow previously observed trends.

1) Quantity of Motion (figures 2 & 4): As expected, the QoMs for *sadness* were typically lower than for other emotions. The exception was the acted walking sequences, where the fear sequences had a very much lower QoM than all the other sequences, while the pride and sadness sequences had comparable QoMs.

The QoMs for *joy* in both acted and danced sequences are typically amongst the highest. The two exceptions were the acted interacting sequences, where some QoMs were in the middle of the range while a single anger sequence had a mean QoM much larger than the other sequences, and the danced sequences, where the QoMs for "fleeing" fear were larger.

Anger sequences were expected to give high QoMs, and although this was sometimes the case (particularly in the acted walking sequences and in the acted interacting Anger 4, and in the dance sequences, where the QoMs for anger are comparable to those for joy) in other cases we had low QoMs (particularly acted standing Anger 5 and acted interacting Anger 1). This illustrates the differences between expressions of hot (high QoM) and suppressed (low QoM) anger.

For *fear*, the acted sequences showed very low QoM, as expected. However the danced sequences showed a high

 http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6681414

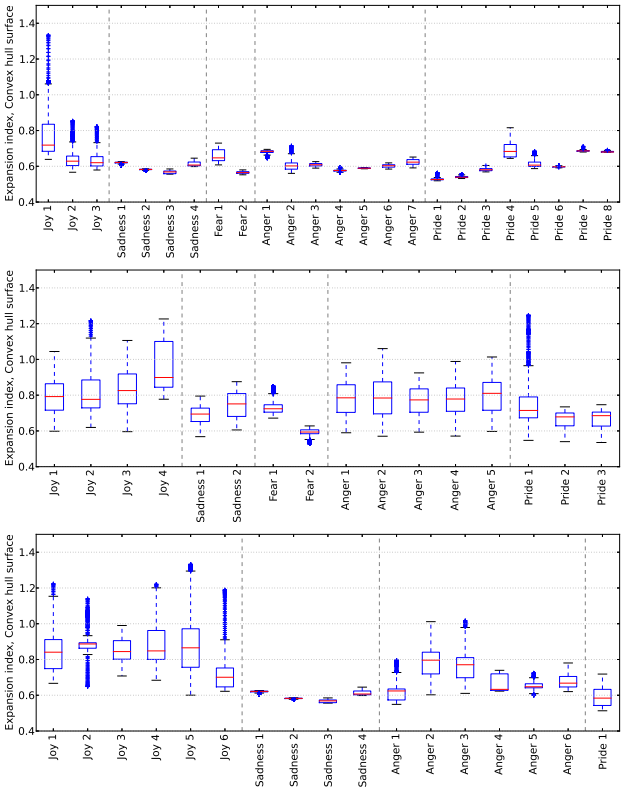

Fig. 3. EIs for acted sequences: standing (top), walking (center) and interacting (bottom).

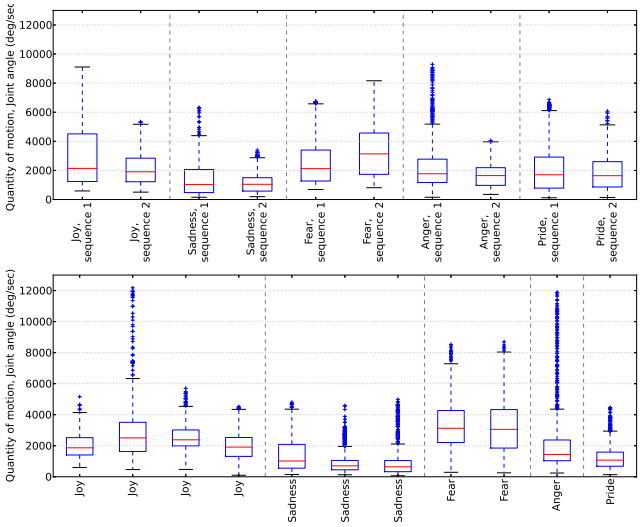

Fig. 4. QoMs for pre-choreographed (top) and free (bottom) dance sequences.

QoM, larger than both joy and anger. Recall that the emotion labeled "fear" in our dance sequences was not described using that word during data collection, but was instead invoked by a scenario. The scenario in this case was "the building is collapsing" and the dancer responded by choosing frantic music and performing in a panicked or fleeing manner resulting in a high QoM. The actor was conversely given a scenario for fear which the director described saying "it's dark" and "there is a killer in the room" resulting in extremely cautious, slow

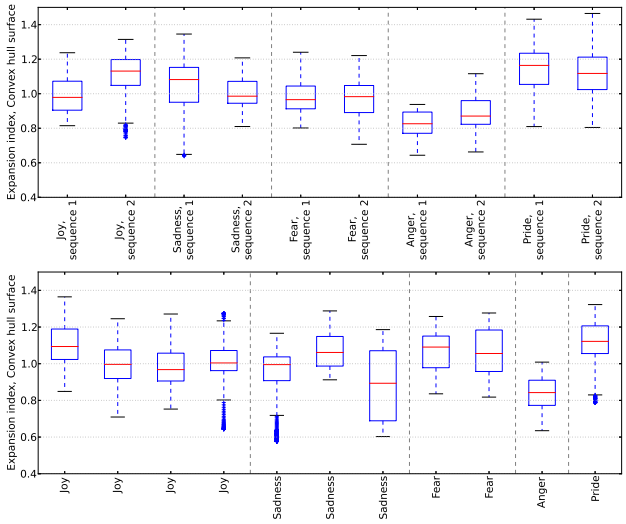

Fig. 5. EIs for the pre-choreographed (top) and free (bottom) dance sequences

movements.

For *pride*, the results showed a degree of variety, with some QoMs comparable to those of joy (acted standing Pride 4 and 5, and acted interacting Pride 1, and in the choreographed dance sequences) and some to sadness (acted walking and free dance sequences).

2) Expansion Index: In the *danced* sequences (figure 5) the clearest result was a sharp reduction in the EI for the *anger* sequences. This was caused by the dancer holding his arms rigid, and closer to his body than normal. However, in contrast to this, for the acted anger sequences (figure 3) the EI was typically higher than in the negative-valence emotions of fear and sadness. For the dance sequences portraying *pride*, there was generally a larger EI than for other emotions. In the acted sequences there were EIs in the lower range for walking and interacting pride, but a wide range of values for standing pride. There was no clear pattern for EIs in other dance sequences, with EIs for the *joy* and *sadness* sequences at both the upper and lower end of the central range.

For the *acted* sequences (figure 3), the EI of *joy* was often higher than other emotions. This trend was strong in five out of the six interacting sequences, but only marginal in the standing sequences, where there was more similarity in EIs across the emotions, probably due to the necessarily similar posture. In the case of *sadness*, the standing/interacting sequences (recall the same data were used for these) are lower than for other emotions. In the case of walking the first sequence has a lower EI, the second, showing residual sadness is closer to the other emotions. There was little clear pattern for EIs in the acted or danced *fear* sequences, with values in the high, middle and low ranges in different situations.

IV. Discussion

A. Problems with the general observations

While we were able to make a number of general observations above, based on a classification of our data into discrete—and for the most part basic—emotions, in many cases these observations were limited or may be misleading. We will now

 http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6681414

highlight a few situations in our data where use of the QoM and EI might result in misidentification of the emotion. It should be noted that with another actor or dancer we may have identified a different set of specific problem cases. This is not a limitation of the study but an inherent feature of emotion expression—individual differences—that highlights the fact that the use (and abuse) of discrete emotion categories with associated prototypical expressions might be misleading.

1) The importance of context: The first observation to make is that for the acted sequences, the similarities within each group (standing, walking, interacting) can overwhelm even strong differences between different emotion categories. The pre-choreographed dance sequences served to provide examples in which the context was to some extent controlled, as the dancer was performing the same steps for each emotion. In social situations, a human will take context into account when assessing the emotion of another agent. This context can include functional actions such as sitting, walking, holding an object, or their focus of attention. It may also include knowledge of the agent's recent states, and any stimulus which may have caused an emotional response.

2) Shortcomings of emotion labels: We have already noted that our actor and dancer "fear" performances were very different, due to the different induction methods used. In most of the studies that we have found, induction is either based on the use of emotional terms, or somehow "shaped" by the experimenter; for example, Roether et al. [4] requested a "fear" response from their actors, but they note that "if an actor first spontaneously chose fast movements, we further instructed him or her to induce a mood that matched slow movements". These differences in interpretation of an emotion label are an example of a more general problem. As Bänziger and Scherer [2] put it: "the use of emotion categories that are too broad and unspecific is detrimental to progress in the field. Basic emotional categories (anger, fear, sadness, etc.) are rather unspecific." When this is combined with a forced-choice design it can lead to a false impression of the accuracy of automatic recognition—see [9].

The use of a few emotion categories should simplify things for automatic recognition. However it is a crude approximation to the real world. On the other hand, using finer graded emotional categories brings its own problems, as the interpretation of subtle emotional labels can differ between individuals. In an international research community this problem is made worse by the use of multiple languages. For an extensive treatment of these issues, we refer the reader to the large-scale cross-cultural study of Scherer et al. [10].

3) Hidden and suppressed emotions: In our acted scenarios, we had situations where anger was hidden (standing next to someone in a lift—standing Anger 6 and 7), where sadness was suppressed (walking Sadness 2, after a devastating event) and where pride was felt, but the character wanted the other person to guess what had happened (interacting Pride 1). As noted above, in our danced sequences the performer chose to dance (unlabeled) anger with the arms held tightly, closer to the body than in other performances. This lead to a very low EI for the whole sequence. The determined control of his own body, combined with sudden, sharp movements, can be considered similar to the acted sequence of suppressed anger in the lift.

In human interaction, the expression of certain emotions (e.g. anger) or extreme emotions may not be socially acceptable based on cultural norms. In addition, an individual may not wish to (directly) share their feelings with someone else. This results in "hidden"[4] emotions. Related to this, a person may wish, even when alone, to suppress negative emotions such as sadness, in an attempt to hold back the unpleasant feelings. Although very frequent in real life, from a categorial perspective these behavioral manifestations would be considered "atypical" expressions of emotions that, in a forced-choice perceptual test could easily be forced into the wrong category.

4) Assumption of monotonicity: In our acted performances we had examples of the same emotion category with different intensities. For example, from the standing selections, Sadness 1 and 3 were high levels of sadness (can be thought of as despair) while Sadness 2 was not so intense. If sadness is generally associated with less movement then we might expect that greater levels of sadness would lead to a smaller QoM. However this was not observed. Here increasing sadness to despair lead to increased agitation and sobbing motions.

In another scenario, the actor was portraying a happy child on her birthday who sees her present, a dog (figure 6, top left). Initially walking, swinging her arms, the first reaction (at around the 5 second mark) was to stop suddenly in joyous surprise and, after an initial open gesture, to bend forward to interact with the dog. This resulted in a drop in the mean QoM and EIs, yet the portrayed emotion was an increase in the intensity of happiness.

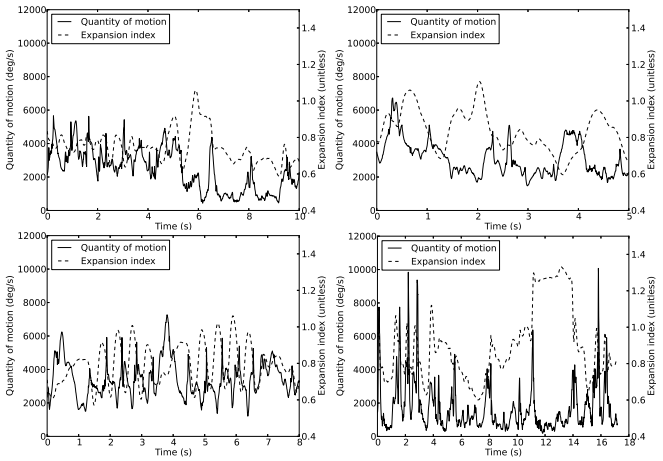

Fig. 6. QoM/EI from acted: "sees a dog" Joy scenario (top left); walking Joy 3 (top right); walking Anger 2 (bottom left); and interacting Joy 5 (bottom right)

5) Easily confused emotion categories: As it is well known in the literature, bodily expressions of joy and hot anger are not easily distinguished by automatic recognition systems. An example in our acted sequences can be seen comparing Joy 3 and Anger 2 walking sequences. These have very similar QoM and EI distributions. Camurri et al. [3] tackled this problem by looking at *motion bells*, essentially looking at the width of the peaks of QoM. We can see this too in the graphs in figure 6,

[4]Although we use the term "hidden", the social purpose may not be to completely conceal the emotion, but rather to modulate it and communicate it in a socially more acceptable manner.

 http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6681414

top right and bottom left, comparing how our metrics vary over time, where there is a clear difference between the two, with sharper, more frequent peaks for anger. However, the motion bell approach does not always work. For example in the QoM and EI for the acted Joy 5, figure 6, bottom right. In this scenario the actor had "won the lottery". The initial sequence includes a number of sharp peaks similar to Anger 2. Informal testing showing a skeleton animation based on this sequence suggests that people also have difficulty recognizing this short sequence as joy without sound or other clues.

B. Validating performances

In contrast with common practice, we did not collect data from multiple actors and dancers in order to provide a statistical average or to find examples of "good" portrayals of each emotion. Rather, we are interested in any and all portrayals, and therefore all examples are interesting in their own unique way. In interactions in real life we do not correct other people: "Hang on! That wasn't a good expression of anger. Do it again!". People simply express emotions the way they do, with rich individual, cultural and contextual variations. Successful social robots will need to interact with individuals, not with averages or gold standards.

In the same vein, after data collection, a common research procedure stemming from emotion recognition studies is to "validate" data by testing with human observers to see if they can recognize the emotion. We again chose not to do this for the following reasons. Firstly, in social interaction, we need to consider how common it is to see emotions that are expressed unambiguously in the short timescales that are used in our analyses. In practice, emotions might be of low intensity, and we have already noted that their expression may be suppressed, or masked by other functional movement. If we are creating an artificial social agent it needs to be able to detect and respond to such subtle signals, as well as being able to produce them in a convincing manner. Secondly, the process of validating motion sequences for recognition rejects data that are ambiguous or misleading. However, in ambiguous situations a human might use contextual clues, knowledge of the person, or might interact/observe further to clarify (this interaction could be as simple as asking "Are you OK?"). Interactions may take place over a period of minutes, hours or years. Experimental data stripped of contextual clues, interaction history, or the opportunity for further interaction is not a realistic model for social interaction. A human may also be poor at accurately recognizing emotions given such artificially impoverished information, as suggested above in the discussion of the confusion of joy and anger. By taking away the social elements we are destroying the most powerful tool we have for learning how to interact with others. Thirdly, selection may also introduce an unintended bias into the sample, where only certain types of clear-cut or stereotypical expressions are retained for each emotion. Even if we have a system for automatically recognizing emotions which outputs probability distributions, as opposed to just selecting one emotion, these probabilities should not be calculated based on a misleading sample. In conclusion, if we are interested in data that is representative of what would typically be seen in social interaction, the selection should not be made on the basis of accurate human recognition from limited cues.

V. Conclusion

Although modeling emotional expression (and recognition) in (by) robots in terms of discrete categories presents advantages such as ease and clarity of interpretation, this approach also poses a number of problems beyond the obvious repetitious nature of the expressions and its negative impact on believability and engagement. The main problems relate to the loss of subtle expressions and feelings, individual features, context, and social interaction elements that are present in real life, not only in human-human interaction but also in interaction with robots even in constrained settings, as we have observed in children interacting with our robot [11]. In agreement with Coulson [5] and Glowinski et al. [12], we need to rethink the way in which we conceptualize and model emotions beyond categories and dimensions. In the case of human-robot interaction we need to focus on what is important for the *interaction*. For a social robot "recognizing" or classifying the affective state of a human that is, say, very still or moves very little, would not be as important as producing a behavior appropriate to the interaction in that situation.

Acknowledgments

We are grateful to Aryel Beck and Luisa Damiano for helping collect and prepare the data, and to Shooting Fish Theatre Company for their collaboration. This research is funded by the EC ALIZ-E project (FP7-ICT-248116). The opinions expressed are solely the authors'.

References

[1] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, 2013.

[2] T. Bänziger and K. R. Scherer, "Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus," in *ACII, Proc. 2nd Int. Conf.* Springer, 2007, pp. 476–487.

[3] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *Int. J. Hum.-Comput. Stud.*, vol. 59, no. 1–2, 2003.

[4] C. L. Roether, L. Omlor, A. Christensen, and M. A. Giese, "Critical features for the perception of emotion from gait," *Journal of Vision*, vol. 9, no. 6, pp. 1–32, 06 2009.

[5] M. Coulson, *Expressing emotion through body movement: A component process approach.* John Benjamins Publishing Company, 2008.

[6] A. Beck, B. Stevens, K. Bard, and L. Cañamero, "Emotional body language displayed by artificial agents," *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 1, March 2012.

[7] M. Rehm, "Experimental designs for cross-cultural interactions: A case study on affective body movements for HRI," *Proc. Humanoids*, 2012.

[8] M. Davies, *Movement and Dance in Early Childhood.* Cambridge: SAGE Publications Ltd, 2003.

[9] A. Winters, "Perceptions of body posture and emotion: A question of methodology," *The New School Psychology Bulletin*, vol. 3, no. 2, 2005.

[10] K. Scherer, H. Wallbott, and A. Summerfield, Eds., *Experiencing emotion: A cross-cultural study.* Cambridge University Press, 1986.

[11] M. Nalin, I. Baroni, I. Kruijff-Korbayová, L. Cañamero, M. Lewis, A. Beck, H. Cuayáhuitl, and A. Sanna, "Children's adaptation in multi-session interaction with a humanoid robot," *IEEE 21st RO-MAN*, 2012.

[12] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *T. Affective Computing*, vol. 2, no. 2, 2011.

