**General comments**

>>*Notable among these are making clearer the links to existing work in the literature (R1, R2, R3)*,

7 new references were added, plus a comparative table from the works studied. These new works include studies using Laban’s features and other approaches.

>>*providing a more comprehensive analysis of the results, including longitudinal (R1,R3)*

>>*providing a clearer link between the series of case studies (R1, R2)*,

As it was suggested by R1 the results are showed after each case study, which makes clear the connection among all the case studies. Additionally, images for each case study setup was added, which will help to give a better idea. Also, it was added an image showing the features used in all the case studies.

>>*providing clarifications regarding the methodology (R2).*

A brief explanation about how the sequences were generated was added. With the additional images, it is expected to give a better understanding on how it was done the case study.   
  
**Reviewers' comments:**  
  
***Reviewer #1:*** *My comments on the manuscript:*

>> (R1) *The state of the art is rather incomplete.*

>>(R1) *Because existing work in robotics are not discussed it is difficult to understand the manuscript's contribution.*

>> (R1) *Finally, trying to use Laban's theory for robots to express emotion has been proposed many times before.*

7 new references were added, some of them using Laban’s approach. Also a comparative table was added.  
   
>>*I would suggest the authors to present meaningful result after each study. There seems to be no connection between the different cases. After reading the paper, I am unable to tell what was learnt from case study 1 or 2 or 3 or 4...* *Moreover, there seems to be no continuity, I was expecting the work to build on each case study before moving to the next one but this is not the case.* *I would suggest the authors to summarize the finding/lesson learned after each case study rather than grouping everything.  
Additionally, because they are all grouped together, the results are not really helpful.*

The suggestion was taken into account and after introduced each case study was also given the results. Also, it was added images showing each case study setup.  
  
  
***Reviewer #2:***   
  
*>> (R2) When talking about the limitations of existing work I expect to see references. For example, in the introduction, which papers showed that expressions based on limbs and faces are not accepted by people? does the authors' work improve on the acceptance?*

It was corrected the sentence in the introduction, because our idea was not to say that faces and limbs are not accepted by people, but rather that not all robots need them to perform their task.

>>(R2) *The expressions in this paper are more abstract. I do not expect them to be more acceptable. At least, I did not find proof from the paper. I suggest the authors to focus on the key motivation, 'designing expressions for robots lacking of humanlike features', instead of other aspects.*

It is true that human like features are need it to convey and increase the perception of emotions, as it shown in our case studies. However not all robots will need to have human like features to share spaces with people. But still, these platforms will be required to somehow interact with people, which could require nonverbal communication, such as emotions.  
>> (R2) *At the end of the Section 'Related Work - Human studies', it is unclear to me what the authors meant by 'drawbacks are the use of video recorded sequences' and 'misses physical contact'. It is too concise to talk about the drawbacks and thus difficult to compare with the contributions of this paper.*

There have been some works that have studied the impact that have an interaction with a real body. One of their found is the impact that has the embodiment in human-robot interactions. We think that a similar effect could be obtained in this kind of studies. However a further study is required to identify the difference.

>>(R2) *The description of the hardware evolution of the robot platform deviates from the topic. It is sufficient to describe the state of the robot when the experiment was carried out.*

Agree, it has been deleted irrelevant information that could deviates from the topic, but information about the platform’s measures is given.

>>(R2) *It would be nice to describe more about the 'emotional enrichment system' instead of just referring to previous work, since it is a key element of this paper.*

Unfortunately given further information about the emotional enrichment system could deviates reader from the topic, so it has been removed from the article and briefly mentioned.  
*>>(R2) The description of the pilot study is out of the picture. It does not add much to the main contribution of the paper. I suggest to remove it.*

As it was suggested, the pilot was deleted and additional information was given in the other case studies.  
  
*>>(R2) Four case studies are described, but the connections between each other are missing. I suggest adding paragraph to explain what the contribution of each case study on top of the previous ones is. In fact, the four case studies are not well organized in the paper. For example, the authors start to say the results of the first study are poor before even describing the results of the first study. The beginning of the fourth case study is similar. If the first study was just a failure and the second one is the replacement, just remove all parts of the first study from the paper! Moreover, in the results section, the authors said themselves the results of the four case studies are not comparable. How can they be connected in one paper then?*

To make the connection among all better, it was added the results after each case study. This will help the reader to understand the situation after each case study and create a mental map about their relationship. Also were included phrases were these relation is clear.   
*>> (R2) The procedure and design of those case studies are not detailed. For example, the first case study seems to be a between-subject study, as the paper says 'different emotions were shown to each group'. How the emotions were distributed to the groups is not mentioned.*

The description about how the sequences were generated has been added.  
  
*>> (R2) The participants of the case studies have a large range of age. Even 4 or 5 years old child was included. I would not consider the data from them as valid and stable.*

It is true that their cognitive abilities are not the same as older child. However, it is important to consider even young children because even they are going to interact with the robots. Their exclusion could lead to misleading results

*>>(R2) Did the authors filter out these data from analysis? Do the case studies not have to follow ethical standards?*

We did not put children in any case of dangerous situation nor asked for information (e.g. name and last name) that could let us to trace back any participants, including child. The only information asked to the participants was sex and age.   
  
*>>(R2) Overall, it is difficult to judge the validity of the results presented in this paper, due to lack of details in experiment design, procedure, and analysis,* *And due to the fact that all case studies were uncontrolled experiments.*

It is important to highlight that the studies presented in the paper are not experiments. It is true that in case studies is impossible to control all the variables, and that is the reason it is not suitable for all the cases. For our purpose, doing a case study was suitable, mainly because we were trying to understand people’s perception to specific movements.

>>*The use of the blue color to eliminate color inference is questionable, as blue is a typical cold/calm color. Why not use a more neutral color like white?*   
  
  
**Reviewer #3**:   
  
>>(R3) *In order to emphasize this aspect, the authors may provide more details as to the measurements of the robot, length of the inner arms, and so on.*

Agree. information about the robot is given in the new images, where the components distribution and measures are given for each platform’s version.

>>*As many other studies are done with existing commercial platforms, the authors could also provide best-practice hardware design suggestions for a non bio-inspired robot, based on their experience iterating with prototypes.*

Thanks for the suggestion, however due to space limitation it was not possible to include them.  
  
>> (R3) *The authors should also be careful not to overclaim. For example, on page 18, the authors write: "this lesson opens the door to the inclusion of emotion to robotic platforms [without] humanoid characteristics" cannot be substantiated, since Saerbeck et al. already studied emotion in the Roomba as early as 2010.*

Agree, this reference has been added in the related work. This and other over claims have been corrected.

>> (R3) *1. Since the paper does not provide in-depth experiments, the authors could provide subjective data such as quotes or comments from observers, if available. Photos of the robots in context of the studies could also help understand the scenarios. For example, it is unclear how isolated the robot was in the museum/open lab contexts: an isolated robot vs. a robot nearby walls or humans could have an effect on recognition rate.*

It is important to highlight that the studies presented in the paper are not experiments. Images describing the robot movements in each case study, as well an image describing the features used are included in the article.

>> *(R3 )2. The results section should be expanded with analysis of the tables. For example:  
- pg. 16, line 19. The paragraph explains the contents of table 4, but what is their significance? What can you conclude? For example, why is happiness well-recognized? What about the confusions between anger and happiness?*

A Fisher's exact test and Holm-Bonferroni correction were done to verify confusion among the different implementations and also verify if there was any difference between the results obtained in the first and second case study.

>>*The references section is messy. For example, "Saerbeck et al .. Ro-man" vs. "Embgen et al... Proceedings of...(Ro-man'12)". Please standardize.*

Thanks for noticing this. All the necessary changes have been done.