GENERAL CONSIDERATIONS OF THE AUTHORS

We greatly appreciate the close readings of the four Reviewers and their pertinent comments that have been used to revise and improve the text. Below we reply in detail to the Reviewers’ concerns: Reviewers’ comments are written in italic and our replies follow in plain text. Changes in the revised manuscript are also highlighted in yellow.

REVIEWER #2

General considerations

This manuscript presents an experiment to compare virtual reality and reality in terms of neuromuscular organization during upper limb rehabilitation. The experiment is explained clearly. Most of the paper is focused on data analysis. The results show similarity between real and virtual rehabilitation activities.

Authors’ reply

We thank the Reviewer for the suggestions to improve the quality and readability of the paper. Issues have been solved and reported in the following.

Comments

R#2.1. The researchers did not consider the effect of the weight of the ball on the muscle activities. Will the result be different when comparing throwing a real heavy ball with a virtual ball?

Authors’ reply

We thank the Reviewer for pointing out this pertinent aspect. We agree with the Reviewer that a ball of considerable weight would influence the results between real throw and virtual one. The direct consequence would be a variation of muscle activations between the two tasks. For this reason, we chose to use for the experimental protocol a ball of negligible weight, approximatively 30 g. We added this information to the amended version of the manuscript.

Amended version of the manuscript (page 5, lines 153):

“with a ball (diameter 6.5 cm, weight 30 g)”

R#2.2. As discussed in section 4 and section 1, some related works have been done. Can the researchers explain why the results do not agree with [27] but do agree with [55]?

Authors’ reply

We thank the Reviewer for allowing us to better explain this aspect. In our opinion the two works did not provide conflicting results, but they used muscle activation to study different phenomena. The differences in findings between the work of Sabatini et al. [27] and Kang et al. [28 instead of 55 after revision] are related to the different aims of the two manuscripts. In fact, Sabatini and colleagues studied a natural upper-arm
movement composed of sequential phases of reaching, grasping, and retrieval in the horizontal plane. They have found differences in the planning of the action both for kinematic variables and for muscle activation, with the conclusion that different subjects plan differently the same motor acts. However, in this work, the task was performed in a real environment and the authors did not investigate whether differences were present in muscle activation with the same task performed in virtual reality. We decided to mention this work in the introduction to emphasize the importance of studying muscle activation to gather further information with respect to the kinematic analysis.

On the other hand, Kang et al. compared the muscle activity of fully immersive VR motions with that of actual archery motions to investigate differences. The authors found that the evaluated muscles did not show any significant difference between real and virtual tasks and therefore, the two groups showed similar muscle activity patterns. These results were totally in line with our results, allowing us to state that virtual reality can be used for rehabilitation as it does not modify the motor organization and muscular activation.

Some content in the discussion section 4 can probably be moved to section 1 to stress the novelty of the paper and make the discussion more coherent.

**Authors’ reply**

We thank the Reviewer for this suggestion. With reference to the comment R#2.2 we decided to add these sentences in the introduction to better address to both comments.

**Amended version of the manuscript (pages 2-3, lines 82-86):**

“On the other hand, Kang et al. compared the muscle activity with fully immersive VR to real archery motions in order to investigate the effects of VR. The authors found no significant difference between real and virtual tasks in terms of muscle activity patterns [28]. These findings encouraged us to question whether virtual reality can be used for rehabilitation, thereby studying different muscle activation by means of muscle synergies.”

A lot of statistical models and skills are used in section 2.3, 2.4, 2.5, and section 3, but the presentation needs to be improved to make it easier to follow. What is the muscle synergy? What is the number of muscle synergies? What is an overall model? In equation (1), is EMG the 12 by 20,000 matrix? If m is the number of muscles, shouldn’t it be 11? It is unclear how to compute W_i and C_i. Also, what is the reconstructed EMG matrix? What is Pearson’s coefficient? What is NNMF?

**Authors’ reply**

We thank the Reviewer for the suggestion to improve the readability of the manuscript. For the question related to the muscle synergies, we decided to add the Figure 4 in the amended version of the manuscript, as also suggested by Reviewer#3 and the Associate Editor.

- What is the number of synergy?
  The number of synergies represents the minimum number of controlling signals for obtaining a good reconstruction of the muscle activation data by using the linear combination. We added this explanation in the amended version of the manuscript.

**Amended version of the manuscript (page 7, lines 214-216):**

“Thus, the NoS represents the minimum number of synergies required to obtain a proper representation of the muscle activation data by using the linear combination in the equation 1.”

- What is an overall model?
  The overall model is when considering all i-th synergies in the model together. We added this explanation to the amended version of the manuscript.
Amended version of the manuscript (page 8, line 234):
“i.e. considering all of the i-th synergies together”

- In equation (1), is EMG the 12 by 20,000 matrix? If m is the number of muscles, shouldn’t it be 11?
We thank the Reviewer for highlighting this mistake. The EMG matrix dimensions are actually 11 by 20,000. We corrected in the amended version of the manuscript.

Amended version of the manuscript (page 6, line 183):
“m was equal to 11”

- It is unclear how to compute Wi and Ci, What is NNMF?
The NNMF is a mathematical algorithm for the factorization of a matrix. This factorization leads to the computation of W and C. For the meaning of W and C please see Figure 4 in the amended version of the manuscript. As regards, NNMF please consider reference [47] for technical details.

- What is the reconstructed EMG matrix?
The reconstructed EMG matrix is the matrix obtained by means of the linear combination of W and C obtained with the minimum number of muscle synergies. We added this information to the amended version of the manuscript.

Amended version of the manuscript (page 6, lines 191-192):
“The reconstructed EMG was obtained by applying the linear combination of Wi and Ci in the output of the NNMF.”

- What is Pearson’s coefficient?
Pearson’s coefficient represents the most used method for assessing the linear correlation among two vectors; in our case it is used to verify the goodness of reconstruction, i.e. the correlation between the acquired EMG and the reconstructed one. We added this information to the amended version of the manuscript.

Amended version of the manuscript (page 7, lines 207-208):
“Pearson’s coefficient allowed us to quantify the linear correlation between two vectors [49].”

R#2.5. What is Shapiro-Wilk test? What is ANOVAs? What is Bonferroni’s test?

Authors’ reply
We thank the Reviewer for the comment. The three reported statistical tests are commonly used in scientific field when looking for statistical differences among different groups. More specifically, Shapiro-Wilk test is applied to verify the normality in distribution of data, representing one of the mandatory hypotheses for the applicability of the ANOVA test. ANOVA, acronyms of Analysis of Variance, allows to find statistical differences when the testing groups are more than two. The results of ANOVA permit to only establish if statistical differences are present among groups without assessing where these differences occurred. Finally, Bonferroni’s test allows to understand where the differences are by performing statistical tests for each pair of groups.

In our opinion the addition of technical information on these aspects in the main text of the manuscript would decrease the readability and fluidity of the paper; for this reason, we decided to only add a reference for the technical details of performed statistical tests.

Amended version of the manuscript (page 8, line 261):
“For technical details about the statistical tests performed, please see [49].”
R#2.6. | *It is unclear how to get the D, ND, and Mode values in Tables 1 to 4? What do they stand for? It is highly recommended to add some qualitative explanation after each section of statistical analysis.*

**Authors’ reply**
We thank the Reviewer for the suggestion to better explain the elements in Table 1 to 4. Actually, D and ND represent the results obtained with the Dominant and Non-Dominant side. Dominant side was defined as the one used by the participant for writing. We added this information in the methods.

**Amended version of the manuscript (page 5, lines 156):**
“"The dominant arm was selected as the one commonly used, by the participant, for writing."

For sake of clarity we added the meaning of D and ND, as well as R and VR, also in the Table captions.
As concerns the Mode, it is computed by selecting the mode value across different tasks performed by the same participant as already reported in the lines 216-217.